Research Paper
Computational Methods for Differential Equations
http://cmde.tabrizu.ac.ir
Vol. *, No. *, *, pp. 1-21
DOI:10.22034/cmde.2026.70717.3530

Change-point based hybrid stochastic differential equation models: the case of WTI crude oil prices

Sevda Özdemir Çalıkuşu[1,*] and Fevzi Erdoğan[2]
[1]Department of Computer Technology, Van Yuzuncu Yil University, Van, Turkey.
[2]Department of Econometrics , Van Yuzuncu Yil University, Van, Turkey.

Abstract

This study models the fluctuations of West Texas Intermediate (WTI) crude oil prices between 01.03.2019 and 13.03.2023 using Geometric Brownian Motion (GBM) and Cox–Ingersoll–Ross (CIR) stochastic differential equations (SDEs). The model parameters are estimated using the Quasi-Maximum Likelihood Estimation (QMLE) method, and the compatibility of both models is validated through the chi-square goodness-of-fit test. Due to abrupt price changes, two significant change points (CPs), on May 14, 2020, and February 24, 2022, are estimated using the QMLE method. These CPs correspond to major global events, namely the COVID-19 pandemic and Russia's military intervention in Ukraine. The GBM and CIR models, redefined based on these CPs, are combined into a hybrid SDE model, which demonstrates superior performance compared to other models, as evaluated by MAPE and RMSE criteria. The hybrid SDE model more accurately predicts sudden price fluctuations, thereby improving the model's accuracy. These findings indicate that CP estimation significantly improves the effectiveness of SDE models, enabling more reliable modeling of abrupt changes in economic time series data. Consequently, the integration of CPs contributes to improving the accuracy of economic forecasts and facilitates more reliable predictions of future market behavior, especially in response to geopolitical and economic disruptions.

Keywords. Geometric brownian motion, Change point estimation, Cox–Ingersoll–Ross, Hybrid model, Stochastic differential equation.
2010 Mathematics Subject Classification. 65C20, 65C30, 65C30, 62P05, 62P20.

1. Introduction

Crude oil prices exhibit complex and time-varying dynamics due to their high sensitivity to both economic and geopolitical developments in global energy and financial markets. Supply disruptions, demand shocks, geopolitical tensions, and global crises often induce abrupt changes in price behavior and regime transitions. These features are particularly pronounced for benchmark crude oil prices such as West Texas Intermediate (WTI). Notably, the unprecedented collapse during the COVID-19 period generated structural breaks of historical magnitude in WTI price dynamics [13]. Subsequent geopolitical shocks further intensified regime transitions in oil markets [33], while the persistent effects of these episodes on energy price formation have been documented in recent studies [26]. As WTI serves as a key reference price for other crude oil grades and is supported by a reliable and consistently reported data infrastructure from the U.S. Energy Information Administration, it provides a suitable empirical setting for investigating regime changes in energy markets [46].

Stochastic differential equations (SDEs) offer a well-established and powerful framework for modeling energy price dynamics in continuous time. In SDE models, price evolution is governed by the interaction between a drift component, representing systematic economic forces, and a diffusion component, capturing random fluctuations and uncertainty. The tradition of diffusion-based continuous-time modeling in oil markets is well rooted in the literature. The foundational framework for an economically interpretable and realistic continuous-time representation of oil prices was introduced by Gibson and Schwartz [19] and subsequently refined by Schwartz [44]. This approach was later extended

Received: 14 December 2025; Accepted: 10 September 2026.
∗ Corresponding author. Email: sevdaozdemir@yyu.edu.tr.

through multifactor specifications and applied across various energy markets [14], with further empirical validation provided in subsequent studies [15]. Collectively, this body of work establishes the analytical and empirical relevance of modeling oil prices within continuous-time stochastic frameworks [39]. Within the SDEs literature, there also exist theoretical studies addressing the stability properties of stochastic processes, which contribute to the general understanding of diffusion dynamics [2, 8]. In addition, studies focusing on numerical schemes and stochastic modeling approaches in financial contexts can be considered within different research objectives, and may be viewed as complementary to this theoretical background [3, 41].

Beyond random fluctuations, crude oil prices are also characterized by time-varying structural behavior and regime shifts. Ignoring structural breaks may lead to misspecified dynamics and distorted inference, particularly in volatility and risk-related measures [36]. The empirical consequences of such breaks have been examined in detail for oil markets [42]. Approaches relying on predetermined break dates demonstrate that oil price dynamics differ markedly across subperiods [31], while data-driven methods reveal that parameter changes in the price-generating process can be captured more flexibly when breakpoints are estimated directly from the data [6]. Recent evidence further indicates that multiple structural breaks may arise in oil markets during periods of heightened geopolitical tension [40]. In the context of risk and spillover analysis, explicitly incorporating structural breaks has also been shown to enhance measurement accuracy [32], with further support provided by more recent studies [12]. Accordingly, the core issue is not merely price volatility, but rather the time-varying parameter structure of the price process and the appropriate modeling of such changes.

Within this context, change-point (CP) analysis serves as a fundamental tool for identifying abrupt shifts in the parameters of stochastic processes without requiring prior knowledge of break locations [11]. An early statistical framework for detecting parameter changes in diffusion-type processes was developed by Kutoyants [30]. Subsequent contributions demonstrated that single and multiple CPs in diffusion volatility can be identified using quasi-maximum likelihood–based procedures [17]. More general estimation strategies for SDEs with regime shifts were later proposed by Iacus [21], and formalized into a systematic framework allowing joint changes in both drift and diffusion parameters by Iacus and Yoshida [23]. Despite these methodological advances, empirical studies that systematically integrate SDE modeling with CP estimation and construct regime-dependent hybrid diffusion structures remain relatively scarce, particularly in crude oil markets.

This study approaches energy markets not by coincidence but based on the observation that several tools developed within the SDE literature can be naturally and coherently applied to oil price dynamics. Euler–Maruyama (EM)–based trajectory and sample-path analyses provide fundamental insights into the behavior of stochastic processes [29]. Parameter estimation via quasi-maximum likelihood estimation (QMLE) offers a consistent and practical solution for SDEs when closed-form transition densities are unavailable [23]. CP analysis further enables the direct identification of regime-dependent behavior from the data [17]. Accordingly, the empirical strategy adopted in this study proceeds by first modeling the WTI price series using classical single-regime diffusion SDEs, then estimating a CP to assess regime adequacy, and subsequently exploring the presence of an additional CP through a re-estimation strategy when the single-break structure appears insufficient. The resulting framework facilitates a systematic comparison of models with no CP, one CP, and two CPs.

In this study, CPs are first estimated as a preliminary step, and the hybrid SDE framework is then constructed conditionally on these estimated regimes. Therefore, the proposed hybrid model does not jointly estimate CPs, but rather exploits their presence to model regime-dependent dynamics.

2. Materials and methods

2.1. Stochastic Framework. Crude oil prices are shaped by supply disruptions, geopolitical tensions, speculative activity, and rapidly changing demand conditions, all of which introduce substantial uncertainty into price dynamics [39]. These features make SDEs a natural and widely adopted framework for modeling oil prices in continuous time [19]. By incorporating randomness via stochastic processes, SDEs allow both systematic economic forces and unpredictable shocks to jointly determine price evolution [44]. Within this framework, randomness is introduced through stochastic processes driven by Brownian motion, which enables the model to capture volatility clustering and abrupt market movements [38]. The deterministic component reflects long-run economic tendencies, while the stochastic component represents short-term uncertainty and market noise [27]. A general one-dimensional Itô diffusion process is defined as

$$dX_t = f(t, X_t, \theta)\, dt + g(t, X_t, \theta)\, dW_t. \tag{2.1}$$

where W_t denotes a standard Brownian motion [38]. The function $f(\cdot)$ represents the drift component governing systematic price movements [27], while $g(\cdot)$ denotes the diffusion component capturing stochastic volatility [38]. Under standard regularity conditions, Eq. (2.1) admits a unique strong solution, ensuring well-defined price dynamics over time [27].

2.2. **Diffusion Models.** Empirical evidence indicates that oil prices alternate between periods dominated by persistent trends and periods characterized by stabilization and mean reversion [39]. To capture these distinct dynamics, this study focuses on two canonical diffusion models that are widely used in financial and commodity markets [9].

2.2.1. *GBM Specification.* The Geometric Brownian Motion (GBM) model is specified as

$$dX_t = \theta_1 X_t\, dt + \sigma X_t\, dW_t, \qquad X_0 = x_0, \tag{2.2}$$

where θ_1 denotes the proportional drift parameter [9]. The parameter $\sigma > 0$ represents volatility and determines the magnitude of stochastic fluctuations [34]. The multiplicative structure of the drift and diffusion terms implies that both expected growth and uncertainty scale with the current price level, a property often observed during growth-dominated phases of oil markets [1]. This structure also ensures that sample paths remain strictly positive over time [9]. The mean and variance of the GBM process are given by

$$m(t, x_0) = x_0\, e^{\theta_1 t}, \tag{2.3}$$

$$v(t, x_0) = x_0^2\, e^{2\theta_1 t} \left(e^{\sigma^2 t} - 1 \right). \tag{2.4}$$

These moment expressions describe exponential growth in expected prices and increasing uncertainty over time [21]. In empirical applications, they are frequently used to construct confidence bands and assess the adequacy of the GBM specification [35].

Although GBM is effective in modeling smooth growth and proportional volatility, it lacks a mean-reverting mechanism and may therefore perform poorly during periods of market correction or structural change [34].

2.2.2. *CIR Specification.* To account for mean-reverting behavior commonly observed in oil prices, the Cox–Ingersoll–Ross, (CIR) model is considered [16]. The model is defined by

$$dX_t = (\theta_1 - \theta_2 X_t)\, dt + \sigma \sqrt{X_t}\, dW_t, \qquad X_0 = x_0. \tag{2.5}$$

The parameter θ_2 controls the speed of mean reversion, reflecting market forces that drive prices toward equilibrium levels [16]. The square-root diffusion term introduces state-dependent volatility, ensuring that fluctuations diminish as prices approach zero and increase when prices are high [18]. This structure guarantees non-negativity of the process under standard conditions [16]. The mean and variance functions of the CIR process are given by

$$m(t, x_0) = \frac{\theta_1}{\theta_2} + \left(x_0 - \frac{\theta_1}{\theta_2} \right) e^{-\theta_2 t}, \tag{2.6}$$

$$v(t, x_0) = x_0 \left(\frac{\sigma^2 \left(e^{-\theta_2 t} - e^{-2\theta_2 t} \right)}{\theta_2} \right) + \frac{\theta_1 \sigma^2 \left(1 - e^{-2\theta_2 t} \right)}{2\theta_2^2}. \tag{2.7}$$

These expressions highlight the stabilizing effect of mean reversion and the asymmetric volatility structure inherent in the CIR model [21].

2.3. **Numerical Approximation.** While the CIR model is well suited for capturing equilibrium-seeking behavior, it may struggle to represent large and abrupt price movements caused by exogenous shocks, motivating the use of regime-dependent modeling approaches [18].

$$X_{t_{i+1}} = X_{t_i} + f(t_i, X_{t_i}, \theta)\, \Delta + g(t_i, X_{t_i}, \theta)\, \Delta W_i, \qquad i = 1, \dots, n-1, \tag{2.8}$$

where $\Delta W_i \sim \mathcal{N}(0, \Delta)$ is the sequence of independent increments of the Wiener process [23]. The EM scheme is computationally efficient and straightforward to implement, making it suitable for large-scale simulations and iterative estimation procedures [29]. Building on this well-established numerical framework, recent studies have proposed alternative approaches for modeling and approximating SDEs. In particular, Ince and Senturk [25] and Ince and Shamilov [24] introduce a new methodology that contributes to the growing literature on numerical and estimation techniques for continuous-time stochastic processes. In this study, the EM method is used both to simulate diffusion paths under estimated parameters and to approximate hybrid SDE trajectories whose analytical solutions are unavailable [23]. In order to determine the range over which the respective EM solutions will vary, the confidence interval $m(t, x_0) \pm 2\sqrt{v(t, x_0)}$ constructed for the solutions of stochastic processes is utilized [35].

2.4. **Parameter Estimation.** Let $x_0, x_1, \dots, x_N$ denote discrete observations of the diffusion process. Given the transition density $p(t_k, x_k \mid t_{k-1}, x_{k-1}; \theta)$, the joint likelihood function is defined as

$$D(\theta) = p_0(x_0 \mid \theta) \prod_{k=1}^{N} p(t_k, x_k \mid t_{k-1}, x_{k-1}; \theta), \tag{2.9}$$

which forms the basis for maximum likelihood estimation (MLE) [5].

To facilitate numerical optimization, the negative log-likelihood function is considered:

$$L(\theta) = -\ln(p_0(x_0 \mid \theta)) - \sum_{k=1}^{N} \ln(p(t_k, x_k \mid t_{k-1}, x_{k-1}; \theta)). \tag{2.10}$$

The MLE is obtained by minimizing this function:

$$\theta^* = \arg\min_{\theta} L(\theta). \tag{2.11}$$

In practice, transition densities are rarely available in closed form for diffusion processes, which limits the direct applicability of MLE [23]. To overcome this limitation, this study adopts a quasi-maximum likelihood estimation (QMLE) approach that relies on Gaussian approximations of discretized increments [28].

To construct the QMLE, consider Eq. (2.1) and its Euler approximation. For equally spaced observations, the increment is expressed as

$$\Delta X_i \approx f(X_{t_{i-1}}, \theta)\, \Delta + g(X_{t_{i-1}}, \theta)\, \Delta \eta_i, \qquad \eta_i \sim N(0, 1). \tag{2.12}$$

This approximation yields the one-dimensional quasi-log-likelihood function:

$$l(\theta) = -\frac{1}{2} \sum_{i=1}^{n} \left[\log\left(g^2(X_{t_{i-1}}, \theta)\right) + \frac{\left(\Delta X_i - f(X_{t_{i-1}}, \theta)\Delta\right)^2}{g^2(X_{t_{i-1}}, \theta)\Delta} \right]. \tag{2.13}$$

The QMLE estimator is then given by

$$\hat{\theta} = \arg\max_{\theta} l(\theta). \tag{2.14}$$

2.5. **Change-Point Framework.** The one-dimensional formulation is deliberately adopted, as the empirical analysis focuses on a single crude oil price series, and extending the likelihood to a general multidimensional framework would introduce unnecessary complexity without improving interpretability or estimation accuracy [23]. The regime-specific parameter vectors are defined as

$$\theta(t) = \begin{cases} \left(\theta_f^{(0)}, \theta_g^{(0)}\right), & t < \tau_0, \\ \left(\theta_f^{(1)}, \theta_g^{(1)}\right), & t \geq \tau_0, \end{cases} \qquad \left(\theta_f^{(0)}, \theta_g^{(0)}\right) \neq \left(\theta_f^{(1)}, \theta_g^{(1)}\right). \tag{2.15}$$

Under this assumption, the diffusion process is written as

$$dX_t = f(X_t;\theta_f)\,dt + g(X_t;\theta_g)\,dW_t. \tag{2.16}$$

The regime-dependent representation of the process is given by

$$dX_t = \begin{cases} f\left(X_t;\theta_f^{(0)}\right)\,dt + g\left(X_t;\theta_g^{(0)}\right)\,dW_t, & t < \tau_0, \\ f\left(X_t;\theta_f^{(1)}\right)\,dt + g\left(X_t;\theta_g^{(1)}\right)\,dW_t, & t \geq \tau_0. \end{cases} \tag{2.17}$$

This formulation allows simultaneous changes in drift and diffusion parameters, which is particularly relevant for oil markets, where shifts in trend and volatility often occur jointly [45].

2.6. **Change-Point Estimation.** Let $\{X_{t_i}\}_{i=0}^n$ denote discrete observations with $t_i = i\Delta$ and $\Delta = T/n$. Based on the quasi-log-likelihood defined in Eq. (2.13), change-point (CP) estimation proceeds by segmenting the sample at a candidate break index $k \in \{1,\dots,n-1\}$ [23].

The associated contrast function is defined as

$$l_i(\theta) = -\frac{1}{2}\left[\log\big(g^2(X_{t_{i-1}};\theta_g)\big) + \frac{\big(\Delta X_i - f(X_{t_{i-1}};\theta_f)\,\Delta\big)^2}{g^2(X_{t_{i-1}};\theta_g)\,\Delta}\right]. \tag{2.18}$$

where $l_i(\cdot)$ denotes the quasi-log-likelihood contribution at time t_i [37].

Let $k \in \{1,\dots,n-1\}$ represent a possible break index, implying a candidate CP $t_k = k\Delta$. The corresponding contrast function is

$$\Phi_n(k;\theta^{(0)},\theta^{(1)}) = -\sum_{i=1}^{k} l_i\left(\theta^{(0)}\right) - \sum_{i=k+1}^{n} l_i\left(\theta^{(1)}\right). \tag{2.19}$$

The CP estimator is therefore given by

$$\hat{\tau} = t_{\hat{k}}, \qquad \hat{k} = \arg\min_{k\in\{1,\dots,n-1\}} \Phi_n\left(k;\,\hat{\theta}^{(0)}(k),\,\hat{\theta}^{(1)}(k)\right). \tag{2.20}$$

Consistency and n-rate convergence of $\hat{\tau}$ follow from the asymptotic theory developed for discretely observed diffusion processes [37, 45].

2.7. **Model Validation.** To validate whether each SDE specification adequately represents the observed WTI dynamics, a chi-square–based goodness-of-fit test is applied. In this study, a chi-square goodness-of-fit test specifically developed for SDEs is employed to assess the validity of the model. This test uses Monte Carlo simulations to evaluate the null hypothesis, determining whether the model is consistent with the observed data. The test, which is thoroughly explained by Bak et al. [7], is commonly used in SDE modeling to ensure model validity.

The core idea behind this method is to compare the observed data with simulated trajectories at each observation point, assigning ranks to these comparisons. Based on these ranks, a test statistic is calculated, which is then compared to the chi-square distribution, derived from Pearson's theorem, to determine the goodness-of-fit. If the null hypothesis is rejected, it indicates that the model is inconsistent with the data.

This method has been applied throughout the study to test the model's consistency with the data, and this result has been further supported by graphical representations.

2.8. **Model Comparison.** Several standard accuracy and information criteria are used to compare alternative model specifications. Choosing the best model among the proposed SDE models for the available data is a result that researchers want to achieve. First, the mean absolute percentage error (MAPE), a scale-independent forecast accuracy measure commonly used in the literature [20], is defined as

$$\text{MAPE} = \frac{1}{n}\left(\sum_{i=1}^{n}\frac{|X_i - Y_i|}{X_i}\right)\times 100. \tag{2.21}$$

where n denotes the number of observations, X_i represents the observed value at time i, and Y_i is the corresponding model prediction. Because MAPE is expressed as a percentage, it facilitates direct comparison across datasets with different scales.

Secondly, the root mean squared error (RMSE), another widely used measure of predictive accuracy, is defined as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2}, \tag{2.22}$$

where again X_i and Y_i denote observed and predicted values, respectively [10]. The RMSE penalizes larger errors more severely and is therefore particularly informative when large deviations from observed values are undesirable.

In addition to accuracy metrics, this study incorporates an information criterion to compare model fit while accounting for model complexity. The Akaike Information Criterion (AIC) [4] provides a likelihood-based assessment and is defined as

$$\text{AIC} = 2k - 2\ln(L), \tag{2.23}$$

where k denotes the number of estimated parameters and L is the maximized likelihood value. By penalizing models with a greater number of parameters, AIC helps prevent overfitting. Consequently, among competing specifications, the preferred model is the one with the lowest AIC value.

Similarly, the Bayesian Information Criterion (BIC) is defined as

$$\text{BIC} = \ln(n)\, k - 2\ln(L), \tag{2.24}$$

where n is the number of observations. While both AIC and BIC provide a balance between goodness-of-fit and model complexity, BIC generally imposes a stronger penalty for complexity, making it more conservative than AIC in model selection [43]. These criteria are especially useful when comparing models of different complexities, as they help determine the best-fitting model without overfitting.

Taken together, MAPE, RMSE, AIC, and BIC offer a comprehensive set of complementary criteria for evaluating and comparing SDE models under various structural assumptions, including those involving regime changes.

3. Comparative Evaluation of SDE Models

In this study, it is recognized that accounting for abrupt changes triggered by global events is crucial for accurately modeling price fluctuations. Accordingly, a hybrid model integrating the GBM and CIR SDE frameworks was employed, with the incorporation of CP detection into the modeling structure. The hybrid model, which combines the strengths of both GBM and CIR, offers a more robust approach for predicting future price movements, particularly within the context of large-scale economic disruptions.

Given that the WTI price trajectory resembles the standard Wiener process, the WTI crude oil closing data between 01.03.2019 and 13.03.2023 were analyzed using SDE modeling techniques. The dataset was obtained from the U.S. Energy Information Administration (EIA) website: https://www.eia.gov/dnav/pet/hist/RWTCD.htm [46]. The time series is illustrated in Figure 1.

This specific time range was selected because it captures a historically volatile period characterized by major macroeconomic and geopolitical disruptions—including the COVID-19 pandemic and the Russia–Ukraine conflict—which significantly influenced global oil supply and demand.

Before proceeding with SDE modeling, it is necessary to determine whether the WTI dataset satisfies the assumptions of the Wiener process.

3.1. **Checking Wiener Process Assumptions for WTI Dataset.** Before proceeding with SDE modeling, it is essential to evaluate whether the WTI dataset satisfies the assumptions of normality, stationarity, and independent increments of the Wiener process. As previously discussed, on April 20, 2020, the closing price of WTI crude oil dropped to negative values, which significantly violated the normality assumption. Such extreme movements are highly unusual and indicate a data anomaly. To ensure the robustness of the SDE analysis, an outlier detection process was conducted on the 1012-day dataset, confirming that the price recorded on April 20, 2020, was an outlier. Consequently, this value was removed from the dataset to prevent distortions in the model estimation. Although

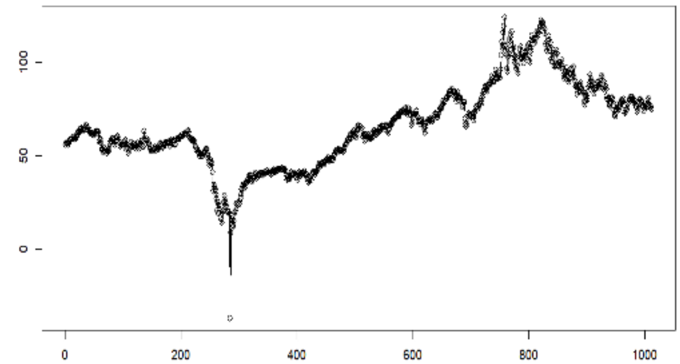


FIGURE 1. WTI crude oil prices in USD per barrel.

(A) (B)

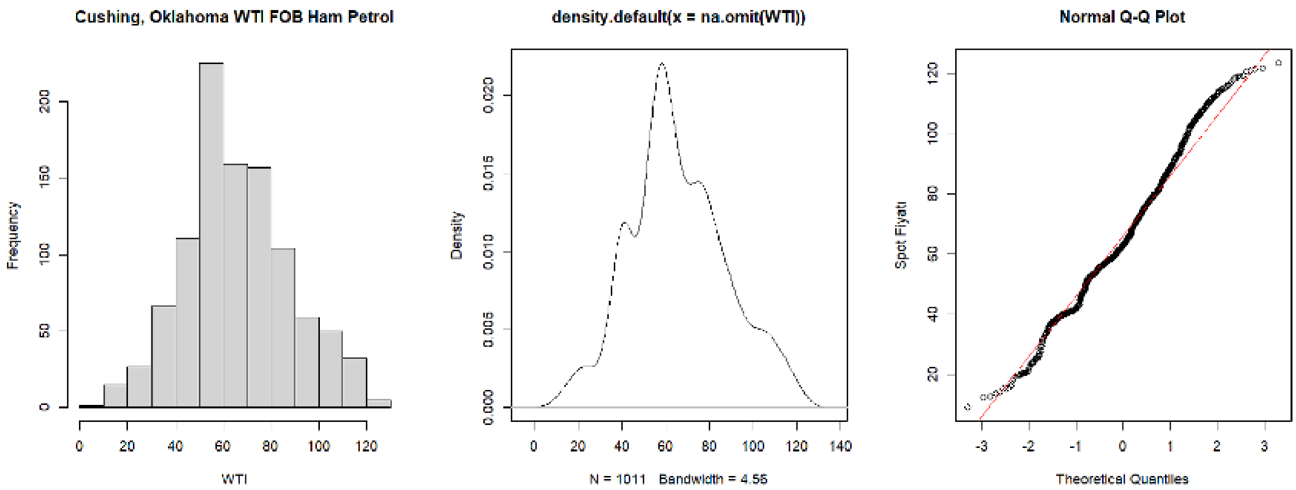
FIGURE 2. WTI crude oil price data: (a) outlier detection, (b) cleaned price series after outlier removal.

the negative oil price observed on April 20, 2020 reflects an exceptional market event, it was excluded from the analysis due to its incompatibility with the underlying assumptions of the SDE framework, particularly log-based transformations and numerical stability. This exclusion is therefore driven by modeling considerations rather than economic interpretation. Figure 2 illustrates this procedure: panel (a) shows the identification of the outlier in the original WTI price series, while panel (b) presents the cleaned WTI price series after the removal of the detected outlier.

3.1.1. *Checking the Normality Assumption.* The normality assumption of the Wiener process for the WTI dataset was evaluated using the Jarque-Bera normality test, implemented through the "tseries" library in RStudio. The test was conducted under the following hypotheses:

H_0 : The increments of the WTI series follow a normal distribution.
H_1 : The increments of the WTI series does not follow a normal distribution.

The test statistic was computed as 4.3262, with a corresponding p-value of 0.115. Given that $p = 0.115 > 0.05 = \alpha$, the null hypothesis cannot be rejected at the 5% significance level. This result suggests that the WTI dataset does not significantly deviate from normality. Furthermore, Figure 3 provides a graphical representation of the dataset, further



(A) Histogram. (B) Density function. (C) Q–Q plot.

FIGURE 3. Histogram, density function and Q-Q plot of WTI data.

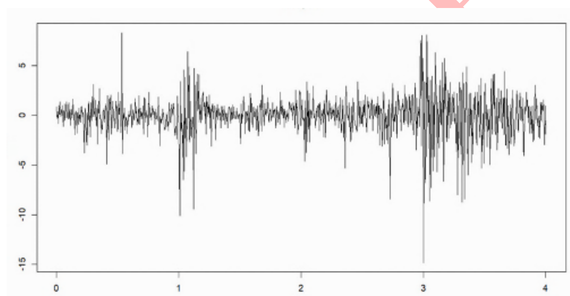


FIGURE 4. The increases of the WTI data set.

supporting the assumption of normality. The visual analysis suggests that the dataset exhibits asymptotic normality, reinforcing the conclusions drawn from the statistical test.

3.1.2. *Checking the Stationarity Assumption.* As the second step, the stationarity assumption of the WTI dataset was tested using the Augmented Dickey-Fuller (ADF) unit root test, implemented through the "urca" library in RStudio. The test was conducted under the following hypotheses:

H_0 : The increments of the WTI dataset are stationary.
H_1 : The increments of the WTI dataset are not stationary.

The test statistic was computed as -23.2192, which is smaller than the critical values of -3.43, -2.86, and -2.57 at the 1%, 5%, and 10% significance levels, respectively. Since the test statistic falls below all three critical thresholds, the null hypothesis cannot be rejected, indicating that the increments of the WTI dataset are stationary. Figure 4 presents the increments of the WTI dataset, where time t spans the interval from $0 = t_0 = 01.03.2019$ to $13.03.2023 = t_n = T$. To standardize the time scale, t was transformed into the interval $[0, 4)$ by setting $\Delta = t_{i+1} - t_i = 1/252$ for $i = 0, 1, 2, \ldots, n$. This time scaling approach is applied consistently throughout the study.

3.1.3. *Checking the assumption of independence.* As the final step, the independence assumption of the WTI dataset was tested using the Box-Pierce test under the following hypotheses:

Table 1. Results of the parameter estimation for Eq. (2.2) and Eq. (2.5) using the WTI dataset.

Model/Parameter	$\boldsymbol{\theta_1}$		$\boldsymbol{\theta_2}$		$\boldsymbol{\sigma}$	
	Estimation	Std. error	Estimation	Std. error	Estimation	Std. error
GBM SDE	0.3373	0.3599	–	–	0.7206	0.0161
CIR SDE	108.8132	42.4974	1.5787	0.6945	4.2037	0.0952

H_0 : The autocorrelation coefficients for all lags of the WTI data are zero,
H_1 : The autocorrelation coefficients are significantly different from zero.

The test statistic was computed as 0.0038642, with a corresponding p-value of 0.9504. Given that $p = 0.9504 > 0.05 = \alpha$, the null hypothesis cannot be rejected at the 5% significance level, confirming that the increments of the WTI dataset are independent. Since the normality, stationarity, and independence assumptions of the Wiener process are satisfied for the WTI dataset, SDE modeling can now be performed with confidence.

3.2. **SDE Modelling for WTI Dataset.** In this section, the WTI dataset is analyzed using the GBM and CIR SDE models, initially defined in Eq. (2.2) and Eq. (2.5). These models serve as the mathematical framework for the proposed hybrid approach. Initially, the GBM and CIR SDE models are applied to the WTI dataset without considering CPs to establish baseline estimations. The parameters of both models are estimated using the QMLE method, and the results are presented in Table 1. All estimations, including parameter and CP detection, are performed using the "yuima" library in RStudio.

If the obtained parameters are substituted into both equations, the resulting formulations are given in Eq. (3.1) and Eq. (3.2) below.

$$dX_t = 0.3373\, X_t\, dt + 0.7206\, X_t\, dW_t, \qquad X_0 = 55.76, \tag{3.1}$$

$$dX_t = (108.8132 - 1.5787\, X_t)\, dt + 4.2037\, \sqrt{X_t}\, dW_t, \qquad X_0 = 55.76. \tag{3.2}$$

Following the parameter estimation, the chi-square goodness-of-fit test —widely utilized in the literature for evaluating SDE models— was conducted to assess the adequacy of the GBM SDE model Eq. (3.1) in representing the WTI dataset.

H_0 : There is no significant lack of fit between the (3.1) GBM SDE model and the WTI data.

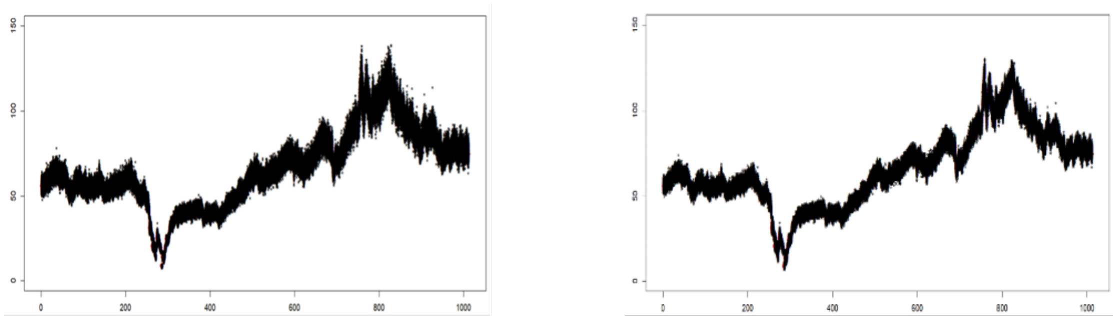
H_1 : There is a significant lack of fit between the (3.1) GBM SDE model and the WTI data.

Under the null hypothesis, $M = 201$ Monte Carlo simulations were conducted for each of the $N = 1012$ observations, and the chi-square test statistic was found to be $\chi_1^2 = 247.55457$. The GBM SDE model, defined by Eq. (3.1), has two estimated parameters, and thus the degrees of freedom are $df = 201 - 2 = 199$. At a 1% significance level ($\alpha = 0.01$), the chi-square critical value is $\chi^2_{199,0.01} = 248.3286$. Since the calculated test statistic χ_1^2 is smaller than the critical value, i.e., $\chi_1^2 < \chi^2_{199,0.01}$, the null hypothesis that there is no lack of fit between the model and the data cannot be rejected.

Similarly, the chi-square goodness-of-fit test was applied to the CIR SDE model Eq. (3.2) to assess whether there was a lack of fit between the data and the model. The hypotheses were:

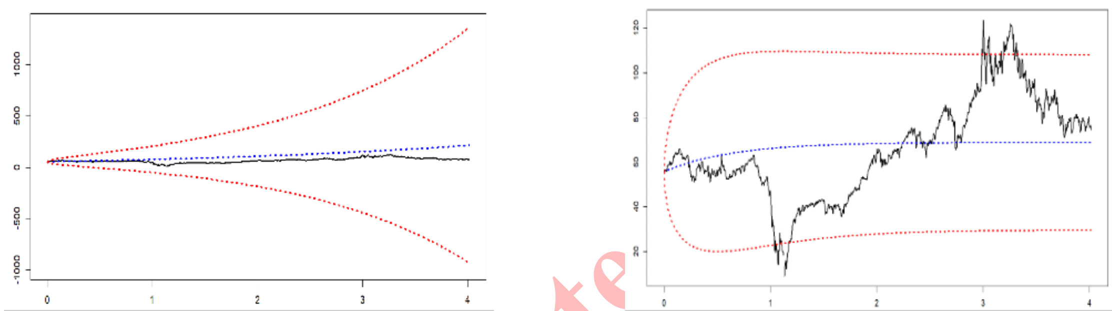
H_0: There is no lack of fit between the model and the data.
H_1: There is a lack of fit between the model and the data.

The test statistic was found to be $\chi_2^2 = 229.9195$. At a 1% significance level, with $df = 198$, the chi-square critical value is $\chi^2_{198,0.01} = 247.2118$. Since the calculated test statistic χ_2^2 is smaller than the critical value, i.e., $\chi_2^2 < \chi^2_{198,0.01}$, the null hypothesis cannot be rejected. Therefore, both the GBM SDE is obtained with Eq. (3.1) and CIR SDE is obtained with Eq. (3.2) models are considered to fit the WTI data well. The results of the goodness-of-fit analysis for both models are visually presented in Figure 5, where the WTI dataset is shown as a red line and the Monte Carlo simulation outcomes are indicated by asterisks ("*"). From this point onward, all figures related to the goodness-of-fit analysis will follow this convention.



(A) (B)

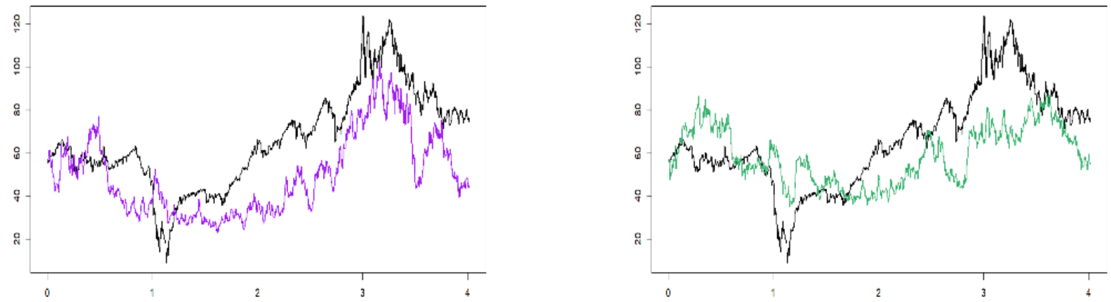
FIGURE 5. Monte Carlo simulations obtained from (a) the (3.1) GBM SDE model and (b) the (3.2) CIR SDE model.



(A) (B)

FIGURE 6. Confidence intervals of the solutions obtained from (a) the (3.1) GBM SDE model and (b) the (3.2) CIR SDE model.

Given that both models exhibit a statistically significant fit to the data, approximate solutions of the derived SDEs are computed to further examine their predictive capabilities. Prior to this, confidence intervals for the solutions of each SDE were constructed. In this context, the confidence intervals for the (3.1) GBM SDE, described in Section 2.2 and calculated using Eq. (2.3) and Eq. (2.4), are visualized in Figure 6(a). Similarly, the confidence intervals for the (3.2) CIR SDE, calculated using Eq. (2.6) and Eq. (2.7), are shown in Figure 6(b). In these figures, the WTI data is shown as a black line, the expected value trajectories of each SDE are depicted by blue dashed lines, and the confidence intervals are represented by red dashed lines. This graphical representation will be maintained in all subsequent figures presenting confidence interval results. All confidence intervals, EM solutions, and graphs were produced using RStudio. A notable finding is that the confidence interval of the (3.2) CIR model is narrower and aligns more closely with the actual WTI data compared to the (3.1) GBM model, suggesting a better fit. This conclusion is further supported by the chi-square test results. The chi-square statistic serves as a useful measure for model comparison. Here, the (3.2) CIR model yields a chi-square value of $\chi_2^2 = 229.9195$, which is lower than the value for the (3.1) GBM model, $\chi_1^2 = 247.5545$. The numerical solutions of both models were computed using the EM method. Figure 7 illustrates the resulting trajectories: the black line represents the observed WTI data, the purple line shows the GBM solution, and the green line denotes the CIR solution. While both trajectories remain within the expected confidence bounds (as seen in Figure 6), Figure 7 reveals that neither model adequately captures the abrupt fluctuations in the WTI prices, particularly the sharp spikes and drops. To address this limitation, a CP analysis was applied within the framework of both models to better capture sudden shifts in the data.



(A) (B)

FIGURE 7. A trajectory of the EM solutions obtained from (a) the (3.1) GBM SDE model and (b) the (3.2) CIR SDE model.

3.3. Change Point Estimation for WTI Dataset.

Two CPs were identified in the WTI dataset using Eq. (2.2) and Eq. (2.5). Using the QMLE method, CPs were estimated independently for the GBM and CIR models, and it was observed that both models detected the same CP positions. In both cases, these points were obtained by analyzing the first 35% and the last 35% of the dataset and were denoted as τ_1 and τ_2. Accordingly, τ_1 was found on May 14, 2020 (303rd data point), while τ_2 was identified on February 24, 2022 (750th data point). The first CP, τ_1, corresponds to the turning point after April 20, 2020, when the COVID-19 pandemic led to a sharp decline in crude oil prices, followed by a subsequent recovery. Similarly, τ_2 marks the moment when Russia's invasion of Ukraine triggered a rapid surge in oil prices. The alignment of these CPs with significant global events demonstrates that the CP analysis was correctly applied within the framework of the SDE models.

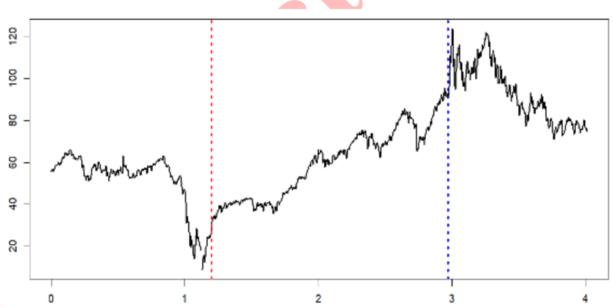


FIGURE 8. τ_1 and τ_2 CP statistics and the ranges $[0, \tau_1)$, $[\tau_1, \tau_2)$, and $[\tau_2, T]$ of the WTI dataset.

The WTI dataset was remodeled by taking into account the CPs τ_1 and τ_2 within the context of SDE modeling. For this purpose, the dataset was examined separately in the intervals $[0, \tau_1)$, $[\tau_1, \tau_2)$, and $[\tau_2, T]$. Figure 8 illustrates these segments, where the black line represents the WTI dataset, the red dashed line indicates τ_1, and the blue dashed line denotes τ_2.

3.4. SDE Modelling for WTI Dataset Considering the CP Estimation.

Following a similar approach to the previous analyses, the dataset was modeled separately for each interval identified in the preceding CP analysis using Eq. (2.2) and Eq. (2.5). The parameters for both models were then estimated using the QMLE method. The estimation results for all intervals are presented in Table 2.

The parameters obtained for each interval mentioned above were substituted into Eq. (2.2) and Eq. (2.5), respectively. Since the CPs τ_1 and τ_2 are considered in this equation, the resulting form is GBM SDE and CIR SDE.

TABLE 2. Parameter estimates for Eq. (2.2) and Eq. (2.5) across all identified intervals using the WTI dataset.

Interval	**Model/Parameter**	$\boldsymbol{\theta}_1$		$\boldsymbol{\theta}_2$		$\boldsymbol{\sigma}$	
		Estimate	Std. error	Estimate	Std. error	Estimate	Std. error
$[0, \tau_1)$	**GBM SDE**	0.4297	0.6915	–	–	0.9524	0.0308
	CIR SDE	203.2234	67.9702	4.2780	1.4942	4.4936	0.1484
$[\tau_1, \tau_2)$	**GBM SDE**	0.7925	0.3389	–	–	0.3789	0.0152
	CIR SDE	147.0962	69.5120	2.2070	1.3747	2.7048	0.1155
$[\tau_2, T]$	**GBM SDE**	-0.0859	0.4549	–	–	0.4691	0.0203
	CIR SDE	397.7344	301.5859	4.4581	3.2809	4.6828	0.2050

$$dX_t = \begin{cases} 0.4297\, X_t\, dt + 0.9524\, X_t\, dW_t, & X_0 = 55.76, \quad t \in [0, \tau_1), \\ 0.7925\, X_t\, dt + 0.3789\, X_t\, dW_t, & X_{\tau_1} = 27.40, \quad t \in [\tau_1, \tau_2), \\ -0.0859\, X_t\, dt + 0.4691\, X_t\, dW_t, & X_{\tau_2} = 92.77, \quad t \in [\tau_2, T]. \end{cases} \tag{3.3}$$

$$dX_t = \begin{cases} (203.2234 - 4.2780\, X_t)\, dt + 4.4936\, \sqrt{X_t}\, dW_t, & X_0 = 55.76, \quad t \in [0, \tau_1), \\ (147.0962 - 2.2070\, X_t)\, dt + 2.7048\, \sqrt{X_t}\, dW_t, & X_{\tau_1} = 27.40, \quad t \in [\tau_1, \tau_2), \\ (397.7344 - 4.4581\, X_t)\, dt + 4.6828\, \sqrt{X_t}\, dW_t, & X_{\tau_2} = 92.77, \quad t \in [\tau_2, T]. \end{cases} \tag{3.4}$$

Similar to the operations in the models of Eq. (3.1) and Eq. (3.2), the chi-square goodness of fit test was applied to assess the alignment between the model and the data for each region. The test was initially conducted for the models within the interval $[0, \tau_1)$.

3.4.1. *Model Fit Evaluation for WTI Dataset.*

Region 1: Interval $[0, \tau_1)$. In this interval, the compatibility of both the (3.3) GBM and (3.4) CIR SDE models with the WTI dataset was assessed using the chi-square goodness-of-fit test. Since there are $N = 302$ observations in the interval $[0, \tau_1)$, $M = 59$ Monte Carlo simulations were generated for each observation in both models, and the resulting stochastic paths were compared to the observed data. The significance level was set at $\alpha = 0.01$ for all tests conducted throughout the regime-wise analysis (i.e., in Regions 1 to 3).

For the (3.3) GBM model, the hypotheses were established as follows:

H_0 : There is no lack of fit between the (3.3) GBM SDE model and the WTI data for $t \in [0, \tau_1)$.

H_1 : There is a lack of fit between the (3.3) GBM SDE model and the WTI data for $t \in [0, \tau_1)$.

The chi-square test statistic was calculated as $\chi^2_3 = 84.0521$. With 57 degrees of freedom, the critical chi-square value is $\chi^2_{(57,0.01)} = 84.7328$. Since the test statistic is below the critical value, the null hypothesis is not rejected, indicating that the GBM model is compatible with the observed data in this region.

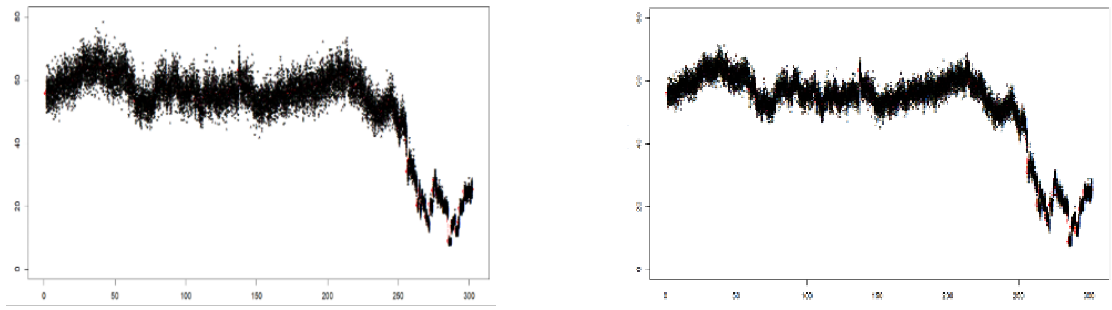
For the (3.4) CIR model, the hypotheses were similarly stated as:

H_0 : The (3.4) CIR SDE model fits the WTI data for $t \in [0, \tau_1)$,

H_1 : The (3.4) CIR SDE model does not fit the WTI data for $t \in [0, \tau_1)$.

The corresponding chi-square test statistic was computed as $\chi^2_6 = 74.5501$. With 56 degrees of freedom, the critical value is $\chi^2_{(56,0.01)} = 83.5134$. As $\chi^2_6 < \chi^2_{(56,0.01)}$, the CIR model is also considered statistically consistent with the WTI data in this region.

These results are demonstrated in Figure 9.



(A) (B)

FIGURE 9. Monte Carlo simulations obtained from (a) the (3.3) GBM SDE model and (b) the (3.4) CIR SDE model for $t \in [0, \tau_1)$.

In this interval, the results indicate that both models exhibit a good fit to the data; however, the (3.4) CIR model demonstrates a slightly better fit due to its mean-reverting characteristic, which aligns more closely with the underlying dynamics of the WTI data compared to the (3.3) GBM model.

Region 2: Interval $[\tau_1, \tau_2)$. During this period, as in Region 1, the goodness-of-fit of both the (3.3) GBM SDE model and the (3.4) CIR SDE model to the WTI dataset was assessed using the chi-square goodness-of-fit test. Within the interval $[\tau_1, \tau_2)$, there are $N = 447$ observations. For each observation, $M = 88$ Monte Carlo simulations were generated, and the resulting stochastic trajectories were compared with the actual data.

For the (3.3) GBM model, the hypotheses were formulated as follows:

H_0 : The (3.3) GBM SDE model is compatible with the WTI data for $t \in [0, \tau_1)$,

H_1 : The (3.3) GBM SDE model is not compatible with the WTI data for $t \in [0, \tau_1)$.

The test statistic was calculated as $\chi^2_4 = 87.6278$. With 86 degrees of freedom, the critical value is $\chi^2_{(86,0.01)} = 119.4139$. Since the test statistic is below the critical threshold, the null hypothesis could not be rejected, indicating that the GBM model is consistent with the observed data in this region.

Similarly, for the (3.4) CIR model, the hypotheses were:

H_0 : The (3.4) CIR SDE model is compatible with the WTI data for $t \in [0, \tau_1)$,

H_1 : The (3.4) CIR SDE model is not compatible with the WTI data for $t \in [0, \tau_1)$.

The corresponding chi-square test statistic was $\chi^2_7 = 86.6255$. With 85 degrees of freedom, the critical value is $\chi^2_{(85,0.01)} = 118.2357$. Since $\chi^2_7 < \chi^2_{(85,0.01)}$, the CIR model also demonstrated statistical consistency with the empirical data in this region. These findings are showed in Figure 10.

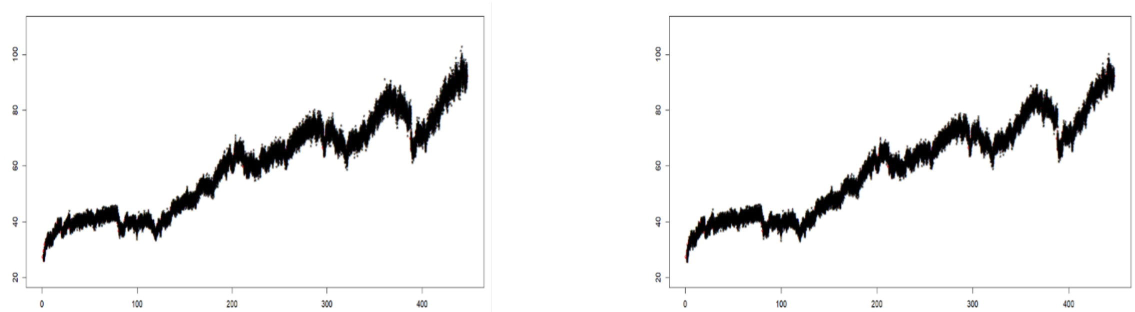
Although both models provide a good fit within this regime, the (3.4) CIR model, owing to its mean-reverting nature and its lower test statistic ($\chi^2_7 < \chi^2_4$), exhibits a closer alignment with the fundamental dynamics of the WTI dataset compared to the (3.3) GBM model.

Region 3: Interval $[\tau_2, T]$. In this final interval, as with the previous regions, the fit of both the (3.3) GBM SDE model and the (3.4) CIR SDE model to the WTI data was evaluated using the chi-square goodness-of-fit test. For the interval $[\tau_2, T]$, there are $N = 263$ observations, and for each observation, $M = 51$ Monte Carlo simulations were generated to compare the resulting stochastic trajectories with the actual data.

For the (3.3) GBM model, the following hypotheses were tested:

H_0 : The (3.3) GBM SDE model is compatible with the WTI data for $t \in [0, \tau_1)$,

H_1 : The (3.3) GBM SDE model is not compatible with the WTI data for $t \in [0, \tau_1)$.



(A) (B)

FIGURE 10. Monte Carlo simulations obtained from (a) the (3.3) GBM SDE model and (b) the (3.4) CIR SDE model for $t \in [\tau_1, \tau_2)$.

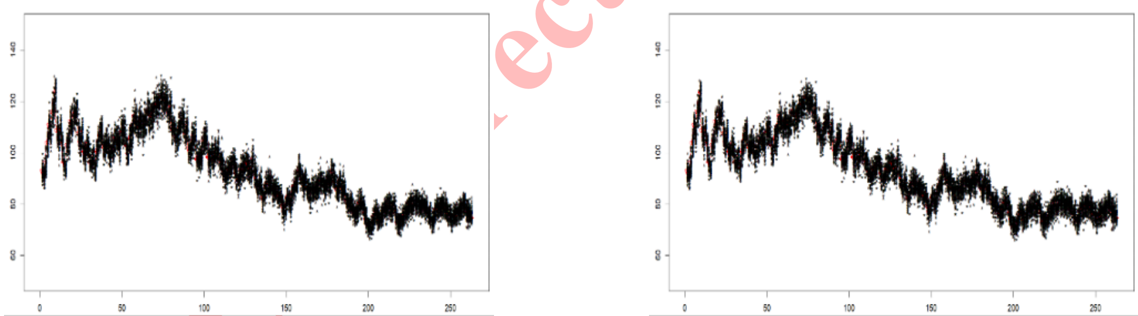
The test statistic was found to be $\chi_5^2 = 38.976$. With 49 degrees of freedom, the critical value is $\chi^2_{(49,0.01)} = 74.9195$. Since the test statistic is less than the critical value, the null hypothesis cannot be rejected, indicating that the (3.3) GBM model is compatible with the observed WTI data in this region.

Similarly, for the (3.4) CIR model, the hypotheses were as follows:

H_0 : The (3.4) CIR SDE model is compatible with the WTI data for $t \in [0, \tau_1)$,

H_1 : The (3.4) CIR SDE model is not compatible with the WTI data for $t \in [0, \tau_1)$.

The chi-square test statistic was computed as $\chi_8^2 = 51.8623$. With 48 degrees of freedom, the critical value is $\chi^2_{(48,0.01)} = 73.6826$. Since $\chi_8^2 < \chi^2_{(48,0.01)}$, the (3.4) CIR model also showed good statistical consistency with the WTI data in this region. These results are illustrated in Figure 11.



(A) (B)

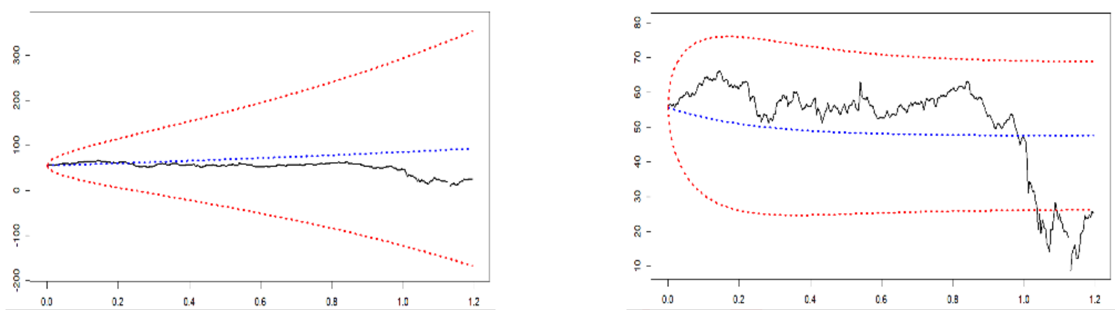
FIGURE 11. Monte Carlo simulations obtained from (a) the (3.3) GBM SDE model and (b) the (3.4) CIR SDE model for $t \in [\tau_2, T]$.

Although both models exhibit a good fit, the GBM model outperforms the CIR model in this interval, as reflected by its lower chi-square statistic.

Given the compatibility of both the (3.3) GBM SDE model and the (3.4) CIR SDE model with the WTI data, the approximate solutions of these models were computed. confidence intervals and EM trajectories for the solutions of both models were calculated separately for each sub-equation, as detailed in the following sections.

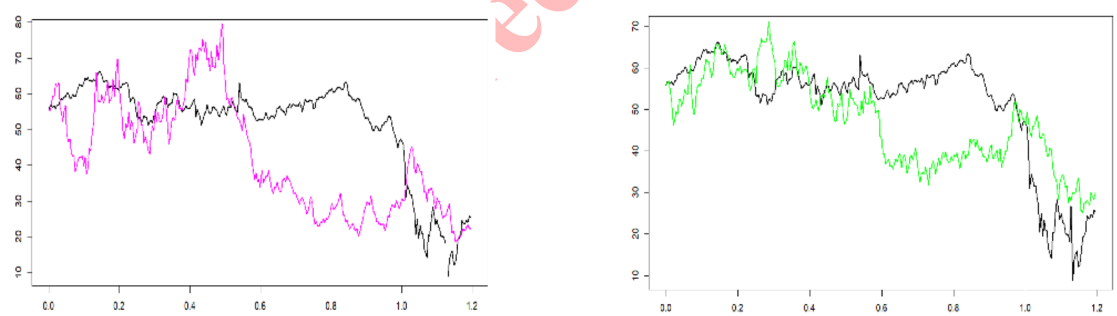
3.4.2. *EM Solutions for the (3.3) GBM SDE Model and the (3.4) CIR SDE Model.* Similar to the analysis in Section 3.2, the EM solutions for both models will be examined region by region, providing a detailed breakdown of the results for each model within the specified intervals.

Region 1: Interval $[0, \tau_1)$. Before proceeding to the EM solutions of the corresponding SDEs, confidence intervals for the solutions of each model were constructed, following a similar approach to that presented in Section 3.2. Accordingly, the confidence intervals, as described in Section 2.2, for the (3.3) GBM SDE were calculated using Eq. (2.3) and Eq. (2.4) and are visualized in Figure 12(a); likewise, those for the Eq. (3.4) CIR SDE were obtained using Eq. (2.6) and Eq. (2.7) and are shown in Figure 12(b), which pertains to this region. For $t \in [0, \tau_1)$, the numerical solutions of the (3.3) GBM and (3.4) CIR SDEs were obtained using the EM method. Representative trajectories corresponding to these solutions are illustrated in Figures 13(a) and 13(b), where the black curves represent the WTI data in the specified interval, while the pink and green curves represent the EM solutions of the GBM and CIR models, respectively. As observed, these trajectories remain within the confidence intervals previously presented in Figures 12(a) and 12(b), respectively. From this point onward, the same color scheme will be consistently applied to the curves corresponding to the same model in subsequent regions (Region 2 and Region 3).



(A) (B)

FIGURE 12. Confidence intervals for the solutions obtained from (a) the (3.3) GBM SDE and (b) the (3.4) CIR SDE in $[0, \tau_1)$.

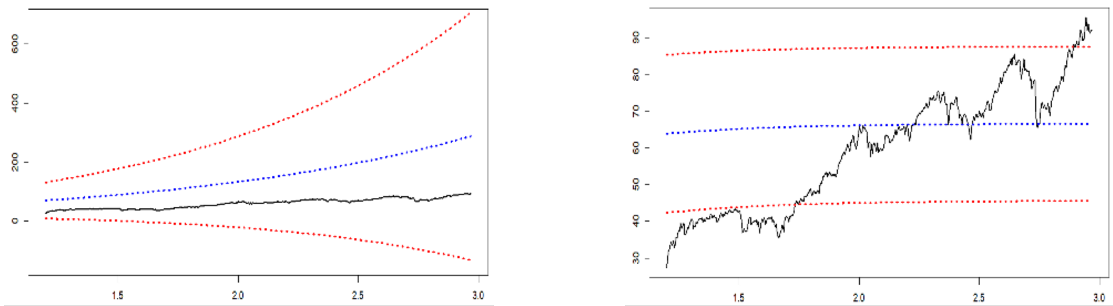


(A) (B)

FIGURE 13. A trajectory of the EM solutions obtained from (a) the (3.3) GBM SDE and (b) the (3.4) CIR SDE in $[0, \tau_1)$.

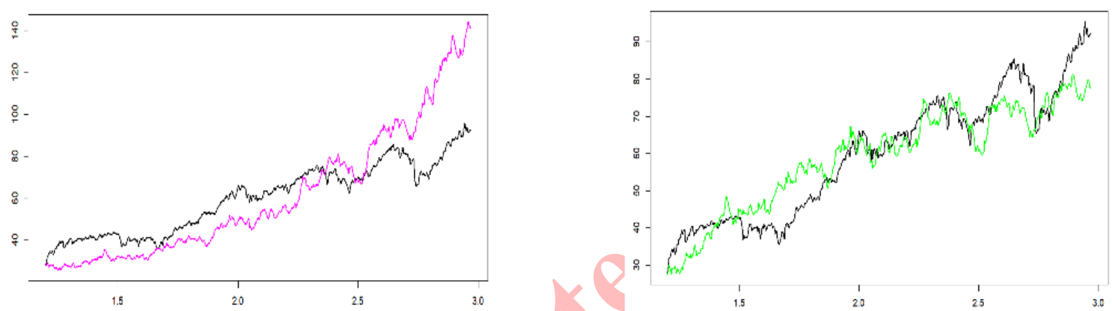
Region 2: Interval $[\tau_1, \tau_2)$. Before proceeding to the EM solutions of the SDEs in this interval, confidence intervals for each model were constructed in parallel with the approach adopted in Region 1. The confidence intervals for the (3.3) GBM SDE are illustrated in Figure 14(a), while those for the (3.4) CIR SDE are depicted in Figure 14(b).

For the interval $[\tau_1, \tau_2)$, numerical solutions for the (3.3) GBM and (3.4) CIR SDEs were obtained using the EM method. The corresponding representative trajectories of these solutions are presented in Figures 15(a) and 15(b). As observed, the computed trajectories lie within the previously established confidence intervals shown in Figures 14(a) and 14(b), respectively.



(A) (B)

Figure 14. Confidence intervals for the solutions obtained from (a) the (3.3) GBM SDE and (b) the (3.4) CIR SDE in $[\tau_1, \tau_2)$.



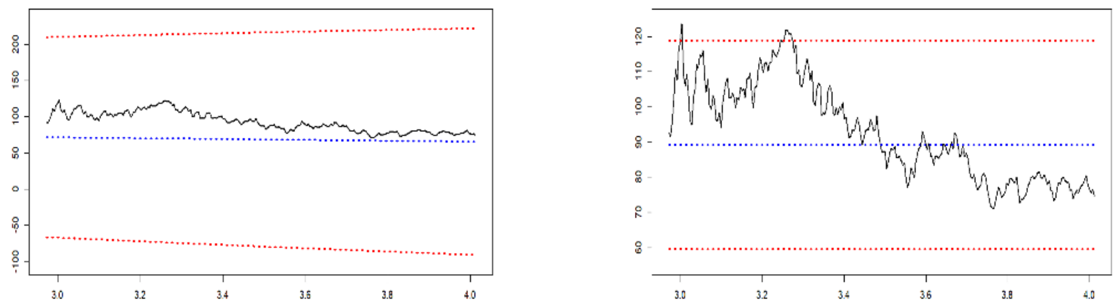
(A) (B)

Figure 15. A trajectory of the EM solutions obtained from (a) the (3.3) GBM SDE and (b) the (3.4) CIR SDE in $[\tau_1, \tau_2)$.

Region 3: Interval $[\tau_2, T]$. In this final interval, confidence intervals were constructed for each model following the methodology outlined earlier in Region 1. Figure 16(a) displays the interval estimates for the (3.3) GBM SDE, while Figure 16(b) illustrates those for the (3.4) CIR SDE. Subsequently, the EM method was employed to compute numerical solutions of both models over $[\tau_2, T]$, and their representative trajectories are shown in Figures 17(a) and 17(b). Notably, these trajectories remain consistent with the confidence bounds previously delineated in Figures 16(a) and 16(b). For a holistic assessment of model performance across the entire time domain, Figure 18 presents the EM-based numerical solutions of the (3.3) GBM SDE and the (3.4) CIR SDE in an integrated manner. The vertical blue dashed lines denote the CPs τ_1 and τ_2, thereby distinctly demarcating the transitions among the three regimes.

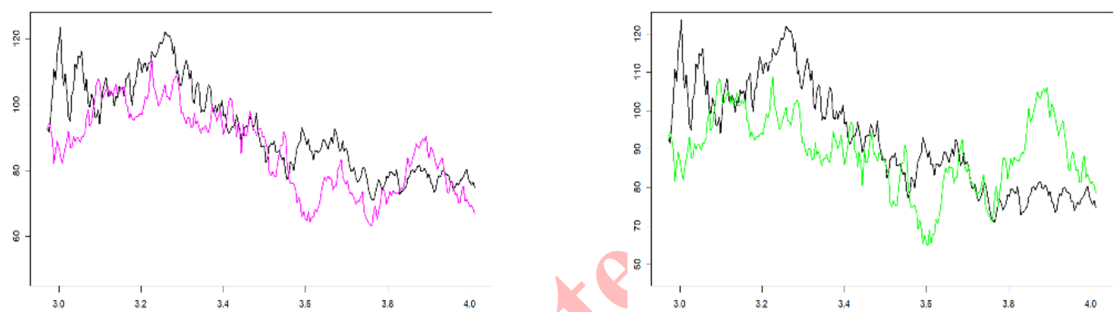
Following the evaluation of the results obtained from the EM-based solutions, the process of model selection for the WTI dataset will commence. In this phase, the performance of different models will be compared, and the most appropriate model will be selected based on statistical criteria.

3.5. **Model Selection for WTI Dataset.** The (3.1) GBM and (3.2) CIR SDE models obtained in Section 3.2 are compared in this section based on the AIC and BIC criteria. The (3.2) CIR SDE model was identified as the most suitable model, demonstrating superior performance in explaining the WTI dataset, as it exhibited lower AIC and BIC values across both criteria. In Subsection 3.4, similar comparisons were conducted for the models, with CP values in the divided intervals taken into account. This led to the selection of the most appropriate model for each corresponding interval. Specifically, the (3.4) CIR SDE model was preferred for the first two regions, while the (3.3)



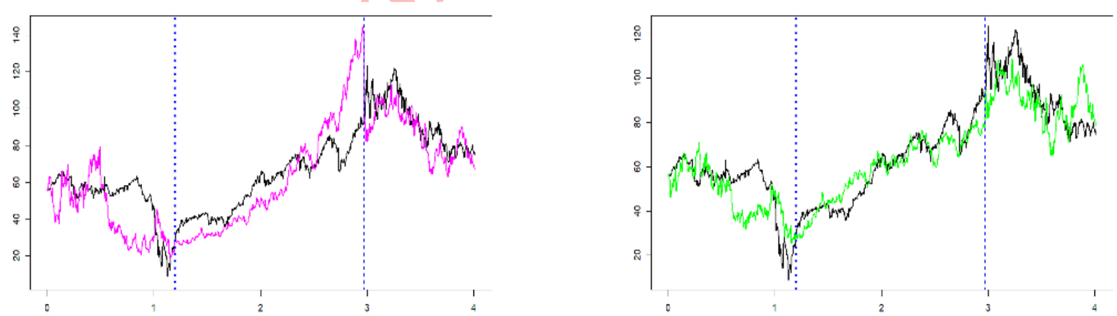
(A) (B)

FIGURE 16. Confidence intervals for the solutions obtained from (a) the (3.3) GBM SDE and (b) the (3.4) CIR SDE in $[\tau_2, T]$.



(A) (B)

FIGURE 17. A trajectory of the EM solutions obtained from (a) the (3.3) GBM SDE and (b) the (3.4) CIR SDE in $[\tau_2, T]$.



(A) (B)

FIGURE 18. A trajectory of the EM solutions obtained from (a) the (3.3) GBM SDE model and (b) the (3.4) CIR SDE model for $t \in [0, T]$.

GBM SDE model was chosen for the third region, owing to their lower AIC and BIC values in their respective ranges. The results are presented in Table 3 and Table 4, respectively.

Table 3. AIC and BIC values of (3.1) GBM SDE and (3.2) CIR SDE established for WTI dataset.

Model	AIC	BIC
(3.1) GBM SDE	4950.618	4960.453
(3.2) CIR SDE	4330.257	4345.010

In the following section, a hybrid model is developed by combining the region-specific models that were identified as the best performers based on model selection criteria. This hybrid approach leverages the strengths of different stochastic models tailored to the distinct statistical characteristics observed in each regime, thereby aiming to enhance the overall modeling accuracy for the WTI dataset.

Table 4. AIC and BIC values of (3.3) GBM SDE and (3.4) CIR SDE established for WTI data in the divided intervals by taking into account the CP values.

Model/Interval	$t \in [0, \tau_1)$		$t \in [\tau_1, \tau_2)$		$t \in [\tau_2, T]$	
	AIC	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
(3.3) GBM SDE	2329.240	2337.579	1012.1363	1019.641	1300.005	1307.187
(3.4) CIR SDE	1972.711	1985.220	993.7283	1004.986	1319.080	1329.853

3.6. **Construction of a Hybrid SDE Model Based on CP Estimation for the WTI Dataset.** Building upon the model selection results discussed in Section 3.5, this section introduces a Hybrid SDE model tailored to the WTI dataset. The proposed hybrid structure integrates the most appropriate models identified for each regime, as determined by CP estimation, across the intervals $[0, \tau_1)$, $[\tau_1, \tau_2)$ and $[\tau_2, T]$. Specifically, the CIR model was selected for the first and second intervals, while the GBM model demonstrated superior performance in the third interval. Accordingly, the hybrid model is constructed by combining the EM-based numerical solutions of the CIR and GBM models over their respective intervals. This approach not only captures the dynamic behavior of the WTI process more effectively but also offers a refined modeling framework that adapts to structural changes in the data. Therefore, the hybrid model that best captures the structural dynamics of the WTI dataset is formally expressed in Equation 3.5. The results are presented in Figure 19, where the black line represents the WTI observation data, the red line corresponds to the EM numerical solution, and the blue dashed lines indicate the CPs τ_1 and τ_2, respectively.

$$dX_t = \begin{cases} (203.2234 - 4.2780 X_t)\, dt + 4.4936\sqrt{X_t}\, dW_t, & X_0 = 55.76, \quad t \in [0, \tau_1), \\ (147.0962 - 2.2070 X_t)\, dt + 2.7048\sqrt{X_t}\, dW_t, & X_{\tau_1} = 27.40, \quad t \in [\tau_1, \tau_2), \\ -0.0859 X_t\, dt + 0.4691 X_t\, dW_t, & X_{\tau_2} = 92.77, \quad t \in [\tau_2, T]. \end{cases} \tag{3.5}$$

3.7. **Model Performance Evaluation for WTI Dataset.** To assess the predictive performance of the stochastic models developed for the WTI dataset, two widely accepted error metrics were employed: the RMSE and the MAPE. The corresponding results for each model are reported in Table 5.

According to the MAPE criterion, the (3.5) Hybrid SDE and the (3.4) CIR SDE models fall within the range $[0.1, 0.2)$, indicating good forecast accuracy. Meanwhile, the (3.3) GBM SDE, the (3.2) CIR SDE, and the (3.1) GBM SDE models fall within the $[0.2, 0.5)$ range, which is generally interpreted as reasonable predictive accuracy in the literature.

Notably, the (3.5) Hybrid SDE model outperforms all others in terms of both RMSE and MAPE, establishing it as the most effective model for the WTI dataset. These results emphasize the utility of CP estimation in improving SDE-based modeling frameworks, especially for time series data characterized by regime shifts and structural volatility.

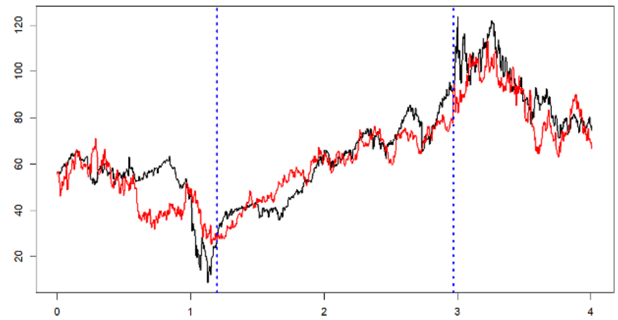


FIGURE 19. A trajectory of the EM numerical solution of the Hybrid SDE model over the full time horizon.

TABLE 5. RMSE and MAPE values of (3.1) GBM SDE, (3.2) CIR SDE, (3.3) GBM SDE, (3.4) CIR SDE and (3.5) Hybrid SDE established for WTI data.

Model	RMSE	MAPE
(3.1) GBM SDE	21.3624	0.3101
(3.2) CIR SDE	19.7637	0.2754
(3.3) GBM SDE	16.4034	0.2138
(3.4) CIR SDE	10.7092	0.1491
(3.5) Hybrid SDE	9.9439	0.1425

4. Discussion and Conclusion

In modeling real-world systems, researchers often face limitations either due to insufficient information about system dynamics or because of the complexity that renders deterministic descriptions inadequate. In such cases, SDEs provide a powerful framework by incorporating the intrinsic randomness observed in many natural and economic processes. This study employed SDE modeling to analyze the daily closing prices of WTI crude oil from March 1, 2019, to March 13, 2023—a period marked by unprecedented global volatility.

Initially, GBM and CIR SDEs were calibrated using the QMLE method, and their compatibility with the dataset was validated via a chi-square goodness-of-fit test specifically adapted for SDEs. The numerical solutions of models were obtained through the EM method, and their trajectories remained within constructed confidence intervals, supporting their robustness.

Recognizing the presence of abrupt structural shifts in oil prices, two significant CPs were identified: the first on May 14, 2020, coinciding with the global oil price collapse during the COVID-19 pandemic, and the second on February 24, 2022, following the geopolitical escalation due to Russia's invasion of Ukraine. These breaks represent critical periods of volatility, and their inclusion significantly enhanced model precision. This structural break detection clearly demonstrates the model's capability to capture crisis-induced volatility regimes—an essential feature for financial modeling under real-world conditions.

The dataset was partitioned into three regimes, and regime-specific models were constructed. The hybrid model, built from the best-performing models in each interval (CIR–CIR–GBM), yielded the lowest RMSE (9.9439) and MAPE (0.1425), establishing itself as the most accurate among the candidate models. The results indicate that incorporating CP estimation into SDE modeling offers substantial gains in model performance and interpretability, especially for financial time series exhibiting non-stationarity and abrupt transitions.

In conclusion, accurately modeling fluctuating crude oil prices enables overcoming the challenges in predicting raw material supply and supports more efficient decision-making to optimize profit margins of refinery companies. Therefore, considering CPs when working with datasets that exhibit sudden changes has a positive impact on the performance of the created model. CP estimation not only strengthens model reliability but also supports timely policy and investment decisions in energy markets under crisis conditions.

Moreover, while this study focuses on the WTI dataset, the methodology is applicable across various scientific and financial domains where structural breaks and stochastic volatility are common. Modeling such fluctuations and predicting future data behavior with the help of historical information is of great importance in many disciplines, including economics, finance, and engineering. The aim of this work is to contribute a foundation for advancing the modeling of more complex and adaptive systems that combine CP detection with stochastic frameworks, facilitating more qualitative modifications in model design and the incorporation of innovative modeling ideas.

References

[1] W. F. Agustini, I. R. Affianti, and E. R. M. Putri, *Stock price prediction using geometric Brownian motion*, J. Phys.: Conf. Ser., *974* (2018), 012047.

[2] D. Ahmadian and O. Farkhondeh Rouz, *Exponential mean-square stability of numerical solutions for stochastic delay integro-differential equations with Poisson jump*, J. Inequal. Appl., *1* (2020), Article 186.

[3] D. Ahmadian, L. V. Ballestra, and N. Karimi, *An extremely efficient numerical method for pricing options in the Black–Scholes model with jumps*, Math. Methods Appl. Sci., *44*(2) (2021), 1843–1862.

[4] H. Akaike, *A new look at the statistical model identification*, IEEE Trans. Automat. Control, *19*(6) (1974), 716–723.

[5] E. Allen, *Modeling with Itô Stochastic Differential Equations*, Springer, New York, 2007.

[6] M. E. H. Arouri, J. Jouini, and D. K. Nguyen, *On the impacts of oil price fluctuations on European equity markets: volatility spillover and hedging effectiveness*, Energy Econ., *34*(2) (2012), 611–617.

[7] J. Bak, H. A. Nielsen, and H. Madsen, *Goodness of fit of stochastic differential equations*, in *21th Symposium i Anvendt Statistik*, Denmark, 1999, 341–346.

[8] L. V. Ballestra, D. Ahmadian, and O. Farkhondeh Rouz, *Stability analysis of split-step θ-Milstein method for a class of n-dimensional stochastic differential equations*, Appl. Math. Comput., *348* (2019), 280–296.

[9] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Polit. Econ., *81*(3) (1973), 637–654.

[10] T. Chai and R. R. Draxler, *Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature*, Geosci. Model Dev., *7*(3) (2014), 1247–1250.

[11] J. Chen and A. K. Gupta, *Parametric statistical change point analysis*, Birkhäuser, 2013.

[12] L. Cheng, L. T. Zhao, Q. Y. Gu, Z. T. Liu, and Z. Y. Li, *Events enhance crude oil price risk measurement: Embedding breakpoints in GARCH-VaR*, Expert Syst. Appl., *296* (2026), 128863.

[13] S. Corbet, Y. G. Hou, Y. Hu, and L. Oxley, *An analysis of investor behavior and information flows surrounding the negative WTI oil price futures event*, Energy Econ., *104* (2021), 105589.

[14] G. Cortazar and E. S. Schwartz, *Implementing a stochastic model for oil futures prices*, Energy Econ., *25*(3) (2003), 215–238.

[15] G. Cortazar, C. Milla, and F. Severino, *A multicommodity model of futures prices: Using futures prices of one commodity to estimate the stochastic process of another*, J. Futures Markets, *28*(6) (2008), 537–560.

[16] J. C. Cox, J. E. Ingersoll, and S. A. Ross, *A theory of the term structure of interest rates*, Econometrica, *53*(2) (1985), 385–407.

[17] A. De Gregorio and S. M. Iacus, *Least squares volatility change point estimation for partially observed diffusion processes*, Commun. Stat. Theory Methods, *37*(15) (2008), 2342–2357.

[18] D. Filipović, *Term-structure models: A graduate course*, Springer, Berlin–Heidelberg, 2009.

[19] R. Gibson and E. S. Schwartz, *Stochastic convenience yield and the pricing of oil contingent claims*, J. Finance, *45*(3) (1990), 959–976.

[20] R. J. Hyndman and A. B. Koehler, *Another look at measures of forecast accuracy*, Int. J. Forecast., *22* (2006), 679–688.

[21] S. M. Iacus, *Simulation and inference for stochastic differential equations with R examples*, Springer, 2008.

[22] S. M. Iacus and N. Yoshida, *Estimation for the change point of volatility in a stochastic differential equation*, Stochastic Process. Appl., *122*(3) (2012), 1068–1092.

[23] S. M. Iacus and N. Yoshida, *Simulation and inference for stochastic processes with YUIMA: A comprehensive R framework for SDEs and other stochastic processes. Use R*, Springer, 2018.
[24] N. İnce and A. Shamilov, *An application of new method to obtain probability density function of solution of stochastic differential equations*, Appl. Math. Nonlinear Sci., *5*(1) (2020), 337–348.
[25] N. İnce and S. Şentürk, *A new method for modelling the stochastic differential equations*, Hacet. J. Math. Stat., *54*(6) (2025), 2380–2398.
[26] H. Jahanshahi, S. Uzun, S. Kaçar, Q. Yao, and M. O. Alassafi, *Artificial intelligence-based prediction of crude oil prices using multiple features under the effect of Russia–Ukraine War and COVID-19 pandemic*, Mathematics, *10*(22) (2022), 4361.
[27] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, 2nd ed., Springer, 1991.
[28] M. Kessler, *Estimation of an ergodic diffusion from discrete observations*, Scand. J. Stat., *24*(2) (1997), 211–229.
[29] P. E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Springer, 1992.
[30] Y. A. Kutoyants, *Statistical inference for ergodic diffusion processes*, Springer, 2013.
[31] J. Lee, J. A. List, and M. C. Strazicich, *Non-renewable resource prices: deterministic or stochastic trends?*, J. Environ. Econ. Manag., *51*(3) (2006), 354–370.
[32] H. Manner, G. Rodriguez, and F. Stöckler, *A changepoint analysis of exchange rate and commodity price risks for Latin American stock markets*, Int. Rev. Econ. Finance, *89* (2024), 1385–1403.
[33] R. E. Mbah and D. Wasum, *Russian–Ukraine 2022 War: a review of the economic impact of Russian–Ukraine crisis on the USA, UK, Canada, and Europe*, Adv. Soc. Sci. Res. J., *9*(3) (2022), 144–153.
[34] R. C. Merton, *Theory of rational option pricing*, Bell J. Econ. Manag. Sci., *4*(1) (1973), 141–183.
[35] T. Mikosch, *Elementary stochastic calculus with finance in view*, World Scientific, 1998.
[36] P. K. Narayan and S. Narayan, *Modelling oil price volatility*, Energy Policy, *35*(12) (2007), 6549–6553.
[37] I. Negri and Y. Nishiyama, *Z-process method for change point problems with applications to discretely observed diffusion processes*, Stat. Methods Appl., *26*(2) (2017), 231–250.
[38] B. Øksendal, *Stochastic differential equations: An introduction with applications*, Springer, 2000.
[39] R. S. Pindyck, *The long-run evolution of energy prices*, Energy J., *20*(2) (1999), 1–27.
[40] I. A. Raifu, T. T. Kumeka, and O. M. David-Wayas, *Testing multiple structural breaks in the oil price–stock price nexus in Asian oil-importing countries during the Russia–Ukraine War*, Asian Econ. Lett., *4*(4) (2023), 2–8.
[41] A. Rathinasamy, D. Ahmadian, and P. Nair, *Second-order balanced stochastic Runge–Kutta methods with multi-dimensional studies*, J. Comput. Appl. Math., *377* (2020), 112890.
[42] A. A. Salisu and I. O. Fasanya, *Modelling oil price volatility with structural breaks*, Energy Policy, *52* (2013), 554–562.
[43] G. Schwarz, *Estimating the dimension of a model*, Ann. Statist., *6*(2) (1978), 461–464.
[44] E. S. Schwartz, *The stochastic behavior of commodity prices: Implications for valuation and hedging*, J. Finance, *52*(3) (1997), 923–973.
[45] Y. Tonaki, Y. Kaino, and M. Uchida, *Estimation for change point of discretely observed ergodic diffusion processes*, Scand. J. Stat., *50*(1) (2023), 142–183.
[46] U.S. Energy Information Administration (EIA), *West Texas Intermediate (WTI) crude oil*, 2020. Available at: https://www.eia.gov/dnav/pet/hist/RWTCD.htm.