

Forecasting value-at-risk of crude oil futures using a hybrid ARIMA-SVR-POT model

Original

Forecasting value-at-risk of crude oil futures using a hybrid ARIMA-SVR-POT model / Zhang, C., Zhou, X.. - In: HELIYON. - ISSN 2405-8440. - 10:1(2024), pp. 1-14. [10.1016/j.heliyon.2023.e23358]

Availability:

This version is available at: 11583/3015850 since: 2026-09-24T01:58:52Z

Publisher:

Elsevier

Published

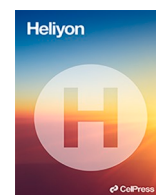
DOI:10.1016/j.heliyon.2023.e23358

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)



Forecasting value-at-risk of crude oil futures using a hybrid ARIMA-SVR-POT model

Chen Zhang, Xinmiao Zhou^{*}

Business School, Ningbo University, 315211, Ningbo, China

ARTICLE INFO

Keywords:

Forecasting risk

WTI

Value at risk

ARIMA-SVR-POT

Kupiec-test

ABSTRACT

Forecasting the value at risk (VaR) of crude oil futures can be a challenging task for investors due to the high volatility of these prices. It is crucial to describe the return in the tail distribution, as extreme values can trigger larger price fluctuations and market risks. In this study, we proposed a hybrid model, ARIMA-SVR-POT, which uses a combination of the autoregressive integrated moving average (ARIMA), support vector regression (SVR), and peak over threshold (POT) method from the extreme value theory. We compared the performance of our hybrid model with three other models, namely ARIMA-EGARCH, ARIMA-SVR, and ARIMA-EGARCH-POT. We demonstrated the effectiveness of our model using crude oil WTI Futures as a sample from June 23, 2016, to September 30, 2022. Our findings show that the ARIMA-SVR-POT hybrid model provides accurate predictions of the returns and volatility. Furthermore, the model performs exceptionally well in capturing the extreme tail of returns and outperforms the other models. We also conducted back-testing in the proposed model and the results show that the ARIMA-SVR-POT model passed the Kupiec test at confidence levels of 95 %, 99 %, 99.5 %, and 99.9 %. Our proposed model provides a more precise reflection of potential losses when estimating VaR. The predicted loss probability is closer to the actual loss occurrence probability, indicating superior performance compared to traditional statistical models, which enhanced crude oil risk management tools and suggested effective measures to manage market risks.

1. Introduction

The role of crude oil in the global economy is crucial as it is one of the most important energy resources worldwide. It provides the foundation for economic and industrial activities in a country or region. With the growing trends of globalization, international trade, and investment, the financial aspects of crude oil have become increasingly significant. The development of the financial derivatives market has made trading crude oil futures and options more active, attracting more investors and financial institutions to participate in the crude oil market. However, the fluctuation of crude oil prices has hindered enterprise investment, increased financing costs, and negatively impacted expected cash flows, and market volatility [1]. This uncertainty poses challenges for investors. Additionally, income redistribution between oil-exporting and oil-importing countries affects inflation and real interest rate expectations [2,3], which has a significant impact on economic operations and poses a threat to national energy security. As a result, investors and policymakers are increasingly concerned about market risks from crude oil price volatility.

Understanding the factors that affect energy prices is crucial because they depend on the balance between supply and demand. Any

^{*} Corresponding author.

E-mail address: zhouxinmiao@nbu.edu.cn (X. Zhou).

change in energy demand can impact the supply and demand ratio, as well as the production levels in producing countries. Several factors significantly impact crude oil prices, such as geopolitical conflicts, policy adjustments, technological advancements, and market expectations [4]. Factors like market and investor sentiment can influence market participants' expectations, leading to investment behavior and market volatility. Accurately predicting such expectations and behavioral changes adds uncertainty to crude oil price forecasts, making it difficult to rely solely on traditional technical analysis methods. Therefore, it is necessary to explore other ways to assess the market risk that investors face due to crude oil price volatility. Value at Risk (VaR) [5] is a risk measure used to determine the maximum potential loss that a financial portfolio or asset may face over a given period. VaR can help investors and traders develop effective investment strategies and assess risk tolerance. Hence, it is essential to understand the VaR of crude oil for macroeconomic policy-making. Using VaR is a promising way to measure the market risk of crude oil prices.

The main objective of this study is to accurately identify market risks that arise due to fluctuations in crude oil prices and take appropriate measures to mitigate them. By analyzing the oil market trends, effective risk management can not only reduce potential losses for oil investors by providing them with smart investment strategies but also help businesses make accurate market predictions and formulate relevant strategic plans. This can help companies control their production costs better and enhance their competitiveness in the market. Furthermore, governments can assess and respond to possible supply disruptions and instability more efficiently by devising appropriate energy security strategies to ensure energy supply reliability. To calculate the VaR, a proper model is needed. This study aims to improve traditional measurement methods by combining them with the machine learning method to propose a hybrid model for measuring the value at risk of crude oil. The empirical analysis provides evidence for the effectiveness of the model. The main contributions of the study are as follows: On one hand, the paper proposes a hybrid model, ARIMA-SVR-POT, which successfully captures market risks associated with fluctuations in crude oil futures prices. On the other hand, the study verifies the superiority of the proposed model in handling extreme events compared to the models in the current literature by using the backtest verification method. Therefore, introducing this model provides a powerful risk management tool for players in the real-world crude oil market. The following sections are arranged: Section 2 introduces the relevant research, section 3 explains the construction of the hybrid model and VaR calculations, section 4 presents the empirical results and analysis, and the last section concludes the study.

2. Related research

The current research on the crude oil market mainly focuses on predicting value-at-risk (VaR) using fundamental information. This highlights the significance of crude oil as a vital component of modern society and the economy. Yin and Wang [6] predicted the skewness risk of crude oil using OLS and revealed a stronger prediction of oil risk during economic recessions. Ji and Guo [7] studied the effect of oil-related events on global oil prices using event study methodology and AR-GARCH models. They discovered a significant negative correlation between the global financial crisis and oil price returns. Bonaccolto [8] examined the causal relationship between the European Monetary Union (EMU), economic policy uncertainty (EPU), oil price returns, and risk of oil. The study found that policy uncertainty is not always positively correlated with oil price volatility, but uncertainty in economic indicators is related to oil price fluctuations during difficult market periods. Ji and Liu [9] conducted a research on crude oil market risk using the time-varying copula approach. They measured Delta conditional value-at-risk (ΔCoVaR) to explore the impact of uncertainty on energy prices. The study revealed that economic policy uncertainty doesn't have a significant effect on energy returns, including clean energy and oil returns. However, financial and energy uncertainty have a higher impact on risk. Liang and Wei [10] conducted a study to determine if crude oil price volatility can be predicted using the global economic policy uncertainty (GEPU) and the Economic Policy Uncertainty Index (EMV) of the U.S. stock market. Their findings showed that GEPU and EMV have a significant impact on predicting oil market volatility. The crude oil market is a highly interconnected market within the global economy, and its prices are influenced by several other economic fields. Monetary and economic policies, such as interest rates, exchange rates, and supply, affect crude oil prices. However, the time difference between implementing policies and their impact on crude oil prices provides people with solutions and time to cope. Therefore, further research is necessary to understand the effect of unexpected events on the price of crude oil.

The spillover effects of crude oil are also an important topic of concern. For instance, there are risks associated with spillover effects between oil prices and exchange rates [11] and macroeconomic fluctuations and crude oil prices [12]. Other studies [13–15] looked into the asymmetric relationship between extreme oil price movements and corporate earnings, suggesting that there are risk spillover effects between oil returns and financial factors. Additionally, investor sentiment is another factor that influences crude oil market risk [16]. Researchers have used various GARCH methods to estimate the volatility and potential asymmetry of crude oil futures [17–19]. Christoffersen and Peter [20] corrected the defect of allowing time-varying jump intensity for crude oil derivatives using large sample data. Ma and Wahab et al. [21] studied how the Economic Policy Uncertainty (EPU) Index can provide greater economic value in the crude oil futures market. Kang and Boda et al. [22] evaluated the economic factors influencing crude oil futures volatility. Wang et al. [23] used the GARCH-EVT-VaR model and continuous wavelet transform to study the nonlinear oil-gold relationship. Fakhfekh and Jeribi [24] investigated cryptocurrency price volatility using GARCH models with different error distributions. In addition, there is the empirical analysis by Athari et al. [25–27] in energy asset management. These literatures provide support for the volatility modeling of crude oil prices and enriches the research on crude oil risks. However, the risk spillover effect of extreme event shocks is insufficient in theory.

Furthermore, Various studies have explored the ability of macroeconomic and financial variables to predict crude oil prices, specifically in the areas of exchange rate volatility [28], news sentiment [29], and geopolitics [30]. Researchers have also started to use mixed combination models to integrate the performance of various models. For example, Shah and Khan et al. [31] used a combination of business simulators, and probabilistic methods to predict oil production increases and associated risks. Alrweili and Fawzy [32] utilized ARIMA-ANN to forecast oil prices, while Liu and Zhang et al. [33] used the GARCH-MIDAS hybrid model to assess the impact

of different Economic Policy Uncertainty (EPU) indices on the volatility of European Union Allowances (EUA) futures prices. Abdollahi et al. [34] created a mixed model incorporating an Adaptive Neuro-Fuzzy Inference System (ANFIS), Autoregressive Fractionally Integrated Moving Average (ARFIMA), and Markov switching models to predict oil prices. Lin et al. [35] employed signal denoising techniques and empirical mode decomposition to transform denoised prices into different Intrinsic Mode Functions (IMFs) and used the long-memory GARCH-M model to forecast different IMFs and residuals. Their study found that the mixed forecasting model WPD-EMD-ARMA-FIGARCH-M produced significant results during extreme events. Finally, Wu et al. [36] proposed a hybrid model called VMD-IGJO-LSTM-ICEEMDAN-LSTM, based on Variational Mode Decomposition (VMD) and Long Short-Term Memory (LSTM), for forecasting crude oil futures prices. However, It should be pointed out that these models are based on the characteristics of historical data to predict the future price returns and volatility of crude oil. Due to the lack of attention to the modeling of extreme cases, these models are not effective enough to deal with extreme events.

Based on the literature survey, the existing studies have evaluated the characteristics and trends of crude oil futures in various aspects. To some extent, this helps to understand the risk characteristics of crude oil futures prices. However, the current literature on measuring extreme risk remains inadequate. The models ARIMA-EGARCH and ARIMA-SVR are already common in the literature, but these models do not focus on modeling extreme tail points. Based on these two models, we hope to study the value at risk of crude oil futures by introducing the POT method, which focuses specifically on the tail extreme point.

3. Theoretical foundation

3.1. ARIMA

The ARIMA model, proposed by Box and Jenkins [37], is commonly used for time series analysis and forecasting. In time series analysis, three parameters are required to be set, the autoregressive order (p), the difference order (d) and the moving average order (q). The general form of $ARIMA(p, d, q)$ is shown in formula (1):

$$\begin{cases} \varphi(B)\nabla^d y_t = \theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_t \varepsilon_s) = 0, s < t \\ \varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \\ \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \end{cases} \quad (1)$$

where: B is the backshift operator; $\nabla^d = (1 - B)^d$ is the higher order difference; $\varphi_i (i = 1, 2, \dots, p)$ and $\theta_j (j = 1, 2, \dots, q)$ are autoregressive parameters and moving average parameters respectively; ε_t is error term and consistent with the normal distribution $N(0, \sigma^2)$. Second and third equations are the autoregressive correlation coefficient polynomial and the moving average coefficient polynomial, respectively.

3.2. SVR

Support vector regression [38] is a machine learning method originally proposed to minimize the upper bound of generalization error based on the principle of structural risk minimization. In the training set $\{x_i, y_i\}_{i=1}^l$, there are a total of l training samples, $x_i \in \mathbf{R}^d$ represents the i th input vector, $y_i \in \mathbf{R}$ represents the i th output vector, Then the nonlinear regression function can be expressed as equation (2).

$$f(x) = \omega \varphi(x) + b \quad (2)$$

in formula (2), ω is the weight vector; φ is the nonlinear transformation function that maps the sample space to the high-dimensional space; b is the bias. The parameters ω and b are obtained by formula (3).

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i - \xi_i^*) \\ \text{s.t.} \quad & y_i - \omega \varphi(\mathbf{x}) - b \leq \varepsilon + \xi \\ & \omega \varphi(\mathbf{x}) + b - y_i \leq \varepsilon + \xi^* \\ & \xi, \xi^* \geq 0, i = 0, 1, \dots, l \end{aligned} \quad (3)$$

In the formula: $\|\mathbf{w}\|^2$ is the model complexity term; C is the penalty factor, which is used to balance the model complexity and the minimum training error; ξ and ξ^* represents the loose variable, which is used to measure the deviation of training samples outside the insensitive loss band. Introducing formula (3) into the Lagrangian function and kernel function, the nonlinear regression function can be expressed as equation (4).

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}_j) + b \quad (4)$$

where: α_i and α_i^* are Lagrangian operators; $k(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function, which can be expressed as the inner product of $\varphi(\mathbf{x}_i)$ and $\varphi(\mathbf{x}_j)$, that is, $k(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \varphi(\mathbf{x}_j)$. This paper uses the radial basis kernel function, namely $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\gamma^2)$, γ are kernel parameters.

3.3. Hybrid model

It is crucial to have reliable predictions of crude oil futures prices for making informed decisions. The ARIMA model extracts linear components and predicts them, while the SVR and POT models handle the nonlinear and extreme data subcomponents. Therefore, the hybrid model in this study, which combines the ARIMA, SVR, and POT models, has the highest prediction accuracy for both linear and nonlinear behaviors. The mixture model $\mathcal{H}_t$ is shown in equation (5)

$$\mathcal{H}_t = \mathcal{L}_t + \mathcal{N}_t + \Sigma_t \quad (5)$$

where $\mathcal{L}_t$ represents the linear part of the rate of return. $\mathcal{N}_t$ and Σ_t represent the nonlinear part and the extreme tail variation part, respectively. Let $\widehat{\mathcal{L}}_t$ be the prediction result of the original return data from the ARIMA model at time t , the residual ε_t , defined as in equation (6).

$$\varepsilon_t = \mathcal{H}_t - \widehat{\mathcal{L}}_t \quad (6)$$

Adjusted ε_t by using the SVR model and denoted as equation (7).

$$\varepsilon_t = \mathcal{F}_{SVR}(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n}) + \Delta_t \quad (7)$$

where $\mathcal{F}_{SVR}$ is a nonlinear expression related to the SVR model and Δ_t is the extreme tail value. It is fitted by the POT model and the VaR value is calculated by equation (8).

$$\Delta_t = \Phi_{POT}(\Delta_{t-1}, \Delta_{t-2}, \dots, \Delta_{t-n}) + \Delta_{re} \quad (8)$$

where Δ_{re} is the random error

Thus, equation (9) can express a hybrid model used in VaR calculation.

$$\widehat{\mathcal{H}}_t = \widehat{\mathcal{L}}_t + \widehat{\mathcal{N}}_t + \widehat{\Sigma}_t \quad (9)$$

where $\widehat{\mathcal{L}}_t$, $\widehat{\mathcal{N}}_t$ and $\widehat{\Sigma}_t$ are the forecast values for linearity, non-linearity, and extreme using the ARIMA and SVR models and POT, respectively.

3.4. VaR calculation

Value at Risk (VaR) is a financial term used by Morgan Bank to define the maximum potential loss an investor could incur by holding a financial asset or investment portfolio over a specified period of time with a given level of confidence in the future [5]. At time t , assuming that the loss variable $L(t)$ obeys the probability density function $f_t(x)$ and the cumulative distribution function is $F_t(x)$,

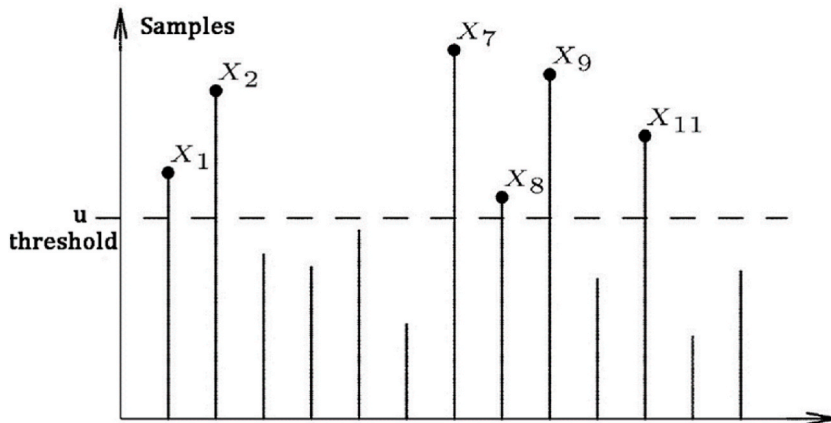


Fig. 1. Diagram of threshold selection.

then given the probability p , we have $\int_{-\infty}^{VaR_{(t,p)}} f_t(x)dx = p$, Then the VaR loss expression of the long asset held can be represented by equation (10).

$$VaR_{(t,p)} = F_t^{-1}(p) \quad (10)$$

If we assume that the distribution of returns satisfies the Gaussian distribution, the $1 - p$ quantile is z_{1-p} , then we can get VaR from the mean and variance as equation (11).

$$VaR_t = \mu + z_{1-p}\sigma_t \quad (11)$$

The distribution of financial assets has the characteristics of ‘spiky peaks and thick tails’, which has been confirmed in existing studies [39–42]. Returns appearing in the tail are considered extreme events. We employ the POT model to calculate the VaR. To calculate the VaR in extreme situation, we use the POT model to specify a threshold value u for the sequence. The selection of the

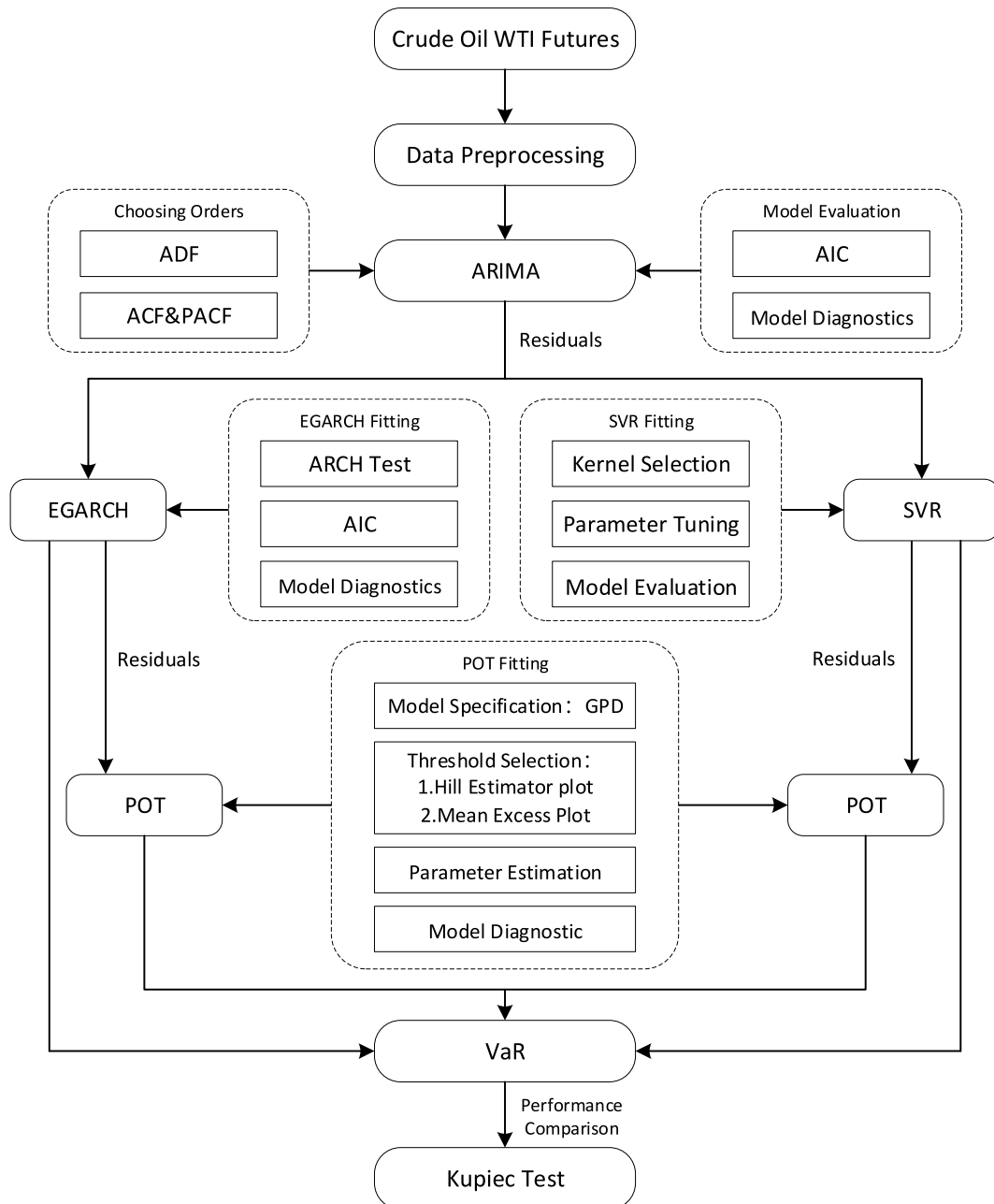


Fig. 2. VaR prediction flow chart.

threshold value affects the estimation of the shape parameter ξ , and it cannot be determined solely based on the statistical model. It often needs to be determined based on the actual situation. Fig. 1 presents the schematic diagram of threshold selection.

Assuming that the returns vector $\mathbf{r}$ are independent of each other, and the cumulative distribution function is denoted by $F(x)$, the range for the change in return vector $\mathbf{r}$ is $[l, k]$. For price return, if we assume that the randomly generated variables are independent and identically distributed samples, we can express the minimum and maximum values of this random variable as equation (12)

$$M_n = \max_{1 \leq j \leq n} \{r_j\}, m_n = \min_{1 \leq j \leq n} \{r_j\} \quad (12)$$

The cumulative distribution function of the minimum value m_n is by equation (13).

$$F_{n,n}(x) = \Pr[m_n \leq x] = \prod_{j=1}^n F(x) = [F(x)]^n \quad (13)$$

Jenkinson [43] showed that the extreme value distribution function of the normalized maximum $x_{(n)}^*$ is by formular (14).

$$F_*(x) = \begin{cases} \exp\left[-(1 + \xi x)^{-1/\xi}\right] & \xi \neq 0 \\ \exp[-\exp(-x)] & \xi = 0 \end{cases} \quad (14)$$

Among them, if $\xi < 0$, $x < -1/\xi$ is established; if $\xi > 0$, $x > -1/\xi$ is established. In the above formula, ξ is the shape parameter.

Smith [44] proposed the POT model of exceeding the threshold. Consider the conditional distribution of $x = r_t - u$ for long positions given $r_t \leq u$. Using the cumulative distribution function in equation (14), then we get the conditional probability of exceeding the threshold as shown in equation (15).

$$\Pr(r \leq x + u | r > u) = \frac{F_*(x + u) - F_*(u)}{1 - F_*(u)} \approx 1 - \left[1 + \frac{\xi x}{\alpha + \xi(u - \beta)}\right]^{-1/\xi} \quad (15)$$

where $x > 0$, and $1 + \xi(u - \beta)/\alpha > 0$. The case of $\xi = 0$ can be regarded as the limit of $\xi \rightarrow 0$. Thus $\Pr(r \leq x + u | r > u) \approx 1 - \exp(-x/\alpha)$. Let $\psi(u) = \alpha + \xi(u - \beta)$. Embrechts [45] proposed the conditional distribution function, which is the generalized Pareto distribution (GPD) showing in formular (16):

$$G_{\xi, \psi(u)}(x) = \begin{cases} 1 - \left[1 + \frac{\xi x}{\psi(u)}\right]^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp[-x/\psi(u)], & \xi = 0 \end{cases} \quad (16)$$

We use the empirical CDF to estimate the cumulative distribution function $F(\eta)$ of returns, then $\hat{F}(u) = \frac{T - N_u}{T}$, where N_u is the number of samples exceeding the threshold u , and T is the sample size. Therefore, The relationship between the cumulative

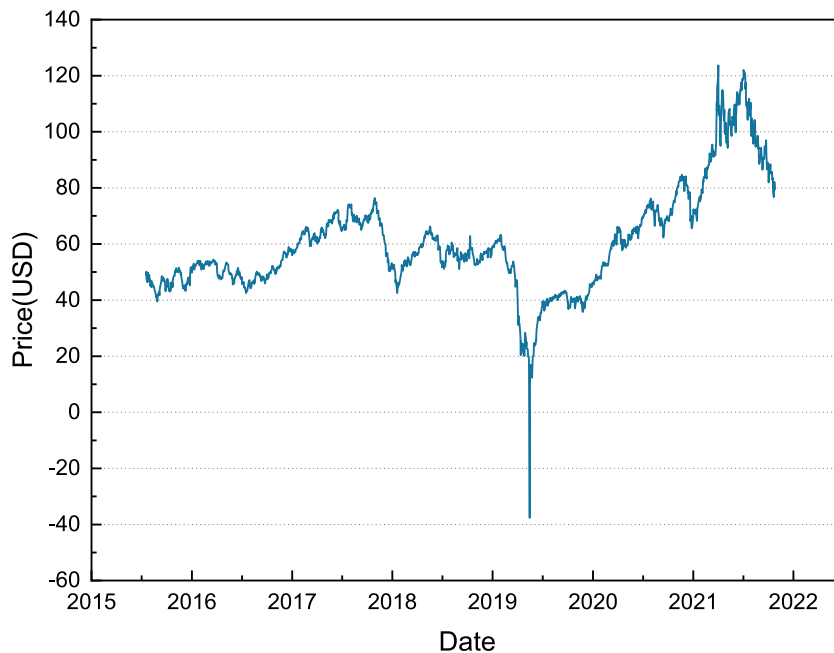


Fig. 3. Crude oil futures price chart.

distribution function $F(y)$ and GPD is shown in equation (17).

$$F(y) = F(u) + G(x)[1 - F(u)] \approx 1 - \frac{N_u}{T} \left[1 + \frac{\xi(y-u)}{\psi(u)} \right]^{-1/\xi} \quad (17)$$

For a given upper tail probability p , by solving for y , the q -quantile VaR of $F(y)$ can be estimated by equation (18).

$$VaR_q = u - \frac{\psi(u)}{\xi} \left\{ 1 - \left[\frac{T}{N_u} (1 - q) \right]^{-\xi} \right\} \quad (18)$$

where the scale and shape parameters of GPD are $\psi(u)$ and ξ , respectively.

After introducing the theoretical foundation, we can use Fig. 2 to illustrate the modeling flow chart of the proposed model.

4. Empirical analysis

4.1. Data sources

Futures markets are known for their efficiency and transparency, which outshines other markets in these aspects. The prices of crude oil futures contracts reflect the collective expectations and viewpoints of market participants, including traders, speculators, and hedgers. These futures prices are highly valued as they provide insight into the market's perception of future trends in crude oil supply and demand. However, the spot prices of crude oil do not reflect this aspect. Therefore, this paper focuses on analyzing the West Texas Intermediate (WTI) crude oil futures prices. The research is aimed at ensuring that the analysis considers extreme events, such as the pandemic and geopolitical events by covering the period from June 23, 2016, to September 30, 2022. A total of 1657 samples were obtained for analysis by excluding holidays. The data used in the study was sourced from [Investing.com](https://www.investing.com).

Fig. 3 depicts a graph that shows the fluctuation of oil prices over the years. Between 2016 and 2018, there was a sudden increase in oil prices followed by a significant decline. In the spring of 2020, oil prices witnessed another sharp decline due to the COVID-19 pandemic and the resulting economic slowdown. It is worth noting that on April 20, 2020, the price of WTI crude oil futures plummeted to a record low of -\$37.63 per barrel. In 2022, international geopolitical conflicts led to an abnormal rise in oil prices. For ease of data processing and analysis, this paper uses the logarithm of the crude oil futures price sequence. The logarithmic return of the crude oil futures on the t -th day can be calculated using the following expression:

$$r = \ln P_t - \ln P_{t-1} \quad (19)$$

where P_t is the closing price on time t , P_{t-1} is the closing price on time $t - 1$, and negative price returns are replaced by adjacent returns.

4.2. Descriptive statistics

The statistical properties of the crude oil futures return series were analyzed in Table 1. The return series shows a slight asymmetry towards positive returns with a skewness value of -0.933 . A kurtosis value of 31.0367 suggests that the return series is leptokurtic, which means it has sharp peaks and thicker tails compared to a normal distribution. Furthermore, The QQ plot in Fig. 4 indicates that the price and return distribution deviate from the standard normal distribution. This suggests that the return distribution does not comply with the assumptions of a standard normal distribution.

4.3. ARIMA-SVR-POT

To calculate VaR with the ARIMA-SVR-POT model, we predict the mean value using ARIMA, the volatility using SVR, and then the POT model is used to capture the extreme values present in the residuals. When analyzing time series data, it's important to ensure the data is stationary. The standard statistical paradigm assumes that the best model for prediction is the one that fits the historical data. In order to determine the stationarity of a selected sample series, we should conduct a unit root test, which is a common method for

Table 1
Descriptive Statistics of data.

	Price	Daily return
Nobs	1657	1656
Minimum	-37.63	-0.3355
Maximum	123.70	0.3196
1.Quartile	48.76	-0.011
3.Quartile	68.49	0.0133
Mean	60.42	0.0003
Median	57.00	0.0022
Stdev	18.59	0.0323
Skewness	0.82	-0.933
Kurtosis	1.59	31.0367

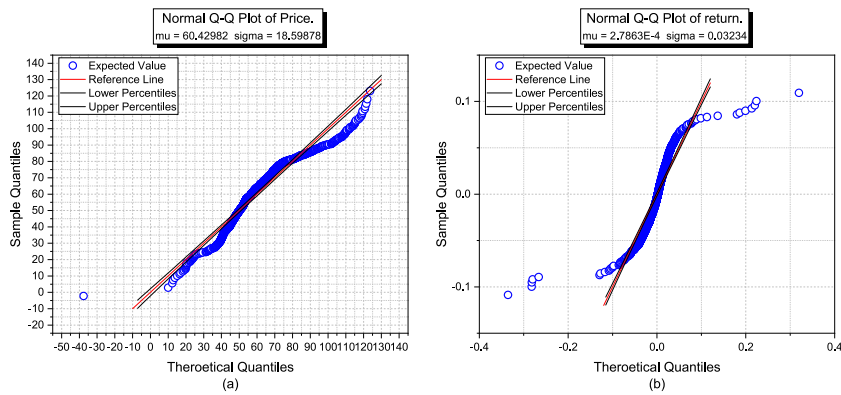


Fig. 4. QQ plot of crude oil future price and return.

evaluating time series samples. Before fitting the ARIMA model to the WTI crude oil futures price regression series, we conducted a unit root test at lag orders of 5, 10, and 15. The results of Table A1 in appendix showed that each statistic had a P -value of 0.01, which is less than 0.05, indicating that the logarithmic return of the price series is a stationary series. Therefore, we concluded that $d = 1$ should be chosen as the order of ARIMA. Then, we determined the order $(p, 1, q)$ of the ARIMA using the Akaike Information Criterion (AIC) [46]. The AIC was used to identify the optimal order by selecting the model with the lowest AIC value. We checked for common p and q values and found the result from Table A2 in appendix A shows that the ARIMA (2,1,2) model had the smallest AIC value of -6714.96 , making it the optimal choice. We move on to fitting the volatility equation. Among the GARCH families, the EGARCH model is suitable for modeling the asymmetry of volatility, which addresses the bias issue of the distribution sequence in the previous descriptive statistics. Therefore, we choose this model to fit the return series' residuals. To ensure that the model we have selected meets the assumptions, we need to verify if the residual sequence has an ARCH effect or if it is simply white noise. We use the square of the sequence residual, notationally α^2 , to test conditional heteroscedasticity. In this research, we use the Ljung-Box statistic $Q(m)$ test to evaluate the series' ARCH effect. The Ljung-Box test is performed on the sequence with lag orders of 5, 10, and 15. The test results of Table A3 in appendix A indicate that for all three lag orders, the P -values were lower than 0.05, suggesting the presence of ARCH effects in the sequence. In order to determine the optimal order of EGARCH model, AIC values of different orders (1,1), (1,2), (2,1), (2, 2) and random perturbations under normal distribution, student distribution, Skew standardized Student-t distribution (sstd) and general error distribution (GED) are compared respectively. According to the test results of Table A4 in appendix A, the minimum AIC is -4.8476 when the sequence is (1,1) and when the sstd is selected. Once we have predicted the results of $\hat{x}_t(1)$ and $\hat{\sigma}_t^2(1)$, we can move the model forward to calculate the Value at Risk of the long position of crude oil futures for different confidence levels (such as 95 %, 99 %, 99.5 %, and 99.9 %). Similarly, when we calculate the VaR of short positions, we only need to take the inverse rate of return. After that, we can carry out the model fitting and calculate it in the same way as for long positions. The results of these calculations are presented in Table A5 of appendix A.

Similarly, We use the ARIMA model to analyze the average component of the price return and the SVR model to fit the non-Gaussian residuals. By predicting one step ahead, we are able to obtain the estimated mean, represented as $\hat{x}_t(1)$, and the estimated volatility, represented as $\hat{\sigma}_t^2(1)$. With these estimates, we can calculate the VaR for both the long and short position of the data in Table A.7 of

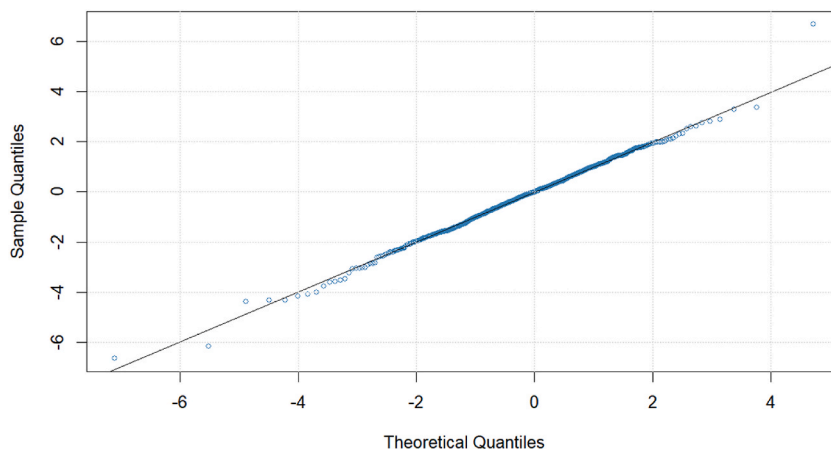


Fig. 5. QQ-Plot of Standardized Residuals after ARIMA-SVR fitting.

appendix A. We evaluate the quality of fit of the residuals by analyzing the QQ plot presented in Fig. 5. Although the overall fit is satisfactory, there are still some extreme values in the tail that do not conform to expectations.

Accurately capturing extreme values is crucial in the ARIMA-SVR and ARIMA-EGARCH models, as they are limited in identifying such values that can lead to significant events. To address this, we propose a hybrid model that fits the extreme tail values of the crude oil return series. The results of the ARIMA-EGARCH-POT model, which follows similar modeling steps to ARIMA-SVR-POT, are presented in Tables A.7 and A.8 in Appendix A. This paper mainly introduces the ARIMA-SVR-POT modeling method, which aims to better capture extreme values in the crude oil return series.

In order to apply the Generalized Pareto Distribution (GPD), an appropriate threshold (u) must be determined. Selecting the appropriate threshold is extremely critical to fitting the GPD and obtaining the correct shape ξ and scale $\psi(u)$ parameters. When analyzing data using the Hill plot, it's essential to consider the gradual upward curve and then its stabilization. It's recommended to focus on the first half of the curve as the data points in the second half may be limited by the sample size, which can make them less reliable. Selecting a threshold point where the starting region is stable can be beneficial. Similarly, when analyzing a Mean Excess Plot, the slope of the curve usually decreases gradually and becomes stable as the excess threshold increases. Observing when the curve starts to become relatively flat can indicate a relatively stable interval for estimating the extreme exponent. To select thresholds, there are various experiences and methods available. Typically, the threshold is determined by observing where the Hill plot stabilizes and approaches an inflection point through the expected excess function. Fig. 6 displays the Hill plot for negative returns, while Fig. 7 illustrates the mean excess plot of negative returns.

Fig. 6 shows the Hill plot of negative returns. It can be observed that the x-axis value stabilizes around 0.0318 for the upper threshold. This means that 142 order statistics exceed the selected threshold. Referring to Fig. 7, we see that the mean excess plot of negative returns increases with this value as the inflection point when the threshold is set around 0.0318. Therefore, the threshold selection is reasonable. Once the threshold value and the order statistics for negative returns are determined, we proceed with the GPD parameter fitting. We use the 'evir' package in the R programming language and set the threshold value to 0.0318 through function commands, and the number of order statistics exceeding the safety threshold to 142. After fitting the GPD distribution, it underwent testing. Four diagnostic plots were generated to test the GPD model, which is shown in Fig. 8. Among these plots, image 8(a) of the excess distribution plot showed that each scatter point closely aligns with the curve, indicating a good fit of the GPD, image 8(b) obtains scatterplots of Residuals, which render random white noise. Image 8(c) displayed the tail distribution of the underlying distribution, showing only two outliers, image(d) of the QQ plot of residuals showed that only a few points in the tail deviate slightly from the straight line, while most points are well-aligned, indicating a satisfactory fit. To calculate the short position, the same procedure is applied by examining positive returns. The GPD fitting parameters for both positive and negative returns are presented in Table 2.

After analyzing the extreme returns both positive and negative, we can use formula (18) with the parameters presented in Table 2 to compute VaR. By choosing different confidence levels, we can calculate VaR estimates for short and long positions of WTI crude oil futures at varying confidence levels such as 95 %, 99 %, 99.5 %, and 99.9 %. The calculated VaR values for different confidence levels are summarized in Table 3.

4.4. Results comparison and test

The effectiveness of Value at Risk (VaR) was measured by four different models at various confidence levels. To compare the models, Kupiec test [47] is a statistical method used to validate the reliability of a result, particularly through the calculation of the LR statistic by $2 \ln \left\{ \left[\left(1 - \frac{N}{T} \right)^{T-N} \left(\frac{N}{T} \right)^N \right] / \left[(1 - \alpha)^{T-N} \alpha^N \right] \right\}$, which follows a $\chi^2(1)$ distribution. Therefore, the LR statistic is considered valid only within the interval $\chi^2_{1-\alpha/2}(1) < LR < \chi^2_{\alpha/2}(1)$, where N represents the number of days of failure, T is the total sample size, α is the predicted probability, and N/T is the actual probability of loss. Tables 4 and 5 summarize the observed failure days for long and short positions, as well as the LR statistics.

Table 4 depicts the number of days where the VaR is less than the actual loss, based on different models and confidence intervals. A failure happens when the predicted VaR is too small to include the actual situation at the given confidence level. As the confidence



Fig. 6. Hill plot of negative returns.

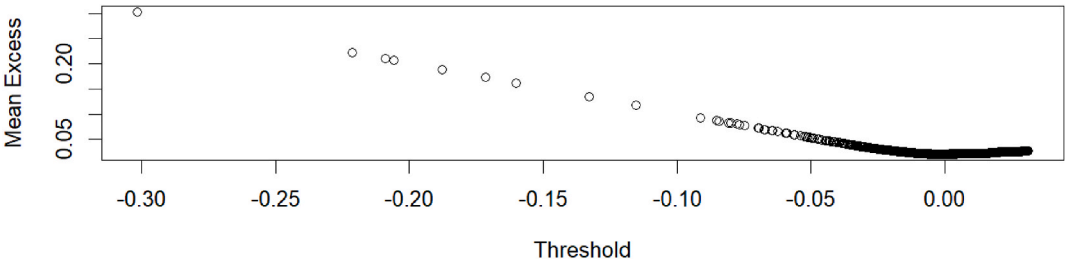


Fig. 7. The mean excess plot of negative returns.

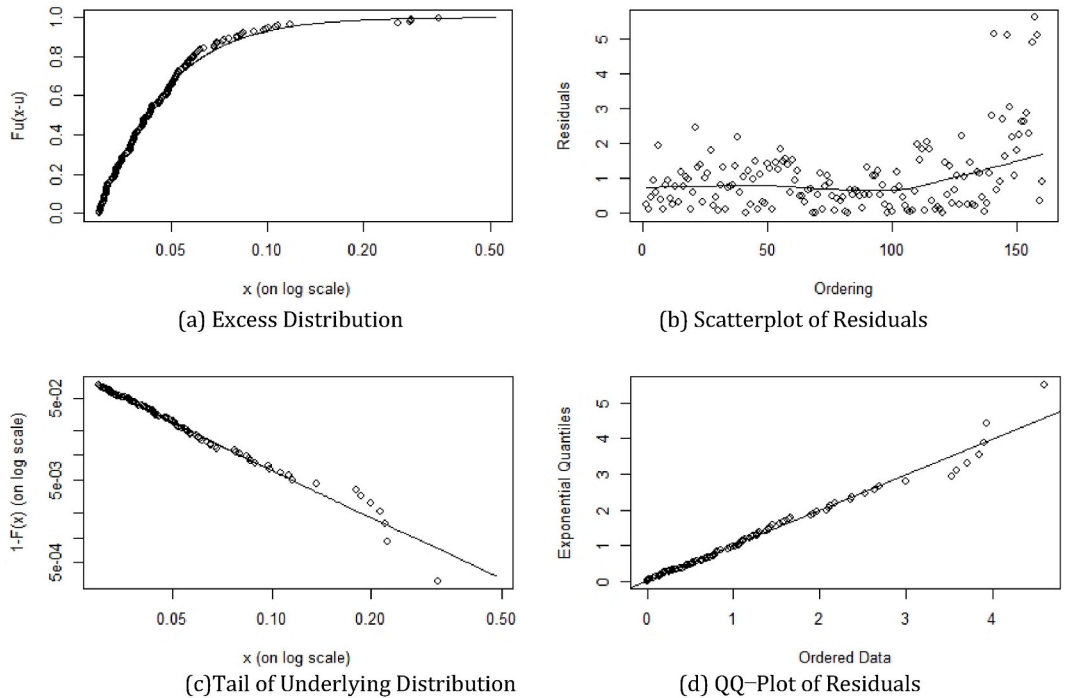


Fig. 8. Diagnostic plots of GPD for negative returns: (a) Check if the residuals follow an excess distribution. (b) Check for heteroscedasticity and nonlinear relationships in the model. (c) Examine the tail distribution of the original data to identify outliers or long-tailed data. (d) Compare the distribution of residuals with the theoretical normal distribution.

Table 2
Parameter estimation of the Generalized Pareto Distribution.

	u	$n. exceed$	Shape ξ	$\psi(u)$
$-r$	0.0318	142	0.4234	0.0154
r	0.0328	134	0.4164	0.0162

Table 3
VaR results under different confidence levels based on ARIMA-SVR-POT.

	95 %	99 %	99.5 %	99.9 %
Long Position	0.0411	0.0857	0.1166	0.2349
Short Position	0.0415	0.0871	0.1183	0.2373

level increases, the VaR value increases, and the number of failure days decreases. This is shown by the decreasing results in the rows of the table. Table 5 is calculated based on the results of Table 4. For long positions at the 95 % confidence level, the LR statistic of the ARIMA-EGARCH model is 13.8199, which falls outside the reference interval. This indicates a large deviation between the probability

Table 4

Failure days summary in long and short position.

	95 %	99 %	99.5 %	99.9 %
Long Position				
ARIMA-EGARCH	52	20	6	5
ARIMA-SVR	95	26	12	4
ARIMA-EGARCH-POT	92	22	10	2
ARIMA-SVR-POT	89	14	7	2
Short Position				
ARIMA-EGARCH	54	22	6	5
ARIMA-SVR	98	25	14	5
ARIMA-EGARCH-POT	95	21	10	3
ARIMA-SVR-POT	92	15	8	1

Table 5

LR Statistics results summary in Long and Short Position.

	95 %	99 %	99.5 %	99.9 %
Long Position				
ARIMA-EGARCH	13.8199	0.6769	0.6982	4.3691
ARIMA-SVR	1.8101	4.6320	1.4739	2.3704
ARIMA-EGARCH-POT	1.0402	1.6364	0.3366	0.0670
ARIMA-SVR-POT	0.4775	0.4219	0.2099	0.0670
Short Position				
ARIMA-EGARCH	11.9601	1.6364	0.6982	4.3691
ARIMA-SVR	2.7811	3.7578	3.2859	4.3691
ARIMA-EGARCH-POT	1.8101	1.1084	0.3366	0.8783
ARIMA-SVR-POT	1.0402	0.1533	0.0096	0.3034
Reference interval	(0.000982,5.02)	(0.000039,7.88)	(0.000010,9.14)	(4e-17,12.11)

of actual loss occurrence and the predicted loss probability, resulting in inaccurate VaR estimates. Under the same confidence level, the LR statistic of the ARIMA-SVR model is 1.8101, which is much smaller than that of the ARIMA-EGARCH model and has passed the Kupiec test. This indicates that the VaR reliability obtained by ARIMA-SVR is higher than that of the ARIMA-EGARCH model. The actual loss probability is closer to the predicted loss probability, and the estimated VaR reflects the potential loss to some extent. However, after the addition of the POT model, both the ARIMA-EGARCH and ARIMA-SVR models have improved their prediction accuracy to some extent. The statistical value of the LR test of the ARIMA -SVR-POT model is relatively lower, indicating that the probability of real risk occurrence in the model is close to the assumed probability. At the 99 % confidence level, the predictive power of the ARIMA-SVR is weaker than that of the ARIMA-EGARCH model for both long and short positions, which can be shown by LR statistics. However, their performance is still not as good as that of the model with POT. The test results are the same as at the 99 % confidence level but change at the 99.9 % confidence. Among the long positions, ARIMA-SVR has performed better than ARIMA-EGARCH. This shows that the traditional econometric model is not stable in predicting the market risk value of crude oil under different confidence levels, sometimes high and sometimes low. At this confidence level, the performance of ARIMA-EGARCH-POT and ARIMA-SVR-POT are similar, but the performance of ARIMA-SVR-POT in the short position is still better. After analyzing the back-testing results, we concluded that the ARIMA-SVR-POT model performed exceptionally well. It passed all the tests, and the difference between the actual loss probability and the predicted probability was minimal. Additionally, its ability to capture extreme risks was superior to the other benchmark models. This further validates the benefits of the proposed model in capturing extreme tail values.

5. Conclusion

Accurately predicting risk Value at Risk of crude oil futures is crucial for investors. It enables them to have effective risk control, which in turn, can improve their confidence in decision-making. By predicting VaR, investors can better understand the potential risks associated with crude oil prices and take appropriate risk management measures accordingly. This study aims to evaluate the market risk of crude oil futures through the creation of a hybrid model called ARIMA-SVR-POT. The model's performance was compared with that of ARIMA -EGARCH, ARIMA -SVR, and ARIMA-EGARCH-POT models. The validity of the risk measurement of these four models was tested using the Kupiec test. The results indicate that the traditional risk measurement model, which estimates the value at risk of WTI crude oil futures, is inaccurate and fails to effectively capture the tail risks. Empirical findings show that the ARIMA-EGARCH model is ineffective in measuring the risk value of long or short positions at a low confidence level. The LR statistics of the ARIMA-SVR and ARIMA-EGARCH models do not consistently pass the Kupiec test at different confidence levels, indicating that their performance in capturing extreme risks is inconsistent. Although the ARIMA-EGARCH-POT model passed the Kupiec test, the LR value was relatively large, suggesting that the measurement's VaR error was high. In contrast, the ARIMA-SVR-POT model successfully passed the Kupiec test at 95 %, 99 %, 99.5 %, and 99.9 % confidence levels, and the LR statistic was close to zero. This indicates that the hybrid model accurately predicts the probability of actual losses and provides a reliable VaR measurement. Therefore, we conclude

that the hybrid model ARIMA-SVR-POT is highly effective in capturing the tail distribution characteristics of WTI crude oil futures, providing highly reliable and accurate VaR estimates.

Based on the above conclusions, we have the following policy suggestion: (1) Crude oil investors should remain vigilant and take steps to prevent extreme events from affecting the international financial market. They should adjust their portfolio positions in a timely manner to avoid extreme portfolio losses caused by the increased linkage effect of extreme risks of various assets under the impact of extreme events. (2) Policymakers should improve market mechanisms and policies, introduce a wealth of risk management tools, guide rational market expectations, reduce the inherent risks of the crude oil market and its sensitivity to extreme events, and avoid excessive impact on the international financial market, which may trigger systemic financial risks. (3) Regulatory authorities should strengthen the regulatory cooperation between transnational financial risk management institutions, stabilize the domestic crude oil financial market, and work together with the financial regulatory authorities in the international market to block the transmission path of extreme risks between the international crude oil financial market and maintain the coordinated development of the crude oil market.

To further improve the risk management ability of crude oil investors and related enterprises to cope with market risks, the ARIMA-SVR-POT method provides a valuable choice and reference for the choice of risk models. However, it is important to note that modeling other financial assets using this model requires further validation. The transmission of financial risk is often influenced by nearby data points. A large sample means that the associated risks are diminished, and data interconnection leads to fewer data samples. Therefore, the further research direction is to develop new models and use existing historical data to accurately capture market risks.

Funding statement

This research was supported by the National Social Science Foundation of China (21BJY002, 20AZD033), National Natural Science Foundation of China (72373078) and China scholarship council.

Data availability statement

Data will be made available upon request or can be found on the [Investing.com](https://www.investing.com).

Additional information

No additional information is available for this paper.

CRediT authorship contribution statement

Chen Zhang: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xinmiao Zhou:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. The results of VaR and parameter calculation from the benchmark model

Table A.1
ADF Test of returns of oil futures.

Lag order	ADF test	P value
5	−15.78	0.01***
10	−12.297	0.01***
15	−9.4365	0.01***

Note: ***, ** and * statistically significant at 1 %, 5 % and 10 % respectively.

Table A.2
AIC results of ARIMA

Order	(1,1,1)	(1,1,2)	(2,1,1)	(2,1,2)
AIC	−6661.95	−6674.41	−6680.54	−6714.96

Table A.3

Ljung-Box statistic test results of residual after ARIMA fitting

Lag order	χ^2	P value
5	48.873	2.356e-09***
10	61.949	1.547e-09***
15	88.749	1.697e-12***

Note: ***, ** and * statistically significant at 1 %, 5 % and 10 % respectively.

Table A.4

AIC results under different distributions in ARIMA-EGARCH

Order	<i>norm</i>	<i>std</i>	<i>sstd</i>	<i>ged</i>
(1 , 1)	−4.7178	−4.8193	−4.8476	−4.8076
(1 , 2)	−4.7236	−4.8180	−4.8465	−4.8077
(2 , 1)	−4.7167	−4.8179	−4.8462	−4.8063
(2 , 2)	−4.7227	−4.8173	−4.8456	−4.8069

Table A.5

VaR results based on ARIMA-EGARCH

	95 %	99 %	99.5 %	99.9 %
Long Position	0.0577	0.0816	0.1184	0.1305
Short Position	0.0526	0.0831	0.1173	0.1227

Table A.6

VaR results based on ARIMA-SVR

	95 %	99 %	99.5 %	99.9 %
Long Position	0.0324	0.0684	0.0862	0.1578
Short Position	0.0305	0.0642	0.0819	0.1603

Table A.7

GPD parameters from ARIMA-EGARCH-POT

	<i>u</i>	<i>n. exceed</i>	Shape ξ	$\psi(u)$
− <i>r</i>	0.0281	134	0.4631	0.0139
<i>r</i>	0.0301	126	0.4363	0.0119

Table A.8

VaR results based on the ARIMA-EGARCH-POT

	95 %	99 %	99.5 %	99.9 %
Long Position	0.0367	0.0771	0.1070	0.2276
Short Position	0.0362	0.0689	0.0923	0.1833

References

- [1] J. Fernández-Villaverde, P. Guerrón-Quintana, K. Kuester, et al., Fiscal volatility shocks and economic activity, *Am. Econ. Rev.* 105 (11) (2015) 3352–3384.
- [2] J. Cunado, F.P. De Gracia, Oil prices, economic activity and inflation: evidence for some Asian countries, *Q. Rev. Econ. Finance* 45 (1) (2005) 65–83.
- [3] W. Qiang, A. Lin, C. Zhao, et al., The impact of international crude oil price fluctuation on the exchange rate of petroleum-importing countries: a summary of recent studies, *Nat. Hazards* 95 (2019) 227–239.
- [4] J. Liu, F. Ma, Y. Tang, et al., Geopolitical risk and oil volatility: a new insight, *Energy Econ.* 84 (2019), 104548.
- [5] D. Duffie, J. Pan, An overview of value at risk, *J. Deriv.* 4 (3) (1997) 7–49.
- [6] L. Yin, Y. Wang, Forecasting the oil prices: what is the role of skewness risk? *Phys. Stat. Mech. Appl.* 534 (2019), 120600.
- [7] Q. Ji, J.F. Guo, Oil price volatility and oil-related events: an Internet concern study perspective, *Appl. Energy* 137 (2015) 256–264.
- [8] G. Bonaccolto, M. Caporin, R. Gupta, The dynamic impact of uncertainty in causing and forecasting the distribution of oil returns and risk, *Phys. Stat. Mech. Appl.* 507 (2018) 446–469.
- [9] Q. Ji, B.Y. Liu, H. Nehler, et al., Uncertainties and extreme risk spillover in the energy markets: a time-varying copula-based CoVaR approach, *Energy Econ.* 76 (2018) 115–126.
- [10] C. Liang, Y. Wei, X. Li, et al., Uncertainty and crude oil market volatility: new evidence, *Appl. Econ.* 52 (27) (2020) 2945–2959.
- [11] D. Wen, L. Liu, C. Ma, et al., Extreme risk spillovers between crude oil prices and the US exchange rate: evidence from oil-exporting and oil-importing countries, *Energy* 212 (2020), 118740.
- [12] Y. Yang, Y.R. Ma, M. Hu, et al., Extreme risk spillover between Chinese and global crude oil futures, *Finance Res. Lett.* 40 (2021), 101743.
- [13] L. Du, Y. He, Extreme risk spillovers between crude oil and stock markets, *Energy Econ.* 51 (2015) 455–465.
- [14] C. Peng, H. Zhu, Y. Guo, et al., Risk spillover of international crude oil to China's firms: evidence from granger causality across quantile, *Energy Econ.* 72 (2018) 188–199.
- [15] W.L. Zhao, Y. Fan, Q. Ji, Extreme risk spillover between crude oil price and financial factors, *Finance Res. Lett.* (2021), 102317.

- [16] Y.J. Zhang, S.H. Li, The impact of investor sentiment on crude oil market risks: evidence from the wavelet approach, *Quant. Finance* 19 (8) (2019) 1357–1371.
- [17] F. Wen, X. Gong, S. Cai, Forecasting the volatility of crude oil futures using HAR-type models with structural breaks, *Energy Econ.* 59 (2016) 400–413.
- [18] Xu Gong, et al., Investigating the risk-return trade-off for crude oil futures using high-frequency data, *Appl. Energy* 196 (2017) 152–161.
- [19] Lu-Tao Zhao, Kun Liu, Xin-Lei Duan, Ming-Fang Li, Oil price risk evaluation using a novel hybrid model based on time-varying long memory, *Energy Econ.* 81 (2019) 70–78.
- [20] Peter Christoffersen, Kris Jacobs, Bingxin Li, Dynamic jump intensities and risk premiums in crude oil futures and options markets, *J. Deriv.* 24 (2) (2016) 8–30.
- [21] F. Ma, M.I.M. Wahab, J. Liu, L. Liu, Is economic policy uncertainty important to forecast the realized volatility of crude oil futures? *Appl. Econ.* 50 (18) (2018) 2087–2101.
- [22] Boda Kang, Nikitopoulos Christina Sklibosios, Marcel Prokopczuk, Economic Determinants of Oil Futures Volatility: A Term Structure Perspective, vol. 88, 2020, 104743. Available at: SSRN 3417706.
- [23] X. Wang, B. Lucey, S. Huang, Can gold hedge against oil price movements: evidence from GARCH-EVT wavelet modeling, *Journal of Commodity Markets* 27 (2022), 100226.
- [24] M. Fakhfekh, A. Jeribi, Volatility dynamics of crypto-currencies' returns: evidence from asymmetric and long memory GARCH models, *Res. Int. Bus. Finance* 51 (2020), 101075.
- [25] Yan L, Wang H, Athari S A, et al. Driving green bond market through energy prices, gold prices and green energy stocks: evidence from a nonlinear approach[J]. *Economic research-Ekonomska istraživanja*, 2022, 35(1): 6479-6499.
- [26] M. Kondo, I. Bora, D. Kirikkaleli, et al., Testing the volatility spillover between crude oil price and the US stock market returns, *Management Science Letters* 9 (8) (2019) 1221–1230.
- [27] S.A. Athari, N.T. Hung, Time–frequency return co-movement among asset classes around the COVID-19 outbreak: portfolio implications, *J. Econ. Finance* 46 (4) (2022) 736–756.
- [28] M. Balcilar, R. Gupta, C. Kyei, et al., Does economic policy uncertainty predict exchange rate returns and volatility? Evidence from a nonparametric causality-in-quantiles test, *Open Econ. Rev.* 27 (2) (2016) 229–250.
- [29] M.W. Brandt, L. Gao, Macro fundamentals or geopolitical events? A textual analysis of news events for crude oil, *J. Empir. Finance* 51 (2019) 64–94.
- [30] D. Mei, F. Ma, Y. Liao, et al., Geopolitical Risk Uncertainty and Oil Future Volatility: Evidence from MIDAS models, *Energy Economics*, 2020.
- [31] M.S. Shah, F. Khan, S. Zendejboudi, et al., A hybrid connectionist enhanced oil recovery model with real-time probabilistic risk assessment, *Geoenvironment Science and Engineering* 227 (2023), 211760.
- [32] H. Alrweili, H. Fawzy, Forecasting crude oil prices using an ARIMA-ANN hybrid model, *J Stat Appl Probab* 11 (3) (2022) 845–855.
- [33] J. Liu, Z. Zhang, L. Yan, et al., Forecasting the volatility of EUA futures with economic policy uncertainty using the GARCH-MIDAS model, *Financial Innovation* 7 (2021) 1–19.
- [34] H. Abdollahi, S.B. Ebrahimi, A new hybrid model for forecasting Brent crude oil price, *Energy* 200 (2020), 117520.
- [35] L. Lin, Y. Jiang, H. Xiao, et al., Crude oil price forecasting based on a novel hybrid long memory GARCH-M and wavelet analysis model, *Phys. Stat. Mech. Appl.* 543 (2020), 123532.
- [36] J. Wu, J. Dong, Z. Wang, et al., A novel hybrid model based on deep learning and error correction for crude oil futures prices forecast, *Resour. Pol.* 83 (2023), 103602.
- [37] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, et al., *Time Series Analysis: Forecasting and control*[M], John Wiley & Sons, 2015.
- [38] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, *Stat. Comput.* 14 (3) (2004) 199–222.
- [39] F. McGroarty, A. Booth, E. Gerding, et al., High frequency trading strategies, market fragility and price spikes: an agent based model perspective, *Ann. Oper. Res.* 282 (1) (2019) 217–244.
- [40] E. Fanone, A. Gamba, M. Prokopczuk, The case of negative day-ahead electricity prices, *Energy Econ.* 35 (2013) 22–34.
- [41] S.R. Thiagarajan, A. Alankar, R. Shaikhutdinov, Tail risk: challenges, mitigation, and research opportunities, *J. Invest.* 24 (2) (2015) 113–121.
- [42] K. Sano, A binary decision model and fat tails in financial market, *Appl. Sci.* 12 (14) (2022) 7019.
- [43] A.F. Jenkinson, The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Q. J. R. Meteorol. Soc.* 81 (348) (1955) 158–171.
- [44] R.L. Smith, Extreme Value Analysis of Environmental Time Series: an Application to Trend Detection in Ground-Level ozone, *Statistical Science*, 1989, pp. 367–377.
- [45] P. Embrechts, C. Klüppelberg, T. Mikosch, *Modelling Extremal Events: for Insurance and finance*[M], Springer Science & Business Media, 2013.
- [46] Y. Sakamoto, M. Ishiguro, G. Kitagawa, Akaike information criterion statistics, *Dordrecht, The Netherlands: D. Reidel* 81 (10.5555) (1986), 26853.
- [47] S.D. Campbell, *A Review of Backtesting and Backtesting procedures*, 2005.