



Do Gas Price and Uncertainty Indices Forecast Crude Oil Prices? Fresh Evidence Through XGBoost Modeling

Kais Tissaoui^{1,3} · Taha Zaghdoudi^{1,4} · Abdelaziz Hakimi² · Mariem Nsaibi¹

Accepted: 25 July 2022 / Published online: 16 September 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

This study examines the forecasting power of the gas price and uncertainty indices for crude oil prices. The complex characteristics of crude oil price such as a non-linear structure, time-varying, and non-stationarity motivate us to use a newly proposed approach of machine learning tools called XGBoost Modelling. This intelligent tool is applied against the SVM and ARIMAX (p,d,q) models to assess the complex relationships between crude oil prices and their forecasters. Empirical evidence shows that machine learning models, such as the SVM and XGBoost models, dominate traditional models, such as ARIMAX, to provide accurate forecasts of crude oil prices. Performance assessment reveals that the XGBoost model displays superior prediction capacity over the SVM model in terms of accuracy and convergence. The superior performance of XGBoost is due to its lower complexity and costs, high accuracy, and rapid processing times. The feature importance analysis conducted by the Shapley additive explanation method (SHAP) highlights that the different uncertainty indexes and the gas price display a significant ability to forecast future WTI crude prices. Additionally, the SHAP values suggest that the oil implied volatility captures valuable forecasting information of gas prices and other uncertainty indices that affect the WTI crude oil price.

Keywords Crude oil price · Gas price · Uncertainty indexes · Complex relationship · eXtreme Gradient Boosting · Shapley additive explanation method

JEL Classification Q47 · Q43 · D81 · C53

1 Introduction

Energy resources are highly valuable to every economy Zhu et al. (2014). Numerous energy researchers, such as Hamilton (2003) and Charles and Darné (2017), suggested that price fluctuations of crude oil may cause significant shocks to the overall economy and equity markets. For example, Jo (2014), Wen et al. (2019) suggested that crude oil price volatility holds the greatest capacity to create economic uncertainty and leads to an increased risk to the energy market in the short term. In addition, the emergence of this risk triggers supplementary energy costs that have damaging effects on the budgets of the country. Thus, market actors should decrease their investments in risky assets such as crude oil commodities.

An agreement among contemporary economists worldwide is to tackle the risk of oil market fluctuations. In response, Charles and Darné (2017) indicated that a better comprehension of the assessment of crude oil prices and their underlying determinants represents the essential first step in dealing with this risk. This shall enable both investors and energy policymakers to establish successful hedging policies and monitor the effects of regulatory measures on the oil market.

In line with this, Ftiti et al. (2020) and Wu et al. (2021) added that an effective assessment of crude oil prices enables traders to improve their confidence and motivate them to increase their investments in risky assets and, hence, realize optimal portfolio management. Globally, this will succeed in achieving equilibrium in global financial markets in the short and long runs. Another method followed by many researchers in energy is to forecast crude oil prices and their fluctuations. This allows traders and policymakers to better determine the precise value of crude oil prices to achieve effective speculation and control of the demand and supply of crude oil at lower costs.

Thus far, numerous studies have developed several tools to forecast crude oil prices. These tools can be subdivided into several categories. The first class focuses on traditional methods, such as autoregressive integrated mobile average models and generalized autoregressive conditional heteroscedasticity (Alquist et al., 2013; Baumeister & Kilian, 2012; Salisu & Fasanya, 2013; Wang & Wu, 2012). This type of tool has the ability to take a linear structure in the relationship between variables. The second class is reserved for artificial intelligence methods such as artificial neural networks (Wang & Wang, 2016). These intelligent models appeared robust against stylized facts in the data such as nonlinearity. The third class comprises heterogeneous autoregressive (HAR) models. Many studies (Chen et al., 2020; Liu et al., 2018) have shown that HAR models emerged effectively when the data used to forecast crude oil prices is of the high-frequency type. The last class uses hybrid tools to forecast crude oil prices (Ftiti et al., 2020; Zhang et al., 2015). This is the result of the combination of two or more tools. For example, Ftiti et al. (2020) proposed a hybrid method composed of two parts: a linear structure (autoregressive model) and a nonlinear structure (polynomial, saturation, and dead zone forms). The authors affirm that hybrid tools are effective in considering the complex characteristics of the connectedness between crude oil prices and their forecasters. Based on the aforementioned tool classifications, different determinants affecting

crude oil prices have been identified; for instance, supply and demand factors (Conrad et al., 2014), political factors (Wu et al., 2021), investor sentiment (Yang et al., 2019), large shocks such as the global financial crisis (Charles & Darné, 2014), uncertainty indexes such as the implied volatility risk index (VIX); economic policy uncertainty (EPU) (Dutta, 2017; Dutta et al., 2021); the Geopolitical Risk index (Nonejad, 2021), and the Equity Market Volatility (EMV) trackers (Dutta et al., 2021).

Extant energy research primarily examines the effects of uncertainty on crude oil prices. This is because of the emergence of various measures of uncertainty related to financial market conditions (CBOE¹ Volatility Index (VIX)), energy market conditions (CBOE Crude Oil Volatility Index (OVX)), and EPU. As mentioned above, researchers have shown that the variation in these different measures of uncertainty can significantly affect crude oil prices. In addition, traders and policymakers observed that this uncertainty risk should be considered more dangerous when it was accompanied by a financial crisis, as in October 1987 and December 2008, or with a pandemic outbreak, such as the COVID-19 pandemic. Furthermore, these conditions lead to a complex association between crude oil price and uncertainty measures.

Other studies have demonstrated that commodity assets such as gas prices play a leading role in explaining and forecasting crude oil prices (Alquist et al., 2013; Cerqueti et al., 2019; Gatfaoui, 2016). In particular, Chai et al. (2018) showed that the crude oil market is specified by the complexity caused by nonlinearity, regime-switching, and the effects of variations in many explanatory variables. Fiti et al. (2020) justified that under normal conditions, the complex association between oil prices and factors such as the gas price can be generated by the demand and supply channels. They suggested that a rising crude oil price motivates demanders to substitute natural gas for oil, thereby leading to a higher demand for gas and, consequently, a price increase. Nevertheless, an upward movement in the price of crude oil due to growing oil demand could lead to an increased price of natural gas in three ways: an increase in gas production, an increase in the production costs of natural gas, and in projects that contribute to the effective utilization of natural gas. The authors also showed that the connectedness between crude oil and gas becomes more complex under abnormal conditions. They cited many events such as the 2001 terrorist attacks, the financial crisis of 2007–2009, the Arab Spring, geopolitical factors (Iran–United States), and the Libyan War. These events play a primordial role in generating an increasingly complex association between gas and oil. In line with this, the change in pricing methods and the emergence of other assets in financial and energy markets, accompanied by an investor's heterogeneity, can also lead to asymmetry and non-linearity; hence, it can represent another source of a complex link between crude oil and gas. In summary, previous studies have attempted to solve the complexity issue in the energy market by providing various forecasting methods and factors. Nevertheless, prediction tools need to be improved to address the disequilibrium caused by uncertainty and complexity.

¹ CBOE refers to The Chicago Board Options Exchange.

This research aims to address the gap in the extant literature by considering the forecasting power of uncertainty indices together with the gas price for the crude oil price. To this end, we used a dataset of the West Texas Intermediate Spot Price (WTI), the United States Economic Policy Uncertainty index (EPU), the Chicago Board Options Exchange Volatility Index (VIX), the Chicago Board Options Exchange Crude Oil Volatility Index (OVX), and Henry Hub Natural Gas Spot Price (GP) from May 10, 2007, to August 09, 2021. In addition, we applied a novel artificial intelligence method called the eXtreme Gradient Boosting (XGBoost) method on a daily dataset. Chen and Guestrin (2016) are the developers of this tool. They indicated that the XGBoost method has a higher ability in terms of learning speed and performance compared to other machine learning (ML) tools. It has become more robust to provide good and accurate forecasts and to reduce complexity, especially when data normalization is applied (Ben Jabeur et al., 2021a, 2021b). The notable motivations behind the use of XGBoost soft-computing tools to forecast crude oil prices are as follows. On the one hand, we refer to Ftiti et al. (2020), who indicated that the crude oil price is sensitive to various factors such as the information advent, behavior of participants in the market, and uncertainty in the financial and energy markets. Therefore, the authors also showed that these events were considered the main sources of complex characteristics (e.g., structural evolution, nonlinear structure, time-varying, and non-stationarity) in crude oil prices. Li et al. (2016) noted that the most important challenge facing researchers today is to apply effective tools to achieve good predictive accuracy for crude oil price series, considering the aforementioned complex characteristics. In our study, detecting these stylized facts motivates us to use computational and intelligent models such as the XGBoost approach for forecasting crude oil prices. On the other hand, the XGBoost tool is characterized by lower complexity and costs with higher accuracy and rapid processing times. This allows the combination of various inputs to generate an improvement in forecasting accuracy, together with applying the SHapley Additive explanation (SHAP) to detect the importance of specific features of crude oil prices.

This study contributes to the literature in several ways. First, to the best of our knowledge, no studies have examined the co-effects of uncertainty indices and gas prices on crude oil price forecasts. This study is the first to explore the combined effects of uncertainty indices and gas prices on crude oil prices. Second, the complex pricing of crude oil motivates us to use the XGBoost method to forecast crude oil prices. Motivated by their lower complexity and costs, higher accuracy, and rapid processing times, this study is the first to use the XGBoost method to forecast the price of WTI crude oil. In addition, to verify the superiority of this novel ML method, we compare it against autoregressive integrated moving average models with exogenous input (ARIMAX (p,d,q) and the Support Vector Machine (SVM). The performance of competitors' models is identified using the performance indicators [root mean square error (RMSE) and mean absolute error (MAE)]. Third, our estimated results offer important implications for policymakers and investors for understanding the link between investments in oil futures contracts and investors' fear, as well as the effect of this association on the stability of the oil market. This

also allows them to discover other forecasters related to uncertainty in crude oil prices using sophisticated ML tools.

Overall, the empirical findings indicate that the proposed XGBoost model outperforms its counterparts (e.g., SVM model and ARIMAX) in accurately forecasting crude oil prices. In addition, the results show that implied volatility dominates gas prices and other uncertainty variables when forecasting crude oil prices. Accordingly, this study's findings can be used by policymakers to achieve effective policy regulation regarding the stability of government spending and economic growth. In addition, traders can profit from accurate forecasting of crude oil prices to realize efficient portfolio diversification and correct hedging against huge fluctuations in oil prices.

The remainder of this paper proceeds as follows. Section 2 presents the literature review. Section 3 presents the data analysis and methodological frameworks. Section 4 provides a discussion on the main findings. A summary of the main concluding remarks and policy implications is presented in Sect. 5.

2 Literature Review

Over the last three decades, both academicians and policymakers have been interested in the relationship between macroeconomic performance and oil price shocks. According to Sreenu (2018), oil prices are considered an important driver of economic development and GDP growth. Prior research has reported that fluctuations in the oil market are a key determinant of recessions (Hamilton, 1983, 2003). Other studies have concluded that the relationship between oil market shocks and economic performance could be causal (Chen & Chen, 2007; Wang, 2013).

Considering the narrow association between oil price fluctuations and the economic performance of nations, forecasting crude oil prices is crucial. An increasing number of studies are focusing on this topic. This study focuses on the role of gas prices and uncertainty as key factors that affect oil pricing volatility.

According to economic theory, crude oil and natural gas prices are reciprocal. These are considered substitutes for each other in consumption or production. Many empirical studies have proven that the connection between the two commodities is significant (Ftiti et al., 2020; Gatfaoui, 2016; Lu et al., 2014). More specifically, other authors (e.g., Brown & Yücel, 2008) proved that the relationship between two variables is linear. However, many researchers have confirmed a nonlinear relationship. For example, Batten et al. (2017) examined the spillover of time-varying prices between natural gas and crude oil markets from 1994 to 2014. The empirical results reveal that the price of natural gas causes the price of crude oil to spill over, with price spillover effects persisting for up to two weeks. Furthermore, the findings indicate a lower price dependency between the two energy commodities after 2006.

Ftiti et al. (2020) used a hybrid model composed of two parts: a linear model represented by an ARX structure and a nonlinear structure represented by many forms, such as dead zone, saturation, and polynomial structures. The empirical

estimates of the Hammerstein-ARX model showed the significant forecasting power of gas prices for crude oil prices.

Less abundant studies have explored the nonlinear link between crude oil and gas prices. For example, Atil et al. (2014) used a nonlinear autoregressive distributed lag (NARDL) model to study the association between oil and gas prices. The empirical findings indicate that there is an asymmetric and nonlinear effect between oil, gasoline, and natural gas prices. However, there are differences in the transmission mechanisms.

Regarding the economic uncertainty-oil price relationship, Baker et al. (2016) and Basu and Bundick (2017) concluded that economic uncertainty is an important factor behind the fluctuations and stability of the macroeconomy. In addition to the macroeconomic dimension, economic uncertainty significantly impacts the financial market (Pastor & Veronesi, 2012) and crude oil market (Van Robays, 2016). Similarly, Yi et al. (2021) tested whether macroeconomic uncertainty could explain and forecast China's INE crude oil futures market volatility. The empirical findings of the GARCH-MIDAS model show that geopolitical risk and economic policy uncertainty from the UK and Japan could be considered stronger predictors of crude oil future volatility in China.

Aloui et al. (2016) used a copula approach to explore the impact of economic uncertainty on crude oil returns. They found that higher economic policy uncertainty indices positively affect crude oil returns only during certain periods. This positive effect explains the positive dependence before the financial crisis and Great Recession. Using both the WTI and BRENT crude oil prices and performing structural equation modeling, Wang and Sun (2017) studied the determinants of oil price changes. The empirical results show that economic activity is the most significant factor that affects oil pricing volatility. The findings also indicate that wars and political tensions explain oil price fluctuations in the largest oil-producing countries. To investigate the effect of crude oil price uncertainty on investment, Phan et al. (2019) used a dataset of more than 33,000 firms located in 54 countries over the period 1984–2015. The authors found that crude oil price uncertainty negatively influences corporate investment. This negative effect differs from that of crude oil producers on crude consumers. The authors also found that this negative effect is dependent on the firm's characteristics and stock market development. Based on WTI and BRENT crude oil price data, Yang (2019) explored the correlation between EPU and oil price shocks in the U.S. Empirical findings of the structural vector autoregression framework support a strong connection between crude oil prices and EPU. The author pointed out that the total association of the Brent crude oil price is more pronounced than that of the WTI crude oil price.

Recently, Lyu et al. (2021) studied the over-time impact of global economic policy uncertainty shocks on oil price fluctuations. Two crude oil pricing benchmarks were used: Brent and West Texas Intermediate crude oil prices. The empirical results indicate a time variation in the impact of economic uncertainty on crude oil price fluctuations. This impact is amplified under extreme market conditions, such as the 2008 international financial crisis and the European sovereign debt crisis of 2010–2012. More recently, Dutta et al. (2021) used a new variable called EMV trackers, which measures uncertainty in the financial market.

Served by quantile regressions, the authors showed that the new-based EMV tracker represents a good forecaster of crude oil volatility. Wang et al. (2021) used geopolitical risk as a proxy for uncertainty. The estimates of the Markov regime-switching-GARCH model proved that geopolitical risk significantly impacts crude oil volatility in terms of forecasting. Referring to the aforementioned literature, we note that researchers are focusing on the predictive ability of uncertainty factors for oil prices. In addition, an in-depth review of previous works is that the trend has shifted toward using methods that are able to predict oil prices with accuracy and are capable of accounting for the complex characteristics of oil prices, such as the structural evolution, nonlinear structure, time-varying, and non-stationarity. Table 1 reports several studies that used advanced ML models to predict crude oil prices. For example, Ben Jabeur et al., (2021a, 2021b) attempted to predict oil prices during the COVID-19 pandemic. To this end, they used advanced ML methods such as LightGBM, CatBoost, XGBoost, random forest (RF), and neural network models. This study is also based on SHapely Additive exPlanations (SHAP) values. The main findings support the superiority of RF and LightGBM over traditional models. The empirical results also indicate that high values of GER and ESG lead to lower crude oil prices. Li et al. (2021) integrated variational mode decomposition (VMD) and random sparse Bayesian learning (RSBL, SBL-based prediction with random lags and random samples), to forecast crude oil prices. The findings indicate that the proposed VMD-RSBL significantly outperforms several state-of-the-art schemes. Li et al. (2021) introduced a novel multiscale hybrid model for crude-oil price forecasting. The variational mode decomposition method is advantageous because it allows the decomposition of the crude oil price into several simple models. The empirical findings confirm that the proposed model can achieve superior forecasting results. Table 1 also considered other studies that use hybrid tools to forecast crude oil prices. Abdollahi (2020) employed a hybrid model that included complete ensemble empirical mode decomposition, support vector machine, particle swarm optimization, and Markov-switching generalized autoregressive conditional heteroskedasticity. The authors demonstrate the dominance of the proposed Hybrid models over their counterpart's models in terms of the accurate forecast of crude oil prices. Similar results are found in Ahmad et al. (2021), who documented that the novel hybrid model consisting of median ensemble empirical mode decomposition and the group method of data handling is superior to statistical and ML models in empirical mode decomposition, artificial neural networks, and ARIMA models.

3 Data and Methodology

3.1 Data Analysis

The data sample contains daily values of the West Texas Intermediate WTI Spot Price (WTI), the United States Economic Policy Uncertainty index (EPU), the Chicago Board Options Exchange Volatility Index (VIX), the Chicago Board Options Exchange Crude Oil Volatility Index (OVX), and Henry Hub Natural Gas Spot Price (GP). Crude oil and gas prices have been downloaded from the Energy

Table 1 Contemporary studies on crude oil price prediction

References	Sample period	Method (s)	Main results
Huang and Deng (2021)	January 1994–July 2018: daily and monthly frequency	Variational mode decomposition (VMD); Long short-term memory (LSTM) network; ARMIA model; Genetic algorithm optimized SVM	The prediction results show the superiority of the variational mode decomposition (VMD), the long term memory network (LSTM) over the ARMIA model, and the SVM optimised by the genetic algorithm
Abdollahi (2020)	June 2015–April 2016: daily frequency	Hybrid model; Complete ensemble empirical mode decomposition; Support vector machine; Particle swarm optimization; Markov-switching generalized autoregressive conditional heteroskedasticity	The results show the predominance of the proposed hybrid models over their counterparts in terms of good fit
Hao et al. (2020)	January 1986–May 2018: monthly frequency	Regression models with regularization constraints	Findings show that the proposed models generate accurate forecasting for the crude oil price
Rubaszek (2020)	January 1984–March 2018: quarterly frequency	Dynamic stochastic general equilibrium; vector autoregression; random walk models	The forecasting results obtained by the dynamic stochastic general equilibrium show more accurate forecasts than those found by the vector autoregression and random walk models
Liu et al (2020)	January 2004–December 2018: monthly frequency	Hierarchical shrinkage model; Autoregression models; Multivariate models	The performance measures show a dominance of the hierarchical shrinkage model in terms of accuracy predictions over the competing models
Wu et al (2020)	January 1986–February 2018: daily frequency	ICEEMDAN-SCA-RVFL (Improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN), sine cosine algorithm (SCA), and random vector functional link (RVFL) neural network); Back propagation neural network; ARIMA; Ensemble empirical mode decomposition	The results indicate that the forecast accuracy is more pronounced using ICEEMDAN-SCA-RVFL than using the back-propagation neural network, ARIMA and the ensemble empirical mode decomposition
Ben Jabeur et al., (2021a, 2021b)	January 2010–April 2020: daily frequency	LightGBM, CatBoost, XGBoost, Random Forest (RF), and neural network models	The main results confirm the superiority of RF and LightGBM over traditional models
Li et al. (2021)	January 2010–December	Variational mode decomposition (VMD) and	The proposed VMD-RSBL system is significantly more

Table 1 continued

References	Sample period	Method (s)	Main results
	2019: daily frequency	random sparse Bayesian learning (RSBL, SBL-based prediction with random lags and random samples)	efficient than many state-of-the-art systems
Ahmad et al. (2021)	February 1989–October 2019: daily frequency	Median ensemble empirical mode decomposition and Group method of Data Handling	The performance measures show that the new hybrid model (Median ensemble empirical mode decomposition and Group method of Data Handling) is superior to traditional models such as empirical mode decomposition, artificial neural network and ARIMA models

Information Administration website.² The EPU indicates the number of articles in newspapers that appear on the financial regulation, economy, monetary and trade policies, and uncertainty in the United States. This variable is downloaded from the EPU website.³ The VIX and OVX data are obtained from the Chicago Board Options Exchange website. Whaley (1993) was the first to develop a VIX to measure uncertainty in the financial market. This is also called “the fear sentiment.” Subsequently, the Chicago Board Options Exchange developed OVX to represent uncertainty in the oil market. The data are downloaded from the CBOE website.⁴ The sample covers the period from May 10, 2007, to August 09, 2021. The choice of the period is justified by the availability and continuity of the data. Thus, we have 3584 observations. The data are separated into two parts. The first part is devoted to model training with 2868 observations. The second part is used for model validation, with 716 observations.

Table 2 plots statistical results for all series data for uncertainty indexes and gas prices. All series have non-normal distribution. This is deduced from the Jarque–Bera statistics that are significant at 1%. Furthermore, the skewness statistics generate values that are different from 0 for all variables. This corroborates the asymmetrical distributions of the series. Additionally, all series have heavier tails than a normal distribution. This is proved by the kurtosis values that are much larger than 3. As well, Fig. 1 presents the correlation matrix between variables. Findings indicate that the oil price (LOP) has a negative and a weaker correlation with all uncertainty indices (LOVX, LVIX, and LEPU). However, the oil price is positively correlated with the gas price (0.60).

² <https://www.eia.gov/>.

³ <https://www.policyuncertainty.com/>.

⁴ <https://www.cboe.com/vix/>.

Table 2 Descriptive statistics

	Mean	Median	Max	Min	SD	Skewness	Kurtosis	Jarque–Bera	Observations
WTI	1.8269	1.8387	2.16229	0.94987	0.1555	-0.5600	3.510	226.2934	3584
GP	0.5412	0.5145	1.37767	0.12385	0.1767	0.79062	3.818	473.3466	3584
OVX	1.5481	1.5320	2.51208	1.16136	0.1591	0.83223	5.531	1370.591	3584
VIX	1.2683	1.2422	1.91745	0.96094	0.1670	0.88404	3.821	567.6611	3584
EPU	1.9978	1.9880	2.90722	0.52113	0.2720	0.03562	3.479	35.14986	3584

Table reports the descriptive statistics including the mean, median, standard deviation (SD), skewness, kurtosis, minimum (Min.), maximum (Max.) Jarque–Bera (JB) and number of observations (Obs.) of the daily innovations of Gas price and uncertainty indexes. The sample period extends from May 10, 2007 to August 09, 2021

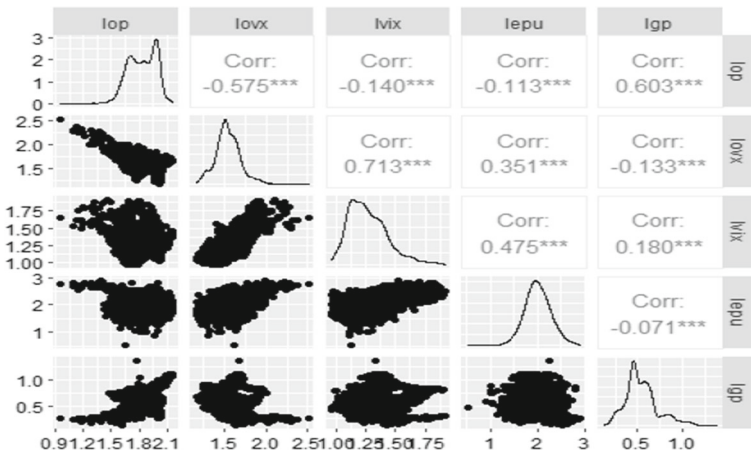


Fig. 1 Correlation matrix

Moreover, LOVX is highly correlated with LVIX (0.713). We then calculate the variance inflation factor to spot the multicollinearity which is reported in Table 3. The results show that the VIF is lower than 5 and 10 and there is no multicollinearity in the dataset.

For unit-root analysis, Fig. 2 plots the autocorrelation function (ACF). We observe that all series have the unit-root problem. It is clear that the ACF decreases in a hyperbolic manner and varies further from zero. Tissaoui (2019) indicates that the existence of a unit-root structure in a series corroborates the long-memory pattern in financial time series. Consequently, the differentiation of all series is important for obtaining stationary data before modeling and forecasting. The differentiation findings are illustrated in Fig. 3. We see that the ACF of the all-time series rapidly declines and varies near zero.

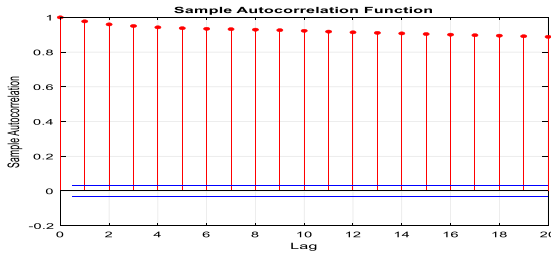
3.2 Methodology

ML involves the ability of computers to learn from a particular set of data and then apply it to another set of data. In this regard, Hao et al. (2020) further added that the ML tool has the potential to capture the hidden non-linear pattern and non-stationarity in the crude oil price series (Li et al. 2016, 2018). Li et al. (2021) reported that ML methods are currently the most popular time-series forecasting methods in the literature. One of the reasons for their attractiveness is that these techniques are more appropriate for continuous variables such as crude oil price time series. From this perspective, this study investigated the potential of several ML-based forecasting algorithms for crude oil prices. Three competitor approaches are used to attain this goal: the XGBoost method, SVM, and ARIMAX (p,d,q).

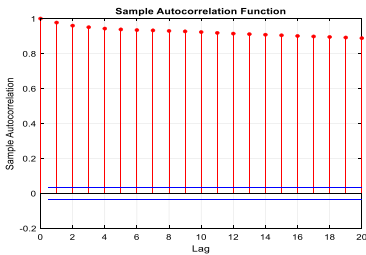
Table 3 Variance inflation factor (VIF)

Variables	VIF
lovx	1.492684
lvix	1.019923
lepu	1.012898
lgp	1.570897

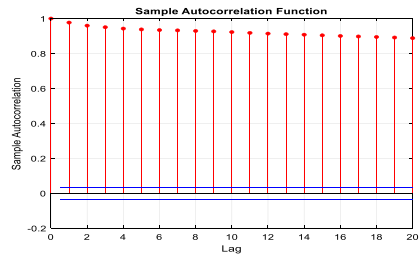
2.1. Autocorrelation of WTI



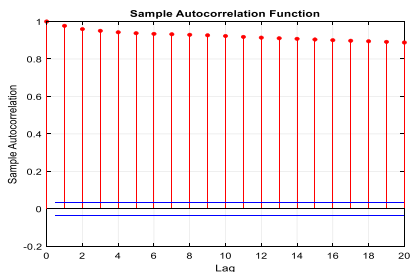
2.2. Autocorrelation of OVX



2.3. Autocorrelation of VIX



2.4. Autocorrelation of EPU



2.5. Autocorrelation of Gas

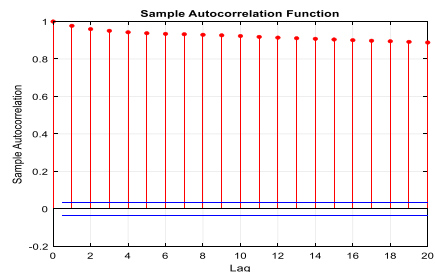
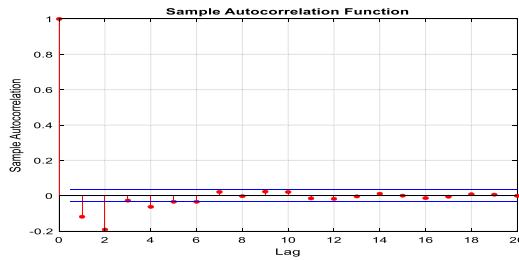


Fig. 2 Autocorrelation functions results of time-series. **a** Autocorrelation of WTI. **b** Autocorrelation of OVX. **c** Autocorrelation of VIX. **d** Autocorrelation of EPU. **e** Autocorrelation of gas

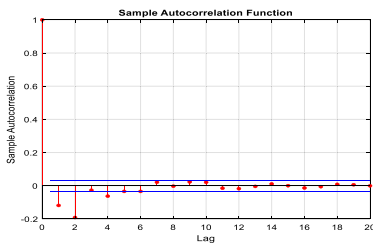
3.2.1 ARIMAX(p,d,q)

In this study, ARIMAX (p,d,q) is considered to be the reference model. The use of this model is motivated by its ability to consider the long-memory behavior shown

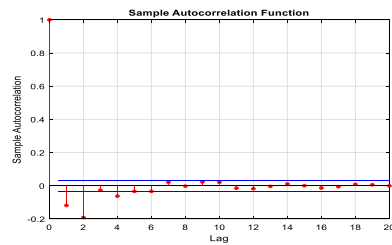
3.1. Autocorrelation of WTI



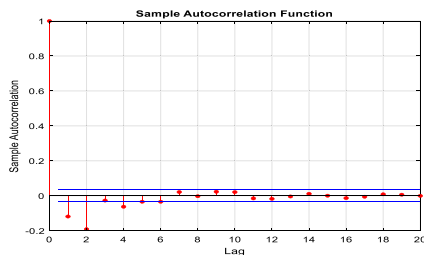
3.2. Autocorrelation of OVX



3.3. Autocorrelation of VIX



3.4. Autocorrelation of EPU



3.5. Autocorrelation of Gas

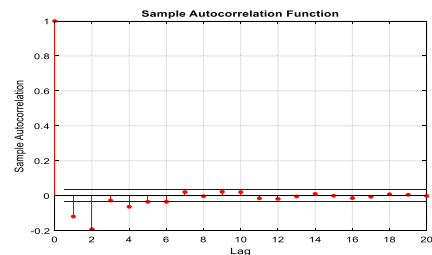


Fig. 3 Autocorrelation functions of one-differenced time-series. **a** Autocorrelation of WTI. **b** Autocorrelation of OVX. **c** Autocorrelation of VIX. **d** Autocorrelation of EPU. **e** Autocorrelation of gas

in the gas price and uncertainty indexes (see Fig. 3). The ARIMAX (p,d,q) model is expressed as:

$$\begin{aligned}
 \Delta^d WTI_t = & \delta + \sum_{i=1}^p \varphi_i WTI_{t-i} + \sum_{i=1}^k \omega_i GP_{t-i} + \sum_{i=1}^k \alpha_i OVX_{t-i} + \sum_{i=1}^k \beta_i VIX_{t-i} \\
 & + \sum_{i=1}^k \gamma_i EPU_{t-i} + \sum_{i=0}^q \sigma_i \varepsilon_{t-i}
 \end{aligned}
 \tag{1}$$

where WTI_t is the crude oil price; WTI_{t-i} are the previous values of the WTI; GP_{t-i} , OVX_{t-i} , VIX_{t-i} and EPU_{t-i} are the previous values of the gas price, OVX, VIX, and EPU, respectively. δ is a constant, and d denotes the order of integration. ω_i , φ_i , α_i , β_i , γ_i and σ_i represent the coefficients, and k , p , and q are the maximum time lags of the forecasters' sequences, output sequence, and residuals, respectively. The

identification of ARIMAX (p,d,q) is achieved using the box and Jenkins method. Subsequently, the Akaike information criterion (AIC) is applied to select the best model from several ARIMAX (p,d,q) models.

3.2.2 Support Vector Machine (SVM)

SVM is a supervised classification method developed by Vapnik (1997), which can also be used for regression using the structural risk minimization (SRM) principle for classification and regression. The SVM method assumes that for training data, $\{x_i, y_i\}_{i=1}^n$ where $x_i \in R^L$ is a vector of L input features, $y_i \in R^L$ is the output target, and (n) is the total number of data patterns. The aim of SVM is to find a function $f(x)$ that predicts the output value whose deviation is less than the insensitive loss parameter (ε) from the desired output y_i for all the training data, and at the same time, is as flat as possible (Smola & Schölkopf, 2004). The linear regression function in the low-dimensional space is mathematically described as follows:

$$f(x) = wx_i + b \quad (2)$$

where x is the weight vector that is normal to the hyperplane and b is the hyperplane bias.

The regression problem is transformed into an optimization problem as follows:

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \zeta_i \\ & \text{subjected} \\ & \text{to } \begin{cases} y_i - \langle \omega(x_i) \rangle - b \leq \varepsilon + \zeta_i \\ \langle \omega(x_i) \rangle + b - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0 \end{cases} \end{aligned} \quad (3)$$

where $\zeta_i, \zeta_i^* \in R$ are the slack variables and C is the penalty coefficient. The Lagrange multiplier is introduced to solve the optimization problem and the regression function takes the following form:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha'_i) k(x_i, x'_i) + b \quad (4)$$

where α_i, α'_i represents the Lagrange multiplier, $k(x_i, x'_i)$ is the kernel function. The Karush Kuhn-Tucker (KKT) conditions are used to compute (b) (Kuhn & Tucker, 1951; Smola & Schölkopf, 2004).

3.2.3 eXtreme Gradient Boosting (XGBoost) Method

XGBoost is an ML technique developed by Ostromski and Birman (2006) that can be used for regression and classification problems. This method has been adopted in different domains, such as healthcare (Singh et al., 2019) and the metal market (Ben Jabeur et al., 2021a, 2021b).

Based on the gradient direction of the loss function, it generates a weak learner at each step and accumulates it in the entire model. An objective function is normalized to prevent overfitting and to make the learning process faster. The model output function is given by the following:

$$\widehat{Y}_i^T = \sum_{k=1}^T f_k(x_i) = \widehat{y}_i^{T-1} + f_T(x_i) \tag{5}$$

where \widehat{y}_i^{T-1} represents the generated tree, $f_T(x_i)$ represents the newly created tree model, and T represents the total number of tree models. In addition, Ma et al. (2020) added that XGBoost is robust in terms of modeling nonlinear associations between variables. It has enormous classification ability. Accordingly, many researchers have indicated that ML is a powerful technique for forecasting time-series data. However, it does not provide interpretable inferences in traditional econometrics. To improve the performance of XGBoost, Lundberg and Lee (2017) proposed a Shapley additive explanation method (SHAP) to interpret the prediction of ML techniques based on game theory advanced by Shapley (1953). The SHAP approach allows us to explain the prediction of a specific input (X) by calculating the impact of each feature on prediction. The key idea of SHAP is to calculate the Shapley values for each feature of the sample to be interpreted, where each Shapley value represents the impact that the feature to which it is associated generates in the prediction. Moreover, ML models usually have a large number of features, where each feature is a discrete or continuous variable, which causes it to be computationally very complicated to calculate the Shapley values for each instance of each feature, and the SHAP method is more suitable for dealing with the issue of our research.

The estimated Shapley value is calculated as follows:

$$\widehat{\phi}_j = \frac{1}{K} \sum_{k=1}^K \left(\widehat{g}(x_{+j}^m) - \widehat{g}(x_{-j}^m) \right) \tag{6}$$

where $\widehat{g}(x_{+j}^m)$ is the prediction for x, but with a random number of feature values.

3.2.4 Performance Metrics

To assess the prediction performance of the different models, we use two criteria: the RMSE and MAE, which are computed as follows:

- $MAE = \frac{100}{N_v} \sum_{t=1}^{N_v} \frac{|\widehat{y}_t - y_t|}{\bar{y}}$ (7)

- $RMSE = \sqrt{\frac{1}{N_v} \sum_{t=1}^{N_v} (y_t - \widehat{y}_t)^2}$ (8)

where \widehat{y}_t is the predicted crude oil price, y_t is the t th current crude oil price, \bar{y}

represents the mean crude oil price, and N_v denotes the number of observations served in the validation phase of forecasting.

4 Empirical Results

4.1 Performance Analysis

In this section, we illustrate the empirical findings generated by the aforementioned models applied to examine the simultaneous impact of gas prices with uncertainty indices on the WTI crude oil price. Figure 4 shows the predicted and current series of crude oil prices, referring to which we compare the linear ARIMAX (p,d,q) model against both the linear SVM and non-linear XGBoost on a validation sample (20% of the sample). As shown in Fig. 4, both forecasting ML tools (e.g., the linear SVM and the nonlinear XGBoost) showed that the curves of the predicted values have almost the same behavior as the curve of the current values for the WTI crude oil. However, Fig. 4 shows that the curve relating to the ARIMAX (p,d,q) model has a very different pattern from the curve of the actual crude oil price. This implies that the forecasting ML method appears robust in terms of the prediction of crude oil prices. In addition, we find it difficult to determine the best-fit model between competitors' ML tools when relying on Fig. 4. Thus, we solve this problem by referring to the performance metrics (RMSE and MAE), which are depicted in Table 4. The forecasting tool with the lowest RMSE and MAE values is selected as the best-fit model. As shown in Table 4, the XGBoost model is dominant in predicting the WTI crude oil price compared with the linear SVM and ARIMAX models. Empirical evidence shows that the values of the performance metrics caused by the XGBoost model (RMSE=0.0581; MAE=0.0392) are the lowest compared to those of the SVM and ARIMAX models. Overall, the ML tool, as a complex model, outperforms the linear model in forecasting crude oil prices with good accuracy.

Fig. 4 Plot of crude oil prices forecast

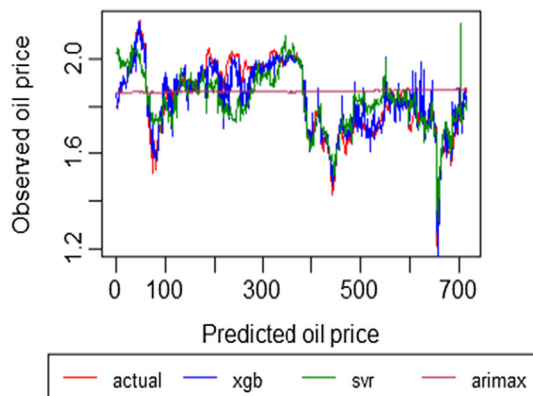
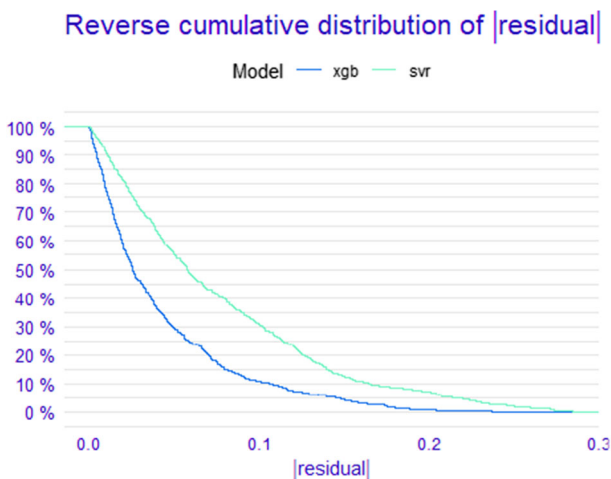


Table 4 Prediction assessment of the candidate models

Models	RMSE	MAE
ARIMAX(p,d,q)	0.2066	0.1599
SVM	0.1001	0.0768
XGboost	0.0581	0.0392

4.2 Feature Analysis

In this section, we focus only on the features of both forecasting machine-learning tools (e.g., linear SVM and nonlinear XGBoost). The ARIMAX model is eliminated from the feature analysis because it generated the worst findings. In addition, the Shapley additive explanation method (SHAP) is used to explain the effect of the gas and uncertainty index variables on the WTI crude oil price. Ben Jabeur et al., (2021a, 2021b) inferred that the Shapley additive explanation method can be used by policymakers and investors to understand ML results, which are characterized by their complexity. Before discussing the feature analysis, it is evident that the convergence of residuals for both linear SVM and nonlinear XGBoost. To achieve this, we use the DALEX R package proposed by Biecek and Burzykowski (2021) to explain the XGBoost and SVM models. The reverse cumulative of the absolute residual from Fig. 5 indicates that there is a lower number of residual in the left tail of the XGBoost residual distribution than the SVM model. The results show that the XGBoost model is more efficient than the SVM model in terms of convergence. These findings support the results of Climent et al. (2019) and Ben Jabeur et al., (2021a, 2021b), who showed the superiority of the XGBoost model over traditional models in gold price forecasting and credit scoring. In addition, our results are in line with Herrera et al. (2019), Wu et al. (2020), and Jabeur et al., (2021a, 2021b),

**Fig. 5** Residual convergence

who revealed that machine-learning tools outperform traditional models in forecasting crude oil prices.

Figures 6 and 7 display the SHAP values for both SVM and XGBoost models by sorting the features by the sum of the magnitudes of the SHAP values over all samples, and using the SHAP values to show the distribution of the impacts of each feature on the model output. The color denotes the value of the feature (high red, low blue). Figures 6 and 7 sort features by the sum of SHAP value magnitudes over all samples, and use SHAP values to show the distribution of the impacts each feature has on the model output.

The order variables are plotted based on their importance in influencing the WTI crude oil price in terms of forecasting. Each row displays a feature. A redder shape indicates that the feature has a superior value, and a bluer shape indicates that the feature has an inferior value. The SHAP values are plotted on the abscissa. In addition, a positive value of SHAP reflects the positive effect of the input on the output, whereas a negative value of SHAP indicates a negative effect of the input on the output. As observed in Fig. 6, LOVX is the most important feature according to the SVM model, which leads to a negative forecast of WTI crude oil. This had the greatest impact on the model. However, gas prices appear in the second position. Gas and oil prices are positively related. This implies that the uncertainty caused by investors' fear in the oil market is a good predictor of the WTI crude oil price, followed by the gas price. Moreover, the results show that LVIX has weaker feature importance in forecasting the WTI crude oil price. This positively impacts crude oil. This means that the uncertainty generated by investors' fear in the financial market is characterized by lower forecasting power for crude oil prices. In addition, the SVM model simulation shows that the LEPU appears to have a less important feature. Therefore, we infer that uncertainty caused by US economic policy has limits as a source of crude oil fluctuations. Similar results are obtained using the XGBoost model. Figure 7 shows that the LOVX is the most important variable that negatively impacts oil prices. More particularly, with a superior LOVX value, the WTI crude oil price may have a smaller probability of decreasing. However, gas price is the most important feature. This indicates that gas prices are an important source for crude oil forecasting. Contrary to the SVM model, the XGBoost model indicates that the forecasting power of LVIX is improved, but it remains less important than the LOVX and gas prices. We also conclude that LEPU exhibits an

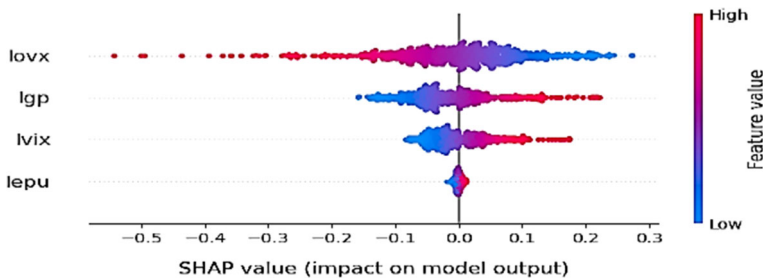


Fig. 6 SVM model: feature importance

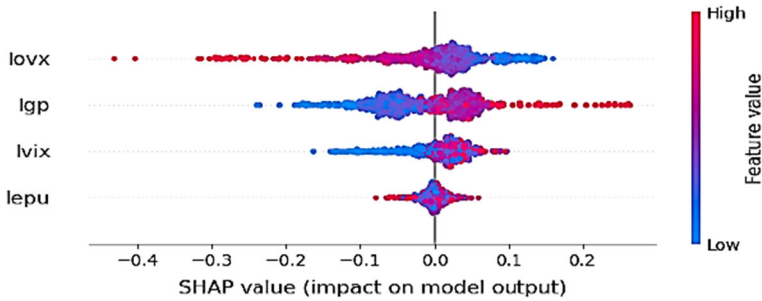


Fig. 7 XGBoost model: feature importance

enhancement in terms of feature importance compared to those shown by the SVM model. However, Figs. 8 and 9 sort the RMSE values to show the impacts that each feature has on the model output. In particular, Figs. 8 and 9 display the width of the interval bands that correspond to variable importance, while the bars indicate the RMSE loss after permutations for the XGBoost and SVM models. The XGBoost model has the lowest RMSE (Fig. 8) compared to the SVM model (Fig. 9). However, as shown previously, the XGBoost model dominates the SVM model in terms of accuracy, performance, and convergence, and the feature importance generated by the XGBoost model is more evident in this study. Thus, we infer that LOVX has a higher ability to predict the WTI crude oil price than the gas price.

5 Conclusion and Implications

This study examines the forecasting power of gas prices and uncertainty indices for crude oil prices. We attempt to compare two models of ML against linear models to determine which one is effective in forecasting the crude oil price using a dataset

Fig. 8 SVM model: RMSE loss after permutations

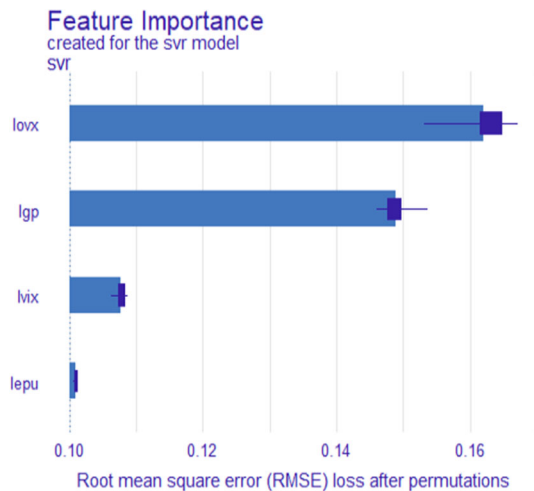
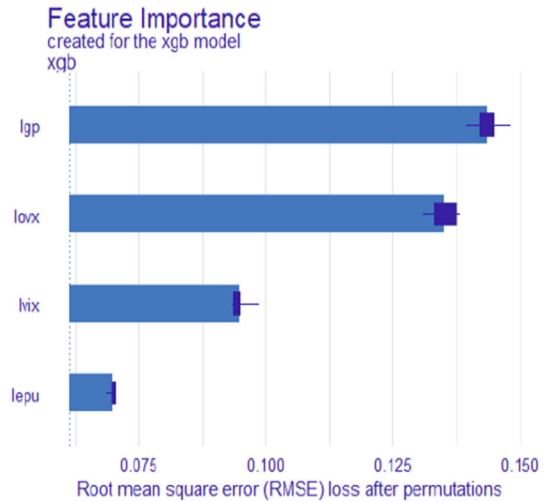


Fig. 9 XGBoost model: RMSE loss after permutations



from May 10, 2007, to August 09, 2021. In particular, we considered the SVM, XGBoost, and ARIMAX (p,d,q) models to examine the simultaneous effects of the uncertainty indices together with the gas price on crude oil price forecasting.

Interesting results are obtained through this study. First, considering the complex relationship between crude oil and its forecasters, the findings reveal the dominance of ML models, such as the SVM and XGBoost models, over traditional models. The performance metrics are the best in the ML models compared with the ARIMAX model. Second, after eliminating ARIMAX from the analysis, the XGBoost model appears to be superior to the SVM model in terms of accuracy and convergence. Third, the feature importance analysis realized by the Shapley additive explanation method shows that the different uncertainty indexes and gas prices display a significant ability to forecast future WTI crude prices. In addition, the SHAP values highlight that the informational content in LOVX dominates that in other uncertainty indexes and the gas price to forecast WTI crude oil. The results of this study have important policy implications for both investors and policymakers.

First, the investors' fear in the oil market (represented by OVX) is shown as the dominant forecaster among the gas price and other uncertainty indices affecting the WTI oil prices. This result shows that increased fear among investors can be a factor in fluctuations in oil prices. Therefore, traders should attach great importance to the main source of this fear, which is mainly investing in oil futures contracts. In other words, many oil futures contracts that are bought and sold in the derivatives markets are not designed as they are now for investment but for other purposes, such as hedging. Thus, the oil contract is a dangerous investment because it is not an investment in the first place and is not guaranteed at the time of settlement and the date of delivery. Therefore, inexperienced traders should not treat these contracts as investment tools and apply investment rules known as other investment tools, such as stocks. In the same context, investors should know that they are speculating in a narrow timeframe, and before the settlement date, they must have gotten rid of the

contract unless they want to receive oil with a place to store it. Indeed, this allows policymakers to be certain that realizing a stable energy market is mainly related to governing and controlling trading in oil futures contracts in order to limit the fear among investors and, consequently, to have a stable oil price. Second, the original approach, namely the XGBoost model, combined with the Shapley additive explanation method (SHAP), has a higher capacity to accurately consider the complex structure in the relationship between crude oil and its forecasters. In other words, the XGBoost tool showed that it has the ability to predict crude oil prices even if the sample used has periods of crisis.⁵ The good fit of this method will induce financial and energy authorities to profit from these complex tools to predict crude oil prices and other risky assets.

Third, the use of the ARIMAX model as a linear model does not succeed in accurately predicting crude oil prices. These limits compel both investors and policymakers to utilize the linear model to model the association between the crude oil price and its forecasters, which are characterized by complexity.

Although this study offers important findings and provides important policy implications for both investors and policymakers, it has some limitations. First, our study did not consider the COVID-19 pandemic as a variable that could affect the relationship between uncertainty indices and oil price fluctuations. Second, the research question was neither addressed before the emergence of the COVID-19 pandemic nor during its spread. The empirical findings of this study can be improved by addressing these limitations. In the future, research can discover new interpretable deep learning algorithms and more predictive uncertainty forecasters.

Acknowledgements This research has been funded by Scientific Research Deanship at University of Ha'il- Saudi Arabia through Project Number RG-20 201.

Declarations

Conflict of interest In this manuscript, all authors declare that we have no conflicts to report.

Ethical Standards All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000.

Human and Animal Rights No animal or human studies were carried out by the authors.

Informed Consent Human subjects were not considered in our manuscript.

References

- Abdollahi, H. (2020). A novel hybrid model for forecasting crude oil price based on time series decomposition. *Applied Energy*, 267, 115035. <https://doi.org/10.1016/j.apenergy.2020.115035>
- Ahmad, W., Aamir, M., Khalil, U., Ishaq, M., Iqbal, N., & Khan, M. (2021). A new approach for forecasting crude oil prices using median ensemble empirical mode decomposition and group method of data handling. *Mathematical Problems in Engineering*. <https://doi.org/10.1155/2021/5589717>

⁵ The April 20, 2020 crisis in which the oil price collapsed below zero.

- Aloui, R., Gupta, R., & Miller, S. M. (2016). Uncertainty and crude oil returns. *Energy Economics*, 55, 92–100. <https://doi.org/10.1016/j.eneco.2016.01.012>
- Alquist, R., Lutz, K., & Robert, V. (2013). Forecasting the price of oil. In G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting* (pp. 427–507). North-Holland.
- Atil, A., Lahiani, A., & Nguyen, D. K. (2014). Asymmetric and nonlinear pass-through of crude oil prices to gasoline and natural gas prices. *Energy Policy*, 65, 567–573. <https://doi.org/10.1016/j.enpol.2013.09.064>
- Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636. <https://doi.org/10.1093/qje/qjw024>
- Basu, S., & Bundick, B. (2017). Uncertainty shocks in a model of effective demand. *Econometrica*, 85(3), 937–958. <https://doi.org/10.3982/ECTA13960>
- Batten, J. A., Ciner, C., & Lucey, B. M. (2017). The dynamic linkages between crude oil and natural gas markets. *Energy Economics*, 62, 155–170. <https://doi.org/10.1016/j.eneco.2016.10.019>
- Baumeister, C., & Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics*, 30(2), 326–336. <https://doi.org/10.1080/07350015.2011.648859>
- Biecek, P., & Burzykowski, T. (2021). Explanatory model analysis: Explore, explain and examine predictive models. Chapman and Hall/CRC.
- Brown, S. P., & Yucel, M. K. (2008). What drives natural gas prices? *The Energy Journal*, 29(2), 45–60. <https://doi.org/10.5547/ISSN0195-6574-EJ-Vol29-No2-3>
- Cerqueti, R., Fanelli, V., & Rotundo, G. (2019). Long run analysis of crude oil portfolios. *Energy Economics*, 79, 183–205. <https://doi.org/10.1016/j.eneco.2017.12.005>
- Chai, J., Xing, L. M., Zhou, X. Y., Zhang, Z. G., & Li, J. X. (2018). Forecasting the wtI crude oil price by a hybrid-refined method. *Energy Economics*, 71, 114–127. <https://doi.org/10.1016/j.eneco.2018.02.004>
- Charles, A., & Darné, A. (2014). Volatility persistence in crude oil markets. *Energy Policy*, 65, 729–742. <https://doi.org/10.1016/j.enpol.2013.10.042>
- Charles, A., & Darné, A. (2017). Forecasting crude-oil market volatility: Further evidence with jumps. *Energy Economics*, 67, 508–519. <https://doi.org/10.1016/j.eneco.2017.09.002>
- Chen, S. S., & Chen, H. C. (2007). Oil prices and real exchange rates. *Energy Economics*, 29(3), 390–404. <https://doi.org/10.1016/j.eneco.2006.08.003>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
- Chen, W., Ma, F., Wei, Y., & Liu, J. (2020). Forecasting oil price volatility using high-frequency data: new evidence. *International Review of Economics & Finance*, 66, 1–12.
- Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. *Journal of Business Research*, 101, 885–896. <https://doi.org/10.1016/j.jbusres.2018.11.015>
- Conrad, C., Loch, K., & Rittler, D. (2014). On the macroeconomic determinants of long-term volatilities and correlations in us stock and crude oil markets. *Journal of Empirical Finance*, 29, 26–40. <https://doi.org/10.1016/j.jempfin.2014.03.009>
- Dutta, A. (2017). Modeling and forecasting oil price risk: The role of implied volatility index. *Journal of Economic Studies*, 44(6), 1003–1016. <https://doi.org/10.1108/JES-11-2016-0218>
- Dutta, A., Bouri, E., & Saeed, T. (2021). News-based equity market uncertainty and crude oil volatility. *Energy*, 222, 119930. <https://doi.org/10.1016/j.energy.2021.119930>
- Ftiti, Z., Tissaoui, K., & Boubakr, S. (2020). On the relationship between oil and gas markets: A new forecasting framework based on a machine learning approach. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-020-03652-2>
- Gatfaoui, H. (2016). Linking the gas and oil markets with the stock market: Investigating the US relationship. *Energy Economics*, 53, 5–16. <https://doi.org/10.1016/j.eneco.2015.05.021>
- Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of Political Economy*, 91(2), 228–248. <https://doi.org/10.1086/261140>
- Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, 113(2), 363–398. [https://doi.org/10.1016/S0304-4076\(02\)00207-5](https://doi.org/10.1016/S0304-4076(02)00207-5)
- Hao, X., Zhao, Y., & Wang, Y. (2020). Forecasting the real prices of crude oil using robust regression models with regularisation constraints. *Energy Economics*, 86, 104683. <https://doi.org/10.1016/j.eneco.2020.104683>

- Herrera, G. P., Constantino, M., Tabak, B. M., Pistori, H., Su, J. J., & Naranpanawa, A. (2019). Long-term forecast of energy commodities price using machine learning. *Energy*, 179, 214–221. <https://doi.org/10.1016/j.energy.2019.04.077>
- Huang, Y., & Deng, Y. (2021). A new crude oil price forecasting model based on variational mode decomposition. *Knowledge-Based Systems*, 213, 106669. <https://doi.org/10.1016/j.knsys.2020.106669>
- Jabeur, S. B., Khalfaoui, R., & Arfi, W. B. (2021b). The effect of green energy, global environmental indexes, and stock markets in predicting oil price crashes: Evidence from explainable machine learning. *Journal of Environmental Management*, 298, 113511. <https://doi.org/10.1016/j.jenvman.2021.113511>
- Jabeur, S. B., Meftah-Wali, S., & Viviani, J. L. (2021a). Forecasting gold price with the xgboost algorithm and shap interaction values. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-021-04187-w>
- Jo, S. (2014). The effects of oil price uncertainty on global real economic activity. *Journal of Money, Credit, and Banking*, 46(6), 1113–1135. <https://doi.org/10.1111/jmcb.12135>
- Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability* (pp. 481–492). University of California Press.
- Li, T., Hu, Z., Jia, Y., Wu, J., & Zhou, Y. (2018). Forecasting crude oil prices using ensemble empirical mode decomposition and sparse Bayesian learning. *Energies*, 11(7), 1882.
- Li, T., Zhou, M., Guo, C., Luo, M., Wu, J., Pan, F., & He, T. (2016). Forecasting crude oil price using EEMD and RVM with adaptive PSO-based kernels. *Energies*, 9(12), 1014.
- Li, R., Hu, Y., Heng, J., & Chen, X. (2021). A novel multiscale forecasting model for crude oil price time series. *Technological Forecasting & Social Change*, 173, 121181. <https://doi.org/10.1016/j.techfore.2021.121181>
- Liu, J., Ma, F., Yang, K., & Zhang, Y. (2018). Forecasting the oil futures price volatility: Large jumps and small jumps. *Energy Economics*, 72, 321–330. <https://doi.org/10.1016/j.eneco.2018.04.023>
- Liu, Y., Wei, Y., Liu, Y., & Li, W. (2020). Forecasting oil price by hierarchical shrinkage in dynamic parameter models. *Discrete Dynamics in Nature and Society*. <https://doi.org/10.1155/2020/6640180>
- Lu, F. B., Hong, Y. M., Wang, S. Y., Lai, K. K., & Liu, J. (2014). Time-varying granger causality tests for applications in global crude oil markets. *Energy Economics*, 42, 289–298. <https://doi.org/10.1016/j.eneco.2014.01.002>
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Lyu, Y., Tuo, S., Wei, Y., & Yang, M. (2021). Time-varying effects of global economic policy uncertainty shocks on crude oil price volatility: New evidence. *Resources Policy*, 70, 101943. <https://doi.org/10.1016/j.resourpol.2020.101943>
- Ma, J., Cheng, J. C., Xu, Z., Chen, K., Lin, C., & Jiang, F. (2020). Identification of the most influential areas for air pollution control using XGBoost and Grid Importance Rank. *Journal of Cleaner Production*, 274, 122835
- Nonejad, N. (2021). Forecasting crude oil price volatility out-of-sample using news-based geopolitical risk index: What forms of nonlinearity help improve forecast accuracy the most? *Finance Research Letters*. <https://doi.org/10.1016/j.frl.2021.102310>
- Ostrowski K., & Birman. K. (2006). Extensible web services architecture for notification in large-scale systems. In *International Conference on Web Services*. IEEE. <https://doi.org/10.1109/ICWS.2006.63>
- Pastor, L., & Veronesi, P. (2012). Uncertainty about government policy and stock prices. *The Journal of Finance*, 67(4), 1219–1264. <https://doi.org/10.1111/j.1540-6261.2012.01746.x>
- Phan, D. H. B., Tran, V. T., & Nguyen, D. T. (2019). Crude oil price uncertainty and corporate investment: New global evidence. *Energy Economics*, 77, 54–65. <https://doi.org/10.1016/j.eneco.2018.08.016>
- Rubaszek, M. (2020). Forecasting crude oil prices with DSGE models. *International Journal of Forecasting*, 37(2), 531–546. <https://doi.org/10.1016/j.ijforecast.2020.07.004>
- Salisu, A. A., & Fasanya, I. O. (2013). Modelling oil price volatility with structural breaks. *Energy Policy*, 52, 554–562. <https://doi.org/10.1016/j.enpol.2012.10.003>
- Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*. <https://doi.org/10.1515/9781400881970-018>
- Singh, N., Singh, P., & Bhagat, D. (2019). A rule extraction approach from support vector machines for diagnosing hypertension among diabetics. *Expert Systems with Applications*, 130, 188–205. <https://doi.org/10.1016/j.eswa.2019.04.029>

- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- Sreenu, N. (2018). The effects of oil price shock on the Indian economy—A study. *The Indian Economic Journal*, 66(1–2), 190–202. <https://doi.org/10.1177/0019466219876491>
- Tissaoui, K. (2019). Forecasting implied volatility risk indexes: International evidence using Hammerstein-ARX approach. *International Review of Financial Analysis*, 64, 232–249
- Van Robays, I. (2016). Macroeconomic uncertainty and oil price volatility. *Oxford Bulletin of Economics and Statistics*, 78(5), 671–693. <https://doi.org/10.1111/obes.12124>
- Vapnik, V. N. (1997, October). The support vector method. In *International Conference on Artificial Neural Networks* (pp. 261–271). Springer, Berlin, Heidelberg.
- Wang, K. H., Su, C. W., & Umar, M. (2021). Geopolitical risk and crude oil security: A Chinese perspective. *Energy*, 219, 119555.
- Wang, J., & Wang, J. (2016). Forecasting energy market indices with recurrent neural networks: Case study of crude oil price fluctuations. *Energy*, 102, 365–374. <https://doi.org/10.1016/j.energy.2016.02.098>
- Wang, Q., & Sun, X. (2017). Crude oil price: Demand, supply, economic activity, economic policy uncertainty and wars—From the perspective of structural equation modelling (sem). *Energy*, 133, 483–490. <https://doi.org/10.1016/j.energy.2017.05.147>
- Wang, Y., & Wu, C. (2012). Energy prices and exchange rates of the us dollar: Further evidence from linear and nonlinear causality analysis. *Economic Modelling*, 29(6), 2289–2297. <https://doi.org/10.1016/j.econmod.2012.07.005>
- Wang, Y. S. (2013). Oil price effects on personal consumption expenditures. *Energy Economics*, 36, 198–204. <https://doi.org/10.1016/j.eneco.2012.08.007>
- Wen, F., Min, F., Zhang, Y. J., & Yang, C. (2019). Crude oil price shocks, monetary policy, and China's economy. *International Journal of Finance & Economics*, 24(2), 812–827. <https://doi.org/10.1002/ijfe.1692>
- Whaley, R. E. (1993). Derivatives on market volatility: Hedging tools long overdue. *The Journal of Derivatives*, 1(1), 71–84.
- Wu, B., Wang, L., Lv, S. X., & Zeng, Y.-R. (2021). Effective crude oil price forecasting using new text-based and big-data-driven model. *Measurement*, 168, 108468. <https://doi.org/10.1016/j.measurement.2020.108468>
- Wu, J., Miu, F., & Li, T. (2020). Daily crude oil price forecasting based on improved CEEMDAN, SCA, and RVFL: A case study in WTI oil market. *Energies*, 13, 1852. <https://doi.org/10.3390/en13071852>
- Yang, C., Gong, X., & Zhang, H. (2019). Volatility forecasting of crude oil futures: The role of investor sentiment and leverage effect. *Resources Policy*, 61, 548–563. <https://doi.org/10.1016/j.resourpol.2018.05.012>
- Yang, L. (2019). Connectedness of economic policy uncertainty and oil price shocks in a time domain perspective. *Energy Economics*, 80, 219–233. <https://doi.org/10.1016/j.eneco.2019.01.006>
- Yi, A., Yang, M., & Li, Y. (2021). Macroeconomic uncertainty and crude oil futures volatility—Evidence from china crude oil futures market. *Frontiers in Environmental Science*, 9, 21. <https://doi.org/10.3389/fenvs.2021.636903>
- Zhang, J. L., Zhang, Y. J., & Zhang, L. (2015). A novel hybrid method for crude oil price forecasting. *Energy Economics*, 49, 649–659. <https://doi.org/10.1016/j.eneco.2015.02.018>
- Zhu, H. M., Li, R., & Li, S. (2014). Modelling dynamic dependence between crude oil prices and Asia-Pacific stock market returns. *International Review of Economics & Finance*, 29, 208–223. <https://doi.org/10.1016/j.iref.2013.05.015>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Kais Tissaoui^{1,3} · Taha Zaghdoudi^{1,4} · Abdelaziz Hakimi² · Mariem Nsaibi¹

✉ Kais Tissaoui
kaistissaoui@yahoo.fr; ks.tissaoui@gmail.com

- ¹ Management Information Systems Department, University of Ha'il, Applied College, PO Box 2440, Hail City, Saudi Arabia
- ² Faculty of Law, Economics and Management of Jendouba, University of Jendouba and V.P.N.C Lab FSJEG, Jendouba, Tunisia
- ³ Faculty of Economic Sciences and Management of Tunis, The International Finance Group, University of Tunis El Manar, Tunis, Tunisia
- ⁴ LR-LEFA, IHEC Carthage, University of Carthage, Carthage, Tunisia