Petroleum Business Review, Vol. 10, No. 1, pp. 43–80, Winter 2026

Enhancing Crude Oil Price Forecasting through Hybrid VMD–SVR Models: Evidence from WTI Futures across Multiple Forecast Horizons

Reza Maaboudi[1*], Mohammad Hassan Fotros[2], and Erfan Babaali[3]

[1] Associate Professor, Department of Economics, Faculty of Humanities, Ayatollah Boroujerdi University, Boroujerd, Iran
[2] Professor, Department of Economics, Faculty of Economic and Social Sciences, Bu-Ali Sina University, Hamedan, Iran
[3] M.A. Student, Department of Economics, Faculty of Humanities, Ayatollah Ozma Borujerdi University, Lorestan, Iran

Highlights

- This study develops a novel hybrid VMD–SVR framework for forecasting West Texas Intermediate WTI crude oil prices across short-, medium-, and long-term horizons.
- Variational Mode Decomposition VMD effectively separates high-frequency noise from underlying structural trends, thereby enhancing the predictive performance of the Support Vector Regression SVR model.
- Empirical results indicate that the hybrid model outperforms conventional machine learning approaches, reducing forecasting errors by 36.8 percent and demonstrating statistical superiority based on the Diebold–Mariano test.
- By capturing complex nonlinear dynamics and filtering volatility, the model offers practical insights for policymakers, investors, and financial analysts, thereby supporting risk management, investment planning, and strategic energy policy formulation.
- Overall, this approach provides a robust and adaptive framework for addressing uncertainty in oil markets and contributes to the advancement of signal-based forecasting methodologies in energy economics.

Received: October 31, 2025; *revised:* February 08, 2026; *accepted:* February 09, 2026

Abstract

West Texas Intermediate (WTI) crude oil serves as a pivotal benchmark in the global energy market, exerting a decisive influence on economic expectations and national macroeconomic policies. As the primary pricing reference on the New York Mercantile Exchange and for numerous energy futures contracts, WTI is characterized by persistent and pronounced price volatility. Such fluctuations, often manifested as abrupt upward or downward shocks, significantly affect key macroeconomic indicators, including inflation, economic growth, trade balances, corporate profitability, production costs, and government budgets. Consequently, variations in WTI prices influence oil and gas markets, financial stability, energy security, and international geopolitical relations. To address these challenges, this study develops a hybrid variational mode decomposition–support vector regression (VMD–SVR) framework to model and forecast WTI crude oil futures prices across short-, medium-, and long-term horizons. Empirical results demonstrate that, across all three horizons, the proposed hybrid model consistently achieves the lowest forecasting errors compared with alternative approaches. Furthermore, the Diebold–Mariano and Wilcoxon tests statistically confirm the superior predictive performance of the hybrid VMD–SVR model. These findings underscore the importance of integrating advanced adaptive signal decomposition techniques, such as variational mode decomposition, with powerful nonlinear learning algorithms, such as support vector regression, for accurate oil price forecasting.

[*] Corresponding author:
Email: 1979maaboudi@gmail.com

The proposed approach not only enhances forecasting accuracy but also provides practical insights for policymakers in managing economic risks, stabilizing budgets dependent on oil revenues, and formulating sustainable energy strategies. In addition, it establishes a foundation for the development of financial forecasting models informed by advanced signal processing methodologies.

Keywords: Crude oil price, Forecasting, Variational mode decomposition, Machine learning models, SVR.

How to cite this article
Maaboudi, R. et al., Enhancing Crude Oil Price Forecasting through Hybrid VMD–SVR Models: Evidence from WTI Futures across Multiple Forecast Horizons, Petroleum Business Review, Vol. 10, No. 1, p. 43–80, 2026. DOI: 10.22050/pbr.2026.556568.1420

1. Introduction

Crude oil, as one of the world's most critical energy resources and strategic commodities, plays a fundamental role in the dynamics of the global economy. Oil-exporting countries derive a substantial portion of their revenues from this resource, whereas importing countries, owing to their heavy dependence on energy for production and economic activity, remain highly exposed to oil price volatility. Crude oil prices are influenced by a complex set of factors, including geopolitical risk, climate change policies, capital flows, high-frequency macroeconomic indicators, and global growth trends. This inherent volatility underscores the need for accurate and reliable forecasting. Policymakers rely on such forecasts to formulate national budgets, guide investment in the energy sector, manage strategic reserves, and design fiscal policies. Similarly, investors and multinational corporations depend on precise price projections to support effective risk management, optimal capital allocation, and long-term strategic planning (Baffes et al., 2015). Consequently, oil prices constitute a critical determinant not only of national economic output (GDP) but also of global inflation, international trade, and financial stability. The development of rigorous scientific forecasting methodologies is therefore essential to enable market participants to identify strategic opportunities and mitigate financial risks effectively.

The accuracy of oil price forecasting is highly sensitive to factors such as the sampling period, data frequency, and the presence of structural breaks, including outliers. Statistical properties such as non-stationarity, noise contamination, nonlinearity, and the inherently stochastic behavior of oil prices further complicate the modeling process. Accurately capturing these dynamics requires advanced and innovative forecasting approaches. The pronounced volatility that has characterized international crude oil markets in recent decades has posed significant challenges to conventional forecasting techniques. Although classical econometric models are grounded in strong theoretical foundations, their performance is often limited under conditions of severe volatility and complex nonlinear interactions. The rapid advancement of artificial intelligence has stimulated increasing interest in machine learning and deep learning algorithms for time series forecasting (Wang et al., 2023; Li et al., 2024). However, these methods also encounter limitations related to interpretability, overfitting, and the handling of highly complex time series structures. In addition, standalone machine learning models often struggle to learn effectively from noisy data, and their forecasts may exhibit substantial errors.

To address these limitations, hybrid forecasting approaches have gained considerable attention. These models integrate machine learning algorithms with signal decomposition techniques to separate an oil price series into components associated with different frequencies. This strategy enables the machine learning model to analyze and forecast each simplified component individually, rather than modeling the raw and highly complex series directly, thereby improving overall predictive accuracy.

This study develops forecasts for West Texas Intermediate (WTI) crude oil prices over the period from May 11, 1987, to September 5, 2025, across short-, medium-, and long-term horizons. As a primary global benchmark, WTI prices play a decisive role in shaping economic expectations and macroeconomic policy worldwide, which underscores the practical importance of accurate forecasting. As the principal pricing reference for the New York Mercantile Exchange and numerous energy futures contracts, WTI exhibits substantial price volatility, making precise forecasting essential for financial market participants and energy policymakers.

The proposed methodology adopts a decomposition-based forecasting framework. First, the oil price time series is decomposed into its constituent components—trend, seasonal, and noise—to isolate meaningful patterns from stochastic fluctuations and facilitate model learning. Next, machine learning algorithms generate forecasts for each component separately according to its frequency characteristics. The individual forecasts are then aggregated to obtain the final price prediction. Finally, forecasting accuracy and model performance are evaluated and compared across all time horizons using standard performance metrics.

The principal contribution of this research is the development of a hybrid framework that integrates time series decomposition with machine learning to construct an effective model for forecasting WTI crude oil prices across multiple time horizons. This study addresses a critical research gap arising from the inability of conventional decomposition methods, such as seasonal-trend decomposition using LOESS (locally estimated scatterplot smoothing) (STL), wavelet transforms, and fast Fourier transform, to adaptively separate variable-frequency components in oil price series, together with the limitations of standalone machine learning models in capturing complex nonlinear patterns in non-stationary and noisy data. This dual limitation often leads to substantial forecasting errors, particularly over medium- and long-term horizons.

2. Literature review

2.1. Oil price forecasting models

The existing literature on oil price forecasting generally classifies models into four categories based on their statistical methodology and structural characteristics: classical time series models, machine learning and artificial intelligence models, hybrid models, and structural or factor-based models (Cheng et al., 2023).

2.1.1 Classical forecasting models

Classical time series models rely exclusively on historical data and seek to provide a theoretically grounded representation of variable dynamics. These models are typically based on strict assumptions of linearity, stationarity, and normality. Prominent examples include the autoregressive integrated moving average model for trend forecasting and the generalized autoregressive conditional heteroskedasticity model for volatility analysis. Although these approaches are valued for their structural simplicity and interpretability, they are often inadequate for capturing the complex nonlinear dynamics inherent in financial markets (Ren et al., 2022).

2.2.2. Machine learning (ML) and artificial intelligence (AI) models

This category comprises a range of algorithms designed to capture nonlinear relationships and complex patterns in data, often achieving superior forecasting accuracy, particularly when large datasets are available. Prominent techniques include support vector machines (SVM), tree-based models such as random forest (RF), and artificial neural networks (ANN). Artificial neural networks are especially

valued for their capacity to model intricate nonlinearities, process high-dimensional data, and maintain robustness in the presence of noise. However, these methods present several challenges, including substantial data requirements, sensitivity to hyperparameter tuning, and a heightened risk of overfitting, all of which may limit their generalizability (Zhang et al., 2023).

Machine learning enables models to learn iteratively from data, thereby improving predictive performance without relying on explicit rule-based programming. These algorithms identify complex patterns within input variables, such as historical oil prices, to generate forecasts and support decision-making. Among the widely used machine learning techniques are support vector machines, random forest, and deep learning models. In particular, support vector regression offers advantages such as strong performance with limited sample sizes, robustness against overfitting, and flexibility through the use of kernel functions. Nevertheless, it is highly sensitive to hyperparameter selection and may face computational limitations when applied to large datasets (Liu et al., 2024).

Random forest, an ensemble learning method based on multiple decision trees, provides high predictive accuracy, inherent resistance to overfitting, and the capacity to handle mixed data types while offering measures of variable importance. Its limitations include limited interpretability due to its black-box structure, sensitivity to the number of trees in the ensemble, and restricted extrapolation capability beyond the range of the training data (Langsetmo et al., 2023).

Deep learning, a subset of machine learning, employs artificial neural networks with multiple hidden layers to learn complex hierarchical data representations. It is particularly effective for high-dimensional and structured data, including long time series and textual information. In the context of oil price forecasting, models such as long short-term memory networks, Convolutional neural networks, and transformer architectures have been applied to capture temporal dependencies and nonlinear patterns (Foroutan et al., 2024).

2.2.3. Structural or factor-based models

Structural or factor-based models explain oil price dynamics by linking market fundamentals, such as global demand, production capacity, inventories, and geopolitical risk, to observed price fluctuations. These models emphasize underlying causal mechanisms, which makes them particularly suitable for long-term forecasting and policy analysis. Recent advances have incorporated approaches such as Bayesian model averaging and structural vector autoregression models to improve variable selection and more effectively capture underlying economic shocks (Kilian and Zhou, 2020; Drachal, 2023).

2.2.4. Hybrid models

Crude oil prices exhibit nonlinear, volatile, and multi-scale behavior. Hybrid models address these characteristics by decomposing the original time series into simpler components, such as trend, short-term cycles, and noise, which are subsequently forecast separately using appropriate machine learning algorithms. The individual predictions are then aggregated through summation or averaging to produce the final forecast. A principal advantage of hybrid approaches is the enhancement of forecasting accuracy through noise reduction and the separation of trend and oscillatory components. Decomposition techniques such as variational mode decomposition and empirical wavelet transform enable machine learning models to train on more stationary and information-rich subseries. Several studies, including Li et al. (2020) and Niu and Zhao (2021), demonstrated that combining variational mode decomposition with long short-term memory networks or kernel extreme learning machines significantly reduces root mean square error and mean absolute error compared with standalone models.

Recent research further supports the effectiveness of multi-stage hybrid frameworks. Zheng et al. (2025) and Kayral et al. (2025) showed that hybrid models employing multi-stage decomposition methods, such as empirical wavelet transform or wavelet-based techniques, enhance forecasting performance under highly volatile market conditions. These approaches decompose complex fluctuations and subsequently integrate the outputs using ensemble learning structures. Another strength of hybrid models lies in their ability to improve the understanding of price dynamics. Signal decomposition facilitates the separation of short-, medium-, and long-term components, thereby providing clearer insights into the underlying structure of market movements. For instance, Peng et al. (2023), by integrating variational mode decomposition with convolutional neural networks and bidirectional Gated recurrent units, demonstrated that decomposition enables more precise analysis of market oscillations. Similarly, Jia et al. (2025), through the development of successive-traction variational mode decomposition, revealed local structures of oil price signals with greater resolution and accuracy.

Hybrid models are particularly well suited to non-stationary and volatile data environments. By integrating decomposition techniques such as variational mode decomposition or wavelet transforms with deep learning architectures, these models achieve high flexibility and predictive accuracy. Niu and Zhao (2021) and Huang et al. (2024) showed that variational mode decomposition- and empirical mode decomposition-based frameworks outperform traditional models under conditions of pronounced volatility. Moreover, hybrid models often demonstrate greater out-of-sample stability due to their reduced reliance on a single predictive algorithm. Boussatta et al. (2023) reported that combining weak regression learners with support vector regression produced hybrid models with more stable performance and lower forecasting errors.

A further advantage of hybrid models is their capacity for multivariate integration. The synergistic combination of decomposition and machine learning enables the effective incorporation of auxiliary variables, such as exchange rates, inventory levels, and geopolitical risk indices, alongside historical price data. Qin and Li (2025) developed a four-module framework consisting of decomposition, prediction, integration, and error correction, demonstrating that such a multi-stage architecture significantly enhances performance by leveraging multivariate information more efficiently.

Finally, hybrid models mitigate overfitting because each decomposed component is modeled separately, which reduces the likelihood that the model will learn noise. Zhu et al. (2025) and Yuan et al. (2025) demonstrated that multifaceted and ensemble-based frameworks improve stability and limit overfitting when forecasting volatile prices. In addition, hybrid frameworks support flexible combinations of algorithms, including long short-term memory networks, convolutional neural networks, gated recurrent units, support vector regression, extreme gradient boosting, and transformer models. Singh et al. (2025) showed that integrating Long Short-Term Memory networks, Convolutional neural networks, and temporal convolutional networks increases model selection flexibility and improves forecasting accuracy relative to individual models.

3. Methodology

The purpose of this research is to develop a modeling framework for forecasting the West Texas Intermediate (WTI) crude oil price. The methodology employs an analytical approach comprising several key stages. First, the oil price time series is decomposed using various signal decomposition techniques. This process separates the original series into distinct intrinsic mode functions (IMFs), each representing unique characteristics, including noise, short-term fluctuations, and long-term trends. Second, during the feature extraction phase, essential statistical and temporal descriptors are computed for each IMF to enhance the representational richness of the data. To preserve causality and prevent data leakage, only lagged components from the decomposed series are used as model inputs. Third, in

the modeling and forecasting phase, the support vector regression (SVR) algorithm with a radial basis function (RBF) kernel is applied to capture nonlinear relationships between the extracted features and future oil prices. Model hyperparameters are optimized via grid search to ensure robust generalization. Forecasts are generated for three horizons: short, medium, and long term. Finally, the decomposed components are forecasted using various machine learning methods, and the predictive accuracy of each model is evaluated statistically using standard error metrics, including root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and R-squared (R^2). In addition, the statistical significance of differences in predictive accuracy is assessed using the Diebold-Mariano test (Coroneo and Iacone, 2025) and the Wilcoxon signed-rank test (Huang and Sen, 2025) to confirm that observed improvements are statistically meaningful and not attributable to random chance.

3.1. Oil price time series decomposition approaches

In the following, conventional signal decomposition methods for the oil price series are presented. The primary objective of applying these techniques is to extract latent signals and hidden features from the data, thereby enhancing the predictive power of the machine learning models through their incorporation into the input feature set.

3.1.1. Seasonal-trend decomposition using LOESS method

The primary objective of the Seasonal-trend decomposition using LOESS method is to decompose a given time series (Y_v) into three key components: seasonal (S_v), trend (T_v), and remainder (R_v). This approach employs LOESS (locally estimated scatterplot smoothing), a nonparametric smoothing technique based on fitting local polynomials to data points. The STL framework consists of two recursive loops: an inner loop and an outer loop. The inner loop iteratively estimates the trend and seasonal components through six consecutive steps. First, a detrended series $\left(Y_v - T_v^{(k)}\right)$ is computed; for the initial iteration, the trend component $\left(T_v^{(0)}\right)$ is set to zero. Second, cycle-subseries smoothing is performed by applying LOESS to the subseries at each seasonal position, yielding the cyclical component $\left(C_v^{(k+1)}\right)$. Third, a three-stage low-pass filter—comprising two moving averages and one LOESS smoother—is applied to $\left(C_v^{(k+1)}\right)$ to obtain the long-term smoothed component $\left(L_v^{(k+1)}\right)$. Fourth, the smoothed subseries is detrended, and the new seasonal component $\left(S_v^{(k+1)}\right)$ is calculated as $C_v^{(k+1)} - L_v^{(k+1)}$. Fifth, the time series is deseasonalized, producing the seasonally adjusted series $\left(Y_v - S_v^{(k+1)}\right)$. Finally, trend smoothing is conducted by applying LOESS to the deseasonalized series to extract the updated trend component $\left(T_v^{(k+1)}\right)$. The outer loop enhances robustness against outliers. When anomalies are detected, the standard LOESS procedure in steps 2 and 6 of the inner loop is replaced with its robust variant. This ensures that outliers have minimal influence on the seasonal and trend estimates, effectively allocating all anomalies to the remainder component (R_v) (Ouyang et al., 2025, p. 3).

3.1.2. Wavelet transform

In this research, the discrete wavelet transform (DWT) is employed as an effective approach for analyzing oil price time series. By decomposing the signal into frequency components across multiple time scales, this method facilitates the identification and extraction of hidden patterns within the data. In contrast to the Fourier transform, which provides information exclusively in the frequency domain,

the wavelet transform preserves both time and frequency characteristics simultaneously, making it a powerful tool for analyzing nonstationary financial signals.

Within the DWT framework, the original signal $f(t)$ is decomposed into two sets of coefficients at different levels *j*: approximation coefficients ((A_j)) and detail coefficients ((D_j)). The approximation coefficients capture the overall trend and low-frequency components of the signal, whereas the detail coefficients represent short-term fluctuations and high-frequency components. The general decomposition formula is expressed as follows.

$$x(t) = A_J(t) + \sum_{j=1}^{J} D_j(t) \tag{1}$$

where $A_J(t)$ denotes the approximation component at the highest decomposition level *J*, and $D_j(t)$ represents the detail component at level *j* (Kayral et al., 2025, p. 5).

3.1.3. Variational mode decomposition

Variational mode decomposition (VMD) is an adaptive, fully nonrecursive signal decomposition technique designed to decompose a real-valued signal into a discrete set of quasi-orthogonal, band-limited intrinsic mode functions (IMFs), each characterized by a specific sparsity property in the spectral domain. Unlike related methods such as empirical mode decomposition (EMD), VMD is grounded in a rigorous mathematical framework and exhibits greater robustness to sampling noise as well as improved resistance to mode mixing. The core of the VMD algorithm involves solving a constrained variational optimization problem to achieve the desired decomposition.

In this study, the VMD algorithm is applied to decompose the original oil price time series, denoted by $f(t)$, into a predefined number (*K*) of modal components, collectively represented as $\{u_k(t)\}_{k=1}^{k}$. VMD performs this decomposition by solving a constrained variational problem in which each intrinsic mode component $u_k(t)$ is modeled as a compact, band-limited signal centered around a specific frequency ω_k. These center frequencies are estimated simultaneously with their corresponding modes during the decomposition process. Specifically, the following objective function is minimized:

$$\min_{\{u_k\},\{w_k\}} \left(\sum_{k=1}^{K} | \partial_t \left(\left[\delta(t) + \frac{\pi}{jt} \right] * u_k(t) \right) e^{-jw_k t} |_2^2 \right) \tag{2}$$

subject to the constraint

$$\sum_{k=1}^{K} u_k(t) = f(t),$$

where * denotes the convolution operator and δ(t) represents the Dirac delta function. The term $\left[\delta(t) + \frac{\pi}{jt}\right] * u_k(t)$ corresponds to the analytic signal of $u_k(t)$, obtained via the Hilbert transform, and is used to estimate the bandwidth of each mode (Qin and Li, 2025, p. 5).

Tuning and optimization of VMD parameters

The performance and effectiveness of the VMD method depend critically on two key hyperparameters: the number of modes (*K*) and the penalty factor (α). Inappropriate selection of these parameters may result in mode mixing or over-decomposition. In this study, to mitigate such risks, the parameters *K* and α are systematically evaluated over a predefined and feasible range. The optimal parameter combination is selected based on a joint criterion that minimizes the reconstruction error while maximizing the forecasting accuracy of the subsequent predictive model.

In the proposed framework, the crude oil price signal serves as the input. Following the decomposition process, the output consists of a set of distinct sub-series representing short-term and long-term frequency components.

3.1.4. Frequency decomposition

The fast Fourier transform (FFT) is an efficient algorithm for computing the discrete Fourier transform (DFT). A direct computation of the DFT has a computational complexity of $O(N^2)$, which is often prohibitive for large datasets. The FFT employs a divide-and-conquer strategy to recursively decompose an *N*-point signal into smaller subproblems, thereby reducing the computational complexity to $O(N \log N)$. Consequently, the FFT enables rapid transformation of time-domain signals into their frequency-domain representation, revealing the amplitudes and phases of constituent components. This capability establishes the FFT as a fundamental tool in digital signal processing, time-series analysis, and filtering applications (Faris Hameed and Jumah Al-Thahab, 2025).

3.2. Machine learning models

Following the generation of lagged and moving average variables and the integration of components derived from the signal decomposition methods, the following models are employed for forecasting.

3.2.1. LightGBM and XGboost

LightGBM and XGBoost are two highly efficient implementations of the gradient boosting framework. These algorithms achieve state-of-the-art predictive performance by iteratively constructing an ensemble of weak decision trees, in which each successive tree is trained to correct the residuals of the aggregated preceding trees.

LightGBM distinguishes itself by significantly accelerating the training process through two principal techniques: gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB). It is a gradient boosting method widely recognized for its computational efficiency and robustness, particularly when applied to large-scale datasets or environments with computational constraints.

From a mathematical perspective, LightGBM constructs an additive model in a forward stage-wise manner. Given a training dataset, the objective is to predict the output γ_i by minimizing a predefined differentiable loss function *L*.

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x) \tag{3}$$

where $h_m(x)$ represents the decision trees at iteration m, γ_m is the corresponding weight of that tree, and M denotes the total number of trees in the ensemble. The model is built in an additive, forward stage-wise manner. At each step m, the algorithm finds the tree weight γ_m that minimizes the overall loss:

$$\gamma_m = \text{argmin}_\gamma \sum_{i=1}^{N} L\big(y_i, f_{m-1}(x_i) + \gamma h_m(x_i)\big) \tag{4}$$

where N is the number of data points, L presents the differentiable loss function (e.g., MAE or MSE), and $f_{m-1}(x_i)$ is the model ensemble up to the previous stage. Furthermore, LightGBM employs the gradient-based one-side sampling (GOSS) technique, which preferentially retains data instances

associated with larger gradients, thereby improving computational efficiency without substantially compromising predictive accuracy (Constantin, 2025, p. 5).

XGBoost (eXtreme gradient boosting), introduced by Chen and Guestrin (2016), represents a significant advancement in gradient boosting techniques, offering substantial flexibility, scalability, and computational efficiency. One of the principal innovations of XGBoost is its explicit regularization mechanism, which mitigates overfitting and enhances the model's generalization capability. This improvement is achieved by incorporating a regularization term into the loss function, thereby controlling model complexity and improving robustness.

In addition, the gain function in XGBoost is used to optimize the selection of splitting points within decision trees, further improving predictive accuracy and computational efficiency. The objective function of XGBoost is defined as follows.

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, y_i') + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

where $l(y_i, y_i')$ is the differentiable loss function quantifying the discrepancy between actual (y_i) and predicted y_i' values. The model incorporates a regularization term $\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} W_j^2$ to penalize complexity and mitigate overfitting, where T is the number of leaves in the tree, w_j indicates the weight (score) of leaf j, and γ and λ are regularization hyperparameters. Predictions for an instance x_i is an additive combination of the outputs from all K trees:

$$y_i' = \sum_{k=1}^{K} f_k(x_i) \tag{6}$$

XGBoost employs a second-order gradient optimization scheme, utilizing both the first-order (g_i) and second-order (h_i) derivatives—corresponding to the gradient and Hessian—of the loss function with respect to the current predictions to guide the construction of the tree structure.

$$g_i = \frac{\partial l(y_i, y_i')}{\partial y_i'}$$
$$h_i = \frac{\partial^2 l(y_i, y_i')}{\partial y_i'^2} \tag{7}$$

To further enhance robustness and generalization, XGBoost incorporates additional techniques, including shrinkage (learning rate) and advanced hyperparameter tuning. This comprehensive framework allows the algorithm to efficiently handle high-dimensional and imbalanced datasets, making it particularly effective for complex forecasting tasks. XGBoost has consistently demonstrated superior performance across diverse domains and is widely recognized as a significant advancement in the evolution of gradient boosting algorithms (Imani et al., 2025, p. 6).

3.2.2. Random forest

The random forest (RF) algorithm is a powerful ensemble learning method that constructs a large number of decision trees during training. For a given input, the final prediction of the random forest is obtained by aggregating the predictions of all individual trees—through majority voting for classification tasks or averaging for regression tasks. This collective decision-making process generally provides superior predictive accuracy and robustness compared to a single decision tree. However, the algorithm's effectiveness can be influenced by the specific characteristics of the dataset.

The core structural strategy underlying random forest is bagging (bootstrap aggregating). The ensemble is built by generating B bootstrap samples from the original training dataset ($(TS_n = \{(x_1, y_1), \ldots, (x_n, y_n)\})$). In each iteration b, a bootstrap sample (X_b, Y_b) of size n is drawn with replacement from TS_n, and a separate decision tree f_b (for either classification or regression) is trained on this bootstrapped dataset. To predict a new instance x', the random forest aggregates the outputs of all B trees. The final prediction $(\hat{f})$ is determined by averaging the individual tree predictions for regression tasks or by majority voting for classification tasks (Imani et al., 2025, p. 5).

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x') \tag{8}$$

3.2.3. Support vector regression

Support vector regression (SVR) is a direct extension of the classical support vector machine algorithm to regression problems. Rooted in statistical learning theory, SVR has gained significant attention for forecasting economic and financial time series. The primary objective of SVR is to identify a function that approximates the underlying relationship in the data such that the deviation between predicted and actual values does not exceed a pre-specified margin of tolerance, denoted by ϵ. Within the SVR framework, prediction errors that fall within this ϵ-insensitive margin are not penalized; only deviations exceeding this threshold contribute to the optimization cost. The data points that lie exactly on the boundaries of the ϵ-insensitive tube—determined through the solution of a constrained optimization problem—are referred to as support vectors (SVs).

A key advantage of SVR over many other machine learning techniques is its formulation as a convex optimization problem. This guarantees a unique global minimum, inherently avoiding convergence to local minima. The SVR model is developed in two distinct phases: training and testing. During the training phase, the algorithm uses the majority of the dataset to identify the support vectors that define the boundaries of the ϵ-insensitive tube. The model's generalization capability is then rigorously evaluated in the testing phase using a hold-out subset of previously unseen data. Additionally, employing cross-validation during model selection provides a robust and generalizable solution that effectively mitigates the risk of overfitting.

Consider a training dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where $x_i \in R^m$ is a feature vector, $y_i \in$ R is the target value, and $i = 1, 2, \ldots, n$. The objective of linear support vector regression is to find a function $y = f(x) = w^T x + b$ by solving the following convex optimization problem.

$$\begin{aligned} &\min\left(\frac{1}{2}|w|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)\right) \\ &\text{subject to} \\ &y_i - (wx_i + b) \le \varepsilon + \xi_i \\ &(wx_i + b) - y_i \le \varepsilon + \xi_i^* \\ &\xi_i, \xi_i^* \ge 0 \end{aligned} \tag{9}$$

Where, ϵ defines the width of the error tolerance band, while ξ_i and ξ_i^* are slack variables that allow for deviations outside this band, controlled by the regularization parameter *C*. Data points that lie within the ϵ-insensitive tube satisfy $\xi_i = \xi_i^* = 0$. Equation 9 represents a convex quadratic optimization problem with linear constraints and admits a unique solution.

The first term in the objective function, $\frac{1}{2}|w|^2$, promotes model flatness, thereby enhancing generalization by penalizing model complexity. The second term, $C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$, regulates the empirical risk by penalizing training errors. A larger value of the regularization parameter C imposes a higher cost on deviations outside the ϵ-insensitive tube, i.e., on data points for which $\xi_i \geq 0$ or $\xi_i^* \geq 0$. The fundamental principle of SVR is to achieve an optimal balance—governed by the parameters ϵ and C—between model complexity (flatness) and fidelity to the training data.

The boundaries of the ϵ-insensitive tube are defined by the support vectors (SVs), which are the data points lying precisely on the margin and are often depicted as filled black dots in schematic diagrams. A penalty (ξ_i) is incurred for any predicted value that deviates from this tube by more than ϵ, with the magnitude of the penalty proportional to the distance from the margin. By applying Lagrange multipliers to Equation 9, the solution is obtained as follows.

$$w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i \ , \ \ y = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i^T x \tag{10}$$

In this formulation, the coefficients α_i and α_i^* are zero for all data points that are not support vectors. Consequently, the SVR model is fully determined by its support vectors.

Real-world data-generating processes are inherently nonlinear, often rendering linear models insufficient to capture their complex dynamics. To address this limitation, support vector methods (SVM/SVR) employ kernel functions. The "kernel trick" offers an efficient solution by implicitly mapping the data into a high-dimensional feature space where linear separation is achievable. This approach requires only the computation of inner products using the kernel functions in the original input space, avoiding the computationally prohibitive task of explicitly calculating coordinates in the transformed space.

By employing a nonlinear kernel function, the SVR model can capture nonlinear relationships in the data. For the empirical study, two kernels were utilized: a simple Linear kernel and the fully nonlinear radial basis function (RBF) kernel. Their mathematical formulations are presented below.

$$\text{Linear } K_1(x_1, x_2) = x_1^T x_2 \tag{11}$$

$$\text{RBF } K_2(x_1, x_2) = e^{-\gamma|x_1 - x_2|^2} \tag{12}$$

where the factor γ represents the kernel parameter (Plakandaras et al., 2017, 7).

3.2.4. Convolutional neural networks (1D-CNN)

Convolutional neural networks (CNNs) have recently demonstrated remarkable performance in time-series forecasting using one-dimensional CNN (1D-CNN) architectures (Kiranyaz et al., 2021). Unlike conventional 2D-CNNs applied in image processing, 1D-CNNs are specifically designed to handle sequential data, capturing local dependencies and shift-invariant patterns through one-dimensional kernels.

The proposed 1D-CNN architecture comprises two primary operations: convolution and pooling. In the convolutional layers, a set of 1D filters slides along the temporal dimension of the input sequence to extract high-level features from the VMD-decomposed components. The convolution operation at a given time step *t* is expressed as follows.

$$y_t = \sigma\left(\sum_{i=0}^{k-1} w_i \cdot x_{t+i} + b\right) \tag{13}$$

Here, x_t represents the input vector, w_i denotes the kernel weights of size k, b is the bias term, and σ is the nonlinear activation function. In this study, the rectified linear unit (ReLU), introduced by Nair and Hinton (2010), is employed due to its computational efficiency and its effectiveness in mitigating the vanishing gradient problem.

Following the convolutional layers, max-pooling layers are applied to down-sample the feature maps (Scherer et al., 2010). This operation selects the maximum value within a specified temporal window, thereby reducing dimensionality and filtering out high-frequency noise while preserving the most salient features for the subsequent regression layers.

3.2.5. Long short-term memory (LSTM)

Long short-term memory (LSTM) networks are a specialized variant of recurrent neural networks (RNNs) designed to overcome the vanishing gradient problem. This architecture is particularly well-suited for modeling crude oil prices, which often exhibit long-term dependencies, volatility clustering, and complex nonlinear dynamics.

The core component of an LSTM unit is the memory cell, which controls the flow of information through three gating mechanisms: the forget gate (f_t), the input gate (i_t), and the output gate (o_t). As described by Gers et al. (2000), the forget gate determines which information should be discarded from the cell state, enabling the model to reset its memory when necessary. The mathematical formulation of the LSTM unit is defined as follows.

$$\begin{aligned}
f_t &= \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \\
i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
\tilde{C}_t &= tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
C_t &= f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \\
o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
h_t &= o_t \cdot tanh(C_t)
\end{aligned} \tag{14}$$

Here, $\left(\tilde{C}_t\right)$ denotes the candidate cell state, C_t represents the cell state (long-term memory), h_t is the hidden state at time t, σ is the sigmoid activation function, and tanh is the hyperbolic tangent function. By adaptively updating the cell state, the LSTM model can effectively capture both short-term fluctuations and long-term trends in the oil market.

In the present study, the proposed algorithms and frameworks for oil price forecasting are implemented using two distinct methodological approaches: the direct approach and the hybrid approach. In the direct approach, machine learning models are trained to extract temporal dependencies directly from the historical price trajectory. The input feature space is constructed using consecutive lags of the observed price series, combined with a lagged moving average to account for trend smoothing. This relationship can be formally expressed as follows.

$$\widehat{y_{d,t+1}} = F_{Direct}(P_{t-1}, P_{t-2}, \ldots, P_{t-L}, \mathrm{MA}_{t-1}, \ldots, \mathrm{MA}_{t-\mathrm{K}}) \tag{15}$$

Here, $\hat{y}_{d,t+1}$ denotes the forecasted price at the next time step, P_{t-i} represents the historical price at lag i, and MA_{t-j} corresponds to the lagged moving average feature at j.

In the hybrid approach, the modeling strategy emphasizes the underlying frequency structure of the data. Following the signal decomposition process—using VMD, STL, wavelet, or FFT—the time series is separated into its constituent components. The forecasting model is then trained on a feature set composed of lagged values of these extracted components. This approach enables the algorithm to capture distinctive patterns ranging from high-frequency fluctuations to low-frequency trends. The formulation of the hybrid predictor is defined as follows.

$$\widehat{y_{h,t+1}} = F_{Hybrid}(\boldsymbol{C_{t-1}}, \boldsymbol{C_{t-2}}, \dots, \boldsymbol{C_{t-L}}, MA_t) \tag{16}$$

In this formulation, C_{t-i} represents the vector of all decomposed components at lag i—for example, the set of all intrinsic mode functions in VMD—allowing the model to exploit the latent spectral information inherent in the time series to forecast its future state.

To ensure the statistical validity and generalizability of the forecasting results, the oil price data are analyzed with respect to key structural and behavioral characteristics, including structural breaks, high volatility and nonlinearity, volatility clustering, short- and long-term temporal dependencies, external and geopolitical shocks, and nonstationarity. Following the framework of Bai and Perron (2003), structural breakpoints are identified, and the overall time series is partitioned into appropriate temporal intervals. Separate modeling and forecasting are then conducted within each interval. For each interval, the data are chronologically divided, with a designated portion used for model training and the remaining portion reserved for testing and evaluating predictive performance.

To maintain a rigorous and causal modeling design, all parameter optimization procedures—including the calibration of signal decomposition hyperparameters (e.g., K in VMD and the decomposition level in wavelet) and machine learning model hyperparameters (e.g., C and γ in SVR)—are performed exclusively on the training dataset. This approach ensures that the configuration of the signal extraction algorithms reflects only patterns observed in historical training data, preventing the selection of parameters biased by future information from the test set. After calibration, the optimized models are employed to generate the features required for forecasting tasks.

4. Results and discussion

4.1. Data-driven sampling strategy and volatility regime detection

In this study, to mitigate fixed-window bias and address concerns regarding arbitrary data selection for modeling oil prices over short-, medium-, and long-term horizons, an endogenous regime detection approach was employed based on the computational framework of Bai and Perron (2003). This approach enables the identification of distinct market regimes based on structural breaks in volatility dynamics rather than simple shifts in price levels. As illustrated in Figure 1, the binary segmentation algorithm successfully detected structural breaks that coincided with volatility clusters, resulting in the definition of three distinct experimental periods.

To evaluate model performance across regimes with distinct volatility structures and economic characteristics, three experimental periods identified through the endogenous regime detection procedure are considered, corresponding to short-, medium-, and long-term market horizons.

The short-term regime encompasses 1,400 trading days from April 27, 2020, to September 5, 2025, reflecting the post-COVID oil market environment. During this period, the average price is relatively high at USD 72.69, with daily volatility of 2.46 percent. Price dynamics in this horizon are driven by rapid market reactions to news, elevated geopolitical risk premiums, and post-pandemic demand recovery. Volatility is therefore highly clustered and shock-driven, characterized by frequent but

relatively short-lived fluctuations that capture transitory uncertainty rather than major structural changes.

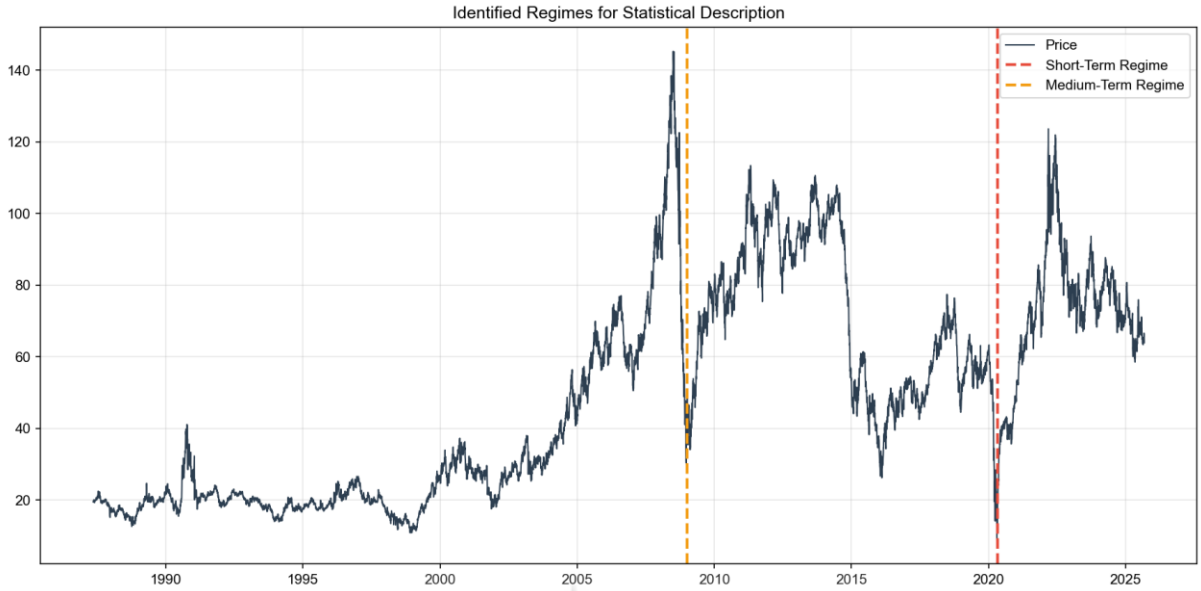

Figure 1

WTI crude oil price time series and the corresponding regimes identified based on their volatility structure.

The medium-term regime spans 4,353 trading days from December 31, 2008, to September 5, 2025, representing a crisis-dominated market environment. This horizon exhibits the highest daily volatility at 2.80 percent and the largest observed daily price decline of 42.36 percent. It includes several major episodes of market disruption, such as the aftermath of the global financial crisis, the 2014 oil price collapse, and the COVID-19 shock. Volatility in this regime is persistent and regime-dependent, arising from structural stress and nonlinear price adjustments rather than short-term information shocks. Consequently, this period provides a challenging environment for testing the robustness of forecasting models under extreme market conditions.

The long-term regime covers 10,000 trading days from May 11, 1987, to September 5, 2025, capturing the secular evolution of the WTI crude oil market across multiple economic cycles. The average price during this horizon is lower at USD 49.15, with daily volatility of 2.61 percent. In contrast to shorter horizons, long-term volatility reflects slow-moving adjustments in demand and supply, technological progress, shifts in market structure, and the increasing financialization of oil markets. This regime embodies broad cyclical and structural dynamics, enabling the model to learn long-run patterns across heterogeneous and historically diverse market conditions.

Overall, the descriptive statistics reported in Table 1 demonstrate that volatility differs not only in magnitude but also in its underlying economic sources across time horizons, highlighting the importance of regime-specific evaluation in oil price forecasting.

4.2. Hyperparameter optimization and experimental design

To ensure methodological rigor, reproducibility, and to mitigate the risk of overfitting or data-snooping bias, all hyperparameters for both signal decomposition techniques and machine learning models were systematically optimized using a structured grid search framework (Seabe et al., 2025). Distinct objective functions were employed at each stage. For signal decomposition, parameters were optimized

to minimize the average correlation among extracted components, thereby ensuring modal orthogonality.

Table 1

Descriptive statistics and economic characteristics of the identified market regimes.

Time horizon	Period	Sample size	Mean	Standard deviation	Max daily drop	Min	Max
Short-term	2020-04-27 to 2025-09-05	1,400	\$72.69	17.66	–12.78%	12.17	123.64
Medium-term	2008-12-31 to 2025-09-05	4,353	\$71.26	20.74	–42.36%	8.91	123.64
Long-term	1987-05-11 to 2025-09-05	10,000	\$49.15	29.37	–42.36%	8.91	145.31

For the variational mode decomposition (VMD) algorithm, the search space was explicitly defined rather than relying on default specifications. The number of intrinsic modes was varied as $K \in \{2,3,4,5\}$, and the penalty parameter as $\alpha \in \{500,1000,1500\}$, allowing an explicit assessment of the trade-off between noise suppression and trend preservation. Similarly, constrained yet representative search spaces were adopted for alternative decomposition methods, including seasonal window lengths $s \in \{7,13,21\}$ for STL decomposition and decomposition levels $L \in \{3,4,5,6\}$ for the wavelet transform, employing both Daubechies (db4) and Symlet (sym5) wavelet families.

In parallel, machine learning models were rigorously tuned using a time series cross-validation strategy to preserve the temporal ordering of the data, with the objective of minimizing the RMSE on the validation set. For the support vector regression model, the regularization parameter was explored over $C \in \{1,10,100,200\}$, while the kernel coefficient was selected from $\gamma \in \{\text{"scale"}, 0.01,0.001\}$. To reduce the risk of learning spurious patterns in ensemble-based models (random forest, XGBoost, and LightGBM), the number of estimators was varied as $N \in \{50,100\}$, maximum tree depth as $D \in \{3,5,10\}$, and learning rate as $\eta \in \{0.05,0.1\}$.

For deep learning architectures, including CNN and LSTM models, the hyperparameter grid included the number of units or filters $\in \{10,16,20,32,50\}$, training epochs $\in \{15,30,45\}$, and a fixed batch size of 32.

4.3. Forecasting implementation

For oil price forecasting, the optimal values of the required parameters are first determined using a structured grid search procedure. The oil price series is then decomposed into its principal components using the signal decomposition methods introduced in this study. Based on the extracted components, oil price dynamics are modeled using the forecasting frameworks described in the methodology section, and model performance is systematically evaluated.

To achieve accurate signal decomposition and prevent information leakage across components, the parameters of the decomposition techniques—such as the number of modes and penalty factor in the VMD algorithm, the decomposition level in the wavelet transform, the seasonal window in STL, and the number of Fourier terms (K) in the FFT method—are optimized using grid search. The objective function for this optimization minimizes the average correlation among the extracted components, thereby ensuring modal independence and that each component conveys unique information. The resulting optimal parameter values obtained via this approach are reported in Table A1 in the Appendix.

Subsequently, the configurations of the machine learning models are tuned. For all forecasting models, including SVR and XGBoost, hyperparameter optimization is performed using grid search combined with time-series cross-validation. This ensures that key parameters, such as the regularization parameter C and kernel coefficient γ in SVR, are selected to minimize generalization error on the validation dataset. Finally, the extracted components are supplied as input features to the optimized models for oil price forecasting. Tables A2–A6 in the Appendix present the estimated parameter settings for the models across different time horizons.

4.4. Short-term forecasting results

The short-term period was modeled using 1,400 samples from the oil price time series, spanning April 27, 2020, to September 5, 2025. Each component extracted from the oil price decomposition reflects distinct underlying economic forces. Visual inspection of the VMD decomposition results shows that the low-frequency component, representing the long-term trend, captures fundamental market dynamics and the gradual recovery of oil demand in the post-COVID period. The medium-frequency components, characterized by regular and quasi-sinusoidal oscillations, correspond to seasonal patterns prevailing in the oil market. In contrast, shock-driven components, identified by abrupt amplitude spikes within intermediate frequency bands—particularly during 2022—effectively isolate the impacts of the Russia–Ukraine war and the intensification of geopolitical uncertainty. The high-frequency component primarily consists of trading noise and speculative fluctuations, which are effectively filtered out by the proposed framework.

Table 2 presents the forecasting performance of both baseline and hybrid models for the short-term period, evaluated using RMSE, MAE, MAPE, SMAPE, R-Squared, and formal forecast comparison tests.

Table 2

Short-term forecasting performance comparison of baseline and hybrid models.

Model	RMSE	MAE	MAPE	SMAPE	R^2	DM Test (p-val)	Wilcoxon (p-val)
VMD + SVR	0.7871	0.5948	0.8671	0.8623	0.9701	-	-
Wavelet + SVR	0.8304	0.6189	0.8997	0.8953	0.9667	0.364	0.291
STL + XGBoost	1.2228	0.8868	1.2775	1.2802	0.9278	0.000	0.000
STL + LGBM	1.2385	0.8714	1.2560	1.2571	0.9259	0.000	0.000
STL + RandomForest	1.2587	0.8723	1.2614	1.2630	0.9235	0.003	0.000
Wavelet + XGBoost	1.2776	0.9558	1.3939	1.3904	0.9212	0.000	0.000
Wavelet + LGBM	1.3448	1.0296	1.5010	1.4978	0.9127	0.000	0.000
STL + SVR	1.3650	0.9261	1.3433	1.3391	0.910	0.000	0.000
FFT + SVR	1.5112	1.1402	1.6507	1.6402	0.8897	0.000	0.000
Wavelet + RandomForest	1.5175	1.1772	1.7236	1.7189	0.8888	0.000	0.000
Direct + XGBoost	1.5330	1.1685	1.6954	1.6851	0.8865	0.000	0.000
Direct + RandomForest	1.5420	1.1737	1.7030	1.6905	0.8852	0.000	0.000
Direct + SVR	1.5432	1.1495	1.6607	1.6520	0.8850	0.000	0.000
Direct + LGBM	1.5699	1.1918	1.7318	1.7195	0.8810	0.000	0.000

Model	RMSE	MAE	MAPE	SMAPE	R^2	DM Test (p-val)	Wilcoxon (p-val)
VMD + XGBoost	1.8939	1.4172	2.0636	2.0671	0.8268	0.000	0.000
VMD + LGBM	2.1039	1.5928	2.3522	2.3319	0.7863	0.000	0.000
VMD + RandomForest	2.2322	1.6863	2.4860	2.4620	0.7595	0.000	0.000
Direct + CNN	3.6561	3.0050	4.4153	4.5066	0.3548	0.000	0.000
FFT + RandomForest	4.0425	3.0276	4.4894	4.5482	0.2112	0.000	0.000
FFT + XGBoost	4.6040	3.1434	4.6554	4.8310	–0.0230	0.000	0.000
FFT + LGBM	5.4555	3.5879	5.3486	5.6085	–0.4364	0.000	0.000
Wavelet + CNN	7.2988	6.160	8.9285	9.1454	–1.5711	0.000	0.000
STL + CNN	8.9515	8.2995	11.9488	12.3398	–2.8673	0.000	0.000
Direct + ARIMA	10.0989	9.1014	13.6277	12.5434	–3.9222	0.000	0.000
VMD + CNN	12.7907	11.2814	16.4248	18.0067	–6.8957	0.000	0.000
FFT + CNN	18.3884	14.5885	21.4858	20.0347	–15.319	0.000	0.000
VMD + LSTM	26.5089	23.6253	34.3515	43.5954	–32.915	0.000	0.000
Wavelet + LSTM	27.3619	24.5662	35.6590	45.7770	–35.132	0.000	0.000
STL + LSTM	37.8821	35.4260	51.4106	72.0295	–68.259	0.000	0.000
Direct + LSTM	47.1600	46.6793	67.2586	102.2020	-106.339	0.000	0.000
FFT + LSTM	47.6592	45.6650	66.4786	84.5549	–108.62	0.000	0.000

The results clearly demonstrate that the proposed VMD–SVR framework is the best-performing model across all evaluation metrics, producing the lowest forecasting errors with an RMSE of 0.7871 and an MAE of 0.5948, while achieving the highest explanatory power with an R^2 value of 0.9701. This superior performance highlights the effectiveness of the VMD–SVR combination in capturing complex nonlinear dynamics and short-term structural shocks inherent in oil price movements. As illustrated in Figure 2, the predicted trajectory generated by the VMD–SVR model closely follows the actual oil price series, with the majority of observations falling within the 95% confidence interval. This strong alignment indicates not only high point-forecast accuracy but also robustness in modeling short-term uncertainty during periods of elevated market volatility.

The wavelet–SVR model ranks second, exhibiting forecasting accuracy very close to that of the benchmark VMD–SVR model. Importantly, the Diebold–Mariano and Wilcoxon signed-rank tests yield p-values of 0.364 and 0.291, respectively, indicating that the difference in predictive accuracy between VMD–SVR and Wavelet–SVR is not statistically significant at the 5% level. This finding suggests that, in the short-term horizon, both adaptive signal decomposition techniques—variational mode decomposition and wavelet transforms—when coupled with support vector regression, are highly effective in extracting relevant nonlinear features from oil price data.

Models based on STL decomposition combined with ensemble learning algorithms (XGBoost, LightGBM, and random forest) occupy the next performance tier. Although these models achieve relatively high coefficients of determination (exceeding 0.92), their error metrics remain noticeably higher than those of the two leading SVR-based hybrid approaches. This pattern indicates that STL decomposition, while effective in capturing smooth seasonal structures, is less capable of

accommodating abrupt shocks and high-frequency volatility that dominate short-term oil price dynamics.

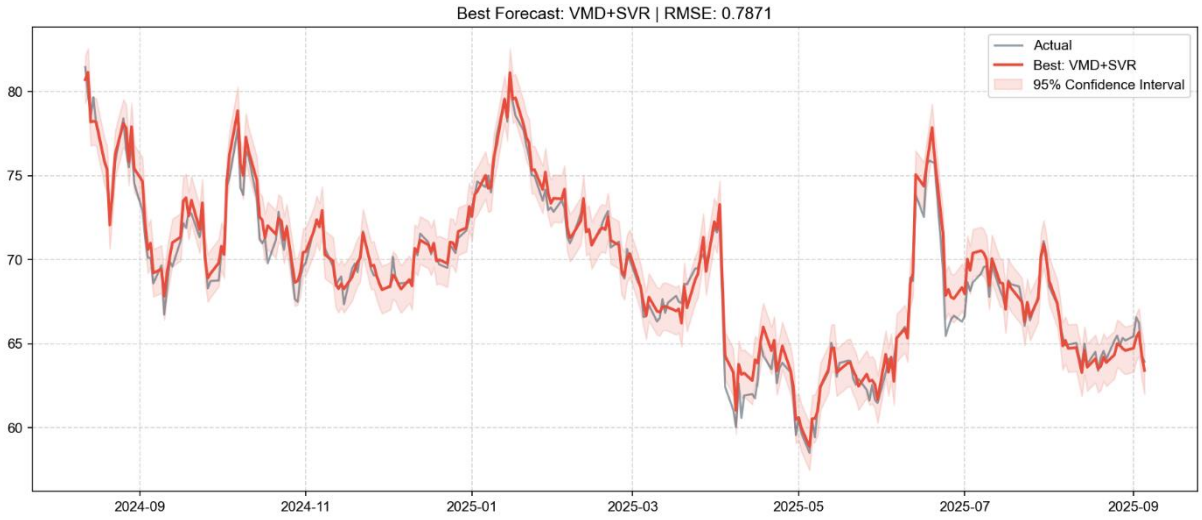

Figure 2

Short-term forecasting performance of the VMD–SVR model for WTI crude oil prices.

In contrast, deep learning architectures such as LSTM and CNN, as well as the classical linear ARIMA model, perform poorly in the short-term forecasting setting, particularly when applied to direct and FFT-based inputs. Strongly negative R^2 values highlight this weakness; for instance, the direct–LSTM model records an R^2 of –106.33, indicating predictive performance substantially worse than a naive mean benchmark. These large forecasting errors are likely driven by the limited effective sample size available for training deep neural networks over short horizons, resulting in severe overfitting and unstable parameter estimation. Additionally, the inability of linear models to capture nonlinearities and asymmetric responses in oil price dynamics further exacerbates poor predictive performance.

Overall, the results in Table 2 demonstrate that hybrid frameworks combining adaptive signal decomposition with parsimonious and robust machine learning models, such as SVR, consistently outperform both deep learning architectures and traditional linear approaches in short-term oil price forecasting. This evidence underscores the importance of aligning model complexity with data availability and market characteristics when designing forecasting systems for highly volatile energy markets.

4.5. Medium-term forecasting results

The medium-term period is modeled using 4,353 trading observations from the oil price time series, spanning December 31, 2008, to September 5, 2025. This extended horizon encompasses multiple business cycles, the global economic recovery following the 2008 financial crisis, and several major structural shocks, including the sharp oil price collapse in 2014 and the COVID-19 pandemic.

The results of the VMD-based decomposition highlight the distinctive effectiveness of this approach for analyzing data over medium-term horizons. Specifically, the trend component, corresponding to the low-frequency mode, successfully captures major structural transformations, such as the gradual recovery from the 2008 global financial crisis and supply-side fluctuations observed over the past decade. The medium-frequency oscillatory components exhibit strong alignment with global business cycles. A key advantage of the VMD method is its ability to accurately isolate large structural shocks—

such as the 2014 oil price collapse and the negative demand shock induced by the COVID-19 pandemic—within intermediate-frequency modes. This prevents these shocks from contaminating the long-term trend or being absorbed into high-frequency noise, enabling a clear and economically meaningful separation of information.

Table 3 summarizes the forecasting performance of all hybrid and benchmark models over the medium-term period, evaluated using RMSE, MAE, MAPE, SMAPE, R^2, and formal forecast comparison tests. The results indicate that the VMD–SVR model consistently outperforms all competing approaches, achieving the lowest errors with an RMSE of 0.7037 and an MAE of 0.5058, while attaining the highest explanatory power with an R-Squared value of 0.9962.

Table 3

Medium-term forecasting performance comparison of baseline and hybrid models.

Model	RMSE	MAE	MAPE	SMAPE	R^2	DM Test (p-val)	Wilcoxon (p-val)
VMD + SVR	0.7037	0.5058	0.6261	0.6256	0.9962	-	-
Wavelet + SVR	0.9373	0.6245	0.7710	0.7713	0.9932	0.000	0.000
STL + SVR	1.6233	1.0338	1.2888	1.2900	0.9795	0.000	0.000
VMD + CNN	1.6416	1.2323	1.6115	1.6176	0.9791	0.000	0.000
Direct + SVR	1.7090	1.2969	1.6359	1.6321	0.9773	0.000	0.000
Wavelet + LGBM	1.7434	1.0340	1.2478	1.2531	0.9764	0.000	0.000
FFT + SVR	1.7529	1.3236	1.6675	1.6614	0.9761	0.000	0.000
Wavelet + RandomForest	1.7555	1.1829	1.4610	1.4634	0.9761	0.000	0.000
STL + LGBM	1.7961	1.1106	1.3485	1.3532	0.9749	0.000	0.000
Wavelet + XGBoost	1.8363	1.0905	1.3212	1.3270	0.9738	0.000	0.000
Direct + RandomForest	1.9091	1.3781	1.7049	1.7047	0.9717	0.000	0.000
VMD + LGBM	1.9286	1.3569	1.6799	1.6827	0.9711	0.000	0.000
Direct + XGBoost	1.9785	1.4152	1.7464	1.7463	0.9696	0.000	0.000
Direct + CNN	2.0205	1.5363	1.9546	1.9439	0.9683	0.000	0.000
Direct + LGBM	2.0398	1.4390	1.7711	1.7721	0.9677	0.000	0.000
VMD + XGBoost	2.0779	1.4406	1.7639	1.7660	0.9664	0.000	0.000
Wavelet + CNN	2.1656	1.6357	2.0689	2.0700	0.9636	0.000	0.000
STL + XGBoost	2.1958	1.2006	1.4244	1.4357	0.9626	0.000	0.000
VMD + RandomForest	2.2145	1.6583	2.0739	2.0701	0.9619	0.000	0.000
STL + RandomForest	2.3900	1.3003	1.5412	1.5558	0.9556	0.000	0.000
STL + CNN	2.4225	1.8077	2.2924	2.2856	0.9544	0.000	0.000
FFT + CNN	2.5651	1.6896	2.0626	2.0373	0.9489	0.000	0.000
FFT + XGBoost	2.8332	1.9099	2.3225	2.3278	0.9377	0.000	0.000
FFT + LGBM	3.0063	2.0009	2.4222	2.4329	0.9298	0.000	0.000
FFT + RandomForest	3.0645	2.2671	2.8262	2.8196	0.9271	0.000	0.000

Model	RMSE	MAE	MAPE	SMAPE	R^2	DM Test (p-val)	Wilcoxon (p-val)
Wavelet + LSTM	8.7653	6.8073	8.8880	9.0270	0.4034	0.000	0.000
VMD + LSTM	11.0154	8.8385	11.4668	12.0794	0.0570	0.000	0.000
Direct + LSTM	14.5279	13.6608	17.4480	19.1127	–0.6390	0.000	0.000
STL + LSTM	16.8527	14.2118	18.1312	18.6725	–1.2056	0.000	0.000
FFT + LSTM	25.0582	23.7202	31.3531	38.2081	–3.8762	0.000	0.000
Direct + ARIMA	31.7945	30.1863	40.8101	33.0588	–6.8503	0.000	0.000

As illustrated in Figure 3, the predicted trajectory closely follows the actual oil prices, demonstrating the model's ability to capture medium-term dynamics and structural patterns. This performance underscores the model's exceptional capability to represent nonlinear dynamics and medium-horizon structural shifts in oil price movements.

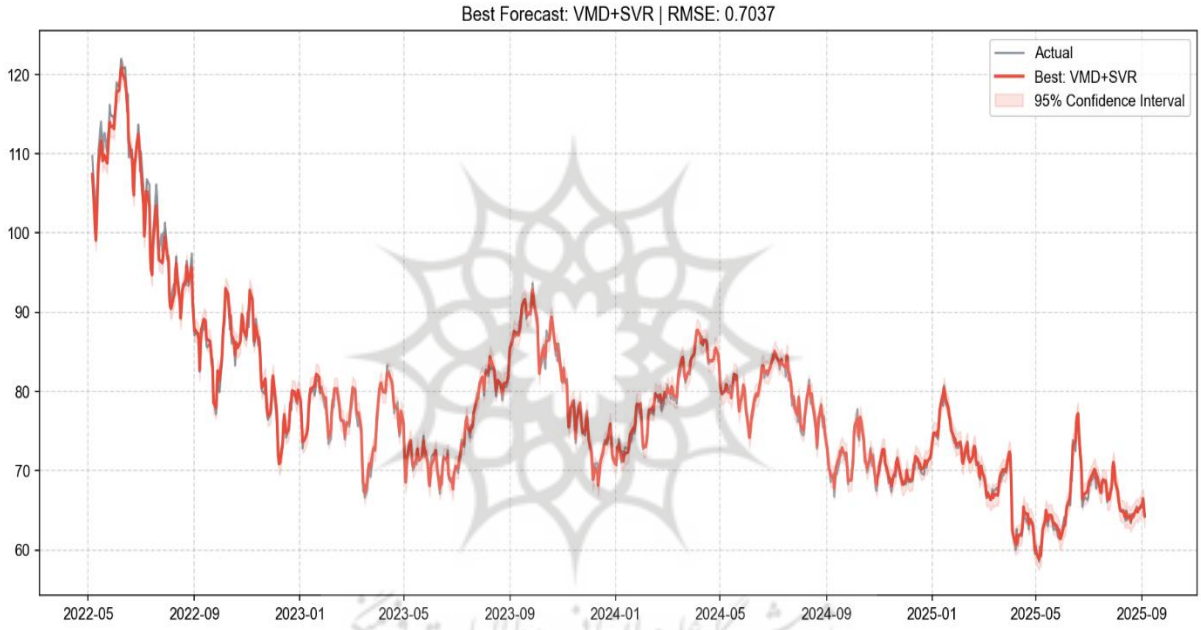

Figure 3

Medium-term forecasting performance of the VMD–SVR model for WTI crude oil prices.

The second-best performing model is wavelet–SVR, with an RMSE of 0.9373, an MAE of 0.6245, and an R^2 of 0.9932. Statistical tests indicate a significant difference between VMD–SVR and wavelet–SVR at the 95 percent confidence level, confirming the superior accuracy of the proposed model over the medium-term horizon. Other SVR-based models, such as STL–SVR and Direct–SVR, occupy the subsequent performance tiers, with R-Squared values exceeding 0.97.

Ensemble learning methods combined with decomposition techniques, such as wavelet with LightGBM, STL with LightGBM, and Direct with random forest, deliver moderate performance. These models effectively capture medium-term trends but demonstrate lower sensitivity to abrupt structural shocks compared to VMD–SVR.

Deep learning architectures, including LSTM and CNN, exhibit markedly weaker performance. For instance, the wavelet–LSTM and VMD–LSTM models record RMSE values of 8.7653 and 11.0154, respectively, with corresponding R^2 values of 0.4034 and 0.0570. Models such as direct–LSTM and FFT–LSTM even produce negative R^2 values, indicating forecasting performance inferior to a simple

mean benchmark. These deficiencies are likely due to the limited effective sample size for training deep networks over medium-term horizons, resulting in overfitting, unstable parameter estimation, and an inability to accurately capture medium-horizon nonlinearities in oil price dynamics.

Overall, the findings reinforce that hybrid models combining adaptive signal decomposition techniques, such as VMD or Wavelet, with robust machine learning approaches like SVR provide superior forecasting accuracy for medium-term oil prices, consistently outperforming both deep learning and classical linear approaches. This underscores the importance of aligning decomposition strategies with appropriate learning models to effectively capture structural patterns and shocks over intermediate horizons.

4.6. Long-term forecasting results

The long-term period, comprising 10,000 daily observations, represents the broadest temporal horizon of the study, spanning from May 11, 1987, to September 5, 2025. This nearly four-decade interval encompasses major geopolitical and economic events, including the Gulf War, the 2008 financial crisis, the shale oil revolution, and the COVID-19 pandemic, providing a rigorous framework for evaluating the long-term memory of forecasting models and their capacity to generalize patterns over multiple decades.

The results of the VMD decomposition underscore the remarkable adaptability of this algorithm in capturing deep structural changes over time. Specifically, the trend or low-frequency component accurately reflects major structural transitions in the energy market, such as the low-oil-price era of the 1990s, demand shocks driven by the emergence of new economies in the 2000s, and price stabilization during the post-shale period. The extracted oscillatory components exhibit a strong correspondence with global business cycles throughout these four decades. A key advantage of the VMD method in this context is its ability to isolate abrupt and extreme shocks—such as the price surge during the Gulf War and the historic market collapse in 2020—within intermediate-frequency modes. This separation prevents transitory shocks from contaminating the main trend and filters high-frequency noise, ensuring robust long-term analysis.

Table 4 presents the forecasting performance of all hybrid and benchmark models over the long-term horizon, evaluated using RMSE, MAE, MAPE, SMAPE, and R-Squared, along with formal forecast comparison tests. The results clearly demonstrate that the VMD–SVR model outperforms all competing approaches, achieving the lowest errors with an RMSE of 1.0829 and an MAE of 0.6308, while attaining the highest explanatory power with an R-Squared of 0.9961. As illustrated in Figure 4, the predicted trajectory of the VMD–SVR model closely aligns with actual WTI crude oil prices, demonstrating its ability to capture long-term trends and structural shifts. This performance highlights the exceptional capability of the VMD–SVR framework to model nonlinear dynamics and long-horizon patterns in oil prices.

Table 4

Long-term forecasting performance comparison of baseline and hybrid models

Model	RMSE	MAE	MAPE	SMAPE	R^2	DM Test (p-val)	Wilcoxon (p-val)
VMD + SVR	1.0829	0.6308	1.1929	1.1430	0.9961	-	-
Wavelet + LGBM	1.3146	0.8389	1.3561	1.3407	0.9943	0.001	0.000
Wavelet + XGBoost	1.3477	0.8863	1.4037	1.3916	0.9940	0.000	0.000
Wavelet + SVR	1.4366	0.6856	1.1465	1.1256	0.9932	0.009	0.956

Model	RMSE	MAE	MAPE	SMAPE	R^2	DM Test (p-val)	Wilcoxon (p-val)
Wavelet + RandomForest	1.6130	1.0447	1.6692	1.6479	0.9914	0.000	0.000
STL + LGBM	1.6317	1.0912	1.7857	1.7403	0.9912	0.000	0.000
STL + XGBoost	1.6554	1.1024	1.7975	1.7584	0.9909	0.000	0.000
Direct + LGBM	1.7050	1.2061	1.9408	1.9261	0.9904	0.000	0.000
FFT + SVR	1.7272	1.2118	1.9484	1.9304	0.9901	0.000	0.000
Direct + SVR	1.7458	1.2170	1.9630	1.9450	0.9899	0.000	0.000
Direct + XGBoost	1.7527	1.2408	1.9918	1.9779	0.9898	0.000	0.000
VMD + LGBM	1.7778	1.2794	2.1674	2.1217	0.9895	0.000	0.000
VMD + XGBoost	1.8490	1.3352	2.2275	2.1783	0.9887	0.000	0.000
Direct + RandomForest	1.8492	1.3222	2.1107	2.0973	0.9887	0.000	0.000
FFT + CNN	1.9003	1.3337	2.2744	2.1890	0.9880	0.000	0.000
STL + SVR	1.9005	1.1055	1.9155	1.8367	0.9880	0.000	0.000
STL + RandomForest	2.0524	1.3331	2.1355	2.0879	0.9860	0.000	0.000
FFT + LGBM	2.2525	1.6336	2.7149	2.6666	0.9832	0.000	0.000
FFT + XGBoost	2.2816	1.6425	2.7658	2.7079	0.9827	0.000	0.000
VMD + RandomForest	2.2819	1.6294	2.7235	2.6768	0.9827	0.000	0.000
Direct + LSTM	2.3854	1.7815	2.7674	2.7465	0.9811	0.000	0.000
Wavelet + LSTM	2.5206	1.6805	2.8581	2.7557	0.9789	0.000	0.000
FFT + RandomForest	2.7269	2.0283	3.3472	3.2929	0.9753	0.000	0.000
VMD + CNN	2.8755	2.1330	3.4829	3.3557	0.9726	0.000	0.000
STL + LSTM	3.2039	2.2403	3.6770	3.5325	0.9659	0.000	0.000
FFT + LSTM	3.2164	2.4125	4.0879	3.8899	0.9657	0.000	0.000
VMD + LSTM	3.2345	2.0911	3.8545	3.5213	0.9653	0.000	0.000
STL + CNN	3.4702	2.6296	4.0403	3.9723	0.9600	0.000	0.000
Wavelet + CNN	3.7062	2.7525	3.9071	3.9442	0.9544	0.000	0.000
Direct + CNN	3.7544	2.9977	4.2977	4.3578	0.9532	0.000	0.000
Direct + ARIMA	18.6465	14.4977	24.2227	22.3373	–0.1536	0.000	0.000

The second-tier models comprise Wavelet–LGBM, Wavelet–XGBoost, and Wavelet–SVR, with RMSE values ranging from 1.3146 to 1.4366 and R^2 values exceeding 0.9932. Statistical analyses indicate that VMD–SVR significantly outperforms these models at the 95 percent confidence level, confirming its robustness over extended forecasting horizons. Other hybrid and ensemble approaches, including STL–LGBM, STL–XGBoost, and Direct–LGBM, demonstrate moderate accuracy, with R^2 values above 0.98. In contrast, deep learning architectures, such as LSTM and CNN, exhibit considerably weaker performance. Specifically, the VMD–LSTM, STL–LSTM, and Direct–LSTM models yield RMSE values exceeding 3.2, with R^2 values decreasing to as low as 0.9653. The direct–ARIMA model performs poorly, with an RMSE of 18.6465 and a negative R^2 of –0.1536, indicating its inability to capture the long-term nonlinear dynamics of oil prices. These results are likely attributable to the

combined effects of model complexity, limited training sample size, and the insufficiency of linear and deep learning methods to extract long-term structural information from the time series.

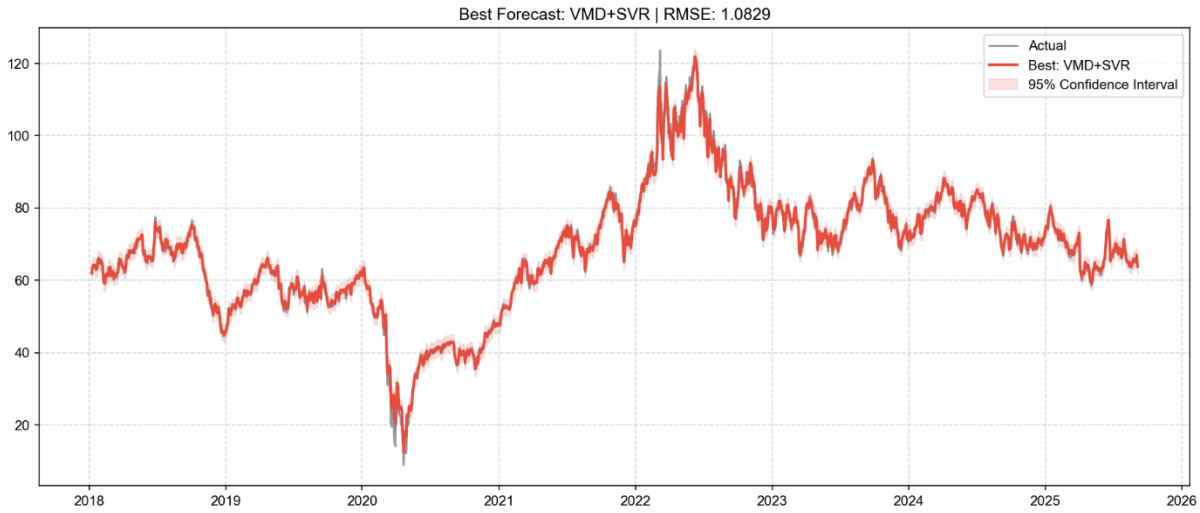

Figure 4

Long-term forecasting performance of the VMD–SVR model for WTI crude oil prices

5. Conclusions

West Texas Intermediate crude oil, as a primary benchmark in the global energy market, plays a central role in determining the value of other crude types and in establishing energy futures contracts. Traded on the New York Mercantile Exchange, WTI prices reflect global supply and demand conditions, the U.S. economic environment, and geopolitical developments influencing the oil market (Baumeister & Kilian, 2016). However, WTI prices exhibit persistent and unpredictable volatility, with fluctuations that have significant implications for macroeconomic variables such as inflation, economic growth, and fiscal stability (Kilian & Zhou, 2020). Consequently, WTI price movements serve not only as a barometer of global energy market dynamics but also as an indicator of overall economic stability (Narayan & Gupta, 2015). Beyond their economic significance, WTI price fluctuations pose challenges for policymaking, risk management, and investment decisions. In particular, heightened volatility transmits uncertainty to other markets, affecting the behavior of economic and financial agents (Wang, Wu, & Yang, 2015). In this context, the development of accurate and intelligent forecasting models for WTI prices is essential for policymakers, oil companies, investors, and financial analysts.

This study presents a comprehensive evaluation of hybrid approaches for daily oil price forecasting, employing an endogenous regime identification strategy based on the Bai and Perron multiple structural break framework. Statistically significant breakpoints are endogenously detected in the oil price series and subsequently used to delineate short-, medium-, and long-term forecasting periods corresponding to distinct market regimes. The results demonstrate that incorporating signal decomposition as a pre-processing step substantially enhances the predictive performance of machine learning models. Among the competing techniques, variational mode decomposition exhibits superior effectiveness due to its adaptive decomposition mechanism, which accommodates the pronounced non-stationarity and volatility clustering inherent in oil price dynamics. By decomposing the original series into intrinsic mode functions, VMD effectively isolates high-frequency noise from underlying structural trends, whereas static decomposition methods, such as FFT and Wavelet transforms, fail to adequately capture regime shifts and structural breaks identified by the Bai–Perron procedure.

Across the 31 evaluated forecasting approaches, the VMD + SVR hybrid model consistently outperforms all competing models in short-, medium-, and long-term periods. This finding underscores the efficacy of integrating adaptive signal decomposition with robust nonlinear learning algorithms for modeling oil price dynamics. Leveraging an RBF kernel, the SVR model successfully captures complex nonlinear dependencies within the VMD-derived feature space, outperforming both deep learning architectures, such as CNNs—which exhibited overfitting due to limited data even in long-term settings—and tree-based boosting methods, including LGBM and XGBoost. Quantitatively, the VMD + SVR framework achieves a substantial reduction in prediction error, ranging from 38.0 percent in the long-term period to 49.0 percent in the short-term regime, peaking at 58.8 percent in the medium-term period relative to the strongest baseline model, direct + SVR. These performance gains are consistently supported by multiple error metrics and formal statistical validation using the Wilcoxon signed-rank test and the Diebold–Mariano test, confirming the statistical superiority of the proposed hybrid model.

These findings are consistent with prior studies (Niu & Zhao, 2021; Huang et al., 2024; Boadi, 2025), which emphasize the potential of hybrid models that integrate signal decomposition with machine learning to enhance predictive accuracy. Such models provide valuable tools for supporting economic, financial, and policy decisions by capturing nonlinear dynamics, filtering noise, and uncovering hidden structures in oil price data. From a policy perspective, this study offers two primary implications. First, the VMD + SVR model can serve as an effective instrument to reduce uncertainty in investment decisions and to improve strategic planning, risk management, and financial analysis. This capability is particularly important for mitigating behavioral biases and decision-making errors among retail investors. Second, given the oil market's sensitivity to factors such as supply-demand imbalances, monetary policy, and climate change, accurate forecasting models like VMD + SVR enable policymakers to anticipate future WTI price movements with greater confidence and to support the development of more resilient energy and fiscal strategies.

Despite the encouraging empirical results, this study has several limitations. First, the VMD algorithm entails higher computational complexity compared with static decomposition methods, which may restrict its applicability in real-time or high-frequency trading environments. Second, the performance of the proposed hybrid framework is sensitive to the optimal tuning of key hyperparameters, specifically the number of modes and the penalty factor. Although a structured grid-search strategy was employed in this study, the implementation of adaptive or data-driven parameter selection mechanisms could further enhance the robustness and stability of the model. Future research may address these limitations by integrating evolutionary optimization algorithms and by extending the proposed framework to other highly volatile commodity markets.

Acknowledgements

The authors sincerely thank the editors and anonymous reviewers for their valuable comments and constructive suggestions, which have substantially contributed to enhancing the quality and clarity of this paper.

Nomenclature

ARIMA	Autoregressive integrated moving average
CNN	Convolutional neural network
Db4	Daubechies 4 wavelet
DWT	Discrete wavelet transform
FFT	Fast Fourier transform

IMF	Intrinsic mode function
LightGBM	Light gradient boosting machine
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
Random Forest	Random Forest algorithm
RMSE	Root mean square error
STL	Seasonal-trend decomposition using LOESS
SVR	Support vector regression
Sym5	Symlet 5 wavelet
VMD	Variational mode decomposition
WTI	West Texas Intermediate
XGBoost	Extreme gradient boosting

References

Baffes, J., Kose, M. A., Ohnsorge, F., & Stocker, M. (2015). The Great Plunge in Oil Prices: Causes, Consequences, and Policy Responses. World Bank Policy Research Note, No. 1. Washington, DC: World Bank. https://econpapers.repec.org/paper/eencamaaa/2015-23.htm.

Bai, J., & Perron, P. (2003). Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics*, 18, 1–22. https://doi.org/10.1002/jae.659.

Boadi, E. (2025). Bitcoin Price Forecasting Based on a Hybrid Variational Mode Decomposition and Long Short-Term Memory Network**.** https://doi.org/10.48550/arXiv.2510.15900.

Boussatta, H., Chihab, M., Chihab, Y., & Chiny, M. (2023). Enhancing Oil Price Forecasting through an Intelligent Hybridized Approach. *International Journal of Advanced Computer Science and Applications*, 14(9), 115–125. https://doi.org/10.14569/IJACSA.2023.0140913.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 785–794. https://doi.org/10.1145/2939672.2939785.

Cheng, X., Wu, P., Liao, S. S., & Wang, X. (2023). An Integrated Model for Crude Oil Forecasting: Causality Assessment and Technical Efficiency. *Energy Economics*, 117, 106467. https://doi.org/10.1016/j.eneco.2022.106467.

Constantin, A. M., Davidescu, E. M., & Manta, A. G. (2025). Time Series Forecasting with LightGBM under Data Scarcity: An Application to Romania's Inland Gas Consumption. Proceedings of the 19th International Conference on Business Excellence 2025, 1518–1531. https://doi.org/10.2478/picbe-2025-0118.

Coroneo, L. & Iacone, F. (2025). Testing For Equal Predictive Accuracy with Strong Dependence, *International Journal of Forecasting*, 41(3), 1073–1092. https://doi.org/10.1016/j.ijforecast.2024.11.003.

Drachal, K. (2023). Forecasting the Crude Oil Spot Price with Bayesian Symbolic Regression. *Energies*, 16(1), 4. https://doi.org/10.3390/en16010004.

Faris Hameed, A., & Jumah Al-Thahab, O., Q. (2025). Comprehensive Review on Fast Fourier Transforms Types and Their Applications. *International Journal of Science and Research Archive*, 16(03), 1146–1152. https://doi.org/10.30574/ijsra.2025.16.3.2663.

Foroutan, P., & Lahmiri, S. (2024). Deep learning Systems for Forecasting the Prices of Crude Oil and Precious Metals. *Financial Innovation*, 10(1), 1–40. https://doi.org/10.1186/s40854-024-00637-z.

Gers, F. A., Schmidhuber, J. & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. *Neural Computer*; 12 (10): 2451–2471. https://doi.org/10.1162/089976600300015015.

Huang, L., Yang, X., Lai, Y., Zou, A., & Zhang, J. (2024). Crude Oil Futures Price Forecasting Based on Variational and Empirical Mode Decompositions and Transformer Mode. *Mathematics,* 12(24), 4034. https://doi.org/10.3390/math12244034

Huang, Z., & Sen, B. (2025). Multivariate Distribution-free Nonparametric Testing: Generalizing Wilcoxon's Tests Via Optimal Transport (arXiv:2503.12236). arXiv. https://doi.org/10.48550/arXiv.2503.12236.

Imani, M., Beikmohammadi, A., & Arabnia, H. R. (2025). Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels. *Technologies*, 13(3), 88. https://doi.org/10.3390/technologies13030088.

Jia, H., Cao, P., Liang, T., Caiafa, C. F., Sun, Z., Kushihashi, Y., Grau, A., Bolea, Y., Duan, F., & Sole-Casals, J. (2025). Short-Time Variational Mode Decomposition. *Computer Science*, arXiv. https://arxiv.org/abs/2501.09174.

Kayral, I. E., Aktaş Bozkurt, M., Bejaoui, A. & Jeribi, A. (2025). Hybrid Wavelet-SVR, Machine Learning, and Deep Learning Models for Predicting Clean Energy Markets Amidst Global Events. *Neural Computing & Applications*, 37, 16781–16823. https://doi.org/10.1007/s00521-025-11345-9.

Kilian, L., & Zhou, X. (2020). The Econometrics of Oil Market VAR Models. Working Paper, No. 8153, Federal Reserve Bank of Dallas. https://doi.org/10.24149/wp2006.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D.J. (2021). 1D Convolutional Neural Networks and Applications: A Survey. *Mechanical Systems and Signal Processing*, 151, 107398. https://doi.org/10.1016/j.ymssp.2020.107398.

Langsetmo, L., Schousboe, J. T., Taylor, B. C., Cauley, J. A., Fink, H. A., Cawthon, P. M., Kado, D. M., Ensrud, K. E. (2023). Advantages and Disadvantages of Random Forest Models for Prediction of Hip Fracture Risk Versus Mortality Risk in the Oldest Old. JBMR Plus, 7(8), e10757. https://doi.org/10.1002/jbm4.10757.

Li, J., Hong, Z., Zhang, C., Wu, J., & Yu, C. (2024). A Novel Hybrid Model for Crude Oil Price Forecasting Based on MEEMD and Mix-KELM. *Expert Systems with Applications*, 246, 123104. https://doi.org/10.1016/j.eswa.2023.123104.

Li, Z., Wang, M., Wang, X., Liu, Z., & Shi, A. (2020). Oil Price Forecasting Based on Variational Mode Decomposition, Relative Entropy and LSTM Neural Network. IOP Conference Series: Materials Science and Engineering, 750(1), 012203. https://doi.org/10.1088/1757-899X/750/1/012203.

Liu, L., Zhou, S., Jie, Q., Du, P., Xu, Y., & Wang, J. (2024). A Robust Time-varying weight combined model for crude oil price forecasting. *Energy*, 299, 131352. https://doi.org/10.1016/j.energy.2024.131352.

Nair, V. & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. Proceedings of the 27th International Conference on Machine Learning, Haifa, 807–814. https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf.

Niu, H., & Zhao, Y. (2021). Crude Oil Prices and Volatility Prediction by a Hybrid Model Based on Kernel Extreme Learning Machine. *Mathematical Biosciences and Engineering*, 18(6), 8096–8122. https://doi.org/10.3934/mbe.2021402.

Ouyang, Z.; Ravier, P.; Jabloun, M. (2021). STL Decomposition of Time Series Can Benefit Forecasting Done by Statistical Methods but Not by Machine Learning Ones. *Energy Proceeding*. 5, 42. https://doi.org/10.3390/ engproc2021005042.

Peng, Z. J., Zhang, C., & Tian, Y.-X. (2023). Crude Oil Price Time Series Forecasting: A Novel Approach Based on Variational Mode Decomposition, Time-Series Imaging, and Deep Learning. IEEE Access, 11, 85431–85444. https://doi.org/10.1109/ACCESS.2023.3301576.

Plakandaras, V., Gogas, P., Papadimitriou, T. & Gupta, R. (2017). The Informational Content of the Term Spread in Forecasting The US Inflation Rate: a Nonlinear Approach. *Journal of Forecasting*, 36(2), 109–122. https://doi.org/10.1002/for.2417.

Qin, Q., & Li, L. (2025). A VMD-based four-stage hybrid forecasting model with error correction for complex coal price series. *Mathematics*, 13(18), 2912. https://doi.org/10.3390/math13182912.

Ren, X., Xu, W., & Duan, K. (2022). Fourier Transform Based LSTM Prediction Model Under Oil Shocks. *Quantitative Finance and Economics*, 6(2), 342–358. https://doi.org/10.3934/QFE.2022015.

Scherer, D., Müller, A., Behnke, S. (2010). Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. In Proceedings of the 20th International Conference on Artificial Neural Networks, 92–101. https://doi.org/10.1007/978-3-642-15825-4_10

Seabe, P. L.; Pindza, E.; Moutsinga, C. R. B.; Aphane, M. (2025). Temporal Attention-Enhanced Stacking Networks: Revolutionizing Multi-Step Bitcoin Forecasting. *Forecasting*, 7(1), 1–28. https://doi.org/10.3390/forecast7010002.

Singh, P., Jha, M., & Patel, H. (2025). Wavelet-Enhanced Deep Learning Ensemble for Accurate Stock Market Forecasting: A Case Study of Nifty 50 Index. IEEE Access, 13, 89945–89960. https://doi.org/10.1109/ACCESS.2025.3568634.

Wang, L., Chen, H., & Kim, J. (2023). Hybrid machine learning models for oil price prediction: A comparative study. *Energy*, 268, 126–643. https://doi.org/10.1016/j.energy.2023.126643.

Yuan, F., Huang, X., Jiang, H., Jiang, Y., Zuo, Z., Wang, L., Wang, Y., Gu, S., & Peng, Y. (2025). An XLSTM–XGBoost Ensemble Model for Forecasting Non-Stationary and Highly Volatile Gasoline Price. *Computers*, 14(7), 256. https://doi.org/10.3390/computers14070256.

Zhang, Y., & Ma, F. (2023). Forecasting Crude Oil Prices with a New Bayesian Combination Approach. *Energy Economics*, 120, 106521. https://doi.org/10.1016/j.eneco.2023.106521.

Zheng, G., Li, Y., & Xia, Y. (2025). Crude Oil Price Forecasting Model Based on Neural Networks and Error Correction. *Applied Sciences*, 15(3), 1055. https://doi.org/10.3390/app15031055.

Zhu, S., Xu, M., Wu, J., Wang, Y., Jiang, X., Huang, Z., Wang, Y., & Zhu, Y. (2025). A Study on Crude Oil Price Forecasting Model Integrating CEEMDAN-VMD Multiscale Decomposition

with CNN-BiLSTM. *Results in Engineering*, 27, 106391. https://doi.org/10.1016/j.rineng.2025.106391.

Appendix

Table A1
Optimal decomposition parameters and their implications

Decomposition method	Parameter	Optimal value (short-term)	Optimal value (medium-term)	Optimal value (long-term)	Technical/statistical implications
VMD	Number of modes	5	4	5	Short/long horizons need more modes due to higher complexity.
	(Penalty factor)	1500	500	1000	Higher penalty reduces noise; lower allows broader trends.
STL	Seasonal window	21	21	21	Consistent 3-week window smooths seasonal variations.
Wavelet	Mother wavelet	sym5	sym5	db4	Symlet preserves phase short-term; db4 handles long-term transients.
	Level	3	3	3	Level 3 balances frequency separation and efficiency.
FFT	Components	5	3	10	More components capture complex long-term cycles.

Table A2
Optimized machine learning model parameters in direct models

Model	Parameter	Short-term	Medium-term	Long-term	Technical implications
SVR	Penalty	200	200	200	High penalty controls noise across all horizons
	Gamma	0.001	0.01	0.001	Higher gamma in medium-term captures local variations.
LGBM	Learning rate	0.1	0.1	0.1	A stable rate ensures convergence.
	Max depth	5	5	5	Shallow trees prevent overfitting.
	N estimators	50	100	100	More estimators for longer-term stability
Random Forest	Max depth	5	5	5	Constrains overfitting
	Min samples split	5	5	5	Ensuring statistically significant splits
XGBoost	Max depth	3	3	5	Slightly deeper trees for long-term
	Learning rate	0.1	0.1	0.1	Stable gradient updates
	N Estimators	100	100	50	Fewer trees in the long term to avoid overfitting
CNN	Filters	32	32	16	Fewer filters for long-term use to reduce redundancy
	Epochs	45	45	45	Converges uniformly across horizons
	Batch size	32	32	32	Balanced gradient stability & memory

Model	Parameter	Short-term	Medium-term	Long-term	Technical implications
LSTM	Units	50	50	50	Stable capacity for temporal dependencies
	Epochs	45	45	45	Adequate learning without overfitting
	Batch size	32	32	32	Consistent weight updates
ARIMA	Order (p,d,q)	(2, 1, 0)	(1, 1, 0)	(2, 1, 0)	Medium-term simpler dynamics

Table A3
Optimized machine learning model parameters in VMD-models

Model	Parameter	Short-term	Medium-term	Long-term	Technical implications
SVR	Penalty	100	200	200	Higher penalty for longer horizons enforces a tighter fit.
	Gamma	0.001	0.001	0.001	Low gamma suits smooth VMD features.
LGBM	Learning rate	0.1	0.1	0.1	Well-separated features allow aggressive learning.
	Max depth	5	5	5	Moderate depth is sufficient after VMD separation.
	N estimators	50	100	100	Moderate depth is sufficient after VMD separation.
Random Forest	Max depth	10	10	10	Deeper trees capture finer details after noise removal.
	N estimators	100	100	50	Fewer trees suffice for long-term trends.
	Min samples split	5	5	5	Ensures stable tree structure.
XGBoost	Max depth	3	3	3	Shallow trees perform well with informative VMD features.
	Learning rate	0.1	0.1	0.1	Feature scaling allows a uniform learning rate.
	N estimators	100	100	100	Consistent boosting rounds optimize performance.
CNN	Filters	32	32	32	VMD modes provide stable feature representation.
	Epochs	45	45	45	Convergence achieved within 45 epochs.
	Batch size	32	32	32	Efficient gradient calculation
LSTM	Units	50	50	50	Capturing nonlinear relationships in IMFs
	Epochs	45	45	45	Stable convergence for all horizons
	Batch size	32	32	32	Efficient training updates

Table A4
Optimized machine learning model parameters in STL-models

Model	Parameter	Short-term	Medium-term	Long-term	Technical implications
SVR	Penalty	200	200	200	High regularization prevents fitting residual noise.
	Gamma	0.001	0.001	0.001	Low gamma ensures generalization over trend and seasonal components.
LGBM	Learning rate	0.05	0.1	0.1	Short-term requires cautious learning to handle high-frequency fluctuations.
	Max depth	5	5	5	Capturing interactions between trend and seasonality without overfitting.
	N estimators	100	100	100	Consistent boosting rounds aggregate STL components effectively.
Random Forest	Max depth	10	10	5	Shallower depth for long-term prevents overfitting residuals.
	Min samples split	5	5	5	Ensuring stable tree growth
	N estimators	100	100	100	Maintaining robustness across horizons
XGBoost	Learning rate	0.1	0.1	0.1	Standard rate leverages separated trend/seasonal signals.
	Max depth	3	5	3	Medium-term depth captures transitional dynamics.
	N estimators	100	100	100	Consistent rounds ensure stability.
CNN	Filters	32	32	16	Fewer filters, long-term focus on dominant trend
	Epochs	45	45	45	Convergence was achieved consistently.
	Batch size	32	32	32	Standard batch size
LSTM	Units	50	50	50	Adequate memory for periodic patterns
	Epochs	45	45	45	Stable convergence
	Batch size	32	32	32	Standard batch size

Table A5
Optimized machine learning model parameters in FFT-based models

Model	Parameter	Short-term	Medium-term	Long-term	Technical implications
SVR	Penalty	200	200	200	High penalty fits spectral peaks accurately.
	Gamma	0.001	0.001	0.001	Low gamma prevents overfitting of spectral noise.
LGBM	Learning rate	0.05	0.05	0.1	Long-term dominant frequencies allow a higher learning rate.
	Max depth	5	5	5	Moderate depth avoids memorizing artifacts.
	N estimators	100	100	100	Consistent trees stabilize FFT feature aggregation.
Random Forest	Max depth	5	5	5	Controlling overfitting due to spectral leakage

Model	Parameter	Short-term	Medium-term	Long-term	Technical implications
XGBoost	Min samples split	5	5	5	Standard split ensures stable nodes.
	N estimators	50	100	100	More trees are needed for medium/long-term variance reduction.
	Learning rate	0.1	0.1	0.1	Stable input from FFT features
	Max depth	3	3	3	Shallow trees are sufficient; frequency components are strong predictors.
	N estimators	100	100	100	Consistent boosting rounds
CNN	Filters	32	16	16	Focusing on dominant frequencies in medium/long-term
	Epochs	45	45	30	Long-term converges faster due to simpler cycles.
	Batch size	32	32	32	Standard batch size
LSTM	Units	50	50	20	Long-term spectral patterns require fewer units.
	Epochs	45	45	45	Stable convergence
	Batch size	32	32	32	Standard batch size

Table A6
Optimized machine learning model parameters in wavelet models

Model	Parameter	Short-term	Medium-term	Long-term	Technical implications
SVR	Penalty	200	200	200	High penalty prevents underfitting Wavelet features.
	Gamma	0.001	0.001	0.001	Low gamma preserves localized frequency information.
LGBM	Learning rate	0.05	0.1	0.1	Short-term slower learning for high-frequency noise; longer horizons allow faster convergence.
	Max depth	5	5	5	Moderate depth avoids overfitting transient wavelet coefficients.
	N estimators	100	100	100	Stable boosting iterations across horizons
Random Forest	Max depth	10	10	10	Capturing a multi-resolution hierarchy of wavelet levels
	Min samples split	5	5	5	Standard pruning
	N estimators	100	100	50	Fewer trees suffice for long-term trends.
XGBoost	Learning rate	0.1	0.1	0.1	Stable gradient boosting
	Max depth	5	3	3	Shallow trees are sufficient for medium/long-term trends.
	N estimators	100	100	100	Consistent boosting rounds
CNN	Filters	32	32	32	Wavelet features maintain stable representations.
	Epochs	45	45	45	Training consistent
	Batch size	32	32	32	Standard batch size

Model	Parameter	Short-term	Medium-term	Long-term	Technical implications
LSTM	Units	50	50	50	Sufficient to model temporal dependencies
	Epochs	45	45	45	Stable convergence
	Batch size	32	32	32	Standard batch size

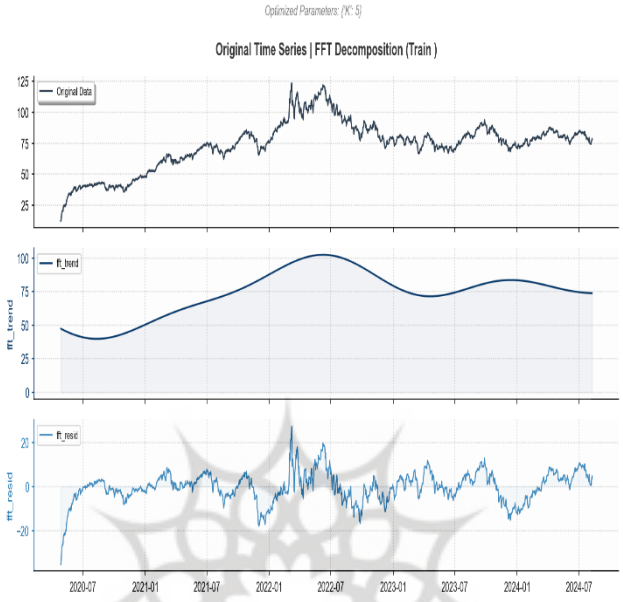

Figure A1

Time series FFT decomposition on the short-term training set

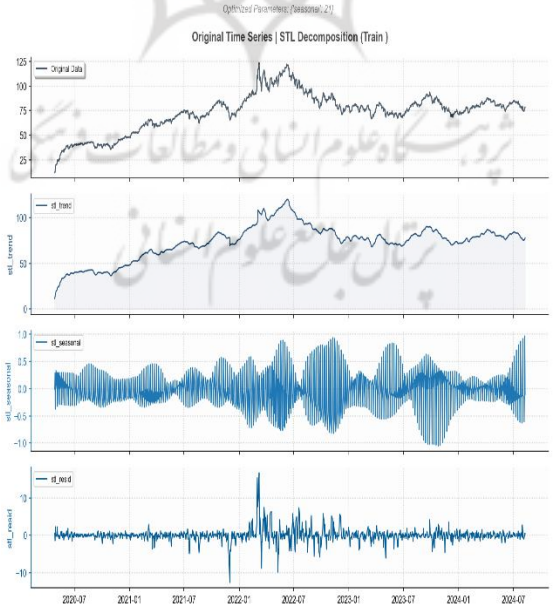

Figure A2

Time series STL decomposition on the short-term training set

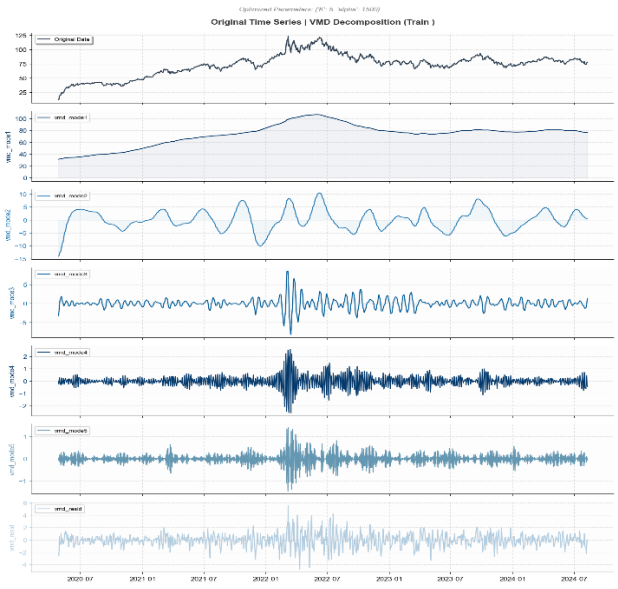

Figure A3

Time series VMD decomposition on the short-term training set

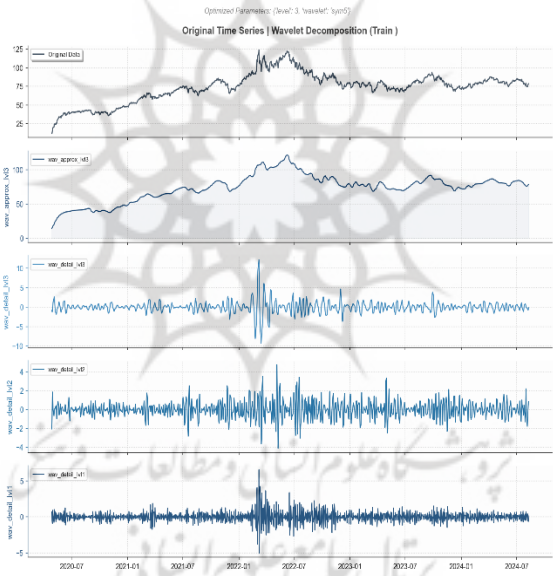

Figure A2

Time series wavelet decomposition on the short-term training set

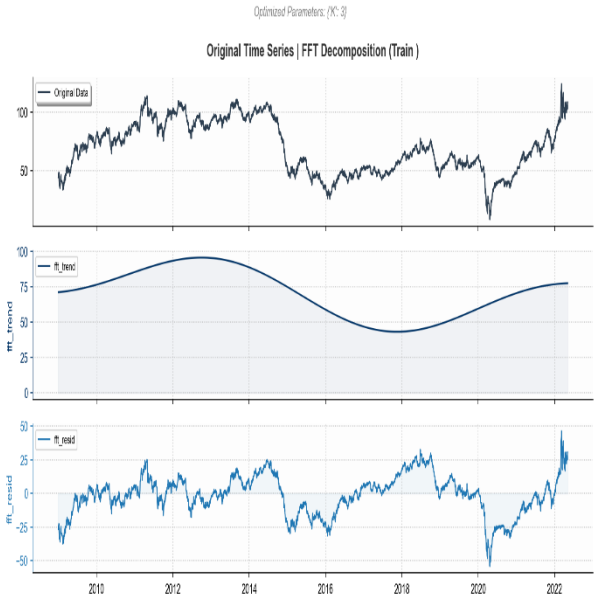

Figure A3

Time series FFT decomposition on the medium-term training set

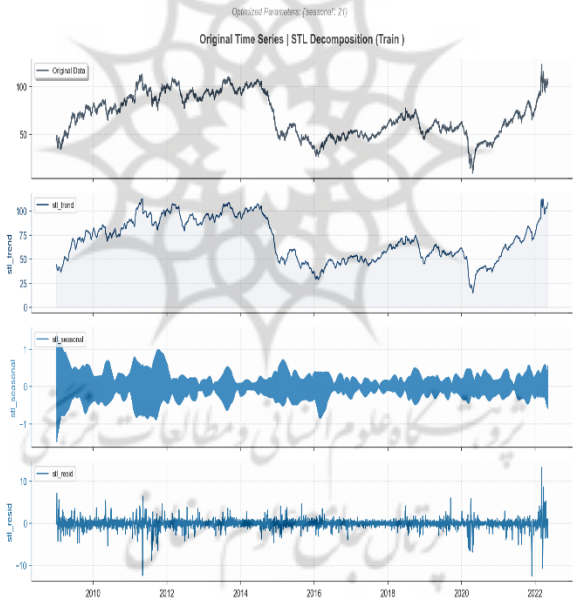

Figure A4

Time series STL decomposition on the medium-term training set

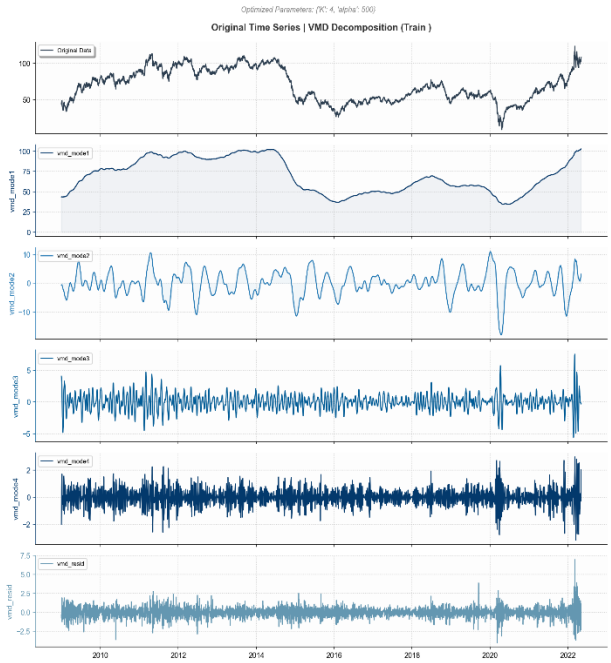

Figure A7

Time series VMD decomposition on the medium-term training set

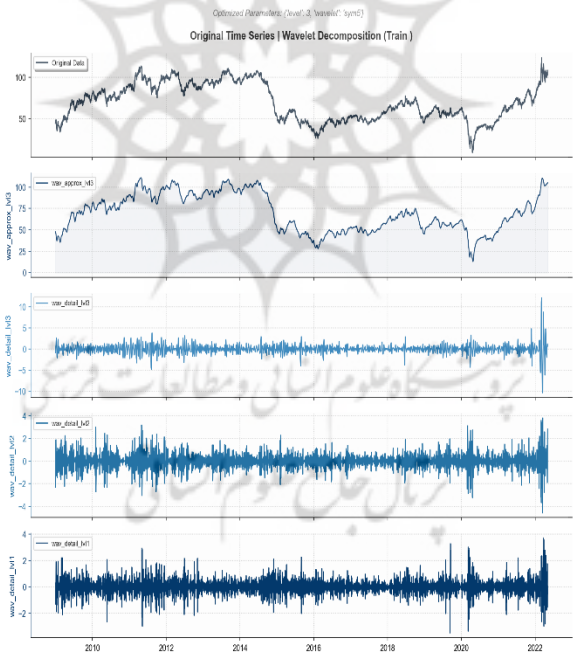

Figure A8

Time series wavelet decomposition on the medium-term training set

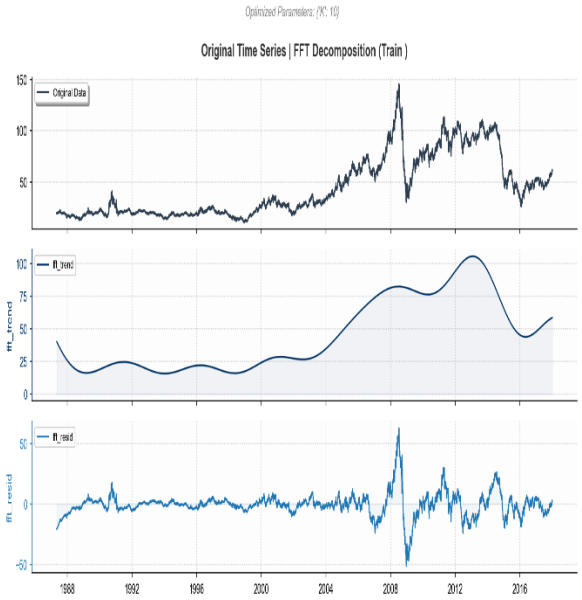

Figure A9

Time series FFT decomposition on the long-term training set

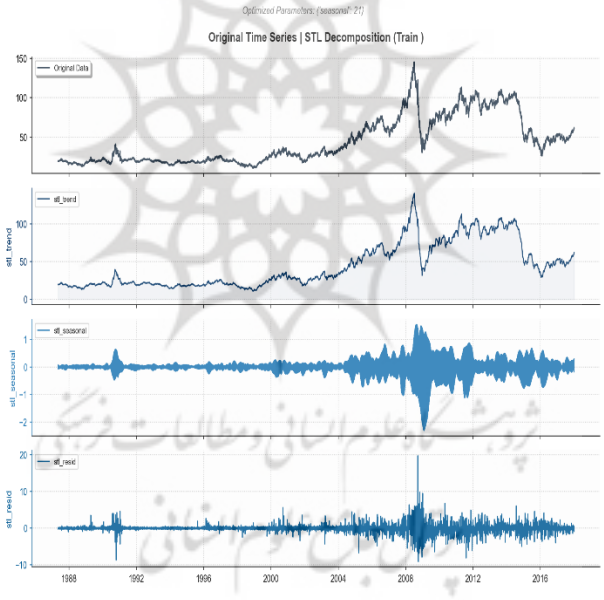

Figure A10

Time series STL decomposition on the long-term training set

Figure A11

Time series VMD decomposition on the long-term training set

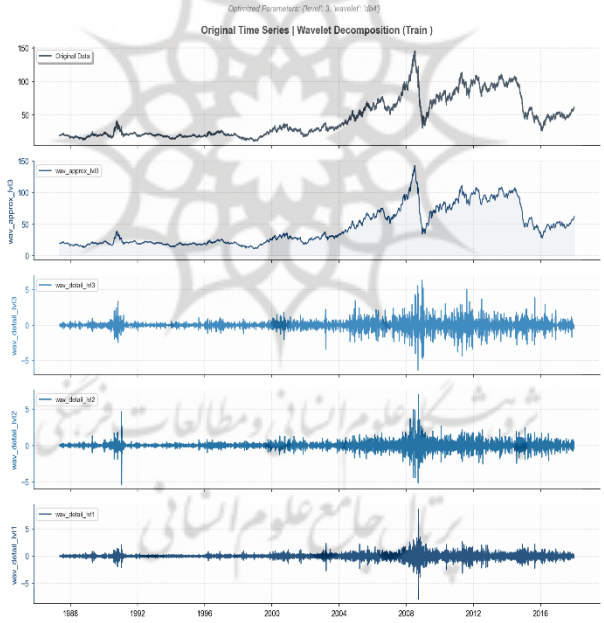

Figure A12

Time series wavelet decomposition on the long-term training set