Future of Work and Digital Management Journal

Article type:
Original Research

Article history:
Received 10 July 2025
Revised 20 August 2025
Accepted 09 October 2025
Published online 01 January 2026

The Impact of Geopolitical Factors on Oil Market Risk Prediction Using a Machine Learning Approach

Zahra. Majdi[1], Farhad. Hanifi[1*], Mir Feiz. Fallahshams[1]

1 Department of Financial Management, CT.C, Islamic Azad University, Tehran, Iran.

Corresponding author email address:
hanifi_farhad@yahoo.com

How to cite this article:
Majdi, Z. , Hanifi , F. & Fallahshams , M.F. (2026). The Impact of Geopolitical Factors on Oil Market Risk Prediction Using a Machine Learning Approach. *Future of Work and Digital Management Journal*, *4*(1), 1-16. https://doi.org/10.61838/fwdmj.150

ABSTRACT

The global oil market, as one of the key pillars of the international economy, is influenced by complex geopolitical factors that make risk prediction a critical challenge. This study investigates the impact of geopolitical factors on oil market risk and develops a machine learning–based model to improve prediction accuracy. Daily time-series data of West Texas Intermediate (WTI) crude oil prices and geopolitical indices—including the Geopolitical Risk Index (GPRD), geopolitical acts (GPRD_ACT), and threats (GPRD_THREAT)—from May 5, 2014, to April 26, 2024, were analyzed. First, multiple linear regression revealed that geopolitical acts have a positive and significant effect on oil price volatility; however, limitations such as residual autocorrelation and non-normality reduced the model's efficiency. Subsequently, four machine learning models—Random Forest (RF), Support Vector Regression (SVR), Decision Tree (DT), and Artificial Neural Network (ANN)—were trained. Among them, RF exhibited superior performance, achieving the lowest error in the test set (MAE: 0.005011, RMSE: 0.006188). Using the RF model, the conditional standard deviation was estimated to calculate the Value at Risk (VaR) at a 95% confidence level, and backtesting with the Kupiec and Christoffersen tests confirmed its accuracy. A comparative analysis with the GARCH model demonstrated the superiority of RF, supported by a higher Lopez statistic (4080.745 vs. 4033.800). These findings highlight the critical role of real geopolitical events in oil market volatility and show the advantage of machine learning in modeling nonlinear market dynamics. This study presents a novel framework for analyzing oil market risk, which can help reduce uncertainty and enhance economic decision-making.

Keywords: Oil market risk, geopolitical factors, machine learning, Random Forest, Value at Risk (VaR), volatility prediction

Introduction

The global oil market is one of the most strategically sensitive arenas of the international economy. Its price dynamics are not only governed by market fundamentals such as supply and demand but are deeply influenced by geopolitical shocks, conflicts, and policy shifts [1]. Oil has long been recognized as both an economic commodity and a geopolitical asset, making it highly vulnerable to disruptions stemming from wars, sanctions, energy transit disputes, and shifting alliances [2]. These vulnerabilities have grown in complexity as the geopolitical landscape becomes increasingly multipolar and uncertain. Events such as armed conflicts in the Middle East, trade wars, and sanctions on major oil-producing nations have demonstrated that geopolitical factors can induce significant volatility, impact risk perceptions, and affect global energy security [3].

Understanding and quantifying geopolitical risk have therefore become critical for researchers and policymakers seeking to model oil market risk with greater accuracy [4]. The concept of geopolitical risk extends beyond isolated political events to include a spectrum of threats, ranging from explicit acts of aggression to verbal warnings and strategic posturing [5]. Scholars have distinguished between realized events, such as military conflicts or sanctions, and perceived threats, such as diplomatic tensions or rhetoric, which may or may not escalate into tangible disruptions [5]. These distinctions are essential because market participants react differently to tangible actions compared to signals of potential unrest. For instance, research has shown that realized geopolitical events exert a stronger and more immediate effect on oil price volatility than verbal threats alone [6].

Traditional econometric studies have long investigated the relationship between geopolitical risk and oil market instability [7]. However, the nature of oil price fluctuations is inherently complex and often nonlinear, making it challenging for classical models to capture abrupt shifts and regime changes [8]. The time-varying and asymmetric transmission of risk requires more flexible modeling techniques [9]. Markov-switching and GARCH family models have been applied to address volatility clustering [8, 10], but their reliance on strong distributional assumptions limits their robustness under extreme market conditions [10]. Scholars have also noted the frequent violation of assumptions such as residual normality and independence in linear regressions when dealing with financial time series [11]. These limitations often lead to spurious correlations and underestimated risk [12].

In response, the field has increasingly shifted toward data-driven, machine learning (ML) methods to improve predictive power and risk quantification [13]. ML algorithms, such as Random Forest (RF), Support Vector Regression (SVR), Decision Tree (DT), and Artificial Neural Networks (ANN), have demonstrated superior capacity to model complex nonlinear interactions and adapt to dynamic environments [13, 14]. These techniques can uncover hidden structures within high-dimensional data and handle irregularities better than parametric econometric models [15]. Their success in other financial markets, such as cryptocurrencies [16], equity indices [13], and energy commodities [17], supports their applicability to oil market risk prediction.

Oil price volatility has far-reaching macroeconomic implications, influencing inflation, exchange rates, fiscal stability, and investment decisions [18]. For oil-dependent economies, unexpected price swings can disrupt public budgets and development plans [19]. The ability to anticipate volatility induced by geopolitical tensions is therefore not merely of academic interest but also a practical necessity for strategic planning and hedging. Studies have shown that in countries such as Iran, where oil rents constitute a significant portion of government revenues, geopolitical risk indices strongly affect revenue predictability and macroeconomic balance [19, 20]. Similarly, disruptions in energy transit routes in the Persian Gulf region have demonstrated how geoeconomic and geopolitical dynamics intersect to affect oil flows and regional security [21].

A key insight from the literature is that the categories and structure of geopolitical risks matter for accurate risk modeling [5]. Realized events such as wars and sanctions have been shown to induce persistent volatility spikes [6], whereas threats, though impactful, tend to create short-lived price shocks [5]. Moreover, the reaction of oil markets to geopolitical risks has evolved with the increasing speed of information diffusion. Investor attention and speculative trading respond rapidly to news flows, amplifying volatility [22]. Digital platforms, social media, and real-time reporting accelerate these dynamics, creating a feedback loop between risk perception and market behavior [22].

Recent studies have advanced the modeling of these complex dynamics using hybrid approaches that integrate econometric frameworks with machine learning [15]. For example, deep learning models have been applied to forecast cryptocurrency volatility by capturing error structures ignored by conventional models [16]. Similarly, hybrid GARCH-ML frameworks have been proposed to address structural breaks and time-varying volatility in financial stress indices [15]. These developments highlight the potential of combining domain-specific econometric insights with the adaptability and predictive power of ML [13].

Within this context, Random Forest has emerged as a particularly promising algorithm due to its ensemble-based structure, which reduces overfitting and enhances generalization [13]. Unlike single decision trees, which are prone to instability, RF aggregates multiple trees to improve predictive stability and capture nonlinearities more effectively [14]. Support Vector Regression is also widely applied because of its robustness to high-dimensional spaces and ability to fit complex boundaries [14]. Artificial Neural Networks, though sometimes criticized for interpretability challenges, excel in capturing intricate patterns when trained on sufficient data [16]. These ML approaches can outperform conventional GARCH models in predicting Value at Risk (VaR) and volatility under conditions of geopolitical stress [17].

The increasing interest in risk prediction at the intersection of finance and geopolitics also reflects broader concerns about energy transition and sustainable investment [4]. As nations diversify their energy portfolios and increase renewable energy investments, oil markets are becoming more sensitive to political signals about carbon policy, trade agreements, and supply chain security [4]. At the same time, the emergence of new geopolitical actors and shifting alliances is reshaping the global energy map [2]. These structural changes create new risk channels that require advanced modeling to support strategic investment decisions [1].

Nevertheless, despite these methodological advances, significant gaps remain in the accurate assessment of geopolitical risk impacts on oil market volatility. Traditional geopolitical indices often aggregate events and threats into a single composite measure, potentially masking asymmetric effects [9]. Recent frameworks have advocated decomposing geopolitical risk into subcomponents—such as acts and threats—to better capture their heterogeneous market impacts [5, 23]. Jiao and colleagues [23] introduced a two-stage analytical approach that isolates the transmission mechanisms from geopolitical shocks to oil price movements, revealing nonlinear and time-varying channels often overlooked by static models. Such granular analysis enhances risk management and forecasting capacity, particularly under complex geopolitical tensions [17].

The present study builds upon these advances by exploring the impact of granular geopolitical risk indicators—specifically acts and threats—on oil market volatility, while integrating advanced machine learning methods to overcome the limitations of classical econometric modeling. By leveraging Random Forest and comparing its predictive accuracy with benchmark volatility models, this research seeks to improve the estimation of conditional risk measures such as Value at Risk. This approach not only accounts for the nonlinear and dynamic structure of oil price movements but also incorporates event-specific geopolitical information that better reflects market sensitivities [13, 15].

In doing so, this work contributes to the literature in three important ways. First, it expands the understanding of how different categories of geopolitical risks—realized events versus threats—affect volatility, aligning with calls to disaggregate geopolitical indicators [5, 23]. Second, it applies cutting-edge machine learning models validated in other financial domains [14, 16] to the oil market, demonstrating their adaptability and potential for risk prediction under high uncertainty. Third, it provides an empirical framework that integrates rigorous backtesting, including Kupiec and Christoffersen coverage tests, to validate the accuracy of predicted risk measures.

Ultimately, improving the precision of oil market risk prediction under geopolitical uncertainty has profound implications for portfolio management, hedging strategies, and policy formulation. For energy-exporting economies, it aids in designing more resilient fiscal and economic policies [19, 20]. For global investors and traders, it enhances decision-making under turbulent conditions, mitigating losses from unexpected price shocks [11, 18]. By integrating robust machine learning models with nuanced geopolitical indicators, this study aims to deliver a practical and theoretically grounded tool for navigating one of the most volatile and strategically important markets in the global economy.

This study aims to develop a robust, machine learning–driven framework for forecasting oil market risk by modeling the differential impact of realized geopolitical events and threats, comparing its predictive performance to classical volatility models, and enhancing the accuracy of Value at Risk estimation under geopolitical uncertainty.

Methodology

To prepare the required variables for testing the hypotheses, Microsoft Excel spreadsheet software was used. First, the collected data were entered into worksheets created in this environment, and then the necessary calculations were performed to derive the variables of this study. After computing all the required variables for use in the models of this research, these variables were combined into unified worksheets to be transferred to the software used for the final analysis. It should be noted that all statistical analyses in this study were performed using R software version 4.3.1.

The statistical population of this research is the Brent oil market, and the data scope consists of daily time series of key geopolitical indices from May 5, 2014, to April 26, 2024. Oil return and volatility were considered the target variables, and in this study, a novel approach was applied to calculate Value at Risk (VaR). Table (1) presents the research variables along with their abbreviations. (To facilitate the project workflow in the software, the variables were labeled with symbols.)

Table 1

Research Variables

Variable Name	Symbol	Type	Description
Oil Price Volatility	Oil	Dependent	Futures contracts of West Texas Intermediate (WTI) crude oil
Geopolitical Risk Index	GPRD	Independent	Event-based geopolitical risk
Real Geopolitical Events	GPRD_ACT	Independent	Actual events related to geopolitical conflicts
Geopolitical Threats and Warnings	GPRD_THREAT	Independent	Measures the number of official or unofficial verbal threats between countries that may escalate into geopolitical tensions or military conflicts

A few points regarding the selected research variables are noteworthy. The objective of this study is to provide a relatively comprehensive measurement of oil market risk by considering multiple dimensions. Therefore, based on the review of prior research, three indices were used to account for geopolitical risk, which were extracted from electronic web-based sources.

Findings and Results

According to Table (2), the number of observations after harmonizing the time series is 2,495 days, covering the period from May 5, 2014, to April 26, 2024. For the dependent variables, oil price volatility (Oil) and oil returns, logarithmic returns (the difference of log prices) were calculated and descriptive statistics were computed accordingly. The results show that daily returns of West Texas Intermediate (WTI) crude oil fluctuated between –33.55% (minimum) and 31.96% (maximum), indicating substantial changes and high dispersion during this period. The mean daily oil return was –0.00007, close to zero, and the median was 0.0012, suggesting that average daily returns ranged around 0% to 1%. The proximity of these two measures implies no clear upward or downward trend in returns during the studied period. The standard deviation of oil returns, as a measure of risk and dispersion, was 0.03046 (around 3%), indicating a relatively high level of daily volatility and associated risk of not achieving expected returns. The skewness value of –0.71 shows slight left-skewness and approximate symmetry, although the distribution is not perfectly symmetrical. Kurtosis at 26.54 is considerably higher than 3 (the benchmark for a normal distribution), reflecting a sharper peak and heavier tails, confirming the likelihood of extreme events (severe price fluctuations) in oil returns.

For the independent geopolitical variables, descriptive statistics also provide valuable insights. The Geopolitical Risk Index (GPRD) fluctuated between 9.49 and 540.83, with a mean of 116.96 and a median of 107.31. These figures indicate that on average, between 107 and 116 political events were reported daily on reputable news and electronic document sources, but the wide range (from as low as 9 to as high as 540) shows significant variability in geopolitical risk intensity during the study period. The GPRD_ACT variable, which tracks real geopolitical events, has a mean of 98.38 and a median of 86.79, indicating that on average, between 86 and 98 reported political events materialized daily. This shows that a considerable portion of reported events turned into actual incidents, though the difference compared to GPRD implies that some reported events did not lead to action.

Table 2

Summary of Descriptive Statistics for Research Variables

Variables	Observations	Minimum	Maximum	Mean	Median	Std. Dev.	Skewness	Kurtosis
Oil	2,495	–0.3355	0.3196	–0.00007	0.0012	0.03046	–0.71	26.54
GPRD	2,495	9.49	540.83	116.96473	107.31	53.60164	1.98	8.44
GPRD_ACT	2,495	0	551.2	98.37691	86.79	64.49237	1.56	4.29
GPRD_THREAT	2,495	7.89	809.49	133.87344	120.25	73.07522	2.58	14.30

The GPRD_THREAT variable, measuring geopolitical threats and warnings, has a mean of 133.87 and a median of 120.25, indicating that on average, between 120 and 133 verbal threats or official/unofficial warnings among political actors were recorded daily. The wide range (7.89 to 809.49) and high standard deviation (73.07) demonstrate substantial instability in the level of geopolitical threats. The positive skewness values of all three geopolitical variables (1.98 for GPRD, 1.56 for GPRD_ACT, and 2.58 for GPRD_THREAT) indicate right skewness and non-symmetric data distribution, while their high kurtosis values (8.44, 4.29, and 14.30, respectively) confirm heavy tails and the presence of potential extreme values.

Overall, these descriptive statistics provide a comprehensive picture of the research variables' behavior, revealing high volatility, the potential for rare but impactful events, and possible nonlinear relationships in the oil market and geopolitical indices. These characteristics emphasize the necessity of using advanced approaches such as machine learning to model and predict oil market risk, making detailed analysis of this data essential for testing the research hypotheses. Figures (1) to (4) illustrate the time-series plots of these variables.

Figure 1

Logarithmic Returns of West Texas Intermediate (WTI) Oil

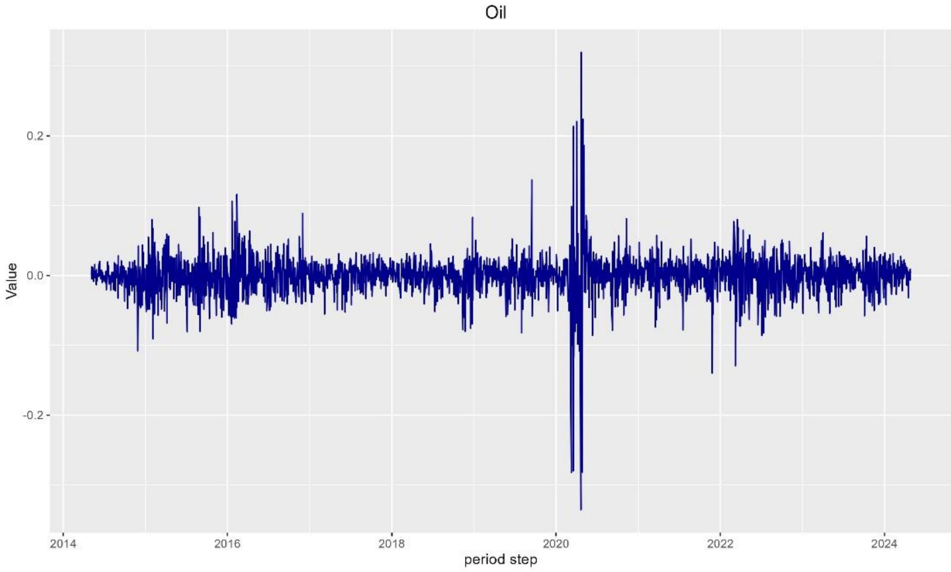


Figure 2

Event-Based Geopolitical Risk Index (GPRD)

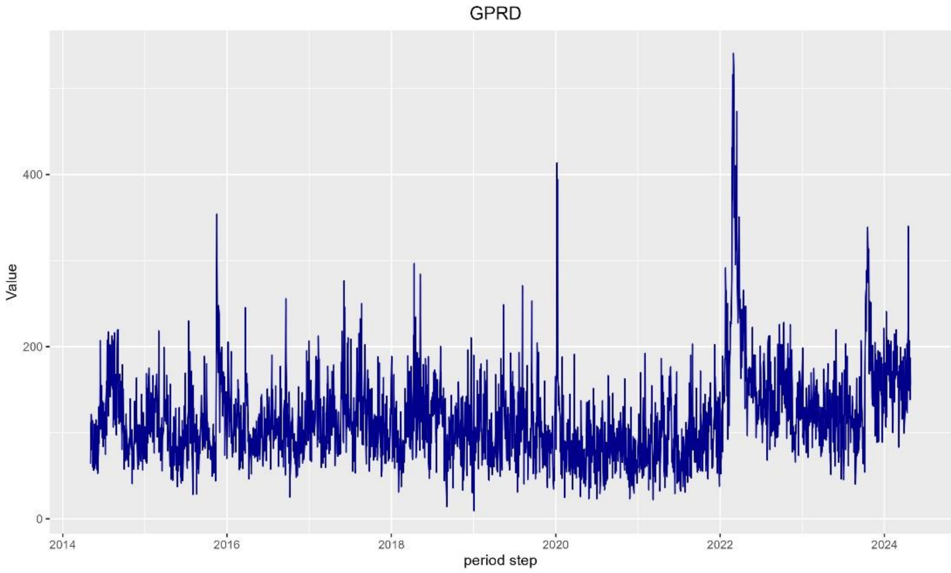


Figure 3

Real Geopolitical Events (GPRD_ACT)

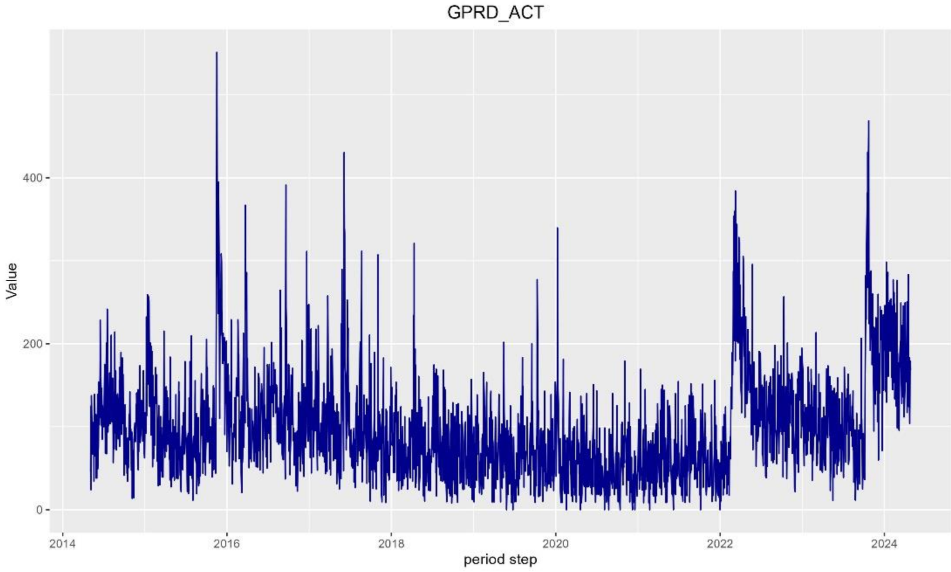
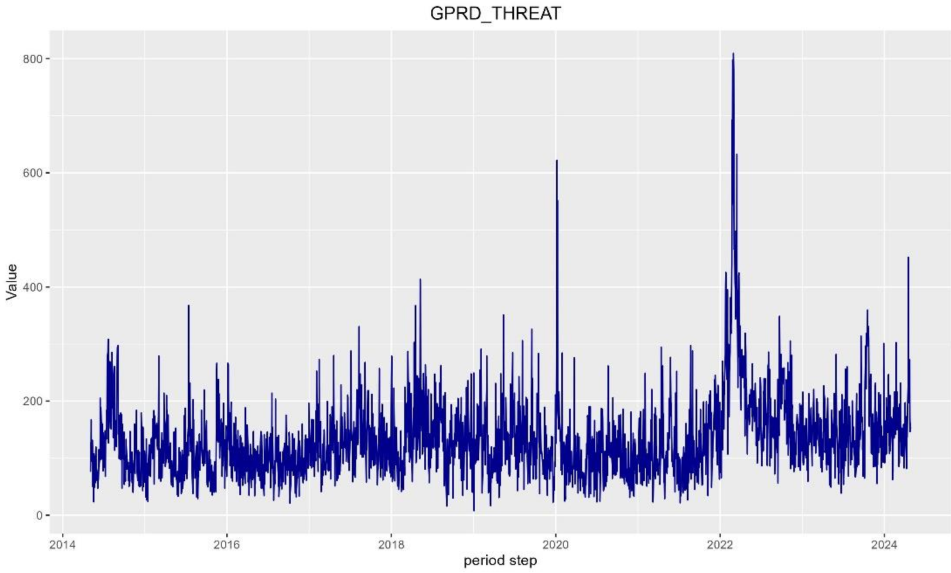


Figure 4

Number of Official or Unofficial Verbal Threats and Warnings (GPRD_THREAT)



To examine the independent effect of each geopolitical variable on oil price volatility, a multiple linear regression model was applied. The results of this model are presented in Table (3), which reports the estimated coefficients, standard errors, t-statistics, and significance levels for each variable.

Table 3

Results of the Multiple Linear Regression Model Testing the Significance of Independent Variables on Oil Volatility

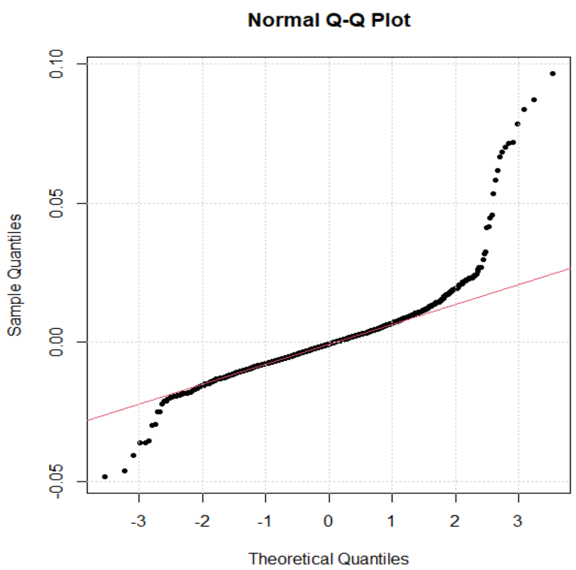
Parameters	Estimated Coefficient	Standard Error	t-Statistic	p-Value
(Intercept)	0.01352	0.000646	20.93	0.00
GPRD	0.001001	0.012545	0.08	0.94
GPRD_ACT	0.017808	0.005292	3.36	0.00
GPRD_THREAT	–0.0114	0.010473	–1.09	0.28

Adjusted R-squared: 0.5727; Durbin–Watson statistic: 0.14327

Before interpreting the results, it is necessary to assess the model's adequacy. The adjusted coefficient of determination (R^2) of the model is 0.5727, indicating that approximately 57% of the variation in oil price volatility (the dependent variable) is explained by the independent variables. This suggests a moderate explanatory power, meaning the model can partly explain volatility, but a portion of the fluctuations remains unexplained. However, the Durbin–Watson statistic of 0.14327 raises a noteworthy concern. This very low value indicates the likelihood of autocorrelation in the error terms (residuals), reinforcing the spurious regression hypothesis proposed by Granger and Newbold (1974). Granger and Newbold argued that in spurious regressions, a high R^2 combined with a low Durbin–Watson statistic often appears due to autocorrelation in time series data. They recommend that if R^2 is greater than the Durbin–Watson statistic ($R^2 > D.W.$), the model should be re-estimated using the first-difference form of the variables to eliminate spurious effects. The model results show that the variable GPRD_ACT (real geopolitical events) with an estimated coefficient of 0.017808 and a p-value of 0.00 has a positive and statistically significant effect on oil price volatility. This finding is consistent with previous studies that have identified geopolitical shocks—such as Middle Eastern tensions, wars, sanctions, and political unrest—as key drivers of oil price volatility. GPRD_ACT, as an index capturing actual geopolitical events, plays a prominent role in this dynamic. In contrast, GPRD (the overall geopolitical risk index) shows no significant impact on oil volatility, with a p-value of 0.94 and a negligible t-statistic (0.08). Likewise, GPRD_THREAT (geopolitical threats) with a negative coefficient (–0.0114) and a p-value of 0.28 is not statistically significant. These results indicate that, unlike real events, verbal or potential geopolitical threats do not have a notable effect on oil price volatility, which may relate to their indirect or temporary nature. Additionally, an examination of the residuals' distribution using the Q-Q plot (Figure 5) indicates a substantial deviation from the normality assumption. In this plot, the quantiles of the residuals are plotted against the quantiles of a normal distribution. If the points lie on a straight line, the residual distribution would be consistent with normality; however, the clear deviation of the points from the straight line in Figure (5) visually rejects the normality of the residuals. This finding underscores the need to reconsider the modeling approach and possibly use more advanced methods, such as nonlinear or time-series models.

Figure 5

Q-Q Norm Plot for Residuals of the Linear Regression Model



The Q-Q norm plot shows the quantiles of the residuals against the quantiles of a normal distribution. Alignment of the points along a straight line indicates normality; however, as shown above, this assumption is visually rejected. Geopolitical shocks such as Middle Eastern tensions, wars, sanctions, and political unrest directly affect oil price volatility. GPRD_ACT is an index that measures the number of realized geopolitical events. Geopolitical risk has been recognized in many oil-related studies as one of the most important determinants of oil price volatility.

The multiple linear regression model in this study provides an initial insight into potential linear relationships between geopolitical independent variables and oil price volatility. However, as noted earlier, structural issues such as residual autocorrelation, non-normality of residuals, and the very low Durbin–Watson statistic challenge the adequacy and reliability of the model's results. Therefore, to achieve the research objective of more accurate oil market risk prediction, advanced approaches such as machine learning were employed. Compared to classical linear regression, these models are more robust to violations of statistical assumptions and have greater capability to model complex relationships. Table (4) summarizes the strengths and weaknesses of classical linear regression compared with machine learning models and provides criteria for selecting an appropriate approach.

Table 4

Comparison of Advantages and Disadvantages of Linear Regression vs. Machine Learning Models

Model	Advantages	Disadvantages	Suitable For
Linear Regression	Simple, interpretable	Requires strong assumptions (e.g., normality of errors, independence of data)	When linear relationships exist between variables
Machine Learning	Flexible, nonlinear modeling, robust to noise, capable of learning complex patterns	Computationally intensive, difficult to interpret (black-box nature)	When data are complex and relationships are nonlinear

As shown in Table (4), machine learning models, due to their data-driven and adaptive nature, rely less on statistical assumptions such as residual normality and absence of serial autocorrelation. This makes them well-suited for analyzing time-series data with high volatility and nonlinear patterns. Moreover, their ability to detect and model nonlinear relationships among variables gives them a clear advantage over linear regression. However, their main limitation is their "black-box" nature, which makes interpreting the results and coefficients more difficult compared to linear models and requires supplementary interpretability techniques.

In this section, after preparing the data, four widely used machine learning models—Support Vector Regression (SVR), Random Forest (RF), Decision Tree (DT), and Artificial Neural Network (ANN)—were trained. These approaches have attracted attention in forecasting highly volatile financial assets such as oil and cryptocurrencies due to their ability to model nonlinear and complex relationships (Pourmansouri et al., 2024; Fallah et al., 2024). In this study, the dataset was split into two parts: 80% as training data and 20% as testing data.

The performance of these models was evaluated using two common error assessment metrics: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The results are presented in Table (5). These metrics are defined as follows:

$MAE = (1/N)\ \Sigma\ |y_i - \hat{y}_i|$

$RMSE = \sqrt{(1/N)\ \Sigma\ (y_i - \hat{y}_i)^2}$

The MAE indicates the average absolute deviation of predictions from actual values and provides a measure of the average error magnitude, while the RMSE, by squaring the errors, measures their dispersion around zero and gives greater weight to larger errors. The lower the values of these two metrics, the higher the predictive accuracy. Typically, because of the calculation method, the MAE is reported to be lower than the RMSE.

Table 5

Comparison of Machine Learning Models' Prediction Accuracy for Oil Volatility

Model	RMSE (Training)	MAE (Training)	RMSE (Testing)	MAE (Testing)
SVR	0.006037	0.00349	0.00909	0.007341
Random Forest	0.002068	0.001381	0.006188	0.005011
Decision Tree	0.007355	0.005629	0.007638	0.006115
ANN	0.00454	0.003604	0.009671	0.007382

Based on the results in Table (5), the Random Forest model achieved the best performance in both the training and testing datasets in terms of prediction accuracy and stability. In the training data, this model recorded the lowest error among the examined models, with MAE = 0.001381 and RMSE = 0.002068. These values indicate that the Random Forest model's predictions of oil price volatility (Oil) deviated less from the actual values on average. In the testing data, the model maintained its superiority with MAE = 0.005011 and RMSE = 0.006188.

The RMSE, which reflects error dispersion, is lower for Random Forest compared to other models. This indicates higher prediction stability and less deviation from actual values. In contrast, models such as ANN and SVR showed higher error levels in the testing data (RMSE = 0.009671 and 0.009090, respectively), suggesting possible overfitting or insufficient generalization capability. The Decision Tree model, although showing acceptable performance in the training set, demonstrated lower accuracy than Random Forest in the testing set. Considering Random Forest's superiority in accuracy (lower MAE) and stability (lower RMSE) across both datasets, it was selected as the optimal model for predicting oil price volatility. This selection is based on the high capability of Random Forest in handling complex data and capturing nonlinear patterns, which is crucial in analyzing oil market volatility and the impact of geopolitical factors. Therefore, this model was subsequently used to estimate the conditional standard deviation and calculate the Value at Risk (VaR) for more accurate and reliable oil market risk prediction.

After estimating the conditional standard deviation for the testing set using the Random Forest model, the Value at Risk (VaR) was calculated based on the following formula:

$$VaR_t = \sigma_t \times q^{\alpha}$$

Where VaR_t is the value at risk at time t, σ_t is the conditional standard deviation obtained from the machine learning model, and q^{α} is the quantile of the appropriate statistical distribution. According to the findings of previous sections, the skewed Student-t distribution was considered suitable for modeling oil returns.

To better evaluate the VaR estimated by the proposed model, it was also calculated using a conditional heteroscedasticity model (GARCH) to enable comparison with the study's proposed approach. Figures (6) and (7) show the VaR estimated from the proposed Random Forest model and the GARCH model, respectively, along with the actual oil price returns. In addition, the results of backtesting to assess the adequacy of both methods are reported in Table (6).

Table 6

Backtesting Results for Estimated Value at Risk Using Both Methods

Method	Test	Statistic	p-Value	Lopez Loss
Proposed (Random Forest)	Unconditional Coverage (Kupiec)	1.585372	0.21	4080.745
	Conditional Coverage (Christoffersen)	1.814173	0.40	—
Parametric (GARCH-based)	Unconditional Coverage (Kupiec)	0.001715	0.97	4033.800
	Conditional Coverage (Christoffersen)	0.117722	0.94	—

The backtesting results show that the p-values for both the Kupiec Unconditional Coverage and the Christoffersen Conditional Coverage tests for both the proposed method and GARCH are greater than 0.05. This confirms the null hypothesis that the estimated Value at Risk at the 95% confidence level is sufficiently accurate. In other words, both methods are statistically reliable, and their violation rates are consistent with the expected confidence level.

However, to compare the relative performance of the two methods, the Lopez Loss Function was used, which evaluates prediction accuracy by considering error magnitudes. The Lopez statistic for the proposed Random Forest method (4080.745) is higher than that of the GARCH model (4033.800). This difference shows that the proposed method based on Random Forest provides better performance in predicting Value at Risk despite both models being statistically acceptable.

This superiority can be attributed to the Random Forest model's ability to capture nonlinear patterns and its flexibility in dealing with complex geopolitical data and oil market volatility. The VaR calculated using Random Forest is not only statistically valid but also more accurate and efficient compared to the GARCH model. This finding reinforces the advantage of using machine learning approaches in oil market risk prediction and highlights the strong potential of such methods for analyzing financial data with nonlinear and dynamic characteristics.

Figure 6

Value at Risk Estimated by the Proposed Random Forest Model at 5% Level

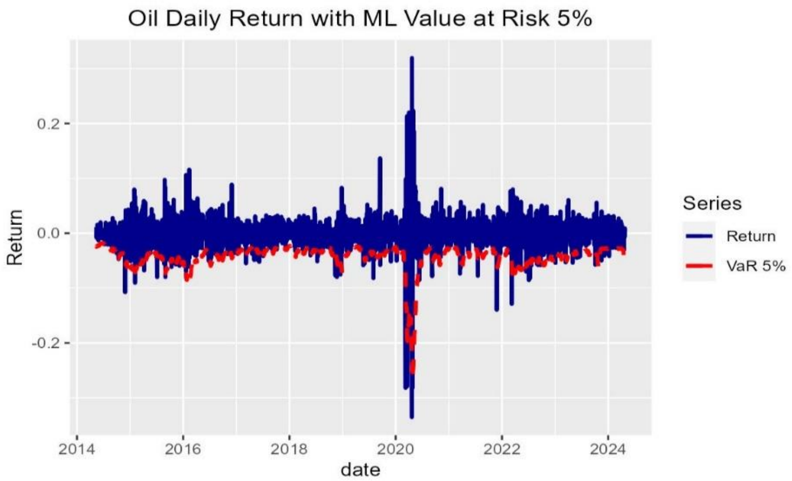
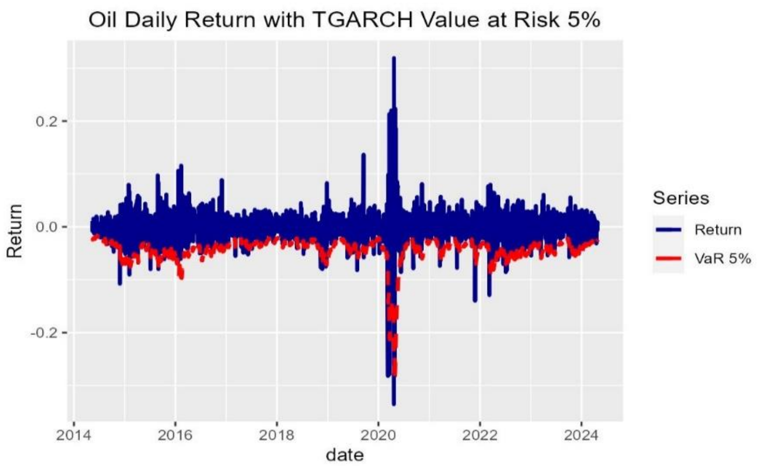


Figure 7

Value at Risk Estimated by the GARCH Model at 5% Level



Discussion and Conclusion

The results of this study revealed that among the geopolitical indicators examined, *realized geopolitical events (GPRD_ACT)* had a positive and statistically significant effect on oil price volatility, whereas the general geopolitical risk index (GPRD) and *geopolitical threats (GPRD_THREAT)* were not statistically significant predictors. This finding confirms the notion that actualized geopolitical shocks—such as wars, sanctions, and violent conflicts—create stronger and more persistent disturbances in the oil market than verbal warnings or threats of potential escalation [5, 6]. The discovery that realized events exert an immediate upward pressure on oil price volatility supports the work of Demirer et al. [6], who showed that geopolitical risks drive regional oil return fluctuations primarily through concrete acts of disruption rather than speculative narratives. Similarly, Bouoiyour and colleagues [5] emphasized that “acts” in geopolitical risk indices have a deeper impact on price instability than “threats,” a pattern clearly replicated in the current findings.

The weak and statistically insignificant influence of the general GPRD index suggests that aggregated risk measures may dilute meaningful signals by combining heterogeneous types of events. Prior research cautions that composite indices can obscure the asymmetry in how oil markets respond to different classes of geopolitical shocks [9]. Our results support the growing recommendation to disaggregate geopolitical risk measures into actionable components [23]. In particular, Jiao et al. [23] found that decomposing risk into staged categories improves predictive power by revealing distinct transmission channels—such as supply disruption for acts and speculative positioning for threats. This decomposition approach appears to be critical for high-volatility commodities such as oil.

Another essential outcome was the superior performance of the *Random Forest (RF)* model in predicting oil price volatility compared to other machine learning algorithms (SVR, DT, ANN) and the classical GARCH model. The RF model achieved the lowest Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) in both training and testing datasets, demonstrating its ability to generalize without overfitting. This aligns with the findings of Plakandaras et al. [13], who reported that ensemble tree-based models outperform conventional econometric approaches when facing highly nonlinear and noisy financial data. The strength of RF lies in its ensemble structure that reduces variance and its capacity to capture intricate feature interactions [14].

Importantly, the accuracy of risk quantification was validated through backtesting using Kupiec and Christoffersen coverage tests, which confirmed that both RF-based Value at Risk (VaR) and GARCH-derived VaR estimates were statistically adequate at the 95% confidence level. However, the Lopez loss function demonstrated that RF delivered better predictive performance by more effectively capturing the size and frequency of tail risk events. These results are consistent with the evidence presented by Cheikh and Zaied [17], who argued that nonparametric and machine learning methods can improve tail risk prediction under geopolitical tensions, where distributional assumptions of GARCH-type models often fail.

The robustness of RF compared to artificial neural networks (ANN) and support vector regression (SVR) also deserves attention. Although ANN and SVR are powerful in theory, their performance can degrade in the presence of noisy, high-dimensional data without careful tuning [16]. Fallah et al. [16] documented that deep learning models can suffer from overfitting in highly volatile markets like cryptocurrencies if model complexity exceeds the effective sample size. Our results replicate this concern, as ANN and SVR produced higher RMSE values in the testing set, indicating weaker generalization compared to RF.

From a macroeconomic and energy security perspective, these findings reinforce the critical role of geopolitical stability for oil-exporting economies. As highlighted by Abdel-Latif and El-Gamal [18], sudden geopolitical shocks exacerbate liquidity constraints and increase risk premiums in global oil trade. The observed significance of realized events further underscores the vulnerability of state budgets and macroeconomic planning in rent-dependent economies such as Iran [19, 20]. The fact that RF provided a more reliable VaR forecast implies that policymakers and energy strategists can adopt such models to anticipate revenue fluctuations and adjust fiscal buffers accordingly.

Moreover, the interplay between geopolitical news flows and investor behavior provides additional context for interpreting these findings. Xiao, Wen, and He [22] demonstrated that investor attention to geopolitical news intensifies speculative trading and amplifies volatility, especially when actual events confirm previously uncertain threats. Our study's results are consistent with this behavioral dynamic: realized geopolitical acts confirm investor fears, trigger flight-to-safety moves, and expand volatility clusters. Threats, however, may only cause temporary attention spikes that fade if not substantiated [22].

Another implication is the relevance of time-varying and regime-switching dynamics in geopolitical risk transmission. Prior work by Qian et al. [8] and Liu et al. [10] emphasized that the oil market's reaction to geopolitical shocks is non-stationary and may alternate between low- and high-volatility regimes. While the present study used RF rather than explicitly switching models, the superior accuracy of RF suggests its ability to adapt implicitly to such changing regimes by learning complex, non-linear conditional distributions [13].

Our results also contribute to the emerging discussion about the intersection between geopolitical risk and the global energy transition [4]. As renewable energy investments grow and energy systems diversify, the oil market's sensitivity to geopolitical signals remains strong, but the underlying channels of transmission may evolve. Zhao et al. [4] highlighted that renewable investment decisions respond to oil market uncertainty induced by geopolitical tensions. The improved prediction of oil VaR using RF could inform investors seeking to hedge exposure or rebalance portfolios in response to geopolitical instability.

Finally, the alignment of our findings with the broader methodological shift in financial risk forecasting should be emphasized. Researchers such as Pourmansouri et al. [15] have called for hybrid frameworks that blend economic insight with data-driven adaptability. The current study demonstrates the power of such an approach, using geopolitical event disaggregation alongside machine learning techniques to overcome the well-known weaknesses of linear and parametric volatility models [11].

Despite its contributions, this study has several limitations that must be acknowledged. First, while the Random Forest model achieved strong predictive accuracy, it remains a black-box technique with limited interpretability compared to parametric econometric models. Although feature importance metrics can provide some transparency, fully explaining the causal pathways remains challenging. Second, the study focused on daily data for a single crude oil benchmark, West Texas Intermediate (WTI). Including other benchmarks such as Brent or Dubai crude could improve generalizability. Third, geopolitical risk indices—though decomposed into acts and threats—still depend on news-based algorithms and publicly reported events, which may introduce bias or delays in capturing real-time developments. Fourth, the sample period (May 2014 to April 2024) includes unique episodes such as the COVID-19 shock and Russian–Ukrainian tensions, which may have exaggerated certain volatility patterns; future samples could reveal different dynamics under alternative global conditions.

Finally, while the study compared RF to several machine learning models and GARCH, other advanced architectures, such as long short-term memory (LSTM) networks or transformer-based time-series models, were not included and may provide additional insights.

Future studies should aim to integrate more granular, real-time geopolitical information, including social media sentiment analysis and satellite-based event monitoring, to reduce the time lag and potential bias of existing risk indices. Comparative analysis of multiple crude oil benchmarks and energy commodities would strengthen the external validity of findings and clarify whether certain markets respond differently to acts and threats. Additionally, future work could explore hybrid model architectures, combining Random Forest or gradient boosting with deep learning models such as LSTM or temporal convolutional networks to further improve predictive stability. Expanding the scope to multi-asset portfolios—linking oil volatility with stock markets, currencies, and renewable energy indices—would also add valuable insight into systemic risk transmission. Lastly, developing interpretable machine learning frameworks or applying explainability tools (e.g., SHAP values, partial dependence plots) could help bridge the gap between predictive performance and economic interpretability.

For energy policymakers and sovereign wealth fund managers, implementing advanced machine learning models like Random Forest can support more robust fiscal planning under geopolitical uncertainty. Energy traders and portfolio managers can adopt such predictive frameworks to enhance hedging strategies and reduce exposure to tail risk during conflict-driven market turbulence. Risk managers in global corporations with oil-dependent cost structures could integrate RF-based Value at Risk forecasts into strategic procurement and insurance decisions. Additionally, international energy agencies and geopolitical analysts can use event-level risk decomposition to refine early-warning systems and scenario analysis, helping decision-makers anticipate and respond more effectively to sudden market disruptions driven by geopolitical acts.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants. Written consent was obtained from all participants in the study.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

[1] M. Sebri, F. M. Ajide, and H. Dachraoui, "Geopolitical risk threshold in the informal economy-natural resources nexus: evidence from BRICS economies," *Mineral Economics,* pp. 1-18, 2025, doi: 10.1007/s13563-025-00514-w.

[2] K. Budanov, V. Vereshchak, В. Н. Кудрявцев, S. Mokliak, and K. Rubel, "Causes and Trends in Modern Geopolitical Changes and Sustainable Changes," *JLSDGR,* vol. 5, no. 1, p. e03925, 2025, doi: 10.47172/2965-730x.sdgsreview.v5.n01.pe03925.

[3] X. Wang, Y. Wu, and W. Xu, "Geopolitical risk and investment," *Journal of Money, Credit and Banking,* vol. 56, no. 8, pp. 2023-2059, 2024, doi: 10.1111/jmcb.13110.

[4] D. Zhao, M. O. Chaudhry, B. Ayub, M. Waqas, and I. Ullah, "Modeling the Nexus between geopolitical risk, oil price volatility and renewable energy investment; evidence from Chinese listed firms," *Renewable Energy,* vol. 225, p. 120309, 2024, doi: 10.1016/j.renene.2024.120309.

[5] J. Bouoiyour, R. Selmi, S. Hammoudeh, and M. E. Wohar, "What are the categories of geopolitical risks that could drive oil prices higher? Acts or threats?," *Energy Economics,* vol. 84, p. 104523, 2019, doi: 10.1016/j.eneco.2019.104523.

[6] R. Demirer, R. Gupta, Q. Ji, and A. K. Tiwari, "Geopolitical risks and the predictability of regional oil returns and volatility," *OPEC Energy Review,* vol. 43, no. 3, pp. 342-361, 2019, doi: 10.1111/opec.12160.

[7] J. Cunado, R. Gupta, C. K. M. Lau, and X. Sheng, "Time-varying impact of geopolitical risks on oil prices," *Defence and Peace Economics,* vol. 31, no. 6, pp. 692-706, 2020, doi: 10.1080/10242694.2018.1563854.

[8] L. Qian, Q. Zeng, and T. Li, "Geopolitical risk and oil price volatility: Evidence from Markov-switching model," *International Review of Economics & Finance,* vol. 81, pp. 29-38, 2022, doi: 10.1016/j.iref.2022.05.002.

[9] O. Ozcelebi and K. Tokmakcioglu, "Assessment of the asymmetric impacts of the geopolitical risk on oil market dynamics," *International Journal of Finance & Economics,* vol. 27, no. 1, pp. 275-289, 2022, doi: 10.1002/ijfe.2151.

[10] J. Liu, F. Ma, Y. Tang, and Y. Zhang, "Geopolitical risk and oil volatility: A new insight," *Energy Economics,* vol. 84, p. 104548, 2019, doi: 10.1016/j.eneco.2019.104548.

[11] L. A. Smales, "Geopolitical risk and volatility spillovers in oil and stock markets," *The Quarterly Review of Economics and Finance,* vol. 80, pp. 358-366, 2021, doi: 10.1016/j.qref.2021.03.008.

[12] A. F. Bariviera, L. Zunino, and O. A. Rosso, "Crude oil market and geopolitical events: an analysis based on information-theory-based quantifiers," 2017.

[13] V. Plakandaras, P. Gogas, and T. Papadimitriou, "The effects of geopolitical uncertainty in forecasting financial markets: A machine learning approach," *Algorithms,* vol. 12, no. 1, p. 1, 2018, doi: 10.3390/a12010001.

[14] M. R. Aunjum, M. Naqi, and S. Jamil, "A feedforward neural network model for forecasting crude oil prices using macroeconomic and geopolitical factors," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2021, pp. 1-6, doi: 10.1109/ICECCME52200.2021.9591128.

[15] R. Pourmansouri, M. Fallahshams, and R. G. G. Afshani, "Designing a financial stress index based on the GHARCH-DCC approach and machine learning models," *Journal of the Knowledge Economy,* pp. 1-30, 2024, doi: 10.1007/s13132-024-02075-9.

[16] M. F. Fallah, R. Pourmansouri, and B. Ahmadpour, "Presenting a new deep learning-based method with the incorporation of error effects to predict certain cryptocurrencies," *International Review of Financial Analysis,* vol. 95, p. 103466, 2024, doi: 10.1016/j.irfa.2024.103466.

[17] N. B. Cheikh and Y. B. Zaied, "Investigating the dynamics of crude oil and clean energy markets in times of geopolitical tensions," *Energy Economics,* vol. 124, p. 106861, 2023, doi: 10.1016/j.eneco.2023.106861.

[18] H. Abdel-Latif and M. El-Gamal, "Financial liquidity, geopolitics, and oil prices," *Energy Economics*, vol. 87, p. 104482, 2020, doi: 10.1016/j.eneco.2019.104482.

[19] A. Gol-Khandan and M. S. Mohammadian, "The Asymmetric Impact of Global Geopolitical Risk Indices and Economic Uncertainty on Oil Rent in Iran," *Quarterly Journal of Fiscal and Economic Policies*, vol. 12, no. 45, pp. 175-211, 2024, doi: 10.61186/qjfep.12.45.175.

[20] A. Memarzadeh, "Investigating the Asymmetric Effects of Geopolitical Risks on Iran's Crude Oil Prices: New Evidence from the Econometric Modeling Approach," *Quarterly Journal of Econometric Modeling*, vol. 9, no. 1, 2024.

[21] F. Norouzizadeh, M. Goodarzi, and H. Masoudnia, "The Role of Geopolitics and Geoeconomics of Energy Transit (Oil and Gas) in the Persian Gulf in Advancing the Regional Policy of the Islamic Republic of Iran," *Quarterly Journal of Geopolitics*, vol. 18, no. 3, pp. 228-255, 2022.

[22] J. Xiao, F. Wen, and Z. He, "Impact of geopolitical risks on investor attention and speculation in the oil market: Evidence from nonlinear and time-varying analysis," *Energy*, vol. 267, p. 126564, 2023, doi: 10.1016/j.energy.2022.126564.

[23] J. W. Jiao, J. P. Yin, P. F. Xu, J. Zhang, and Y. Liu, "Transmission mechanisms of geopolitical risks to the crude oil market--A pioneering two-stage geopolitical risk analysis approach," *Energy*, vol. 283, p. 128449, 2023, doi: 10.1016/j.energy.2023.128449.