Proceedings of CONF-BPS 2026 Symposium: GenAI, Labour Markets, and the Economics of Human and Financial Capital
DOI: 10.54254/2754-1169/2026.LD31510

Forecasting Crude Oil Price Volatility: A Comparative Study of ARIMA, GARCH, and Hybrid Machine Learning Models

Jikai Tang

Antai College of Economics & Management, Shanghai Jiao Tong University, Shanghai, China
2020jd237t@sjtu.edu.cn

Abstract. Predicting the volatility of crude oil prices is of crucial significance, yet it poses challenges because of the intricate interaction process between geopolitical factors and economic factors. In this study, a systematic and comparative analysis is carried out on traditional econometric models and modern machine-learning models for the volatility of WTI crude oil. This study starts by making use of ARIMA and GARCH-family models (including GARCH and APGARCH) to capture the linear patterns and the asymmetric volatility clustering phenomenon. In order to deal with the nonlinear limitations of these parametric models, this study then puts forward a brand-new hybrid framework that combines the APGARCH model with an Artificial Neural Network (ANN). By using the WTI spot price data, the empirical results of this paper show the better performance of the hybrid approach. Specifically speaking, the APGARCH-ANN model reduces the forecast errors by more than 80% when it comes to RMSE in comparison with the standalone APGARCH model. This notable improvement emphasizes the synergy that exists between the econometric feature extraction process and the machine learning's ability to perform nonlinear approximation. The findings offer a robust and advanced forecasting tool for energy firms as well as financial institutions, which can enhance the risk management process and the strategic decision - making process when facing the volatile oil markets.

Keywords: Volatility Process of Crude Oil Price, Process of Volatility Forecasting, Hybrid Forecasting Models of Such Kind, GARCH Family Models, Artificial Neural Networks (ANN)

1. Introduction

1.1. The significance of crude oil price forecasting

Crude oil, which is represented by benchmarks like WTI and Brent, is the most strategically vital commodity in the world. In the process of evaluating global economic stability, energy security and geopolitical risks, its price volatility serves as a key indicator. AK Patidar and other researchers carried out a systematic analysis of the correlation between the occurrence of significant fluctuations in crude oil prices and global events [1]. For major energy corporations, airlines, as well as financial institutions, to conduct an accurate forecast of crude oil volatility has become a core requirement for them to carry out risk management.

DOI: 10.54254/2754-1169/2026.LD31510

1.2. Methodological foundations: from ARIMA to GARCH

1.2.1. Stationarity and ARIMA modeling

Xie and his colleagues made use of the ARIMA model to carry out the task of forecasting the reliability of repairable systems. They put forward the point that the ARIMA model is an effective approach for the process of time series analysis. It has the ability to produce satisfactory outcomes when it comes to the aspect of forecasting performance [2]. Edward Herranz made an indication that linear regressions that involve unit root processes frequently give rise to spurious regression results [3]. Unit root tests, for example the Augmented Dickey-Fuller (ADF) test, carry out the assessment to see whether a time series reverts to a constant mean as time goes by. This is a prerequisite for the process of valid ARIMA modeling. This situation reminded the author that the operation of unit root testing must be carried out on the data before the ARIMA model is applied. In the case where the data fails the test, the ARIMA model cannot be used in a direct manner. Instead, transformations (such as log or Box - Cox transformations) should be applied to the data so as to pass the test. This kind of methodological approach ensures that ARIMA models can generate reliable forecasts when they are applied to the process of a stationary time series.

1.2.2. Capturing time-varying volatility

While the techniques of traditional volatility modeling, for example, ARIMA, have shown their utility, they are unable to deal with the volatility clustering that is commonly witnessed in financial time series. This is because of their assumption that the variance remains constant. Engle carried out an analysis and found that financial time series often show heteroscedasticity. In this situation, the variance of the forecast error changes as time goes on. This situation then led to the introduction of the ARCH (Autoregressive Conditional Heteroscedasticity) model. The purpose of introducing this model is to capture the time - varying volatility as well as the volatility clustering [4]. This kind of innovation addressed a basic limitation of ARIMA. It did so by enabling the volatility to change according to the market conditions. This situation then made the author think about whether it is necessary to make use of the GARCH model to conduct a fitting process on the volatility of crude oil price data.

1.3. Beyond linearity: the need for advanced modeling

Although the models belonging to the GARCH family are able to successfully capture this kind of heteroskedasticity and the leverage effects that are caused by negative shocks, they are parametric models. In the process of trying to describe the nonlinear dependencies in crude oil volatility that are driven by complex and sudden events, they encounter difficulties. Therefore, the necessity of this study is found in the fact that it aims to overcome the limitations of traditional GARCH models when it comes to capturing the nonlinear residual information of crude oil price data. Heba Dhaher Alwan and his colleagues carried out an action of employing the APGARCH and NAGARCH models. They used these models to estimate parameters for the return series of the Iraqi Dinar to U.S. Dollar exchange rate (IQ/USD). The time frame they considered was from July 21, 2011, to July 21, 2021. After that, these parameter estimates were used to generate the data. The identification process was carried out through Ljung-Box and ARCH tests, where there were 1,000 replications. The results indicated the presence of autocorrelation and heteroskedasticity. Moreover, the situation was that these effects became more intense as the sample size increased. More specifically, in the situation where the NAGARCH model was under the normal distribution, it performed the best. On the other hand, the APGARCH model achieved the optimal results when it was combined with the generalized error distribution. These findings put forward the suggestion that when model selection is carried out, both the distributional assumptions and the asymmetric effects should be taken into consideration. This research put forward the suggestion that the author can take actions to employ other models from the GARCH family to perform the process of fitting crude oil price data.

1.4. Research gap and contributions

Most scholars have directed their attention to the prices of other commodities. However, within the existing body of research, there are only a limited number of articles that specifically carry out studies on the price of crude oil. Additionally, in the process of conducting research, the majority of the existing research has employed either traditional econometric models or modern machine - learning models. And then, there are a few studies that engage in the process of comparing these two types of models. In conclusion, in the previous studies, there has not been a comprehensive analysis carried out on the fitting performance of ARIMA, GARCH, and ANN models for crude oil prices. This paper takes on the task of filling this kind of research gap.

First of all, the author carried out the process of performing basic cleaning and then did the task of carrying out imputation on the crude oil data. After that, the author made use of the ADF test to carry out an assessment of whether the data were suitable for the ARIMA model. After this assessment, the author went on to carry out the action of applying necessary transformations to the data. In the situation after fitting the ARIMA model, the ARCH - LM test was used to carry out a determination of whether a GARCH model was necessary to capture the volatility characteristics. Then, in order to conduct an investigation into the asymmetric effects of positive and negative shocks on crude oil prices, the author introduced the APARCH model. At last, the author carried out the combination of traditional econometric models with machine learning models by making use of the GARCH-ANN model to carry out an evaluation of the predictive performance on crude oil data. The commercial value of this kind of research lies in the fact that it provides reliable forecasting tools for quantitative funds and investment banks, so as to let the goal of achieving excess returns be realized. Moreover, the research improves the quality of decision-making regarding national energy security reserves and reduces the adverse impacts of oil price volatility on global supply chains and inflation. Therefore, this study has significant theoretical importance and practical application value.

2. Method

2.1. Data collection and preprocessing

The study makes use of the WTI crude oil spot price data which is obtained from the situation where the U.S. Energy Information Administration (EIA) has published the WTI crude oil spot price data. The data source system is FRED (Federal Reserve Economic Data), and this is an economic database. This database is maintained by the Federal Reserve, and the Federal Reserve aggregates hundreds of data sources coming from U.S. government agencies as well as international organizations. In addition, it provides free and reliable API access. The act of selecting this kind of data offers the following kinds of advantages: the EIA is an official U.S. agency, and this situation ensures that the data is reliable; it provides long - term and continuous data; and it is available on a daily basis, which makes it suitable for the process of conducting analysis across various time

DOI: 10.54254/2754-1169/2026.LD31510

scales. The act of choosing WTI crude oil spot prices as the research subject also has significant economic implications: crude oil prices have a direct influence on inflation, they reflect the health of the economy, and they have an impact on multiple industries, ranging from the chemical industry to the transportation industry.

2.2. Stationarity testing and transformation

For the purpose of making the research process more convenient, the author initially carried out an operation of filling in the missing values. Specifically, the author used the previous value (which is the forward-fill method) to perform this filling-in process. After that, the author took the step of converting the data. This conversion was achieved by taking the average of the data so as to transform it into monthly data. Next, the author conducted an action of carrying out an ADF test on the original data. During this test, the author found that the original data approximated a random walk. This characteristic made the original data not suitable for being directly used in an ARIMA model. After that, the author applied a series of operations to the data. First, the author carried out a logarithmic transformation on the data, and then followed it with a first-order differencing process. Finally, the processed data was put through an ADF test, and it passed this test.

2.3. Model specification and estimation

After the Box-Jenkins methodology had its influence on the research process, an action of conducting model identification took place. This was done by means of examining the ACF plot and the PACF plot of the data. Then, the author went on to select a certain model, which was an ARIMA(1,1,1) model. The selection of this ARIMA(1,1,1) model was subject to further validation. This validation was carried out by using information criteria, which included the AIC as well as the BIC, along with residual diagnostics. And this whole validation process was carried out under the situation where the Box-Jenkins methodology was followed.

However, Engle and other researchers made an observation. They found that, in the process of dealing with time-series data, the volatility within the time series often cannot be captured in an adequate manner just by using ARIMA models on their own. This situation implies that there is a need to carry out an action. That is, to apply an ARCH-LM test so as to make a determination about whether a GARCH model should be employed further [4]. The ARCH-LM test was carried out under the condition where 12 lags were set. The purpose of this test was to detect the persistence of volatility over a time span of one year.

Since the data that has undergone processing incorporates information from previous periods, in other words, during the process of dealing with the data, information from previous periods has been integrated into it. Tim Bollerslev’s paper demonstrates this situation. Specifically, through the process of carrying out an empirical case study on inflation rate uncertainty, it shows that past conditional variances can also be included in the current conditional variance equation when using GARCH-type models [5]. So, the author carried out an action of conducting an ARCH-LM test. Then, based on the results obtained from the test, the author went on to develop a GARCH model.

To carry out the process of modeling the asymmetry that exists between the upward and the downward price trends, the author took action by following the approach put forward by Heba Dhaher Alwan and other scholars. Then, the author carried out the operation of fitting an APGARCH model [6].

DOI: 10.54254/2754-1169/2026.LD31510

2.4. Hybrid model development

Finally, the author made a reference to the findings of Werner Kristjanpoller R and Esteban Hernández P et al. [7]. Their findings demonstrated that hybrid models, which are the models that combine neural networks with traditional econometric models, can bring about an improvement in predictive performance. On the basis of this situation, the author carried out an experiment on a hybrid ARIMA+APGARCH+ANN model. This hybrid approach is one that combines the interpretability of econometric models with the flexibility of machine learning, and it can address both the linear dependencies and the nonlinear dependencies in crude oil volatility. The author also conducted tests on different numbers of hidden layers. Hyperparameter tuning was carried out through the use of grid search along with 5 - fold cross - validation so as to let the optimal network architecture be determined.

3. Result

3.1. Exploratory data analysis

According to Figure 1, the WTI crude oil price data spanning from 2000 to 2025 show the typical characteristics of financial time series. These characteristics are marked by the fact that there is significant structural volatility and event-driven attributes. When this study considers the long-term situation, the price center has gone through a situation where it has made a noticeable upward shift. It started from the early range, where the price was between 20 dollars and 80 dollars, and then demonstrated strong volatility clustering. The series is highly sensitive to major external shocks. It clearly reflects the drastic price fluctuations that are triggered by key events. These key events include the 2008 financial governance crisis, the 2014 shale oil revolution, the 2020 COVID-19 pandemic, and the 2022 geopolitical conflicts. In 2020, there was even a situation where an extreme negative anomaly occurred. These events emphasize the need for modeling approaches that are able to take into account structural breaks and extreme events. The data also shows evident cyclical fluctuations and possible seasonal patterns. However, it is a complex interplay of macroeconomic conditions, geopolitical factors, and supply-demand dynamics that drives its overall trend. So, when this study is in the process of analyzing this series, the paper needs to pay special attention to volatility modeling.

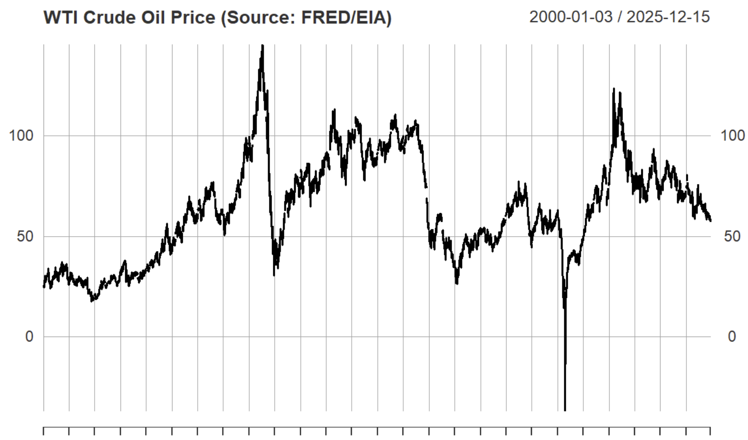

Figure 1. WTI crude oil price (photo credit: origin)

DOI: 10.54254/2754-1169/2026.LD31510

3.2. Stationarity testing and transformation

3.2.1. Unit root test results

According to Table 1, the results of the ADF test show that, in the situation where the significance level is set at 5%, the monthly WTI crude oil price series demonstrates non-stationary characteristics (the p-value is 0.3674 and the Dickey-Fuller statistic is equal to - 2.4958). This kind of situation implies that this series has a unit root. In other words, its statistical properties, for example, the mean and the variance, change as time goes on. It may potentially show random walk or trend components. This type of behavior does not meet the fundamental assumption of stationarity, which is required in the process of time series analysis. If one directly conducts modeling on the non-stationary price series, it can easily result in the occurrence of spurious regression problems. This will make the statistical inferences invalid. Therefore, this study gives a recommendation that the original series should be transformed. Usually, this can be done by taking first differences or calculating logarithmic returns. By doing so, the data can be converted into a stationary form before time series models (such as ARIMA or GARCH-type models) are applied. This kind of approach can ensure that the analytical results are reliable and valid.

3.2.2. Stationarity transformation

Based on the situation this study examines, Table 1 as well as Table 2, and the data that have undergone the transformation process pass the ADF test.

Table 1. ADF test result of original data

Item	Value/Description
Data Series	DCOILWTICO_monthly
Dickey-Fuller Statistic	-2.4958
Lag Order	6
p-value	0.3674
Alternative Hypothesis	The series is stationary.
Test Conclusion	p-value > 0.05, fail to reject the null hypothesis; the series is non-stationary.

Table 2. ADF test result of transformed data

Item	Value/Description
Data Series	returns_log
Dickey-Fuller Statistic	-7.1036
Lag Order	6
p-value	0.01
Alternative Hypothesis	The series is stationary
Test Conclusion	p-value < 0.05, reject the null hypothesis; series is stationary

3.3. ARIMA model estimation

3.3.1. Model identification

According to Figure 2 and Figure 3, the author carried out an observation. In the process of this observation, it was found that both the ACF plot and the PACF plot show a situation where they

DOI: 10.54254/2754-1169/2026.LD31510

exhibit truncation at lag 1. This situation indicates that the ARIMA(1,1,1) model is the most suitable choice among all the available models. Jadhav, V., Chinnappa Reddy, B. V., and their colleagues conducted a validation process. In this process, they validated the price forecasts for major crops such as paddy, ragi, and maize in Karnataka state, India, for the year 2016 [8]. They adopted a method of applying univariate ARIMA techniques. By using this method to generate grain price predictions, they finally found that the ARIMA price forecast results completely demonstrate the effectiveness of this model when it is used as a price forecasting tool [8]. This kind of situation suggests that the author may make an attempt to carry out a fitting process. In this process, the author may try to fit crude oil prices (which are the log returns of crude oil) by using the ARIMA model.

3.3.2. Residual diagnostics

Based on the data presented in Table 3, the author carried out a process of performing a residual analysis on the ARIMA model. During this process of analysis, it was discovered that the residuals can be regarded as being in a state where they are independent and identically distributed (i.i.d.) white noise. This kind of diagnostic result satisfies a crucial assumption within the process of ARIMA modeling. Satisfying this assumption ensures the validity of the inferences that will be made in the subsequent process.

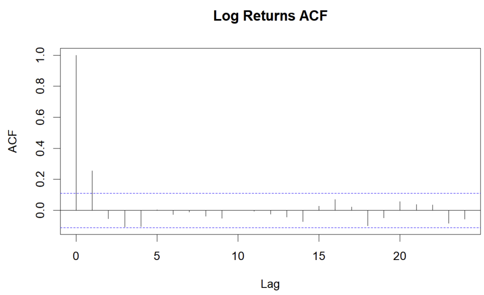

Figure 2. Log returns ACF (photo credit: origin)

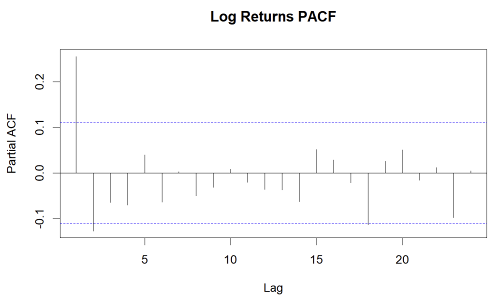

Figure 3. Log returns PACF (photo credit: origin)

DOI: 10.54254/2754-1169/2026.LD31510

Table 3. Residual analysis

Item	Value/Description
Model Type	ARIMA(1,1,1)
Data Series	returns_log
Coefficients	ar1 = 0.259
	ma1 = -1.0000
Standard Errors	s.e. (ar1) = 0.055
	s.e. (ma1) = 0.0097
Model Fit Metrics	sigma^2 = 0.00951
	Log Likelihood = 279.1
	AIC = -552.21
Box-Ljung Test	
Data Source	model_arima_111$residuals
X-squared Statistic	0.2912
Degrees of Freedom	1
p-value	0.5895
Test Conclusion	p-value > 0.05, no significant autocorrelation in residuals

3.4. Volatility modeling

3.4.1. ARCH effects testing

To capture the volatility characteristics that are present in the process of analyzing the data, the author carried out an operation to perform the ARCH-LM test on the squared residuals of the ARIMA model. Based on the situation where the author referred to Table 4, the author discovered that the squared residuals showed a situation where they exhibited significant autocorrelation and then failed the ARCH-LM test.

3.4.2. GARCH model fit

Therefore, the author carried out an employment of a GARCH(1,1) model. In the situation where the author referred to Table 5, it was found that all p-values for the tests conducted on the squared residuals were far greater than 0.05. This kind of situation indicates that the model has managed to capture all autocorrelation and volatility structures in the crude oil return series, and the residuals of this model are white noise.

Table 4. ARCH-LM test of ARIMA model

Item	Value/Description
Test Name	ARCH LM Test
Null Hypothesis	No ARCH effects
Data Source	residuals_arima
Chi-squared Statistic	84.182
Degrees of Freedom	12
p-value	6.536e-13
Test Conclusion	p-value < 0.05, reject the null hypothesis; significant ARCH effects are present

DOI: 10.54254/2754-1169/2026.LD31510

Table 5. Test results of GARCH model

Test Item	Symbol	Statistic / Test Value	p-value
Jarque-Bera Test	R	Chi^2 = 61.2390604	5.040413e-14
Shapiro-Wilk Test	R	W = 0.9767999	6.343412e-05
Ljung-Box Test (Residuals)	R	Q(10) = 7.8688313	0.6416476
	R	Q(15) = 12.7794761	0.6193265
	R	Q(20) = 15.4755610	0.7485803
Ljung-Box Test (Squared Residuals)	R^2	Q(10) = 4.5367452	0.9199033
	R^2	Q(15) = 7.2941241	0.9489681
	R^2	Q(20) = 13.0042719	0.8772007
LM Arch Test	R	TR^2 = 5.5473729	0.9371609
Main Conclusion	Residuals are not normally distributed, but show no autocorrelation or ARCH effects		

3.5. Asymmetric volatility analysis

To carry out the process of modeling the asymmetric impact that price and return fluctuations have between their upward and downward movements, the author made use of the APGARCH (Asymmetric Power GARCH) model. Table 6 shows the values of AIC and BIC for the GARCH model. In the meantime, Table 7 presents the values of AIC and BIC for the APGARCH model. When this study bases the analysis on Tables 6 and 7, this study can find that both the AIC value and the BIC value of the APGARCH model are lower than those of the GARCH model. As a result, the act of introducing the APGARCH model leads to an improvement in the goodness - of - fit situation.

Table 6. AIC and BIC values for the GARCH model

Information Criterion Type	Value
AIC (Akaike Information Criterion)	-2.058067
BIC (Bayesian Information Criterion)	-1.985917
SIC (Schwarz Information Criterion)	-2.058793
HQIC (Hannan-Quinn Information Criterion)	-2.029227

Table 7. AIC and BIC values for the APGARCH model

Information Criterion Type	Value
AIC (Akaike Information Criterion)	-2.120372
BIC (Bayesian Information Criterion)	-2.024172
SIC (Schwarz Information Criterion)	-2.121652
HQIC (Hannan-Quinn Information Criterion)	-2.081920

3.6. Hybrid model performance

3.6.1. Model architecture

Zhao, along with his colleagues, took actions to bridge the gap that existed between econometrics and machine learning. These two fields had, in the past, been developed in an independent manner with very little interaction taking place between them. They achieved this by establishing an equivalence relationship between GARCH family models and their corresponding neural network

DOI: 10.54254/2754-1169/2026.LD31510

models. They put forward an innovative method, which they named GARCH-NN, for the purpose of constructing volatility models that are based on neural networks [9]. Han and his colleagues carried out an integration process where they combined Convolutional Neural Networks with the Generalized Autoregressive Conditional Heteroskedasticity model. Through this process, they discovered that the hybrid model was capable of comprehensively capturing and analyzing the complex features of time-series data across different time intervals. They conducted an evaluation of the model on multiple typical time-series datasets. The results of this evaluation demonstrated that this kind of model showed a higher level of predictive accuracy and stability when compared to traditional methods [10]. Bahareh Amirshahi and his colleagues also engaged in a combination activity where they combined the APGARCH model with ANN. They found that the hybrid model was excellent at the task of forecasting [11]. This situation inspired the author to carry out a hybridization process of the APGARCH model with ANN on crude oil price datasets.

The author initially carried out the process of extracting the linear volatility characteristics from the time series. This was done by making use of the GARCH model. These features that were extracted, together with variables like lagged realized volatility, were subsequently taken and used as inputs. The inputs were fed into a double - hidden - layer neural network that has a structure where there are 4 neurons in the first layer, then 8 neurons in the second layer, then 4 neurons in the third layer, and finally 1 neuron in the output layer. The purpose behind this was to make use of the nonlinear fitting capabilities that deep learning possesses. By doing so, an attempt was made to carry out the task of correcting the prediction errors of traditional statistical models.

3.6.2. Forecasting accuracy

Table 8 showcases a comparative diagram regarding the results in the situation where a comparison is made between the pure APARCH model and the hybrid model that the authors have employed. As can be seen from the figure, in the process of comparing the performance and accuracy, the hybrid model shows a quite remarkable improvement in its performance and accuracy when it is compared to the standalone APGARCH model. This gain in performance can be ascribed to the ability of the ANN to capture those nonlinear dependencies that traditional models fail to capture.

Table 8. Comparative results of APGARCH and hybrid model

	APGARCH	APGARCH+ANN
RMSE	0.022276	0.0037249
MAE	0.008716	0.00214

4. Discussion

4.1. Decoding the advantages of the hybrid firefighting model: the process of achieving synergy and dealing with trade-offs

The empirical results clearly and without any ambiguity serve as evidence that they demonstrate the forecasting capability, which is superior in nature, of the hybrid APGARCH-ANN model. In this section, an exploration is carried out to find out the reasons that lie behind this kind of success. Then, an examination is conducted regarding the inherent trade-offs that are involved in the process. And finally, a reflection is made upon the limitations belonging to traditional econometric methods.

DOI: 10.54254/2754-1169/2026.LD31510

4.1.1. A quantum leap in forecast accuracy

According to what is presented in Table 8, the empirical outcomes show that in the situation where an integration of the Artificial Neural Network (ANN) is carried out, there is a substantial leap in accuracy. The Root Mean Squared Error (RMSE) experienced a sharp drop. It went from 0.022276 and then reached 0.0037249, which means a reduction of more than 83%. At the same time, the Mean Absolute Error (MAE) decreased. It started from 0.008716 and then went down to 0.0021400. This kind of dramatic improvement emphasizes a fundamental limitation of those standalone econometric models. That is to say, their inherent linear (or conditionally linear) structure is not well-suited to capture the complex and high-order nonlinear dependencies that are prevalent in crude oil markets during the periods when there is stress. This obvious contrast highlights the inherent limitations of standalone parametric models when it comes to the process of forecasting during the turbulent market regimes. Although the Asymmetric Power Generalized Autoregressive Conditional Heteroskedasticity (APGARCH) effectively captures volatility clustering and asymmetric leverage effects through its parametric structure, its linear functional form often fails to take into account the high-order nonlinear dynamics that are triggered by geopolitical shocks or sudden policy shifts. Then, the subsequent ANN layer acts as a universal function approximator. It goes through a process of learning to map these refined features and capture the complex and nonlinear dynamics that the APGARCH structure fails to capture.

4.1.2. The synergistic mechanism: the process of feature extraction meets the process of nonlinear mapping

The achievement of success for this hybrid model comes from the situation where there is a synergy that takes place between “econometric feature extraction” and “data-driven learning.”

Feature Engineering: In the process of feature engineering, the APGARCH model carries out a transformation on raw return sequences. It takes these raw return sequences and turns them into high-quality features. These high-quality features include conditional variance as well as standardized residuals. By doing this, it is able to filter out a large amount of the linear noise. This is a two-stage process, and this kind of process effectively decouples the modeling task. Specifically, the APGARCH model takes on the task of handling the linear and asymmetric volatility features that are well-understood. Meanwhile, the ANN specializes in the task of learning the residual and nonlinear patterns.

Nonlinear Augmentation: In the situation where it builds upon these refined features that have been obtained, the artificial neural network (ANN) makes use of its multi-layered neuronal architecture in order to carry out complex nonlinear mapping. The fact that there is a drastic reduction in the root - mean - square error (RMSE) indicates that the underlying structure of crude oil volatility involves significant nonlinear components. These components are only completely expressed when the component of the artificial neural network (ANN) is introduced.

4.1.3. Inherent trade-offs: interpretability vs. predictive power

Even though the hybrid model has an overwhelmingly superior statistical performance, it is necessary to carry out a cautious academic interpretation of this situation.

The “Black-Box” Challenge: In the situation where the prediction error was subjected to an optimization process from 0.022 to 0.003, the internal weights of the ANN are lacking in the direct economic interpretability, which can be found in the APGARCH parameters. There exists an

DOI: 10.54254/2754-1169/2026.LD31510

inherent trade-off relationship between the predictive power of the model and the transparency of the model. This kind of situation is in line with the common critique regarding ML-augmented financial models. In such models, the gains in accuracy are achieved at the cost of economic interpretability. This fact makes it necessary to carry out careful validation of the model and stress testing of the model in the process of practical applications.

Robustness Concerns: The high precision that has been observed is in a situation where it is sensitive to the process of tuning the architecture of the hidden layer. If the network is overly complex, it will be at risk of experiencing the problem of overfitting to the training noise. So, in order to let the improvement of 83% in precision be achieved, it is necessary to carry out rigorous out - of - sample testing as well as regularization techniques. The purpose is to make sure that the model can maintain its stability when it is in the situation of facing unprecedented market environments, for instance, the governance of liquidity crisis or the occurrence of structural breaks.

4.2. The econometric ceiling: constraints faced by parametric volatility models in the process of their application and development

The empirical findings further bring to light an important subtlety. That is to say, in the situation where the APGARCH model is made to achieve a statistically better in - sample fit than the standard GARCH (this is shown by the fact that the AIC is lower), the corresponding improvement in out - of - sample forecast accuracy (RMSE) turns out to be marginal. This kind of phenomenon can be ascribed to a number of factors.

First of all, in contrast to equity markets, where a distinct “leverage effect” exists, crude oil, which is a commodity, frequently shows a more symmetric volatility response when it comes to price shocks. Geopolitical tensions or supply disruptions have the potential to trigger massive volatility without considering the direction of the price move. This makes the asymmetric parameter γ in APGARCH become less effective.

More importantly, from the perspective of forecasting, in the case where the APGARCH has its flexibility increased through asymmetry terms, at its core, it still remains a linear parametric model. The kind of incremental refinements that take place within a linear framework is limited in its nature when it comes to their ability to deal with the higher-order nonlinear dynamics that are driven by sudden shifts in market sentiment or policy shocks. Moreover, as the BIC penalty for additional parameters indicates, the marginal gains in information criteria imply that the GARCH-family models have arrived at what can be called an “econometric ceiling.” The fact that parametric models are not able to further reduce forecasting errors highlights the necessity to make a transition to non-parametric approaches.

This context places emphasis on the significance of the hybrid approach. This approach, in a strategic manner, takes the parametric models and uses them for the process of structured feature extraction. At the same time, it combines these parametric models with the non-parametric ANN. The non-parametric ANN is utilized for the purpose of nonlinear pattern recognition. By doing all of these things, this approach is able to go beyond this kind of ceiling.

5. Conclusion

This study carried out a comprehensive and detailed comparative analysis of the volatility forecasting models for WTI crude oil. The scope of these models extends from the traditional econometric approaches, which include ARIMA, GARCH, and APGARCH, to a brand-new hybrid machine learning framework. In the process of this research, the central and main aim is to conduct an investigation to figure out whether a synergistic combination has the ability to overcome the inherent limitations that exist in the standalone models.

5.1. Summary of key findings

This empirical analysis leads to three main conclusions. First of all, and most notably, in the situation where a comparison is made, the hybrid APGARCH-ANN model shows a substantial and extremely large advantage in the process of predictive accuracy over all the standalone econometric benchmarks. The decreases in RMSE (which is approximately 83%) and MAE (which is approximately 75%) are not just small increases but represent a kind of change in the paradigm. This validates the core hypothesis that ANN can effectively capture the high-order nonlinearities that are missed by parametric GARCH-family models. Second, this paper discovers that within the family of linear parametric models, in the process where there is a progression from GARCH to the more flexible APGARCH, it only brings marginal improvements in the out-of-sample forecasting performance, even though it has a better in-sample fit. This implies the existence of an "econometric ceiling" for such models when they are applied to complex, event-driven series, for example, oil prices. Third, the success of the hybrid model depends on a careful architectural design. In the experiments of this study, it is indicated that an optimal balance is achieved when there is a specific dual - hidden - layer network. This highlights the importance of a rigorous hyperparameter tuning process to avoid overfitting while making use of the nonlinear predictive power.

5.2. Contributions and implications

This research makes contributions in two different aspects. From a theoretical perspective, in the process of conducting research offers strong empirical evidence to show the superiority of hybrid modeling when it comes to the task of forecasting financial volatility. It proposes a viable methodological path that can be used to go beyond the limitations of traditional econometrics. In practical terms, after the development of the high-precision forecasting framework it provides tangible value for stakeholders. For energy firms, airlines as well as financial institutions, it allows them to carry out enhanced risk management, perform more accurate derivative pricing, and develop improved hedging strategies. As a result, it makes contributions to maintaining market stability and enhancing operational efficiency.

5.3. Limitations and future research directions

This study is by no means free from limitations. In the case where the "black - box" nature of the hybrid model exerts its influence, it leads to a reduction in interpretability. This is a kind of common trade-off situation that occurs with neural networks. Moreover, in the process of the model's development, it was subjected to training and then testing on historical data. When it comes to the situation where there are unprecedented structural breaks, the robustness of this model needs to be further probed. For future research work, it should set its focus on several promising typical paths. Firstly, it involves the act of incorporating relevant exogenous variables, such as geopolitical risk indices and global inventories, into the hybrid framework. Secondly, it requires exploring other advanced machine learning architectures, for example, Long Short-Term Memory (LSTM) networks and transformer models, for the purpose of sequence prediction. Thirdly, it is necessary to conduct rigorous real-time and out-of-sample back-testing so as to fully let the practical utility of the model in a live trading or risk-management environment be validated.

DOI: 10.54254/2754-1169/2026.LD31510

References

[1] Patidar, A.K., Jain, P., Dhasmana, P. and other scholars (2024). The impact that global events have on the crude oil economy: a comprehensive review that is carried out on the geopolitics of energy and economic polarization. GeoJournal, 89: 50.

[2] Ho, S. L., and Xie, M. (1998). The utilization of ARIMA models in order to carry out reliability forecasting and then analysis. Computers and Industrial Engineering, 35(1 then 2), 213 then 216. https: //doi.org/10.1016/s0360-8352(98)00066-7

[3] Herranz, E. (2017) carried out a study on the topic of unit root tests. The study was published in the WIREs Comput Stat, and the publication details show that it was in volume 9, specifically corresponding to e1396.

[4] Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987-1007.

[5] Dhaher Alwan, H. and Najm Abdulla, S. (2022). An analysis is carried out on models (that is, the NAGARCH and the APGARCH) through the process of using simulations. Journal of Economics and Administrative Sciences, 28(132), pages from 62 then 73.

[6] Bollerslev, T. (1986). Generalized autoregressive and then conditional heteroskedasticity. Journal of Econometrics, 31(3), 307 and then 327.

[7] Kristjanpoller, W.D., along with Hernández, E. (2017). In the situation where the volatility of main metals is being considered, it is forecasted by a hybrid ANN - GARCH model that has regressors. This research is published in Expert Syst. Appl., and specifically, in the 84th volume, the content ranges from page 290 to page 300.

[8] Jadhav, V., Chinnappa Reddy, B. V., and Gaddi, G. M. (2017). Application of the ARIMA Model for Forecasting Agricultural Prices. Journal of Agricultural Science and Technology, 19(5), 981-992.

[9] Zhao, P., Zhu, H., NG, W. S. H., and Lee, D. L. (2024). In the process of transitioning from the GARCH model to the neural network for the purpose of conducting volatility forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 38, number 15, ranging from 16998 to 17006.

[10] Han, H., Liu, Z., Barrios Barrios, M. and other scholars (2024). In the process of carrying out research, they developed a time series forecasting model for the extraction of non - stationary series patterns. This development was achieved through the utilization of deep learning and the implementation of GARCH modeling. The relevant research results were published in the journal Cloud Computer, specifically in volume 13, issue 2.

[11] Amirshahi, B., & Lahmiri, S. (2023). Models of the hybrid deep learning type and GARCH - family models which are of the kind for the purpose of making forecasts regarding the volatility of cryptocurrencies. Machine Learning with Applications.