ATHENS UNIVERSITY OF ECONOMICS AND BUSSINESS

SCHOOL OF INFORMATION SCIENCES AND TECHNOLOGY

Master of Science in Data Science

Forecasting Energy Market Series Using Econometric Models and Machine Learning Techniques

Spyridon Mastrodimitris Gounaropoulos

Ioannis Vrontos

Athens

February 2024

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my beloved parent for their stout emotional and material support.

Furthermore, I want to express my deepest thanks to my supervising professor, I. Vrontos, not only for the countless hours spent improving my understanding of mathematical and theoretical concepts but also for the enthusiasm that inspired me to go above and beyond in pursuing research on this subject.

Περίληψη

Η παρούσα μελέτη επικεντρώνεται στη σύγκριση διαφόρων τεχνικών μοντελοποίησης και πρόβλεψης που εφαρμόζονται στις αγορές ενέργειας. Ειδικότερα στις τιμές και την μεταβλητότητα του αργού πετρελαίου στην Αμερικανική και του φυσικού αερίου στην Ευρωπαϊκή και Ασιατική αγορά. Επιπλέον, ενσωματώσαμε μοντέλα μηχανικής μάθησης όπως random forest, XGBoost, δίκτυα long short-term memory και temporal fusion transformers στη διαδικασία πρόβλεψης και αξιολογήσαμε τις επιδόσεις αυτών έναντι πιο παραδοσιακών γραμμικών οικονομετρικών τεχνικών όπως μοντέλα παλινδρόμησης με αυτοσυσχετιζόμενα σφάλματα και γενικευμένη αυτοπαλινδρόμηση υπό όρους ετεροσκεδαστικότητας. Μέσω των δοκιμών εκτός δείγματος επιβεβαιώσαμε τη σημασία των εξωγενών γεωπολιτικών μεταβλητών, τη σημασία που έχει η διάρκεια των σοκ στις προβλέψεις αλλά και την πολλά υποσχόμενη ακρίβεια μοντέλων Self-attention στις χρονοσειρές των αγορών ενέργειας.

Λέξεις-κλειδιά: Αργό Πετρέλαιο, Φυσικό Αέριο, Χρηματοοικονομικές Χρονοσειρές, GARCH, Μηχανική Μάθηση, Random Forest, Gradient Boosting, Long Short-term Memory Network, Temporal Fusion Transformer, Self-attention

Abstract

This study is dedicated to comparing different modeling and forecasting techniques applied to energy markets, specifically crude oil for the American and natural gas for European and Asian market prices and volatility. Furthermore, we are incorporating machine learning models such as random forests, gradient boosting, long short-term memory networks, and temporal fusion transformers in the prediction process, evaluating the impact of those over more traditional linear econometric techniques like regression models with autocorrelated errors and conditional heteroscedasticity components. Through out-of-sample testing, we confirmed the importance of event indicators, the consistency of shocks in model prediction, and the promising accuracy of self-attention models in energy market time series.

Keywords: Crude Oil, Natural Gas, Financial Time Series, GARCH, Machine Learning, Random Forest, Gradient Boosting, Long Short-term Memory Network, Temporal Fusion Transformer, Self-attention

1. Introduction

Energy plays a vital role in our economic and social life, from illumination and heating to fueling vehicles, marine shipping, the global air transportation system, and powering nationwide industries. Even after years of decarbonization efforts, petroleum products and natural gas remain the base of our economic system, as stated by the Organization for Economic Co-operation and Development OECD (2022).

Hydrocarbon price uncertainty is widely accepted to affect actual economic activity, and its volatility has an essential influence on economic stability, social stability, and national security, e.g., Elder and Serletis (2010), Jo (2014), Wu and Zhang (2014). More specifically, many empirical studies have documented that high oil prices negatively affect economic growth (Kilian, 2008; Kilian and Vigfusson, 2011). Also, high natural gas and electricity prices, of which the demand in the short term is highly inelastic, have had significant impacts on inflation and living standards, as seen in OECD (2022).

Inflation pressure induced by a positive oil price shock will likely trigger tighter monetary policies that hinder economic growth. The effect is so prominent that oil prices can be used to predict the economic growth rate Narayan et al. (2014). Moreover, the oil price is a key signal in business activities and financial markets. Narayan and Sharma (2011) and Phan et al. (2015) find that firm returns across different sectors in the stock market tend to correlate significantly to oil price shocks. When the oil price reaches US$100 or above per barrel, the psychological barrier effect lowers firm returns (Narayan and Narayan, 2014).

Examples of volatility spillover from oil to the stock market have also often been observed. For instance, Arouri et al. (2011, 2012) found significant volatility spillover between oil and sector stock returns. Furthermore, they noted that the spillover is usually unidirectional - from oil markets to stock markets - in Europe but bidirectional in the United States. Masih et al. (2011) documented the dominance of oil volatility on actual stock returns and, more importantly, emphasized how this effect has increased over time.

Measuring and predicting the volatility of asset returns, in general, is vital for risk management in general, asset allocation, and option pricing. Risk management is, to a large extent, about measuring potential future losses of a portfolio. The underlying asset's future volatility is a crucial parameter (see, for example, Wennström, 2014). Volatility changes can affect the risk exposure of producers and industrial oil consumers, which may change their respective investments in oil-producing assets or reserves, inventories, and facilities. Volatility also determines the value of commodity-based contingent claims. Thus, understanding volatility dynamics is vital for decisions regarding derivative valuation, hedging, and investments in oil.

Evidence suggests that volatility in equity markets is asymmetric, implying that returns and conditional volatility are negatively correlated. Christie (1982) explains this volatility asymmetry based on the leverage hypothesis and argues that a drop in the value of the stock leads to increases in financial leverage, making the stock riskier and thereby increasing the underlying volatility. Because oil price volatility profoundly affects the investment time horizon, firms must adjust their risk management procedures accordingly. Wang and Wu (2018) examined asymmetric volatility spillovers between oil and international stock markets using a vector autoregression framework and a directed acyclic graph technique (see also, Liu et al., 2022). This is because the accurate prediction of crude oil prices is beneficial to perfect the plans of corresponding production, marketing, and investing, regulate market risks, and enhance future gains of the oil-related industries; see, for example, Zhao et al. (2017). Also, oil price volatility is the core of asset pricing, asset allocation, and risk management.

In the not-too-distant past, several theoretical models assumed constant volatility; see Merton (1969) and Black and Scholes (1973). Today, it is a well-known fact that the volatility of asset returns is time-varying and predictable, see Andersen and Bollerslev (1997). Volatility also has some commonly seen characteristics. Three crucial stylized facts about volatility are that volatility exhibits persistence, volatility is mean reverting, and innovations have an asymmetric impact on volatility (Engle and Patton, 2001).

Examining the existence of external changes that can cause serious disturbances of the market equilibrium is of great interest to actors concerned about how well they can manage the risks when associated with frequent changes in energy markets, as noted by Lee and Lee (2009), Lee et al. (2010) and Salisu and Fasanya (2012). The study by Chen et al. (2015) used a modified iterated cumulative sums of squares (ICSS) algorithm to identify structural breaks before using an asymmetric GARCH model to measure the news impacts on oil price volatility accurately.

The subject of volatility in oil prices is increasingly gaining interest in theory and practice; there are two main reasons for this development. First, oil price data are now available at a high frequency. Therefore, there was increasing evidence of statistically significant relations between observations that previously seemed independent and too far apart. Second, in connection with the high frequency of oil price data, there is the possibility of time-varying volatility, referred to as conditional heteroscedasticity (see Harris and Sollis, 2005). Correctly modeling volatility in oil prices is essential for building accurate pricing models, as predicting volatility will further our understanding of the broader financial markets, the energy industry, and the overall economy (Ewing and Malik, 2017).

Since oil volatility can affect general market activities through the channel of the macroeconomy (Kilian and Vega, 2011), oil volatility risk is viewed as a kind of systemic risk, meaning that almost no economic activity is independent of it. For the aforementioned multiple reasons, the study of the price and volatility of hydrocarbons are of utmost importance for multiple actors: policymakers, market investors, academics, and, of course, all oil-related industries, among others (Liu et al., 2022). However, crude oil prediction is always challenging in practice, as stated by Yu et al. (2017).

One reason is that numerous information factors usually affect crude oil prices, including the fundamental supply-demand relationship (Monge et al., 2017) and external uncertainty factors (Jones, Leiby, and Paik, 2005). Exogenous shocks can significantly impact price and volatility, sometimes briefly and sometimes more persistently, bringing a new set of conditions. These socks can be an act of piracy in a major marine trade route, a terror attack, a natural disaster, a global pandemic, a geopolitical or diplomatic maneuver of an oil-producing county or OPEC+ itself, or a regional war leading to the destruction of strategic energy infrastructure along with monumental changes in country energy strategy and global sanctions from major energy consuming countries. These factors expand the prediction results' uncertainty and lower the prediction accuracy. Crude oil prices have the characteristics of highly non-linear, irregular, and complex. Simple econometric models cannot effectively extract these features. Campbell and Hentschel (1992) argued that news brings higher current volatility, increasing future volatility as volatility is highly persistent.

2. Problem Statement

Modeling and forecasting hydrocarbon price and volatility is paramount for multiple actors. As oil and macroeconomic data became more readily available at higher frequencies, along with the exponentially rising computational strength of artificial intelligence methodologies that can capitalize on big data and deal effectively with non-linear and non-stationary problems, a new approach became possible. Some of the attempts that have been made are Li and Ge (2013) using support vector regression (SVR), Yu et al. (2016) employed least squares support vector regression (LSSVR), and Lahmiri (2017) used artificial neural networks (ANNs).

Through this analysis, we try to approach price and volatility prediction using linear regression-type models with autocorrelated errors, non-linear model specifications, and quantile regression models (Liu et al., 2022; Ewing and Malik, 2017; Meligkotsidou and Vrontos, 2008), along with machine learning techniques (Vrontos et al., 2021). Furthermore, we will use a temporal fusion transformers architecture as proposed by Lim et al. (2020) and implemented by Johannes (2023) to determine if the latest machine learning methods, implemented already in multiple kinds of problems, can also offer significant benefits in energy series.

Heteroskedasticity models (Engle, 1982 and Bollerslev, 1986) will also be considered, together with the conditional mean specifications, for the analysis of the underlying series since there is considerable evidence that energy market return series and volatilities are expected to respond differently and asymmetrically to changes in economic and financial market conditions, especially during periods of high uncertainty caused by random events and shocks, such as geopolitical crises, for example, local conflicts, and global pandemics. A by-product of the analysis is to investigate the immediate impact of the war in Ukraine and its effects (e.g., the Destruction of Nordstream) in the energy sector, specifically for oil and gas.

3. Importance of Study

Implementing an up-to-date, explainable, and effective oil and natural gas price and volatility model capable of dealing ably with exogenous shocks is essential for a market that, as time passes, proves to be more volatile and less interconnected. Countries, corporations, firms, and everyday people could massively benefit from stable and predictable prices. As stated by the International Energy Forum’s report (IEF upstream oil and gas investment outlook, 2023), in the current situation, nobody wins, the current energy price volatility harms consumers, investors, businesses, and governments.

4. Literature review

In the realm of time series analysis, a variety of models are employed to decipher and forecast patterns over time. Auto-Regressive (AR) models predict future values by analyzing past trends, while Moving Average (MA) models smooth out fluctuations using past forecast errors. The ARMA model combines these approaches and is adept at handling complex data structures. Another method is the Vector Auto-Regression (VAR) model, which can capture the interplay between multiple series. One other important component is ARCH and its evolution, focusing on volatility predicting, which is essential in improving predictions and assessing financial risk.

In terms of volatility, the Auto-Regressive Conditional Heteroscedasticity (ARCH) model, which was proposed by Engle (1982) and later developed further by Bollerslev (1986) and Taylor (1986) into the highly praised Generalised AutoRegressive Conditional Heteroscedasticity (GARCH) model, incorporating past variances into current estimates. GARCH has been a point of reference for more than three decades for any research into volatility.

From the simple GARCH, research has come a long way. Due to the nature of many financial problems characterized by fat tails and asymmetries, the models needed to evolve. In the next 20 years, new models will be proposed. Some of them are EGARCH by Nelson (1991), APARCH by Ding, Engle and Granger (1993), GJR GARCH by Glosten et al. (1993), and Andersen et al. (2003), TARCH by Zakoian (1994). Also, Engle and Kroner (1995) further expanded the model to handle multiple financial series, leveraging historical data dependencies. In the same year, FIGARCH was introduced by Baillie et al. (1996), followed three years later by CGARCH (Engle and Lee, 1999).

The GARCH model family includes regression with one variable and multiple regression interpretations Bauwens et al. (2006). Furthermore, different techniques have been slowly incorporated, such as a rolling window, tree structures, even Bayesian methods, and Markov chains (Audrino and Buhlmann, 2001; Audrino and Barone-Adesi, 2006; Dellaportas and Vrontos, 2007). At the same time, different GARCH types keep coming up even in later years, i.e., MRS-GARCH by Haas et al. (2004), MVGARCH by Alizadeh et al. (2008), and ZD-GARCH proposed by Zhang et al. (2018).

As for time-series analysis of petroleum and, more specifically, petroleum futures, significant work has been done by Sadorsky (2006), who used GARCH(1,1) and GJR GARCH and proved that for daily prices BEKK model showed inferior performance than the standard GARCH(1,1) model. Other examples are GARCH models used by Yang et al. (2002) for the US oil market, by Oberndorfer (2009) for the oil market of the Eurozone, and by Hwang et al. (2004) for the major industrialized countries.

Upon the GARCH implementation for petroleum price forecast, Nomikos and Pouliasis (2011) followed up by implementing a double GARCH and MRS GARCH for optimal results. Some academics have proposed that switching exponential to model with Student-t distributed error terms tends to capture better the asymmetric effects of the oil market. Moreover, Chkili et al. (2014) concluded that non-linear volatility models can better describe the volatility of oil prices.

The need to capture non-linearities and improve upon existing models has led to incorporating deep learning methods in combination with existing theory. Hybrid approaches to deep learning techniques, initially successful in image and speech recognition, are increasingly applied to volatility modeling. Architectures like the recurrent neural network (RNN) and their variants like long short-term memory (LSTM), bidirectional recurrent neural network (BRNN), and gated recurrent unit (GRU) are particularly suited for this task because they model sequential data. Recent innovations aim to add flexibility and randomness to these networks, enhancing their forecasting power (Hochreiter and Schmidhuber, 1997; Mike Schuster and Paliwal, 1997; Cho et al., 2014).

Often treated as "black boxes," these models are regarded as non-linear and nonparametric techniques in which no a priori assumption concerning the model's mathematical form (equation) is formulated. The function mapping input data into output signals (forecasts) is formed at the stage of training the model, implemented based on a learning set including historical quotations. Over recent years, in data analysis, researchers and practitioners have increasingly used dynamic artificial neural networks equipped with the ability to remember and process information from some recent period. Attempts have been made to integrate deep-learning tools with the GARCH methodology, and research on this type of hybrid model has been undertaken in many works.

In particular, Kristjanpoller and Minutolo (2015) have applied hybrid models based on feed-forward back-propagation neural network and GARCH to predict the volatility of gold and oil prices; see also Kristjanpoller and Minutolo (2016). Hu, Ni, and Wen (2020) developed a hybrid deep learning method combining GARCH with neural networks and applied it to forecasting the volatility of copper prices. Finally, Liu and So (2020) incorporated a GARCH model into an LSTM network to improve stock volatility prediction. Michańków, Kwiatkowski, and Morajda (2023) analyzed the predictive effectiveness of hybrid GARCH-GRU models compared to 'pure' GARCH models, mainly to examine the synergistic benefits of the former. The results precluded the conclusion of the advantages of such 'hybridization'.

For example, Hu, Ni, and Wen (2020) attempted to use GARCH models to forecast volatility, such as in copper returns, and these forecasts were then used as input features for LSTM networks, combining the strengths of both methods. Swarup and Kushwaha (2017) forecasted Nickel prices based on LSTM and GRU and proved that both methods could provide significant improvements in accuracy and computation rate over previous ones.

Some other approaches are a model that combines ARCH-type and Stochastic Volatility (SV) models, leveraging both autoregressive and stochastic volatility features for more accurate predictions (see, for example, Shephard, 2005) and a GARCH-Markov Switching model (Hamilton, 2020) are used to capture both autoregressive patterns and regime-switching behavior in financial time series data.

Other researchers combine the Neural Stochastic Volatility Model (NSVM) and multi-layer perceptrons (MLPs). The Neural Stochastic Volatility Model (NSVM) is combined with multi-layer perceptrons (MLPs) to compute means and variances of distributions. In this probabilistic model, observable variables depend on previous observables and latent variables. In contrast, latent variables only depend on their past values, and the distributions are modeled as normal with specified mean and variance parameters (see, for example, Luo et al., 2018).

Hosker et al. (2018) recently compared three established financial models used in equity market volatility forecasting to various machine/deep learning supervised regression methods. Their analysis specifically forecasts the one-month VIX futures contract three and five days ahead. They conclude that ML methods based on RNN and LSTM generate improved results over linear regression, Principal Components Analysis (PCA), and ARIMA models.

Furthermore, many researchers attempted to test and compare different architectures to each other. One important work was by Vrontos, Galakis, and Vrontos (2021), who found that the application of specific machine learning techniques in implied volatility prediction is strongly encouraged, as it significantly improves the accuracy of out-of-sample forecasts based on numerous statistical evaluation measures over more mainstream econometric methods.

Moreover, using penalized likelihood methods, such as the Elastic Net regularization technique, is strongly advised, as it can narrow down the number and identify the significant predictors. As anticipated, there is a certain degree of differentiation in the predictive accuracy of the underlying machine-learning models. Model accuracy is inconsistent across all models; however, some techniques exhibit superior relative performance (Naive Bayes, Ridge Deviance, Adaptive Boosting, Discriminant Analysis). Their analysis employs an elaborate set of predictive variables containing 31 macroeconomic and financial market-related indicators, most of which have been widely followed and used in the existing literature on the predictability of asset returns. All these authors demonstrated that hybrid models can overcome the performance of traditional GARCH models in stock volatility forecasting.

Recurrent neural networks or long short-term memory networks, which store information about previous time steps in memory, have been established as state-of-the-art methods for learning long-term dependencies on sequential data. However, Vaswani et al. (2017) propose a model architecture with an encoder-decoder structure called a transformer, which is no longer based on recurrent units but on the self-attention mechanism.

The use of this deep learning mechanism helps the model to dynamically decide how much attention to give to each part of the input sequence by assigning specific weights. In order to consider the order of the sequential data, the transformer network uses a particular strategy called positional encoding. Transformers changed existing machine learning approaches to natural language processing (NLP) since they achieved excellent results and proved more computationally efficient than the previous models due to parallelizable training (Dai et al., 2019).

There are three significant blocks that characterize this neural network architecture. First is a multiheaded temporal attention block that aims to identify long-range patterns present in the time series. Second, is a Sequence-by-sequence LSTM encoder-decoder whose function accounts for the short-range patterns present in the time series. Third, network blocks are used to eliminate unimportant inputs to the neural network. One essential feature of Temporal Fusion Transformer (TFT) is the ability to determine the weight or contribution of explanatory variables to the network's performance, akin to defining statistical significance in econometric regression models, with the important difference that no distributional assumption is required (as seen at Swarup and Kushwaha, 2023).

Lim et al. (2021) introduced the Temporal Fusion Transformer (TFT) model, which combines attention mechanisms and LSTM structures in its architecture and is able to process different types of data. They show that their new model performs well in various applications, including financial applications. Numerous variations of transformer models for time series forecasting now exist, showing promising results, such as Wu et al. (2020), Wen et al. (2022), and Farsani and Pazouki (2021). TFTs are well-suited for applications that require the projection of multiple time steps ahead into the future, and they can handle impressive amounts of data. They have been shown to achieve state-of-the-art performance on a significant variety of time series forecasting tasks and forecasting horizons.

Johannes (2023) supports that TFT also significantly outperforms two machine learning benchmarks in S&P volatility forecasting, namely LSTM networks and random forests when using sectoral and overall pooling. The use of pooling methods generates a much larger data set for the model but also leads to growing heterogeneity in the data. They found that sectoral and overall pooling improves the prediction performance for all machine learning models.

His results confirm that machine learning algorithms are beneficial for forecasting financial volatility and that the novel and flexible TFT approach might become helpful in asset pricing and risk management. On the other hand, Swarup and Kushwaha (2023) analysis showed that models such as TFTs are highly non-linear and computationally intensive and do not consistently outperform traditional methods. This observation agrees with the phenomenon observed by Makridakis et al. (2020).

These methods are ever developing and have the benefit of being able to generate batch forecasts for time series, something challenging with traditional methods that require a combination of strong technical, theoretical expertise and artistic intuition. Also, there is evidence that the good fit of an in-sample model does not certify a similar out-of-sample behavior. The TFT model does not seem to sidestep this dichotomy.

Finally, one crucial variable yet undiscussed is research on event indicators. Political, social, or economic events could trigger these changes in volatility. Lamoureux and Lastrapes (1990) and Wilson et al. (1996) show that volatility persistence is often overrepresented and overestimated when standard GARCH models are applied to a series with underlying changes in variance. Mikosch and Starica (2004) agree on something, adding that ignoring shifts in the unconditional variance will result in higher estimated volatility persistence within a GARCH model. In later years, Ewing and Malik (2010) found that oil shocks dissipate quickly but have a strong initial impact once.

If the break-point is known, we can perform conventional unit root tests on the sub-samples defined by the break-point to check the persistence for each time series. However, it is unwarranted to assume the date of the break-point in advance. Unit root testing is a standard method to detect structural breaks; unit root testing can be executed using a Dickey-Fuller test, created by Dickey and Fuller (1979), an ADF test proposed by Phillips and Perron (1988), a PP test by Elliott et al. (1996) or a KPSS test published by Kwiatkowski et al. (1992). New methods have been developed to test the unit root, given the existence of structural breaks; see, for example, Narayan and Popp (2010). Their results are consistent, whereas the conventional tests become distorted by structural changes. CUSUM is the primary methodology, and many researchers have focused on its application (as seen in Brown et al., 1975 and Zeileis, 2004). Although the CUSUM of the squares-based tests of Leybourne et al. (2007) provides accurate results if the structural break in persistence occurs too early or too late in the time interval, these methods may not detect it.

Models that allow for structural breaks are often appealing for modeling returns since they can capture the changes in risk exposures and benefit hedge fund portfolio management, which naturally follows market shifts. They can also capture non-linearities and changes in the variance across time; assuming different segment-specific variances is an alternative way of modeling time-varying volatilities, rather than modeling the volatility directly (for example, using GARCH models as in Giamouridis and Vrontos, 2007 and Vrontos et al., 2008).

5. Data

For our oil analysis, we use monthly and quarterly price data of West Texas Intermediate (WTI), the benchmark for all North American Markets. The monthly data spans over 22 years from October 2001 to October 2023 and the quarterly data spans 27 years from July 1996 to July 2023.

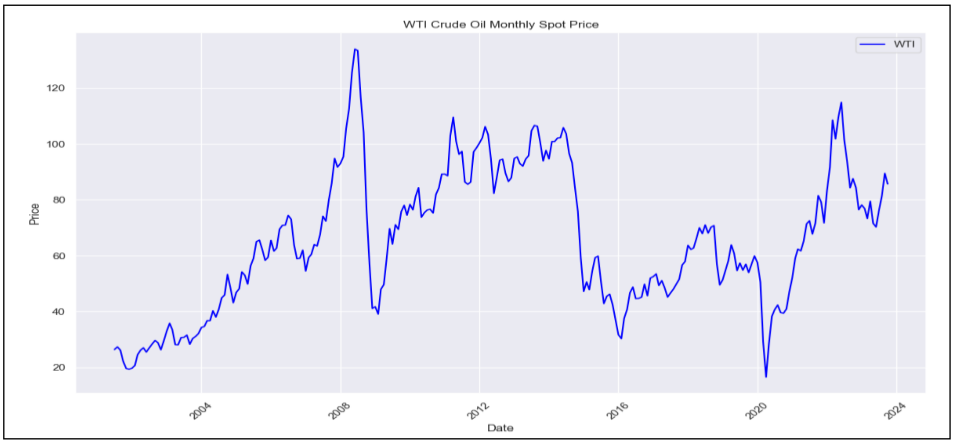


Figure 1.1: Western Texas Intermediate Monthly Spot Price

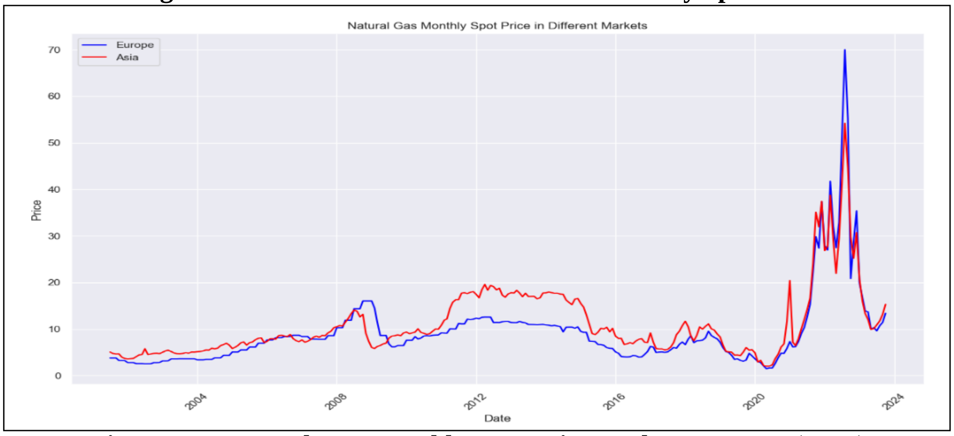


Figure 1.2: Natural Gas Monthly Spot Price on the European (BLU) and Asian Market (RED)

As for the natural gas, we selected the monthly and quarterly data for the European natural gas and Asian natural gas. The reason behind this choice was to compare how global events influence the same commodity in different market conditions and different international markets. The European gas market operates independently, with its own supply and demand dynamics, infrastructure, and regulatory frameworks. Pricing mechanisms in Europe tend to be more strongly influenced by geopolitical factors, regional supply and demand imbalances, and, as of late, the increased dependency on Liquified Natural Gas (LNG) imports after the start of the Russian-Ukrainian war. The Asian market has unique features, including a historic, very high dependence on LNG imports, long-term supply contracts, and regional variations in pricing mechanisms. Demand in Asia, particularly from emerging economies, can be more dynamic and also more sensitive to factors like economic growth, industrial development, and environmental policies.

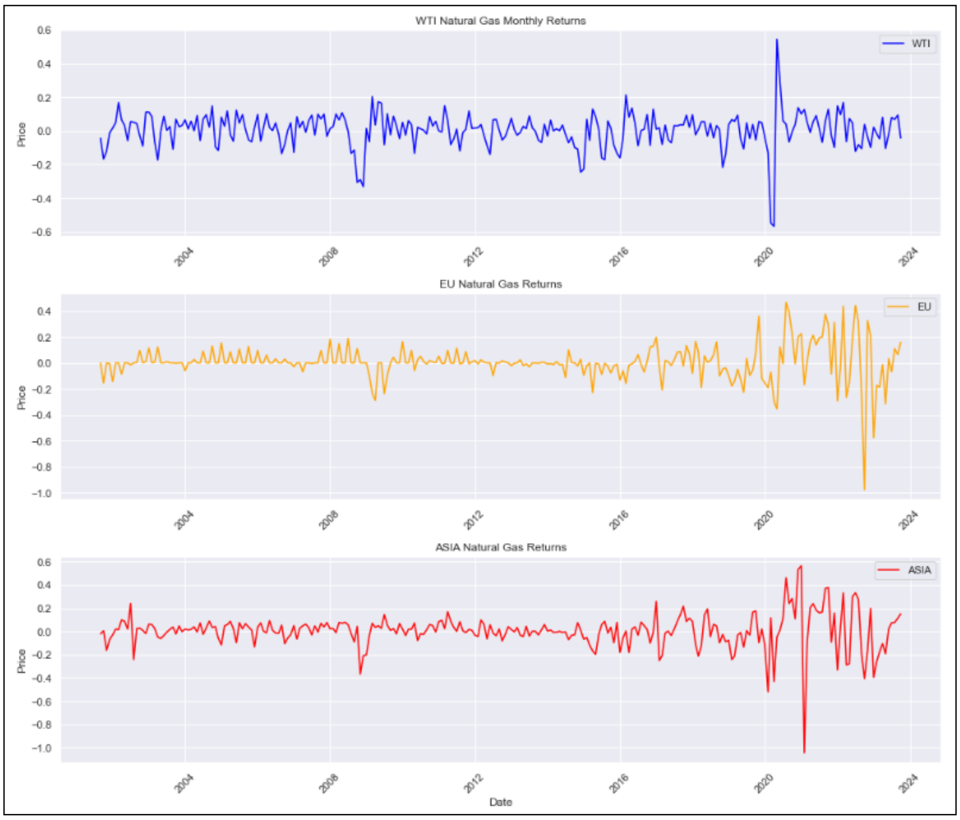


Figure 2: Monthly Returns for Oil (BLU), European Natural Gas (ORANGE) and Asian Natural Gas (RED)

	MONTHLY DATA			QUARTERLY DATA		
	Americal Oil	**European Gas**	**Asian Gas**	**Americal Oil**	**European Gas**	**Asian Gas**
SAMPLE SIZE	265	265	265	109	109	109
STANDARD DEVIATION	0,11	0,15	0,15	0,16	0,22	0,22
SKEWNESS	-0,95	-1,16	-1.19	-1,14	0,03	0,38
KURTOSIS	7,77	11,05	10,44	3,15	2,85	2,58

Table 1: Oil and Gas Returns Data Characteristics

Beyond the crude oil and gas prices, other variables have been selected for the analysis as predictive variables. The variables selected for further analysis were the Equity Market Volatility Tracker, the Economic Policy uncertainty Index for Europe (Germany, the United Kingdom, France, Italy, and Spain), the Economic Policy uncertainty Index for the USA, the Current General Business Conditions for Ney York, the 3-Month Treasury Bill Secondary Market Rate, the Infectious Disease Tracker, and the Global price of Nickel. The data were collected from the Federal Reserve's Economic Database. As predictive variables for the quarterly data were selected, the US Gross Domestic Product, the EU19 Gross Domestic Product, the Japanese Gross Domestic Product, the Equity Market Volatility Tracker, the Economic Policy uncertainty Index for Europe (Germany, the United Kingdom, France, Italy, and Spain), the Economic Policy uncertainty Index for the USA, the 3-Month Treasury Bill Secondary Market Rate, the Infectious Disease Tracker. All previously mentioned data were obtained from the Federal Reserve's Economic Database.

To measure the impact of exogenous shocks and the importance of including qualitative categorical variables in our models, we introduced two new predictive variables acting as event indicators. The first was the COV19 annotating the start of the lockdowns, and the second was the “RUWAR” annotating the beginning of the Russia-Ukraine War. Both variables were binary, becoming and staying one at the start of the event and zero everywhere else.

The forecast horizon will be two years, translating into a horizon of 24 one-step ahead predictions for the monthly data and eight one-step ahead predictions for the quarterly data. As a result, the start of the prediction splits the training and the testing datasets. The reason for the lack of a validation dataset was that since the comparison between machine learning and Regression models with autocorrelated errors is made in a way where every new prediction is made based on knowledge of the previous lag, a validation set for tunning would be very time-sensitive. Along with the fact that the observations are not independent but carry long memory, any attempt to use a sequence of data far from the present would decrease the model’s accuracy, and so since there is no fear of overfitting for the aforementioned reasons. That is why no attempt at a validation set was made. Perhaps an exemption is the Temporal Fusion Transformers, but that would be further analyzed in a separate section.

6. Methodology

6.1 Autoregressive Model (AR)

The AR(1) model is given as:

$$y_t = a_0 + a_1 \cdot y_{t-1} + \varepsilon_t \tag{1}$$

where ε_t is a white noise series with mean of 0 and variance σ^2. The general function for an AR process, AR(p), is as in the equation:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \varepsilon_t \tag{2}$$

6.2 Moving Average Model (MA)

The general form for the MA(1) model is:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t \tag{3}$$

where μ is the mean and $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ is white noise.

The general form of an MA process, the MA(q) model, is:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{4}$$

6.3 Autoregressive Moving Average (ARMA)

An ARMA model is the combination of the AR and the MA processes. The AR represents the effects of previous observations. The MA part represents the effects of previous random shocks (errors):

$$y_t = \mu + \sum_{j=1}^{p} a_j \, y_{t-j} + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} \tag{5}$$

where p representing the autoregressive terms and q moving-average terms.

6.4 Autoregressive Conditional Heteroskedasticity (ARCH)

The ARCH model is a statistical model used to analyze and model the volatility of asset returns. It addresses the issue of heteroskedasticity - where the conditional variance of a time series changes over time - by allowing the variance of the current error term to be a function of the actual sizes of the previous period's error terms. The key function of the ARCH model is to model the variance of the current period's error term as a linear function of the squares of the previous periods' error terms. This is based on the insight that in many financial time series, large changes in prices are usually followed by large changes, and small changes tend to be followed closely by other small changes. This phenomenon is known as volatility clustering.

The ARCH(q) model for the series ε_t is given as:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 \tag{6}$$

where σ_t^2 is the conditional variance of the error term at time t, $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ are the error terms from the previous periods, $a_0, a_1, \dots, a_q$ are the parameters to be estimated, q is the order of the ARCH model, and the number of is the number of lagged squared error terms included.

6.5 Generalized Autoregressive Conditional Heteroskedasticity (GARCH)

It was developed to overcome ARCH model challenges. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, being an extension of ARCH, is designed to capture the volatility dynamics in financial time series data more efficiently. It overcomes an important inefficiency of ARCH in modeling highly volatile series that require a large number of parameters.

The core function of the GARCH model is to model the current period's variance as a function not only of the squares of the previous periods' error terms, like in ARCH, but also of past period variances. This inclusion of lagged variances allows the GARCH model to capture the persistence of volatility over time with fewer parameters than would be required in a purely ARCH model.

The GARCH(p, q) model can be formally expressed as:

$$\sigma_t^2 = a_0 + \sum_{j=1}^{q} a_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} b_j \sigma_{t-j}^2 \qquad (7)$$

where σ_t^2 is the conditional variance of the error term at time t, $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ are the error terms from the previous periods, $a_0, a_1, \dots, a_q$ and $b_1, b_2, \dots, b_p$ are the parameters to be estimated, q is the number of lagged squared error terms included, p is the number of lagged conditional variances included.

6.6 Exponential GARCH Model (EGARCH)

Models of the GARCH family assume that the main factor in determining future volatility is the magnitude and not the positivity or negativity of expected excess returns. According to Nelson (1991), the EGARCH model was developed to cater to the shortcomings of the GARCH models. Unlike GARCH, which could imply non-negative coefficients to ensure that the conditional variance is positive, the EGARCH model directly models the logarithm of the variance—allowing for the accommodation of asymmetric effects of positive and negative shocks on volatility.

The EGARCH(p, q) model can be formally expressed as:

$$\log(\sigma_t^2) = a_0 + \sum_{j=1}^{q} a_j g(Z_{t-j}) + \sum_{j=1}^{p} b_j \log(\sigma_{t-j}^2) \qquad (8)$$

where σ_t^2 is the conditional variance of the error term at time t, Z_{t-j} is the standardized error term at time $t-i$, $g(Z_{t-j}) = \theta Z_{t-j} + \gamma\left(|Z_{t-j}| - E(|Z_{t-j}|)\right)$ allowing for the modeling of asymmetry, $a_0, a_q, b_p, \theta, \gamma$ are the parameters to be estimated, q is the number of lagged squared error terms included, p is the number of lagged conditional variances included.

6.7 The integrated GARCH model (IGARCH)

The Integrated Generalized Autoregressive Conditional Heteroskedasticity (IGARCH) model represents a significant evolution in modeling financial time series, particularly in capturing the persistence of volatility shocks. Unlike the standard GARCH model, which assumes that past shocks' impact on volatility decays over time. In the IGARCH model, the sum of the autoregressive and moving average coefficients is constrained to equal one, implying that shocks to the conditional variance have a permanent effect on future volatility. This characteristic makes the IGARCH model particularly suited for modeling time series with long memory in volatility, where the impact of past shocks does not diminish over time.

The IGARCH(p, q) model can be formally expressed in a similar way to the GARCH, but with the constraint that the sum of the autoregressive b and moving average a parameters equal to one:

$$\sigma_t^2 = a_0 + \sum_{j=1}^{q} a_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} b_j \sigma_{t-j}^2 \quad (9)$$

where:

$$\sum_{j=1}^{q} a_j + \sum_{j=1}^{p} b_j = 1 \quad (10)$$

and σ_t^2 is the conditional variance of the error term at time t, $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ are the error terms from the previous periods, $a_0, a_1, \dots, a_q \; and \; b_1, b_2, \dots, b_p$ are the parameters to be estimated, q is the number of lagged squared error terms included, p is the number of lagged conditional variances included.

A critical aspect of the IGARCH model is its implication for the unconditional variance of the series, which becomes undefined under the unit root condition. This fact poses challenges for theoretical analysis and practical application, particularly in risk management and derivative pricing, where accurate volatility forecasts are crucial. Despite this, the IGARCH model's ability to capture the long-term dependencies in volatility makes it a valuable tool in financial econometrics.

The empirical performance of the IGARCH model has been the subject of extensive study in the literature. Studies comparing the IGARCH model to other GARCH variants generally find that the IGARCH model provides a better fit for financial time series that exhibit high persistence in volatility. However, the model's assumption of permanent shocks to volatility is also a possible limitation, as it may not be appropriate for all financial time series; this has led to the development of fractional GARCH models, which aim to capture the long memory in volatility more flexibly.

6.8 Glosten-Jagannathan-Runkle GARCH Model (GJR-GARCH)

Glosten et al. (1993) proposed modifying the original GARCH model by introducing a dummy variable in order to capture asymmetric effects in financial time series. The key feature of the GJR-GARCH model is the introduction of an additional term in the variance equation to model the asymmetric response of volatility to positive and negative shocks. This allows the model to distinguish between these shocks, capturing the observation that negative shocks (e.g., unexpected asset price decreases) typically lead to a larger increase in volatility compared to positive shocks.

The GJR-GARCH(p, q) model can be formally expressed as:

$$\sigma_t^2 = a_0 + \sum_{i=1}^{q} a_i \varepsilon_{t-1}^2 + \sum_{l=1}^{p} b_l \sigma_{t-l}^2 + \sum_{k=1}^{q} \gamma_k \varepsilon_{t-k}^2 I_{t-k} \quad (11)$$

where σ_t^2 is the conditional variance of the error term at time t, $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_{t-q}$ are the error terms from the previous periods, I_{t-k} is an indicator function, it is equal to 1 if $\varepsilon_{t-k} < 0$ and 0 otherwise, a_0, a_q, b_l, γ_q are the parameters to be estimated, q is the number of lagged squared error terms included, p is the number of lagged conditional variances included.

This model allows for different volatility responses to positive or negative shocks and supposes there must be no news to observe the minimum volatility.

6.9 Random Forest

The Random Forest by Breiman (2001) represents an alternative machine learning technique that can generate improved predictive accuracy compared to other classification or regression tree model techniques. Its main advantage for time series analysis is its ability to capture complex nonlinear relationships and interactions between lagged variables without the need for explicit model specification as in traditional econometric models.

A Regression Random Forest for time series forecasting during the training phase constructs multiple decision trees. Each tree is built on a different bootstrap sample from the original data, and at each split in the tree, a random subset of predictors is considered. This randomness ensures that the trees are de-correlated, enhancing the model's ability to generalize by reducing overfitting, a very useful condition. The final prediction is made by averaging the predictions from all the trees, which helps to improve accuracy and robustness.

For a Regression Random Forest, the ensemble prediction at a new point x is an average of the predictions from all the individual regression trees in the forest. Mathematically, the prediction of a Random Forest model can be expressed as:

$$y(x) = \frac{\sum_{b=1}^{B} T_b(x; \Theta_b)}{B} \tag{12}$$

where $y(x)$ is the predicted value for the input x, B is the number of trees in the forest, $T_b(x; \Theta_b)$is the prediction of the b-th tree for the input x, Θ_b represents the random vector of parameters chosen independently for each tree at its construction.

6.10 Extreme Gradient Boosting (XGBoost)

In 2016, Chen and Guestrin (2016) proposed the XGBoost, a tree-based ensemble technique that is particularly good at accomplishing classification and regression tasks. It is based on the Gradient Boosting Machines (GBM) algorithm, which learns by combining multiple weak models (in this case, decision trees) to form a more robust model (as seen in Friedman, 2001). As an ensemble learning method, it operates by combining predictions from multiple weak models to produce a more robust prediction.

XGBoost has claimed its place as one of the most favored and widely used machine learning algorithms because of its ability to handle large datasets and achieve state-of-the-art performance in a great variety of machine learning tasks such as regression and classification. Its wide adoption as a go-to tool for multiple Kaggle competitions is proof of its success.

The objective function that requires minimization is the following:

$$\mathcal{L}^{(t)} = \sum_{i=l}^{n} l(y_i \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{13}$$

Where y_i is representing the training data.

In a way, it can be interpreted as a function of functions (i.e., $\mathcal{L}$ is a function of CART learners, a sum of the current and previous additive trees), and as Chen and Guestrin (2016) refer cannot be optimized using traditional optimization methods in Euclidean space. For that reason, the use of the Taylor approximation is necessary. Because we need to transform the original objective function to a function in the Euclidean domain in order to be able to use traditional optimization techniques, we can write the objective (loss) function as a simple function of the newly added learner and thus apply Euclidean space optimization techniques.

So, if we decide to take the second-order Taylor approximation, we have:

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2} f''(a)(x-a)^2 \tag{14}$$

$$\mathcal{L}^{(t)} \approx \sum_{i=l}^{n} \left[l\left(y_i \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \tag{15}$$

Being at iteration t, to build a learner that achieves the maximum possible reduction of loss, the following questions arrive. Is it possible to find the optimal next learner, and if is there any way to calculate the gain (loss reduction) after adding a specific learner.

There is a way to measure the quality of a tree structure q, as the authors refer, and the scoring function is the following:

$$\tilde{\mathcal{L}}^{(t)}(q) = -\frac{1}{2} \sum_{j=1}^{T} \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{16}$$

In practice, in order to build the learner, we start with a single root (contains all the training examples), iterate over all features and values per feature, and evaluate each possible split loss reduction:

Gain = loss (father instances) - (loss (left branch) + loss (right branch))

The gain for the best split must be positive. Otherwise, we must stop growing the branch. Summarizing the GBT model, which is a sum of CART (tree), learners will attempt to minimize the log loss objective and the scores at the leaves, which are, in actuality, the weights. The importance of the weights carry across all the trees of the model and are constantly adjusted in order to minimize the loss.

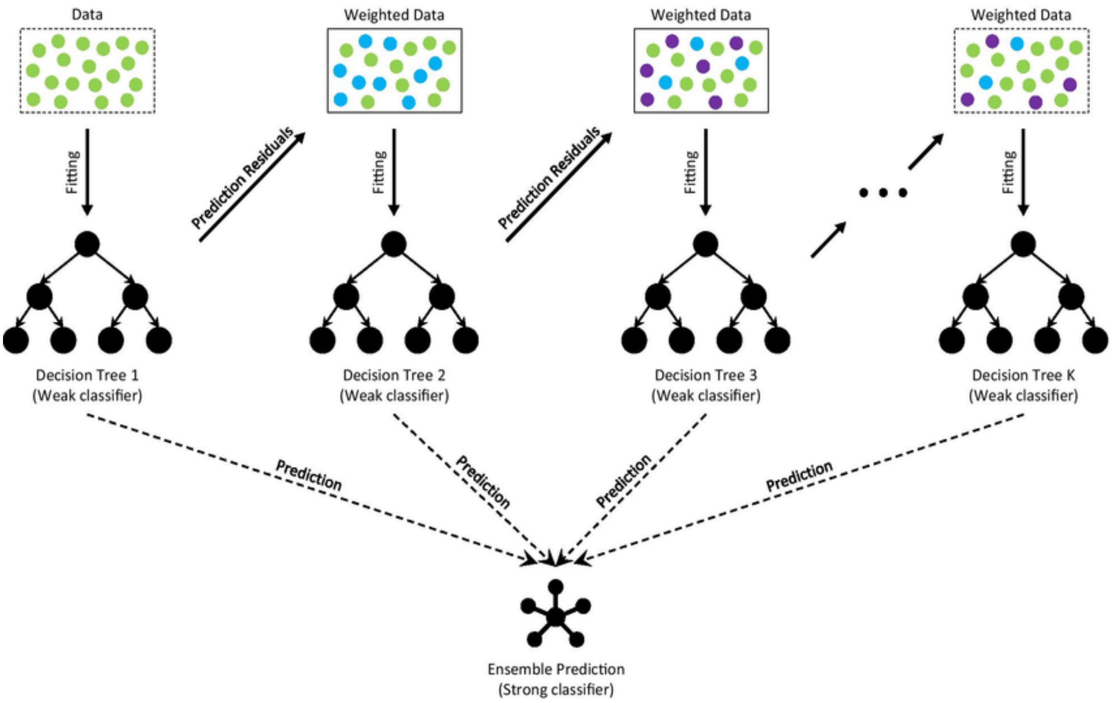


Figure 3: Example of the gradient-boosting decision tree methodology (Deng, Zhou, Wang, 2021)

6.11 Long Short-Term Memory (LSTM)

Long Short Term Memory (LSTM) neural networks are an extension of recurrent neural networks (RNNs), proposed by Rumelhart et al. (1986). RNNs are a particular type of neural network that introduces recursion by allowing the use of sequential, autocorrelated data. The sequence (or observed time series) is accompanied by a hidden input, a memory state that stores information provided with previous time steps. The following input in the sequence is predicted using this recursive hidden state:

$$h_t = g(W_x x_t + W_x h_{t-1} + b_h) \tag{17}$$

where $g(\cdot)$ is an activation function (e.g., logistic sigmoid, hyperbolic tangent, or Rectified Linear Unit (ReLU)), $x = (x_{12} x_2, x_3, \dots, x_T)$ is the sequence of observed time series of length T, while $h = (h_1, h_2, h_3, \dots, h_T)$ represents a random vector hidden state of the same length T. W_x and W_h are weight matrices (parameters) of the neural network, corresponding to x and h respectively, and b_n is a bias vector.

Such an equation assumes that the sequence can be of infinite length or at least an arbitrarily large number T, but due to computational obstacles (such as the problem of vanishing or exploding gradients), the sequence length T is practically limited to only a few timesteps.

The problem mentioned above is practically solved by the introduction of LSTM by Hochreiter and Schmidhuber (1997). LSTMs expand the idea of hidden states by introducing gating mechanisms, which tell whether to preserve or ignore the input from the hidden state. LSTMs can "remember" or "forget" particular timesteps if deemed necessary, building the long-term dependency parameter matrix. In detail, there are three gates: forget, input, and output. The following equations calculated iteratively build up the LSTM network:

$$i_t = g(W_{ix}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i) \tag{18}$$

$$f_t = g\left(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f\right) \tag{19}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh\left(W_{cx}x_t + W_{ch}h_{t-1} + b_c\right) \tag{20}$$

$$o_t = g(W_{ox}x_t + W_{oh}h_{t-1} + W_{oc}c_{t-1} + b_o) \tag{21}$$

$$h_t = o_t \odot h(c_t) \tag{22}$$

$$y_t = W_{yh}h_t + b_y \tag{23}$$

where W terms denote weight matrices (e.g.: W_{ix} is a matrix of weights from the input gate to the input x), and the b terms denote bias vectors (e.g. b_i is the input gate bias vector), $g(\cdot)$ and $h(\cdot)$ denote sigmoid and hyperbolic tangent activation functions respectively here, i, f and o denote input, forget, and output gates respectively, c_t Another hidden state vector, specifically named cell activation vector (responsible for activating specific gates). The output of the neural network is to be of any distribution $p\left(y|x_j\theta\right)$.

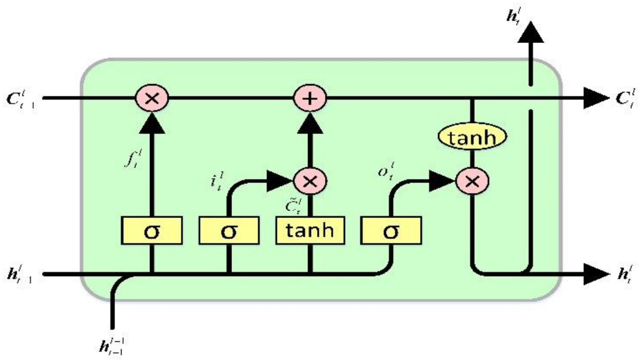


Figure 4: Architecture of a single LSTM network cell

Source: Jacobs (2017)

6.12 Temporal Fusion Transformers (TFT)

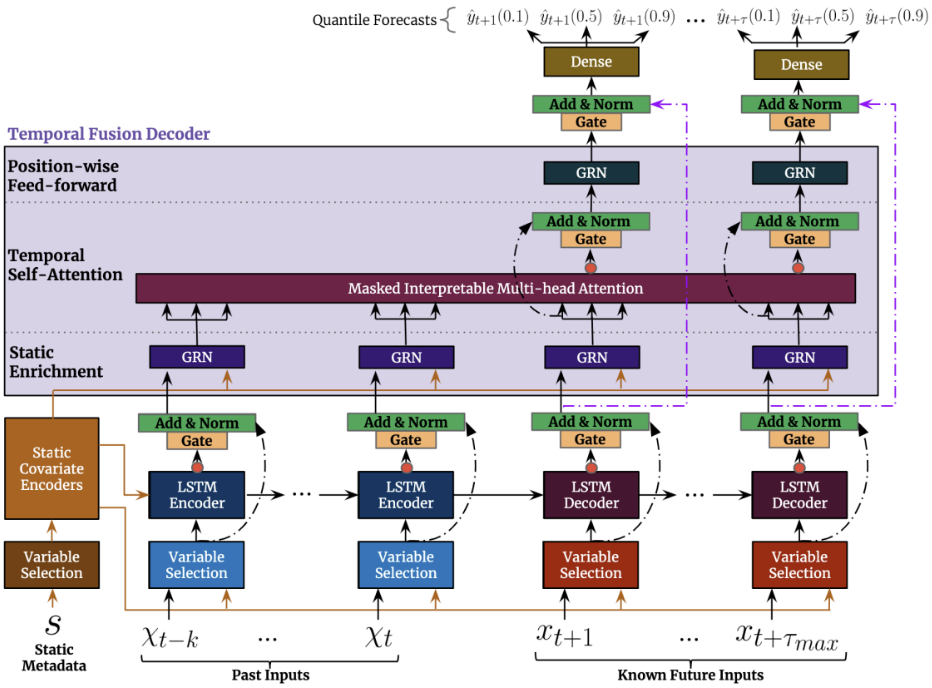


Figure 5: TFT architecture

Source: Lim et al. (2020)

TFT models are deep neural networks (DNN) for multi-horizon forecasting based on the idea of a self-attention mechanism. This architecture allows for a set of categorical variables $c_n \in R^{m_c}$ and time-varying covariates $X_{\eta,t} = [Z_{n,t}^T]$, where $z_{n,t} \in R^{m_z}$ are known only up to the present and $x_{n,t} \in R^{m_x}$ are a priori known into the future. The corresponding target variable is given by $RV_{n,t} \in R^+$. Let t denote an arbitrary time step with a given lookback window k and a maximum step-ahead window T_{max}. The model takes the described input variables and computes the T-step-ahead forecast of RV for stock n. The TFT architecture consists of many different building blocks to filter relevant information from the features. The following five main units are used for this purpose:

1. Gated residual network (GRN): A special gating mechanism, consisting of several layers, selects unneeded parts of the model architecture to reduce complexity.

2. Variable selection network (VSN): For each time step t, this component chooses meaningful features and assigns corresponding weights to them.

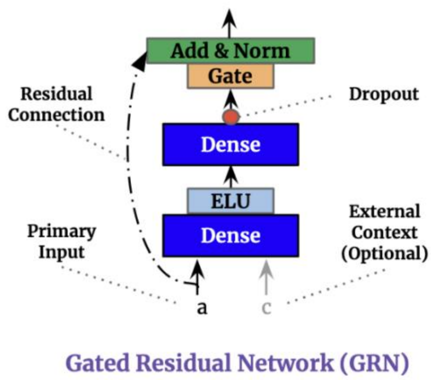


Figure 6: GRN architecture

Source: Lim et al. (2020)

3. LSTM encoder-decoder structure: This layer allows the model to recognize the temporal ordering of the data and, thus, the input variables' short and long-term dependencies. For the categorical features, the TFT network uses separate encoders.

4. Interpretable multi-head attention mechanism: This novel approach provides feature interpretability by adjusting the common mechanism.

5. Quantile regression forecasts: The prediction of several q^{th} percentiles enable the calculation of forecast intervals on top of point forecasts (Lim et al., 2020).

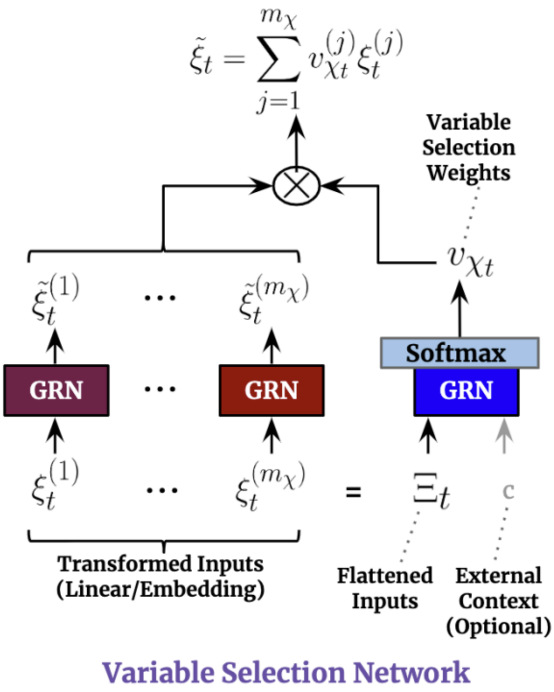


Figure 7: TFT Method of Variable Selection

Source: Lim et al. (2020)

7. Forecasting Performance Metrics

This work adopts three commonly used criteria to examine the robustness and superiority of the model from different aspects. These are mean absolute error (MAE), mean absolute percent error (MAPE), and mean square error (MSE). As the errors decrease, the more accurate the model becomes in forecasting correctly the returns of all energy products on different timescales:

$$MAE = \frac{\sum_{t=1}^{T} |y_t - \hat{y}_t|}{T} \tag{24}$$

$$MAPE = 100 \cdot \frac{\sum_{t=1}^{T} \left| \frac{y_t - \hat{y}_t}{y_t} \right|}{T} \tag{25}$$

$$MSE = \frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{T} \tag{26}$$

where T is the data length, Y_t is the actual data, and $\hat{Y}_t$ is the predicted data. Also, it is very uncommon for a model to achieve the lowest error rate in all metrics. This means that the ranking process becomes more complex. As a result, we decided to judge our models independently for each statistic in our analysis and then attempt to extract conclusions.

8. Data Exploration-Feature Retention

As previously mentioned, multiple variables were tested for possible predictive ability over the prices of oil and gas. To the already mentioned variables obtained by the Federal Reserve Economic Database, 12 new variables were also added, those being the lagged values. For the monthly data, we introduced the lagged values of the previous month. As for the quarterly data were for the four previous quarters.

From the monthly data, the Economic Policy Uncertainty Index for Europe and for the quarterly data, the EU19 Gross Domestic Product, the Japanese Gross Domestic Product, the Economic Policy Uncertainty Index for Europe, and the Economic Policy Uncertainty Index for the USA were rejected as showed high collinearities as detected by variance inflation factor (VIF) tests.

As always practiced in time series analysis, the data were differentiated. The Equity Market Volatility Tracker, the 3-Month Treasury Bill Secondary Market Rate, and the Current General Business Conditions for New York were passed a simple differencing, while for the rest of the Variables, a logarithmic differentiation was used.

After the variable differentiation, two statistical tests were used to provide evidence that the new series were indeed stationary and could be used in our analysis. The tests were the Augmented Dickey–Fuller (ADF) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test.

9. Empirical Results

Throughout this work, ten different methods were used and tested as possible solutions for the problem of predicting future returns of different commodities in different global markets. The first two methods selected were two linear regressions, the first using only the lagged variables of the returns, and the second was a multiple linear regression that added other predictive variables. Those two methods are used as the baseline to compare the rest of the methodologies. All the methodologies can be categorized as statistical-econometric (linear regression, autoregressive, moving averages, conditional heteroskedasticity models) or computational-machine learning (random forest, XGBoost, LSTM, temporal fusion transformers). This creates an interesting dynamic of old and new as well as tackling the same problem from different perspectives.

The sequence of presenting the results is based on a concept of continuity and gradual improvement with increasing complexity in the underlying model, starting with the simplest linear regression with a single variable and ending with the highly complex and labyrinth structure of the TFT.

As it will become apparent in the next pages, a more complicated model does not assure a better fit to reality, and some of those simpler models will outperform their more complicated counterparts. Also, as the models become more complicated, they require more precise tuning to the data, risking some overfitting or an inferior model if the tuning is not done properly. One more thing is the need for certain conditions, so the use of the model is warranted and is correctly implemented. All those conditions create a highly challenging environment for the comparison of multiple models of different philosophies and require both patience and aptitude in order to show reliable results.

9.1 Predictive Variables

One significant part that is important to analyze is the predictive variables used in our research. More specifically, the goal of this work was not only to compare different methods but also, if possible, to narrow down sources of information that their inclusion in our models as predictive variables could offer advantages in the prediction of the future movement of the market.

As mentioned before, the process of acquiring, filtering, and finally incorporating those predictive variables had not been without effort. Many were found to be interdependent, and thus, their inclusion was redundant to the model, while others failed to show strong predictive power. Also, in the implementation of the regression with autocorrelated errors, some variables had to be abandoned to keep the residual errors within bounds and the models robust. The following tables analytically show the use of the predictive variables per model and per case.

WTI MONTHLY

	Regression	GARCH	IGARCH	EGARCH	RF	XGBOOST	LSTM
WTI_LAG	X	X	X	X	X	X	X
EQUITY MARKET VOLATILITY							
ECONOMIC POLICY UNCERTAINTY EUROPE	X		X	X	X	X	X
NY BUSINESS CONDITIONS							
3-MONTH TREASURY BILL							
INFECTIOUS DISEASE TRACKER	X				X	X	X
NICKEL							
COV19							
RUWAR							

Table 2: Predictive Variable Incorporation per Model for WTI Monthly Data

As expected, the use of the previous return values is used by all models used for the monthly predictions, with all showing great statistical significance (Tables 2, 4, and 6). As for other important variables to be considered concerning the prediction of the American crude proved to be the economic policy uncertainty index and the Infection disease tracker used, the first used by the vast majority and the second by the multiple regression and all machine learning models as seen in Table 2.

WTI QUARTERLY						
	Regression	GARCH	IGARCH	RF	XGBOOST	LSTM
WTI_LAG1	X	X	X	X	X	X
WTI_LAG2						
WTI_LAG3					X	
WTI_LAG4						
USAGDP	X			X		X
EQUITY MARKET VOLATILITY		X				
3-MONTH TREASURY BILL						
INFECTIOUS DISEASE TRACKER	X			X	X	X
COV19				X	X	X
RUWAR				X		X

Table 3: Predictive Variable Incorporation per Model for WTI Quarterly Data

Regarding the models implemented for the prediction of the quarterly prices of American crude oil, we detect a lack of statistical significance of distant lags for the returns while other predictive variables are more commonly selected. The infectious disease tracker is as common as in the monthly prediction selected by the multiple regressor, the random forest, the XGBoost, and the LSTM. Another important factor was the US gross domestic product, while less often selected variables were the event indicators exclusively used by the machine learning models.

EUNG MONTHLY							
	Regression	GARCH	IGARCH	EGARCH	RF	XGBOOST	LSTM
EUNG_LAG	X	X	X	X	X	X	X
EQUITY MARKET VOLATILITY							
ECONOMIC POLICY UNCERTAINTY EUROPE				X			
NY BUSINESS CONDITIONS							
3-MONTH TREASURY BILL							
INFECTIOUS DISEASE TRACKER							
NICKEL							
COV19	X				X	X	X
RUWAR	X				X	X	

Table 3: Predictive Variable Incorporation per Model for EUNG Monthly Data

Table 3 shows that the natural gas presents significant differences in comparison to the crude oil (Table 2). While the forecasting models majorly selected the Economic policy uncertainty in Europe and the Infectious disease tracker as predictive variables, the models for the European natural gas preferred the event indicators. An interesting find perhaps caused by the inelasticity of demand for natural gas (e.g., home heating) in comparison to the close connection of oil to economic activity (e.g., lockdowns). Also, there is a hypothesis to be made here that perhaps the infectious disease tracker might not be accurately portraying the conditions on the European continent with the desired precision as it is a global statistic.

EUNG QUARTERLY						
	Regression	GARCH	IGARCH	RF	XGBOOST	LSTM
EUNG_LAG1	X	X	X	X	X	X
EUNG_LAG2	X	X		X		X
EUNG_LAG3	X			X		X
EUNG_LAG4						
USAGDP	X			X		X
EQUITY MARKET VOLATILITY						
3-MONTH TREASURY BILL		X	X			
INFECTIOUS DISEASE TRACKER					X	
COV19				X	X	X
RUWAR				X	X	X

Table 4: Predictive Variable Incorporation per Model for EUNG Quarterly Data

Overall, the differences between oil and gas are less pronounced concerning the quarterly data. The main difference is the stronger predictive abilities of lagged return prices, as seen in Table 11. While lagged returns of the second, third, and fourth previous quarters were considered unimportant for oil, for natural gas EUNG_LAG2 and EUNG_LAG3 are selected for their predictive ability 57% and 42% of the times. Only XGBoost consistently found them unhelpful to our prediction.

ASIANG MONTHLY							
	Regression	GARCH	IGARCH	EGARCH	RF	XGBOOST	LSTM
ASIA_LAG	X	X	X	X	X	X	X
EQUITY MARKET VOLATILITY							
ECONOMIC POLICY UNCERTAINTY EUROPE	X		X	X	X	X	X
NY BUSINESS CONDITIONS							
3-MONTH TREASURY BILL							
INFECTIOUS DISEASE TRACKER	X				X	X	X
NICKEL							
COV19							
RUWAR							

Table 5: Predictive Variable Incorporation per Model for ASIANG Monthly Data

ASIANG QUARTERLY						
	Regression	GARCH	IGARCH	RF	XGBOOST	LSTM
ASIA_LAG1	X	X	X	X	X	X
ASIA_LAG2						
ASIA_LAG3					X	
ASIA_LAG4						
USAGDP	X			X		X
EQUITY MARKET VOLATILITY		X				
3-MONTH TREASURY BILL						
INFECTIOUS DISEASE TRACKER	X			X	X	X
COV19				X	X	X
RUWAR				X		X

Table 6: Predictive Variable Incorporation per Model for ASIANG Quarterly Data

Tables 5 and 6 present the close connection of the Asian energy market to the American. Based on the selection of predictive variables, more common ground is found between the different commodities of Asian gas and American crude than the European natural gas, the same asset on a different market. The economic policy uncertainty and the infectious disease tracker take center stage again, and as of the quarterly data, Asian Gas and American oil share the exact same selection of factors.

More specifically, the returns of Asian natural gas are less dependent on previous returns, and the future returns depend more on the volatility of the American stock market and less on the current bond rates in comparison to the European Market, where the inverse trend is present. Also, even though important events play an important part in the forecasting process at a macro level for quarterly returns, the same events are not as important for predictions on a monthly basis. Overall, the use of macroeconomic variables in the modeling of monthly time series proves significantly more difficult than quarterly time series. As a result, five out of nine predictive variables were never used. If we take away the lagged returns, only three were used, and they were just below 40% of the examples. Those three somewhat successful variables were the infectious disease tracker and the two event indicators that we created, COV19 marking the start of lockdowns at the start of 2020 and RUWAR marking the start of the special military operation in 2021. The two last came as a posteriori knowledge, so not really an effective way to predict the future, but on the other hand, we have again proved the deep scars created in the energy market by those events.

As for the quarterly prediction, a completely different image was revealed. Only the past four quarter returns were never used a stark difference. While almost all variables were included at some point, some had a stronger presence than others. First of all, the COV19, the lagged returns of the previous quarter and the previous eight months, and RUWAR were used in half or all examples. Secondary importance had the infectious disease tracker (28% of the cases), the USA GDP (16% of the cases), the 3-month treasury bills (10%), and lastly, the equity market volatility (5%). Some more conclusions were that the infectious disease tracker was commonly used by the models for the US and Asian markets but not for the EU cases, that the RUWAR was less likely to be used for the US market compared to the rest of the world and that the lagged return values for the quarterly data for WTI showed less predictive ability compared to the EU and Asia. Here, it is important to state that this might have been either because of the market conditions or because we are comparing oil and gas, which are indeed related as products but do not always follow the same behavior, especially in the last decade.

In conclusion, it is more effective to leverage macroeconomic effects for larger timeframe predictions, and there are exogenous predictive variables that can improve our predictions, such as event indicators, the infectious disease tracker, or relevant variables connected to the expectations of the economy, GDP growth, and bond rates. In the end, we again proved that the energy market is highly volatile and can be easily destabilized by factors outside of the market, almost always categorical, that are impossible to model before making their appearance.

QUARTERLY	WTI						EUNG						ASIANG					
	Regression	GARCH	IGARCH	RF	XGBOOST	LSTM	Regression	GARCH	IGARCH	RF	XGBOOST	LSTM	Regression	GARCH	IGARCH	RF	XGBOOST	LSTM
LAG1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
LAG2							X	X		X		X						
LAG3					X		X			X		X					X	
LAG4																		
USAGDP	X			X		X	X			X		X	X			X		X
Equity Market Volatility		X												X				
3-Month Treasury Bill								X	X									
Infectious Disease Tracker	X			X	X	X					X		X			X	X	X
COV19				X	X	X				X	X	X				X	X	X
RUWAR				X		X				X	X	X				X		X

MONTHLY	WTI							EUNG							ASIANG						
	Regression	GARCH	IGARCH	EGARCH	RF	XGBOOST	LSTM	Regression	GARCH	IGARCH	EGARCH	RF	XGBOOST	LSTM	Regression	GARCH	IGARCH	EGARCH	RF	XGBOOST	LSTM
LAG	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Equity Market Volatility																					
Economic Policy Uncertainty Europe	X		X	X	X	X	X				X				X		X	X	X	X	X
NY Business Conditions																					
3-Month Treasury Bill																					
Infectious Disease Tracker	X				X	X	X	X							X				X	X	X
Nickel																					
COV19								X				X	X	X							
RUWAR								X				X	X								

Table 7: General Summary of All Selected Predictive Variables Over Different Models and Test cases

9.2 Model Results

9.2.1 Regression

As the first method and most commonly used for simple problems, linear regression is the simplest tool that we used for this analysis as the baseline that all other methods should prove to be superior to. As it is considered important to avoid repetition, instead of showing all results for all six cases, the results for Quarterly Returns for European Natural Gas will be used as a representation of the process, while the rest of the analytic findings can be found in the Appendix.

OLS Regression Results

Dep. Variable:	EUNG	R-squared:	0.322
Model:	OLS	Adj. R-squared:	0.253
Method:	Least Squares	F-statistic:	4.650
Date:	Sat, 20 Jan 2024	Prob (F-statistic):	2.04e-05
Time:	12:32:26	Log-Likelihood:	33.842
No. Observations:	109	AIC:	-45.68
Df Residuals:	98	BIC:	-16.08
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.0206	0.026	-0.790	0.432	-0.072	0.031
EUNG_LAG1	0.2291	0.114	2.018	0.046	0.004	0.454
EUNG_LAG2	-0.1750	0.105	-1.669	0.098	-0.383	0.033
EUNG_LAG3	0.1810	0.102	1.773	0.079	-0.022	0.384
EUNG_LAG4	0.0982	0.114	0.863	0.390	-0.128	0.324
USAGDP	1.7833	1.601	1.114	0.268	-1.393	4.960
Equity Market Volatility	0.0023	0.003	0.698	0.487	-0.004	0.009
3 Month Treasury Bill	-0.0298	0.053	-0.568	0.572	-0.134	0.075
Infectious Disease Tracker	-0.0641	0.041	-1.554	0.123	-0.146	0.018
COV19	0.2595	0.077	3.379	0.001	0.107	0.412
RUWAR	-0.4635	0.135	-3.424	0.001	-0.732	-0.195

Omnibus:	18.637	Durbin-Watson:	2.043
Prob(Omnibus):	0.000	Jarque-Bera (JB):	69.032
Skew:	-0.379	Prob(JB):	1.02e-15
Kurtosis:	6.824	Cond. No.	586.

Table 8: General Summary of all Variables in Linear Regression for Quarterly Prediction for European Natural Gas

As clearly seen, not all variables have a significant impact on the commodity return prediction. As referenced in Chapter 9.1, the variable selection changes case by case of model and time series. After careful step-by-step selection, the multiple regression model for the European natural gas uses the lagged returns up to lag three and the event indicator variable COV19. It is important to note that the R squared is low, noting the lack of information provided to explain the complexity of the problem.

```
                            OLS Regression Results
=======================================================================================
Dep. Variable:                   EUNG   R-squared (uncentered):                   0.426
Model:                            OLS   Adj. R-squared (uncentered):              0.402
Method:                 Least Squares   F-statistic:                              17.96
Date:                Sat, 20 Jan 2024   Prob (F-statistic):                    4.56e-11
Time:                        12:32:45   Log-Likelihood:                          57.707
No. Observations:                 101   AIC:                                     -107.4
Df Residuals:                      97   BIC:                                     -96.95
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
EUNG_LAG1      0.6419      0.096      6.713      0.000       0.452       0.832
EUNG_LAG2     -0.4439      0.107     -4.137      0.000      -0.657      -0.231
EUNG_LAG3      0.3209      0.097      3.318      0.001       0.129       0.513
COV19          0.2368      0.058      4.069      0.000       0.121       0.352
==============================================================================
Omnibus:                       12.475   Durbin-Watson:                   2.087
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               24.860
Skew:                          -0.422   Prob(JB):                     4.00e-06
Kurtosis:                       5.279   Cond. No.                         2.50
==============================================================================
```

Table 9: Report Summary for Multiple Linear Regression for Quarterly Data

One hypothesis here is that the existence of more exogenous factors not included in this model, e.g., geopolitical factors, override the importance of those variables included, or their relationship with the returns is not constant throughout the dataset. As a result, the architecture of this linear model is not capable of understanding and capitalizing on these ever-changing relationships.

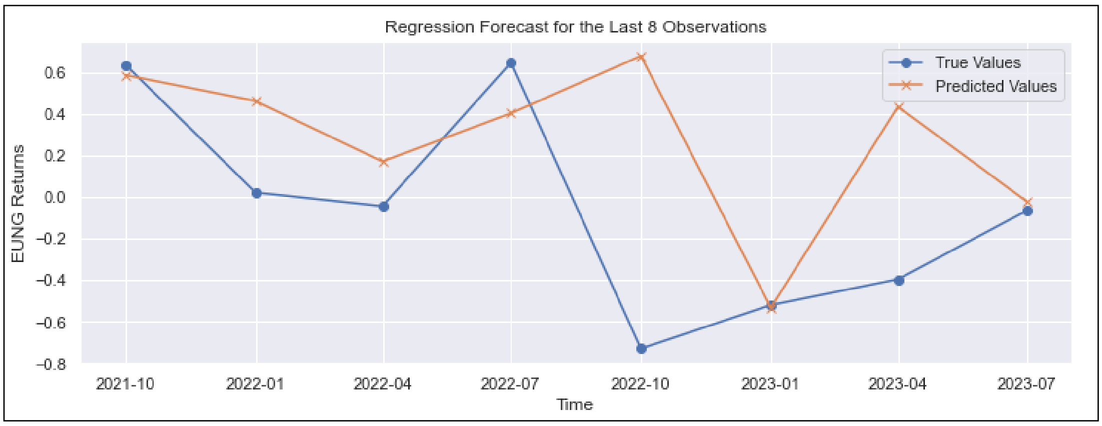


Figure 8: Multiple Linear Regression Forecasting for Quarterly Data

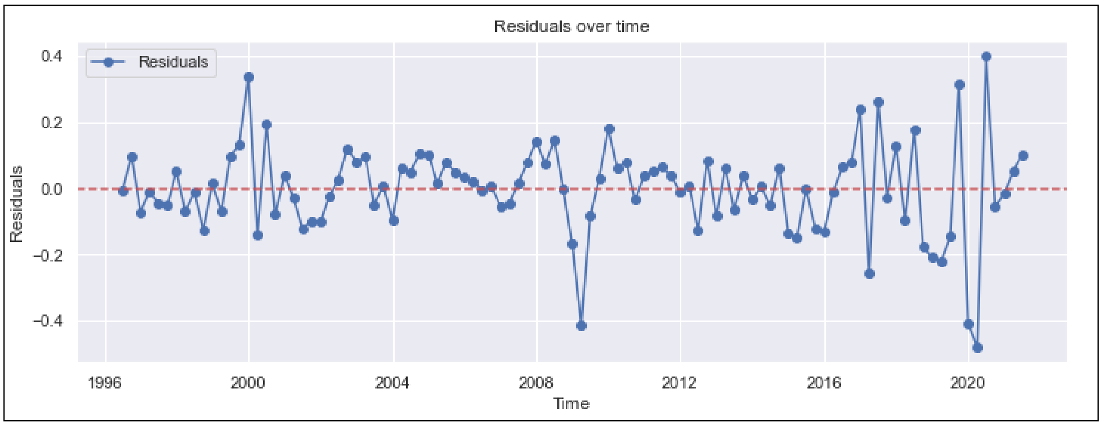


Figure 9: Residuals of Multiple Linear Regression over time

In Figure 9, we can vividly observe that autocorrelation shows that the residuals do not seem to behave like white noise. They do not have a mean close to 0 and seem to include trends. This means that within the residuals, information exists that we failed to draw and incorporate into our prediction.

As mentioned before, for every model, certain rules and conditions must be present to model the data effectively, as in most financial time series, previous observations are not independent, and the model residuals are prone to autocorrelation. While the ACF residuals are within bounds, the PACF steps out of bounds and understates an inadequate adjustment for autocorrelation that needs the use of an AR() component.

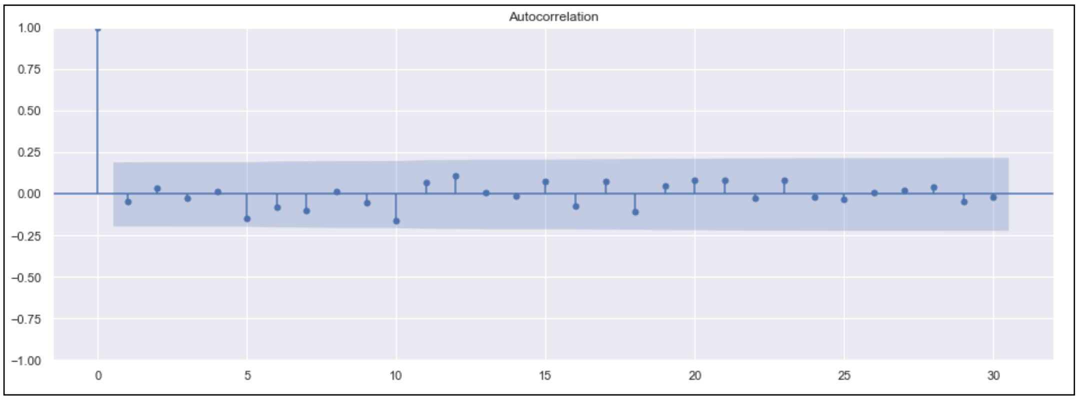


Figure 10: ACF of Residuals for the Multiple Linear Regression

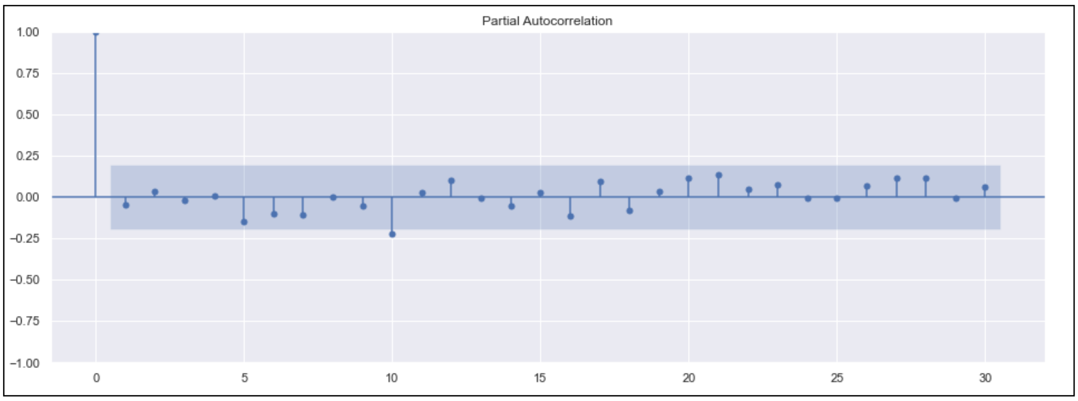


Figure 11: PACF of Residuals for the Multiple Linear Regression

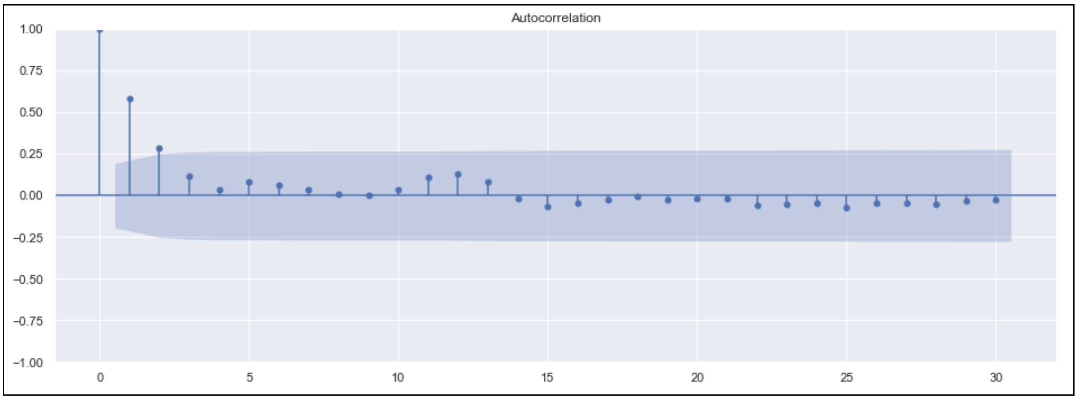


Figure 12: ACF of Squared Residuals for the Multiple Linear Regression

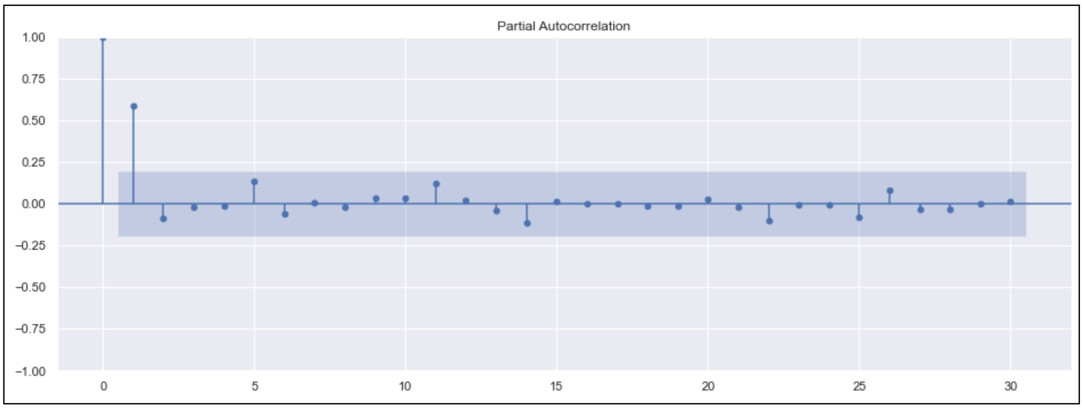


Figure 13: PACF of Squared Residuals for the Multiple Linear Regression

Figures 12 and 13 reveal the main problem of the multiple regression approach. They indicate a significant autocorrelation in the squared residuals. More specifically, we see in Figure 12 autocorrelation in the variance of the residuals, pointing towards volatility clustering or heteroscedasticity. This suggests that the model's residuals have patterns of varying volatility that the model has not captured. Figure 13 also shows out-of-bounds squared residuals in a PACF plot, suggesting the existence of specific lagged relationships in the variance that are not accounted for by our model.

Heteroscedasticity in the residuals can lead to inefficient estimates and can affect the reliability of some standard tests of significance. When the squared residuals show significant autocorrelations, it suggests that the model might be failing to capture some aspect of the data's volatility or structural changes over time. The many heteroskedastic problems push us to the implementation of GARCH components to improve the model fit by capturing this lagged effect in variance.

Lastly, one more violated assumption is the normality of residuals. The residuals do not follow a normal distribution, which is seen through a simple quantile-quantile plot in Figure 14 but also confirmed in our code by using a Shapiro-Wilk test.

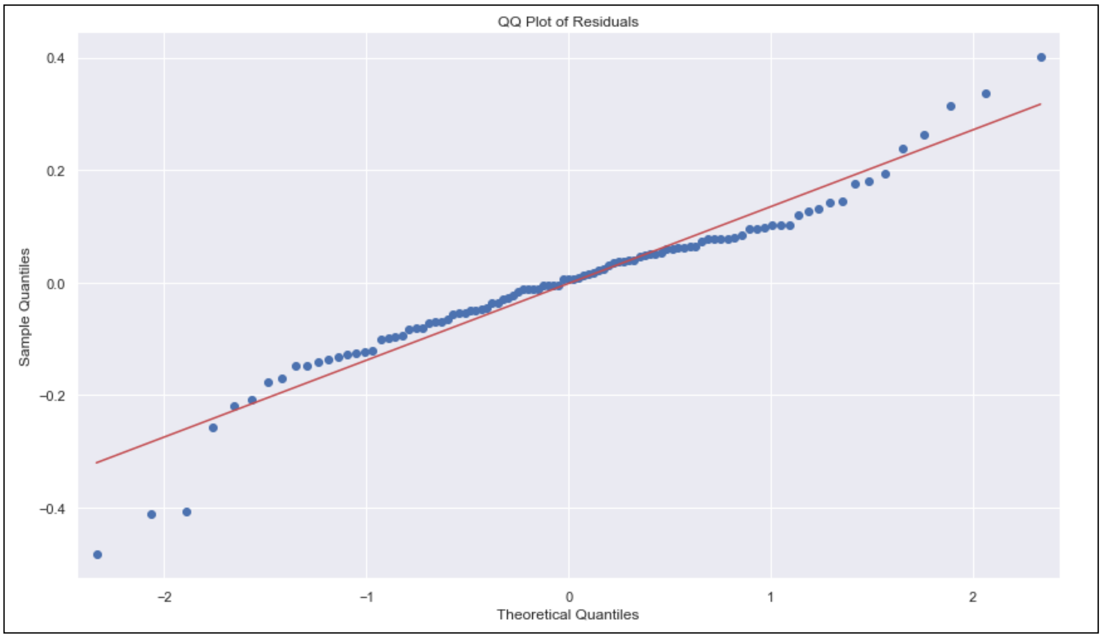


Figure 14: Quantile Quantile Plot for Residuals of Multiple Linear Regression

The pattern analyzed for the European natural gas quarterly data repeats each case study, and this fact is a strong indicator that Regression is a method truly unfit for this peculiar problem. The violation of assumptions, the lack of independence of the residuals, the presence of heteroskedasticity, and the violation of normality led us to the use of autoregressive, moving averages, and GARCH components in order to effectively utilize the nature of our data.

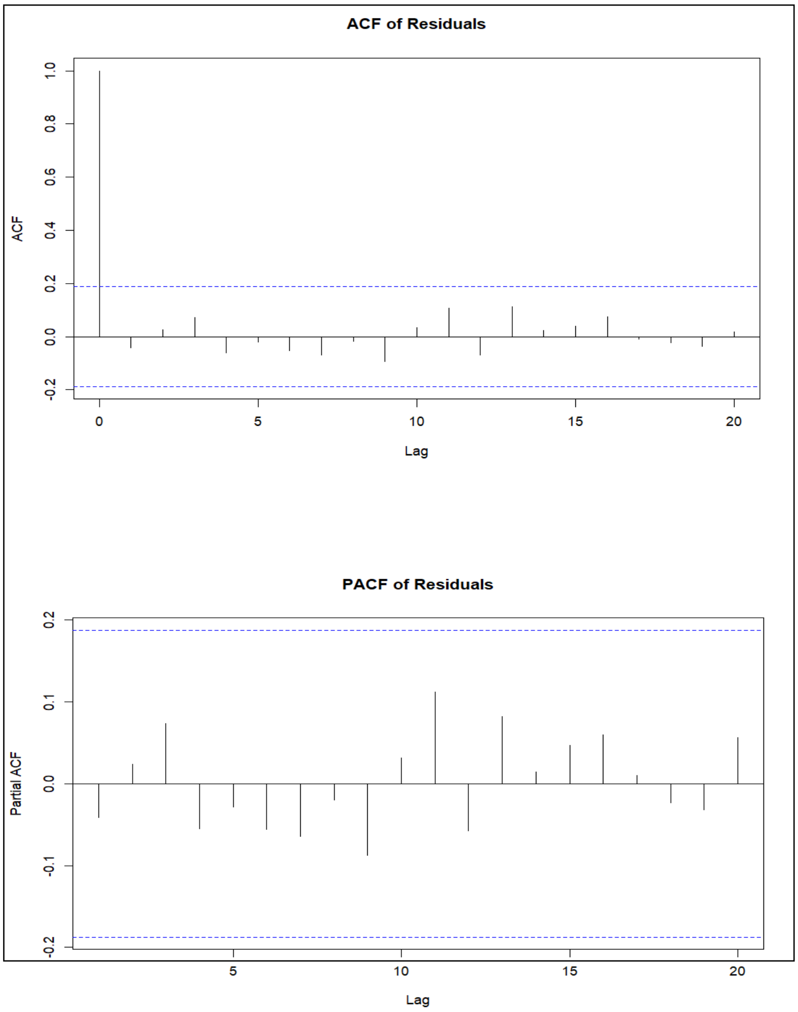
9.2.2 Regression model with autocorrelated errors and GARCH components

As mentioned above, the nature of the problem is most fitting for the implementation of Autoregressive, Moving Averages, and Autoregressive Conditional Heteroskedasticity models. For the six cases we are testing, those being the monthly and quarterly data of the American crude, European Natural Gas, and Asian Natural Gas, four different implementations of the GARCH Family were used—the simple GARCH, IGARCH, EGARCH and GJR-GARCH.

To prove the beneficial effect of the AR, MA, and GARCH upon the existing regression model in chapter 9.2.1, we fitted MA(2) and GARCH(1,1) to determine if the detected problems of autocorrelation and heteroskedasticity are terminated effectively.

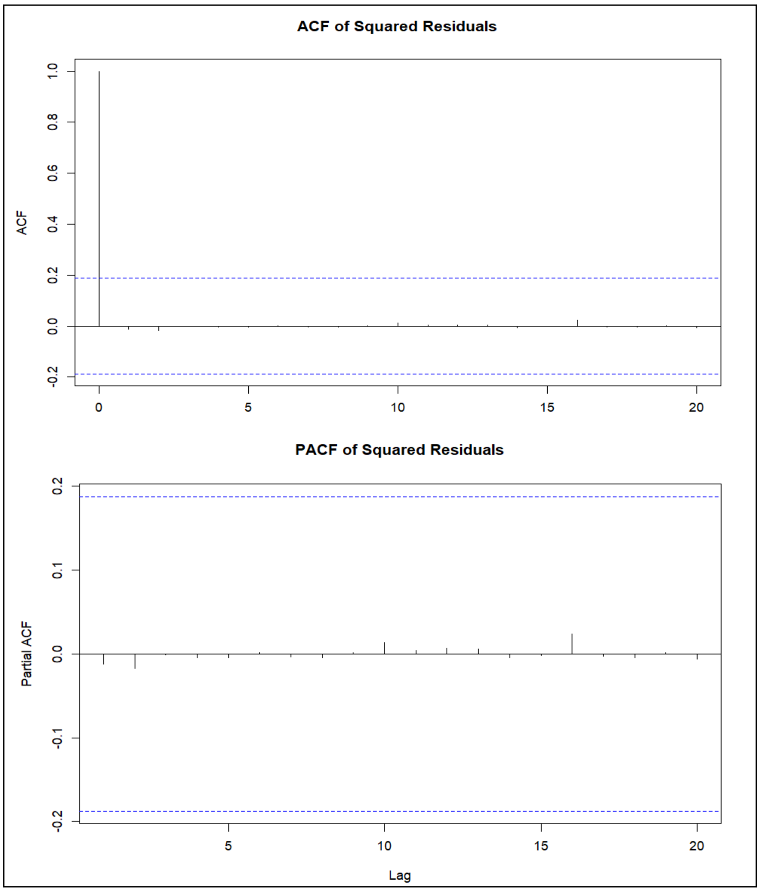
	Estimate	*Standard Error*	*t Value*	*Pr(> | t |)*
Mu	0.008704	0.005631	1.5457	0.122179
Ma1	0.027460	0.005553	4.9450	0.000001
Ma2	1.082013	0.001151	939.457	0.000000
Mxreg1	0.627333	0.001322	474.4285	0.000000
Mxreg2	-1.120604	0.001297	-863.7411	0.000000
Mxreg3	0.732686	0.002042	358.8770	0.000000
Mxreg4	0.197767	0.052101	3.7959	0.000147
Omega	0.006918	0.005327	1.2986	0.194084
Alpha1	0.759536	0.294401	2.5799	0.009882
Beta1	0.239464	0.114318	2.0947	0.036196
Shape	3.037048	1.212899	2.5040	0.012281

Table 10: Robust Standard Errors of the Regression after applying Moving Average and GARCH components



Figures 15, 16: ACF and PACF Residual plots after applying Moving Average and GARCH components

As can be observed in Figures 15 and 16, any autocorrelation problems have been solved. With the use of the MA(2), we captured an important relationship of shocks over six months, previously detected in Figure 11 as out of bound for lag 10.



Figures 17, 18: ACF and PACF Squared Residual plots after applying Moving Average and GARCH components

The use of the Generalized Autoregressive Conditional Heteroskedasticity of order (1,1) solved any heteroskedastic issues seen in Figures 12 and 13, which were significant. The lack of autocorrelations in the squared residuals suggests that the GARCH(1,1) model adequately captures the conditional volatility present in the time series data. This means that the model is successful in explaining the volatility clustering. Moreover, the fact that incorporating a GARCH(1,1) component leads to no visible autocorrelations in squared residuals confirms the presence of GARCH effects in the data. This is an important diagnostic feature, as it justifies the use of GARCH models over simpler models that assume constant volatility. Lastly, the absence of significant autocorrelations in the squared residuals implies that there is little to no leftover pattern in the volatility structure that the model has failed to capture, indicating an efficient utilization of information in the time series, leading to improved forecasting accuracy.

One yet unsolved problem is the residual distribution that remains non-normal.

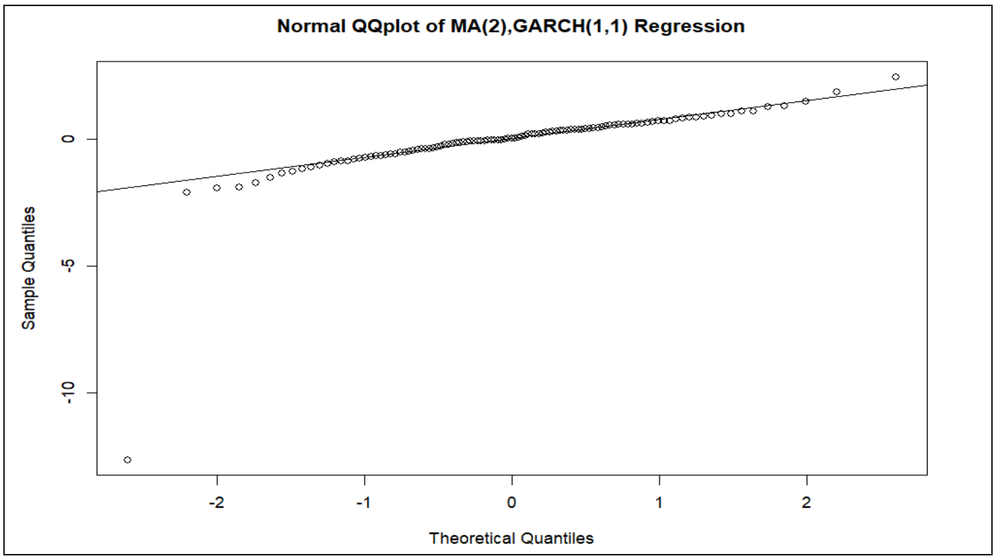


Figure 19: Quantile Quantile Plot for Residuals of Multiple Linear Regression with MA(2) and GARCH(1,1)

JARQUE BERA TEST		
DATA	Degrees of Freedom	P-value
RESIDUALS	2	2.2e-16
	SHAPIRO-WILK TEST	
DATA	W	P-value
RESIDUALS	0.5437	2.2e-16

Table 11: ARIMA GARCH Robust Standard Errors

The lack of Normality was significant, and based on the Akaike Information criterion, we gradually made further changes to our model, such as including the 3-month Treasury Bill rate, to exclude the lagged returns of the EUNG_LAG2 and thus the MA(2), implement an AR(1) component and use the skewed student as our distribution.

	Estimate	*Standard Error*	*t Value*	*Pr*(> | *t* |)
Mu	0.004537	0.007236	0.62705	0.530628
Ar1	-0.315865	0.133247	-2.37052	0.017763
Mxreg1	0.993242	0.168889	5.88104	0.000000
Mxreg2	-0.427169	0.170122	-2.51097	0.012040
Mxreg3	0.045497	0.013652	3.33264	0.000860
Omega	0.001253	0.000775	1.61553	0.106195
Alpha1	0.638190	0.100720	6.33627	0.000000
Beta1	0.360810	0.143643	2.51185	0.012010
Shape	9.593404	8.125839	1.18060	0.237760

Table 12: ARIMA GARCH Robust Standard Errors

The robust standard error, as provided in Table 5, prove the statistical significance of the three predictive variables and the importance of all the components mentioned above.

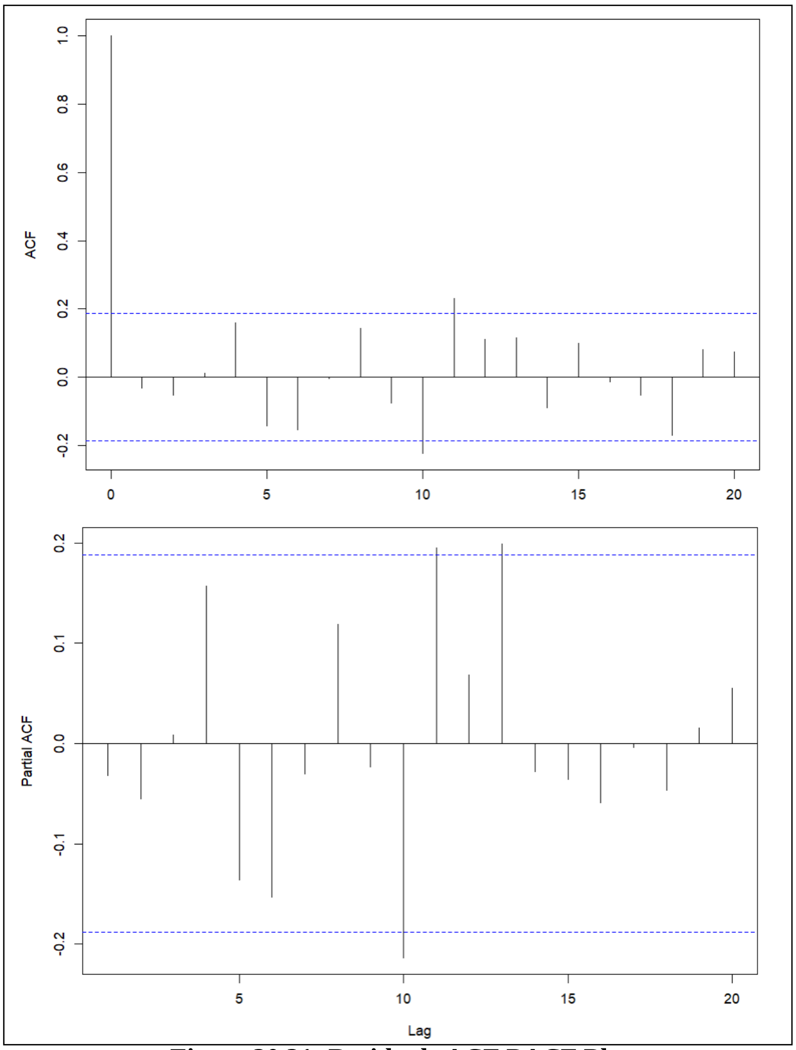


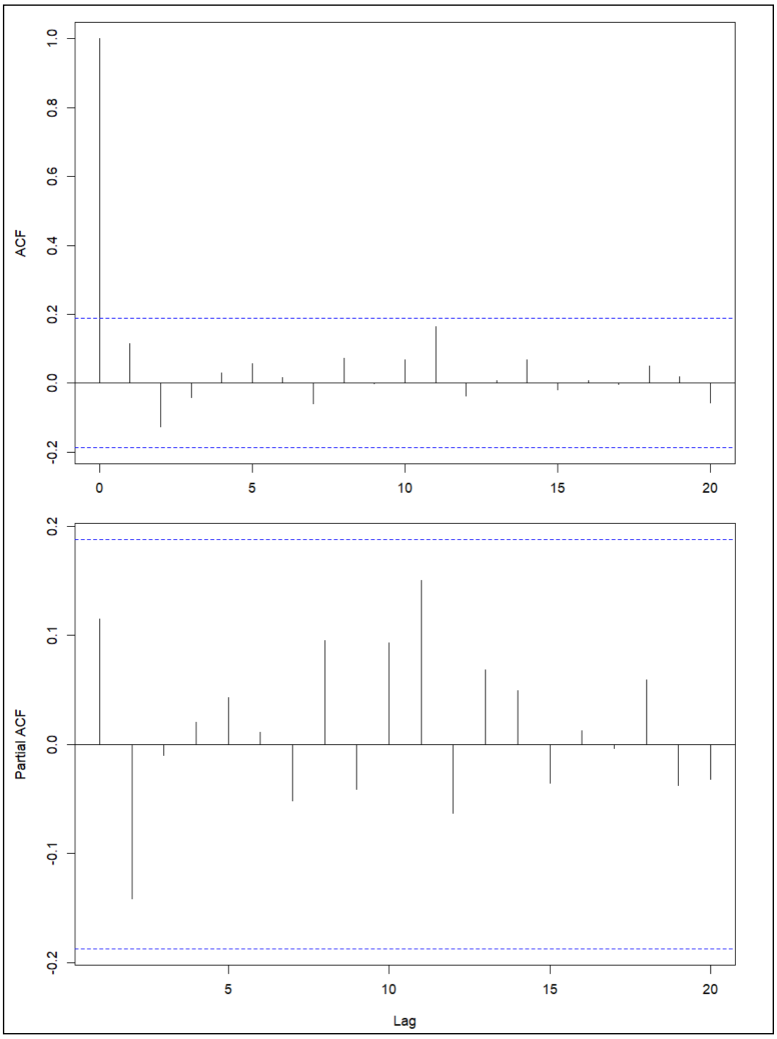
Figure 20,21: Residuals ACF PACF Plot

Because for some lags(10,11,13), it is observed that correlation values exceed the significance thresholds, we confirmed that there is no significant autocorrelation by conducting Ljung-Box tests that proved that there was not enough statistical evidence of a proven autocorrelation. As a result, we can say that the values falling outside the confidence bands are noise.

LJUNG-BOX TEST

TEST STATISTIC	P-value	Degrees of Freedom
17.90378	0.056608	10

Table 13: Ljung-Box test results for lag 10



Figures 22, 23: Squared Residuals ACF PACF Plots

In conclusion, the autocorrelation and partial autocorrelation errors have vastly improved in comparison to the Linear regression, and even if not as perfect as the previous model, the Akaike Information criterion was smaller, thus indicating a better-fitting model.

Lastly, by correctly identifying the true distribution of our data, we made great progress in the distribution of our residuals as well (seen in Figure 24 and Table 13). Since the residuals are normally distributed, it is ensured that the model has captured the underlying data generation process well, leaving only random noise that cannot be predicted or explained by the model and proving a goodness-of-fit.

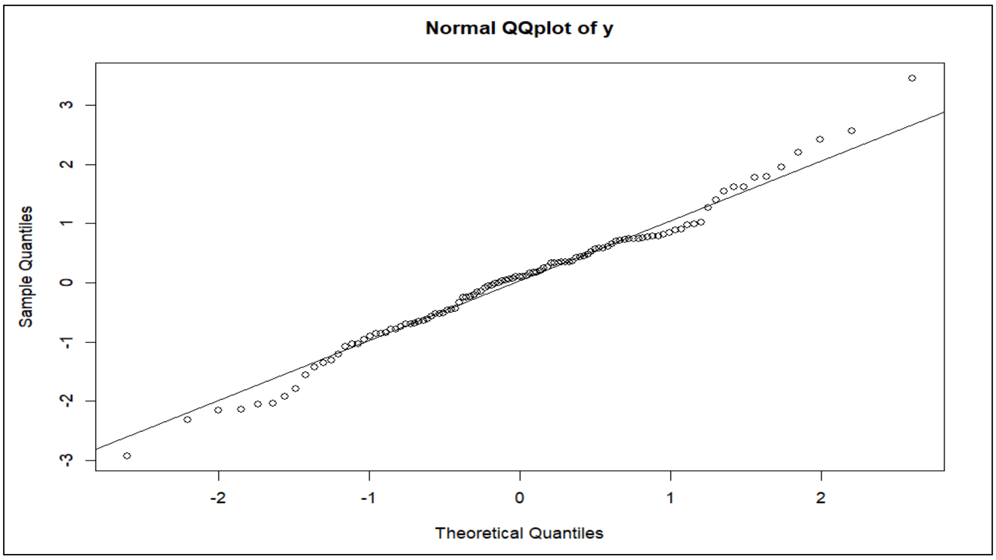


Figure 24: Quantile Quantile Plot for Residuals of Multiple Linear Regression with Skewed Student Distribution

JARQUE BERA TEST		
DATA	Degrees of Freedom	P-value
RESIDUALS	2	0.3991
SHAPIRO-WILK TEST		
DATA	W	P-value
RESIDUALS	0.98615	0.3229

Table 13: ARIMA GARCH Robust Standard Errors

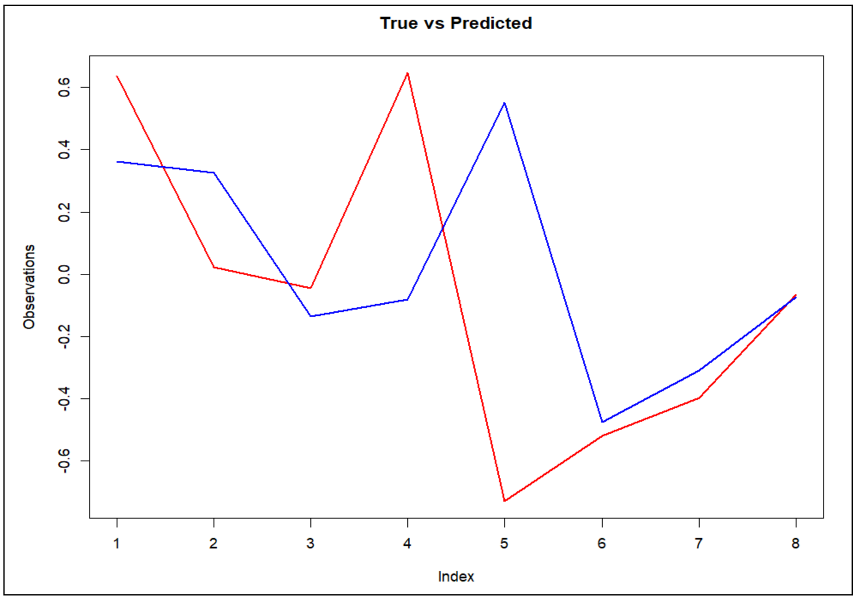


Figure 25: Forecasting for Quarterly Data

The Generalized Autoregressive Conditional Heteroskedasticity model of order (1,1) proves its reputation for reliability as it closely follows the true observation even at a delay. Based on the MAE, MAPE, and MSE metrics, GARCH(1,1) finds itself to be an above-average performing methodology. That observation is repeated in all six examples as it moves between the mean and the second-best solution. Also, it was the best of the GARCH family of models in terms of finding the statistical significance of components and predictive variables, as well for lags not exceeding permissible limits. In conclusion, it is the perfect baseline for a competitive solution.

9.2.3 Regression model with autocorrelated errors and IGARCH

For the IGARCH model, the predictive variables selected were the previous quarter's returns and the past quarter's 3-month Treasury Bill rate. The model uses AR(3), MA(2), IGARCH(1,1), and a generalized error distribution.

	Estimate	***Standard Error***	***t Value***	***Pr(> | t |)***
Mu	-0.005079	0.000035	-145.23802	0.000000
Ar1	-0.270080	0.002255	-119.76035	0.000000
Ar2	-1.165850	0.004721	-256.96938	0.000000
Ar3	-0.064657	0.001245	-51.91332	0.000000
Ma1	0.192975	0.003208	60.15064	0.000000
Ma2	1.210538	0.004824	250.92081	0.000000
Mxreg1	0.613641	0.021218	28.92081	0.000000
Mxreg2	0.047577	0.003111	15.29440	0.000000
Omega	0.001761	0.005938	0.29660	0.766770
Alpha1	0.837132	0.211709	3.95417	0.000077
Beta1	0.162868	NA	NA	NA
Shape	2.008758	5.456869	0.36812	0.712787

Table 15: IGARCH Robust Standard Errors

The IGARCH model's residual errors are presented in Figures 26 and 27. It is clear that for lags, 13 in the ACF and 10 in PACF come outside of the preferred bound, but because we conducted two Ljung-box tests that didn't showed statistical significance so our conclusion is that they are result of noice carrying no significant information and thus degreesing the prediction accuracy.

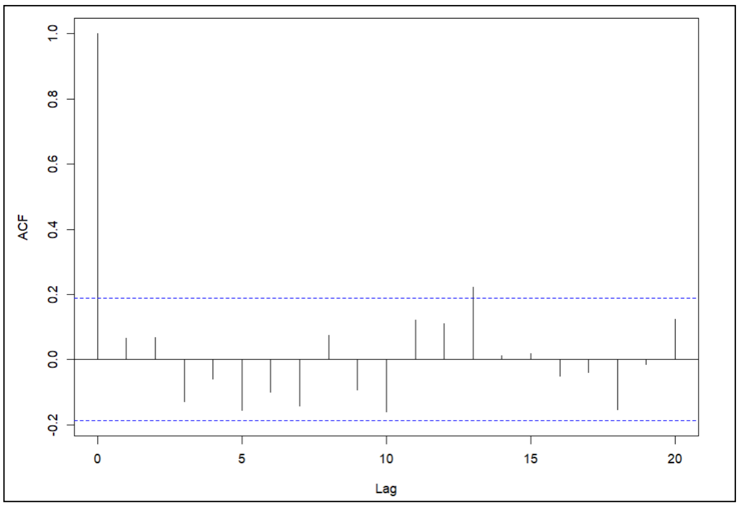


Figure 26: IGARCH Residuals ACF PACF Plots

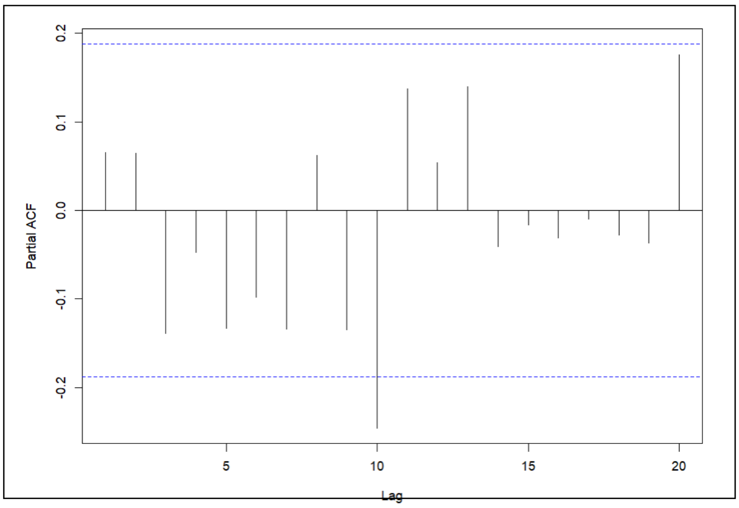


Figure 27: IGARCH Residuals ACF PACF Plots

The squared residual errors are also very good and stay within bounds. The lack of autocorrelations in the squared residuals notifies us for a successful inclusion of conditional volatility present.

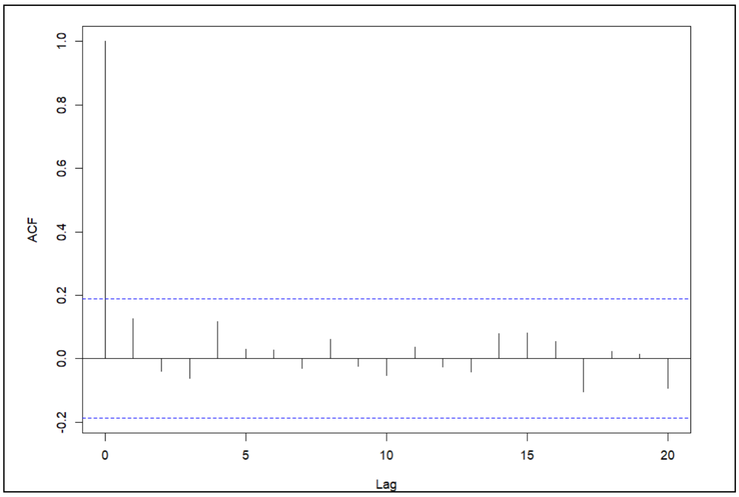


Figure 28: IGARCH Squared Residuals ACF Plot

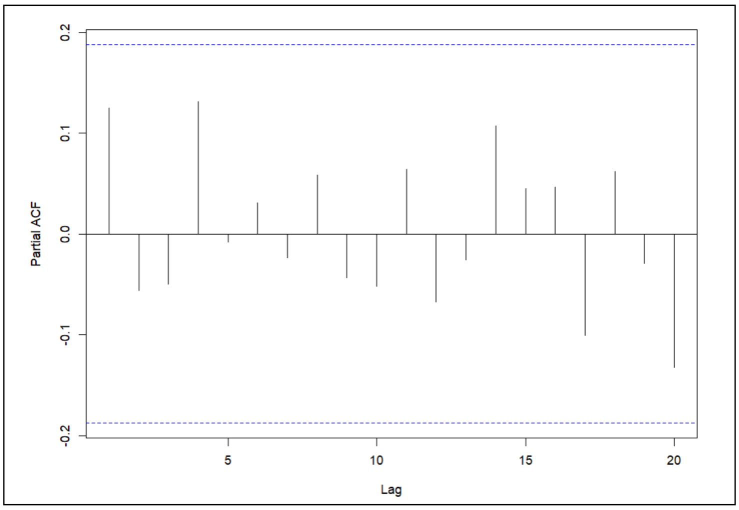


Figure 29: IGARCH Squared Residuals PACF Plot

The IGARCH showed to be a great fit due to the high and long memory volatility present in this type of commodities. Moreover it achieved to pass our baseline, set by the GARCH, at almost every case and statistic. If we only analyzed monthly data, the IGARCH sub-model would be the best, with a large margin for accurate prediction. While concerning the quarterly predictions, it was not the best still performed comparably to the TFT. One of the few drawbacks of this model was its significantly worse flexibility compared to the simple GARCH. Leading in far greater effort in choosing, testing, and incorporating autoregressive components, moving average components, and predictive variables. Nevertheless, it was far more applicable than the two remaining heteroscedastic methodologies.

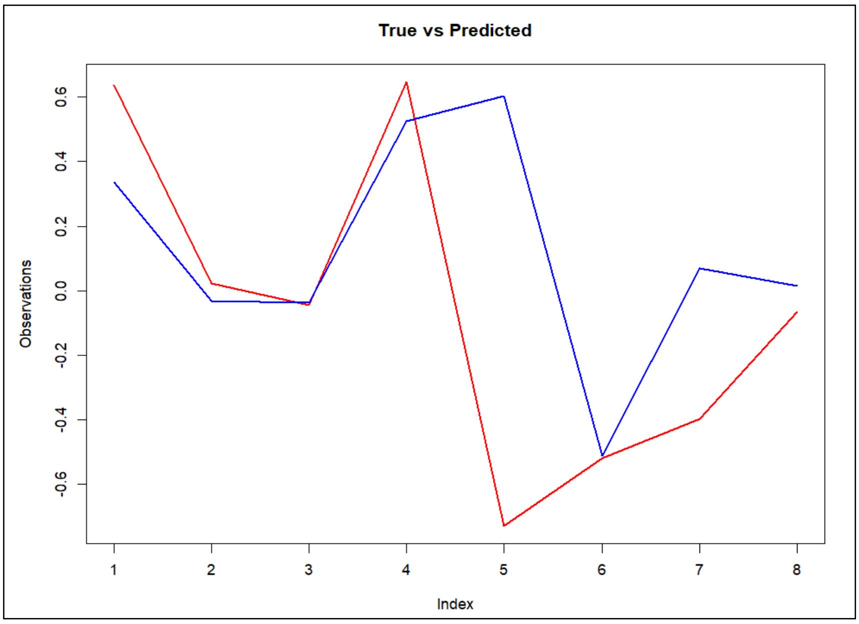


Figure 30: IGARCH Forecasting for Quarterly Data

9.2.4 EGARCH

This model has an illustrious reputation for implementation in financial data; unfortunately, its use on this particular problem had more mixed results. Specifically for European Natural Gas, no combination of components and variables was capable of producing a robust fit, and this was not the only example. Although it performed adequately for monthly returns in all three examples with average accuracy. EGARCH was very particular in the use of multiple predictive variables, and as a result, in most cases, only the lagged values of y were used. However, the EGARCH was not the weakest member of the GARCH family.

9.2.5 GJR-GARCH

The GJR-GARCH model proved completely inadequate for this problem. Out of all six test cases, there was only one robust implementation (Natural Gas for the Asian markets). More interestingly, it was the example where the ML models performed the worst, so GJR-GARCH managed the third-best accuracy result.

In conclusion, the models that are specifically set to solve issues of leverage and asymmetric impact of shock failed to make a great fit. In contrast, IGARCH, with a focus on the lasting impact of shock, and GARCH, with its more generalized and flexible philosophy, were easier to use and more effective in making predictions.

9.2.6 Random Forest

The first step in attempting to utilize the Random Forest Regressor is to select the statistically important explanatory variables. The first three lagged returns and the event indicators were selected. Then, the data were split for training and testing. Also, for the hyperparameter tunning, a randomized search was implemented. For every case, a different tuning was executed separately for each commodity. The difference between monthly and quarterly frequencies was important, with more variables selected by the quarterly models in general. Also, some variables appeared more often than others. The findings concerning the predictive macroeconomic variables will be discussed in detail later.

For the randomized search, the range of the possible hyperparameter combinations was a scale of magnitude higher. However, after achieving a satisfactory result was reached, the number of possible hyperparameters was dropped for better execution performance due to limited time and computation resources.

Random Forest Selected Hyperparameter for European Natural Gas Quarterly Data

Estimators	Samples For Internal Node Split	Samples Required for Leaf Node	Maximum Tree Depth	Calculation Method For Split	Bootstrap
500	12	6	10	auto	YES

Table 16: Random Forests Selected Hyperparameter after Random Search

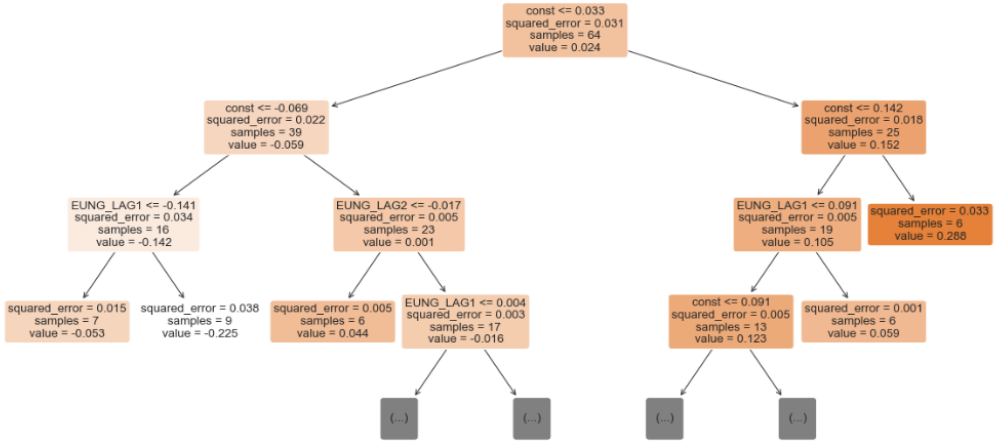


Figure 31: Example of a Tree

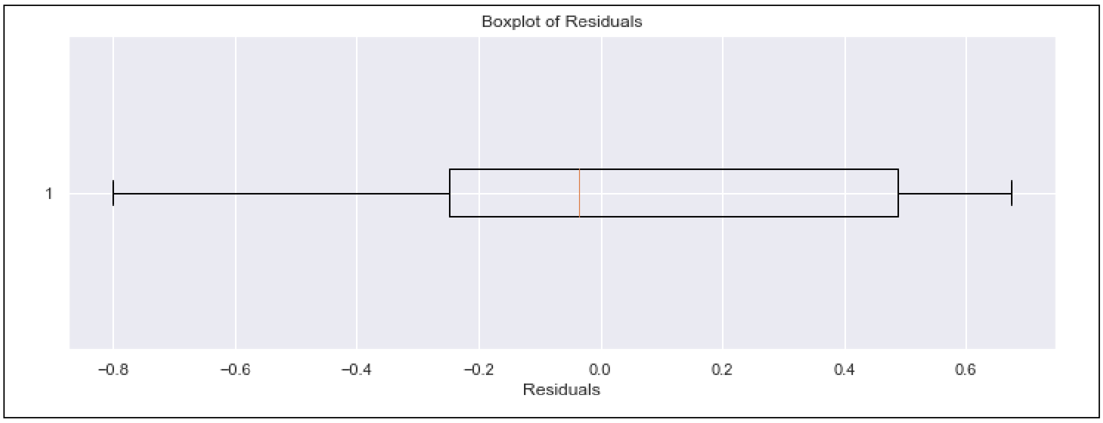


Figure 32: Random Forest Residuals Boxplot

The resulting residuals of the random forest show a mean close to zero but a wide spread. As a result, we can assume that we effectively deal with any strong trends in the data, but at the same time, there is a strong possibility that strong shocks pass through our net and we lose information.

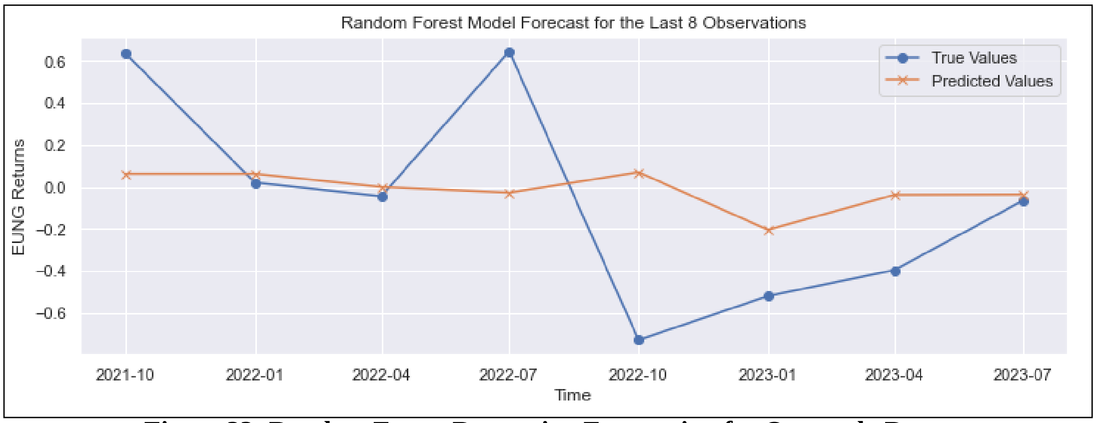


Figure 33: Random Forest Regression Forecasting for Quarterly Data

The resulting prediction is, one might say, very conservative and only moves under extreme market conditions and still only slightly before returning to an equilibrium. This is unusual in comparison to the rest of the cases, but overall, Random Forest finds itself a reliable option, achieving accuracy close to the GARCH and, in general, an above-average performance.

9.2.7 XGBOOST

For the XG Boost, the data preparation was very similar to the Random Forest as they have similar architecture to a degree, the difference being the selection of a Grid Search instead of a Random Search. Here, the hypothesis is that an extensive grid search would offer a higher accuracy prediction, but again, under time pressure and with limited resources, a short random search was selected. Again, each Commodity and frequency had separate searches. In comparison to the Random Forest, the hyperparameters showed a lesser tendency to choose different results and, as such, had a tighter spread. As for the final results, XGBoost seems to underperform in comparison to the Random Forest and, overall, achieve below-average accuracy. So, the question that arises is if the problem is in the model or is in the insufficient tunning.

XGBoost Selected Hyperparameter for European Natural Gas Quarterly Data

Estimators	Subsample ratio	Learning Rate	Maximum Depth	L1 Regularization	L2 Regularization	Random Samples Rate
250	0.7	0.2	2	1	1	0.5

Table 17: XGBoost Hyperparameters

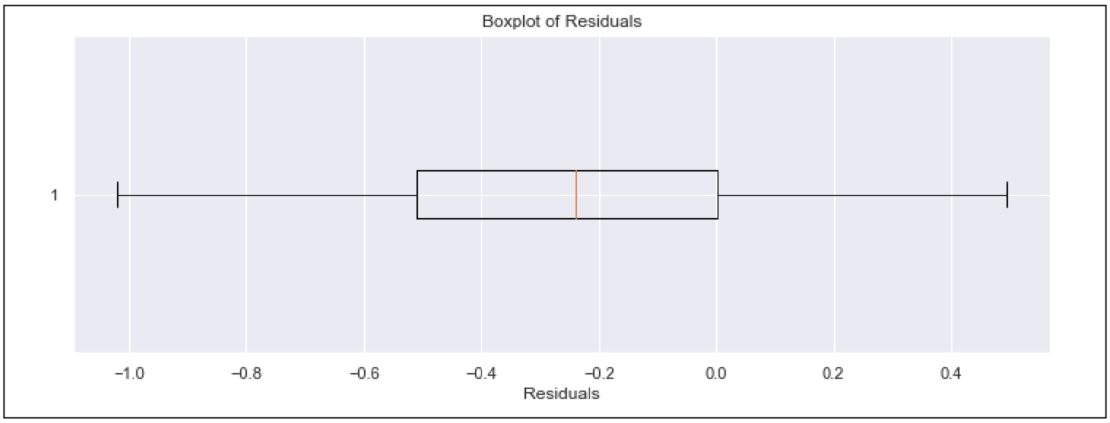


Figure 34: XGBoost Residuals Boxplot

Between Figures 34 and 32, we can observe significant differences in the mean and the spread. While the mean of the residuals was close to zero for the XGBoost, the residuals as all the distribution shifted to the negatives. This is a fact strongly supporting the existence of bias in our model, something that is confirmed by Figure 35.

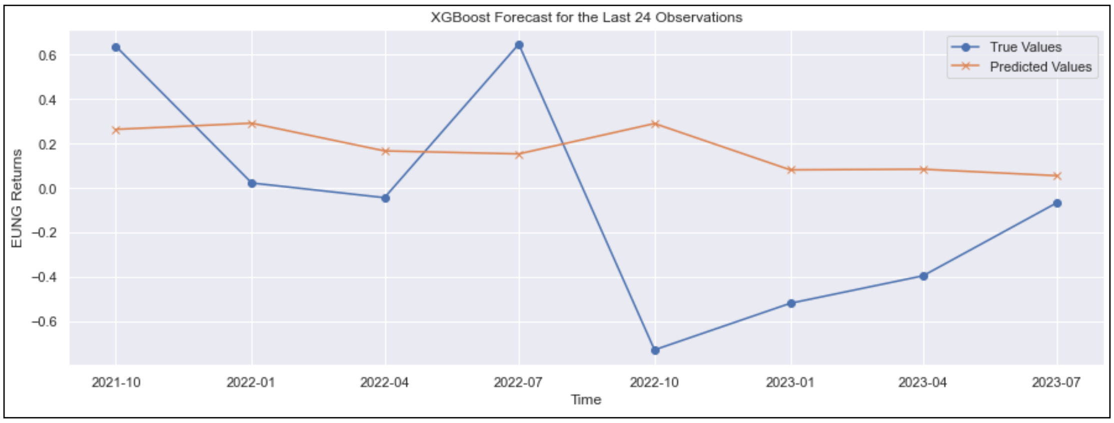


Figure 35: XG Boost Regression Forecasting for Quarterly Data

As the architecture is similar to the Random Forest, the predictions show some similarities to the Random Forest prediction. The mean of predictions is higher than the true variables, and at no point, it predicts negative returns, a not-very-positive result, especially when analyzing economic data. We must state at this point that this positive bias does not repeat for the test cases of the monthly data, where XG Boost proves far more effective. As an example, we offer the XGBoost prediction for the Asia Natural Gas (Figure 36).

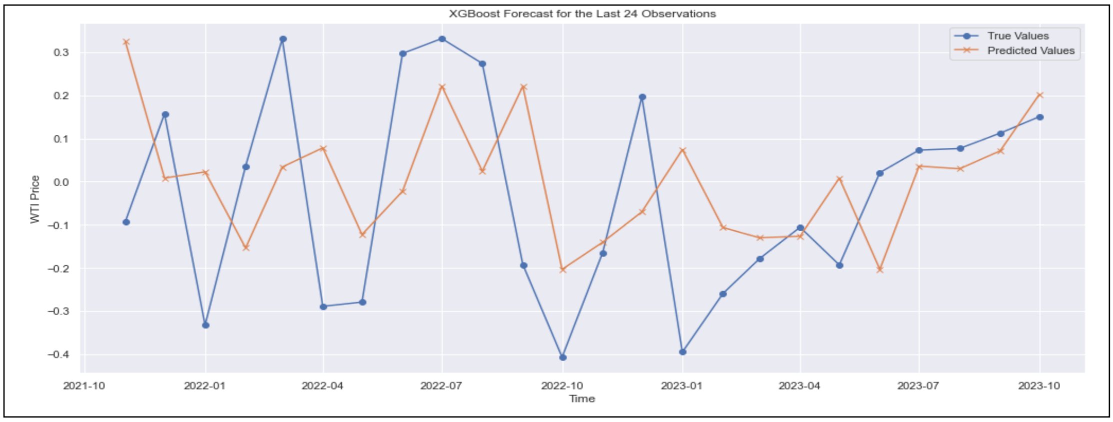


Figure 36: XG Boost Forecasting for Asia LNG Price for Monthly Data

9.2.10 LSTM

Perhaps the most different of other models used in this thesis are the LSTMs. Not only are they not modified and specifically planned for time-series problems as the Temporal Fusion Transformers that would be later presented, but they also lack the basic regression philosophy and tree-like structures, both already implemented. Even without prior knowledge, this becomes apparent from the empirical results as they have entirely different behaviors and show great volatility with extreme results.

The LSTM layers and hyperparameters were set and tested by hand due to a lack of computation resources to run hundreds of tests over a validation set. Again, the layer, number of nodes, and learning rate were different for each problem, meaning that for each occasion, the model must be tuned for the specific time series, and no single setup is capable of providing satisfying results for all combinations. Also, this is the first model whose input was not only differentiated to be stationary but also needed to be scaled between 0 and 1 so the machine learning model can be used optimally.

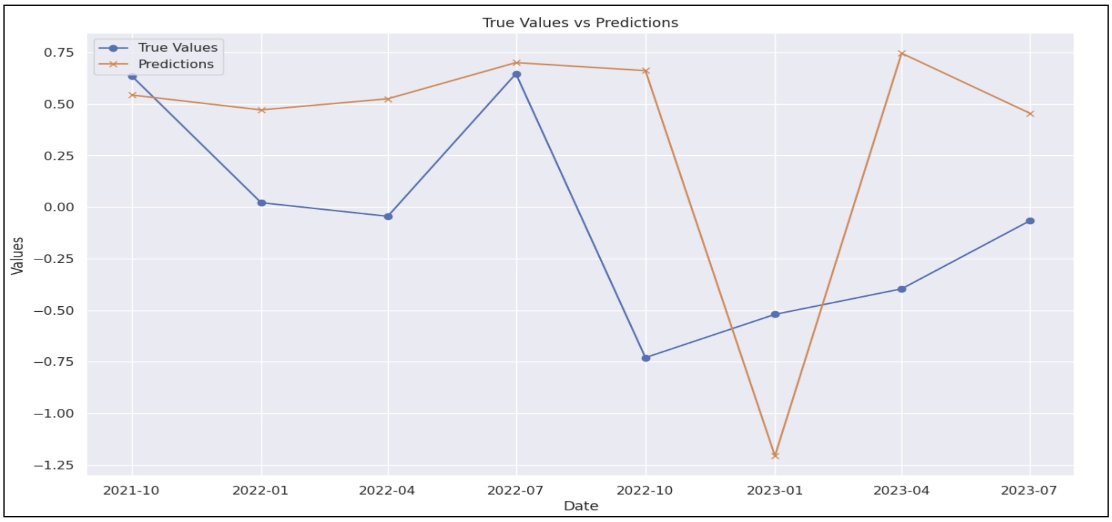


Figure 37: LSTM Forecasting for European Gas Quarterly Data

The extreme values are repeated on quarterly and monthly predictions, and many times, the model fails completely to capture the true trend, making the LSTM the weakest predictor of this paper. Even though a more extensive modification of layers, learning rate, number of nodes, and learning rate would perhaps improve the predictions, I do not find the model as necessary fitting for this kind of problem. It might be used in hybrid Grach combinations, but by itself, I am convinced that it would be overshadowed by almost any other model that has been created specifically for time-series problems. A great example of a model architecture crafted especially for time series and incorporating LSTM is the Temporal Fusion Transformers.

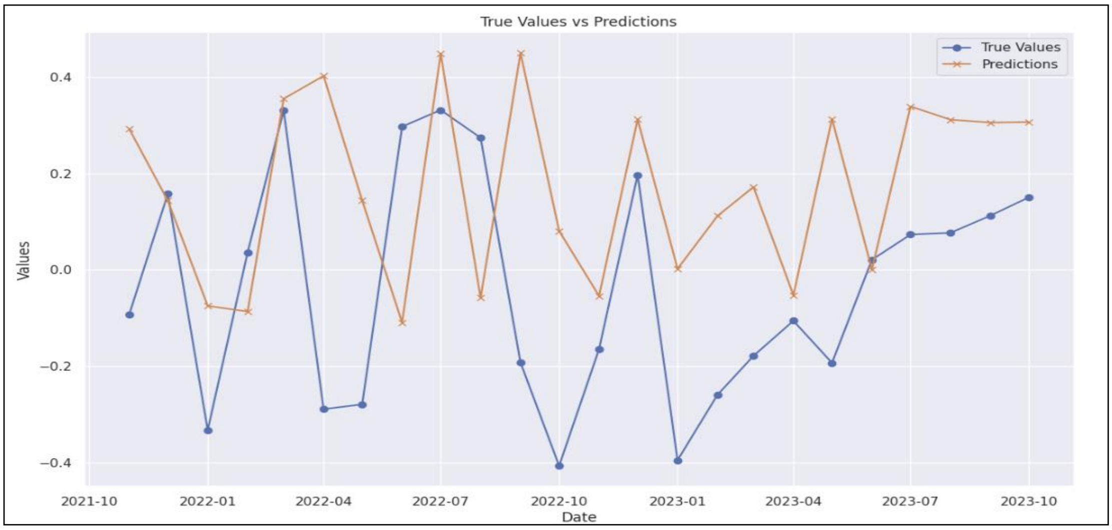


Figure 38: LSTM Forecasting for Asian LNG Monthly Data

9.2.11 TFT

The Temporal Fusion Transformers is, in a way, an amalgamation of multiple different philosophies previously expressed in models in the previous pages. In a way, its predictions often look like a combination of ARIMA and LSTM. For the implementation, the library Darts was used, and beyond the scaling, it was also used in the LSTM, which also required the calculation of covariates as the model uses covariates for the prediction. The main difficulty faced by this model, even if proposed as fairly light, was its long execution time, being the heaviest of all previous models used. That meant that a random search or grid search was indeed out of the question. This leads to the tunning of all hyperparameters such as hidden layers size, number of LSTM layers, number of attention heads, dropout rate, batch sizes, and number of epochs to be done by hand, so in no way optimally selected.

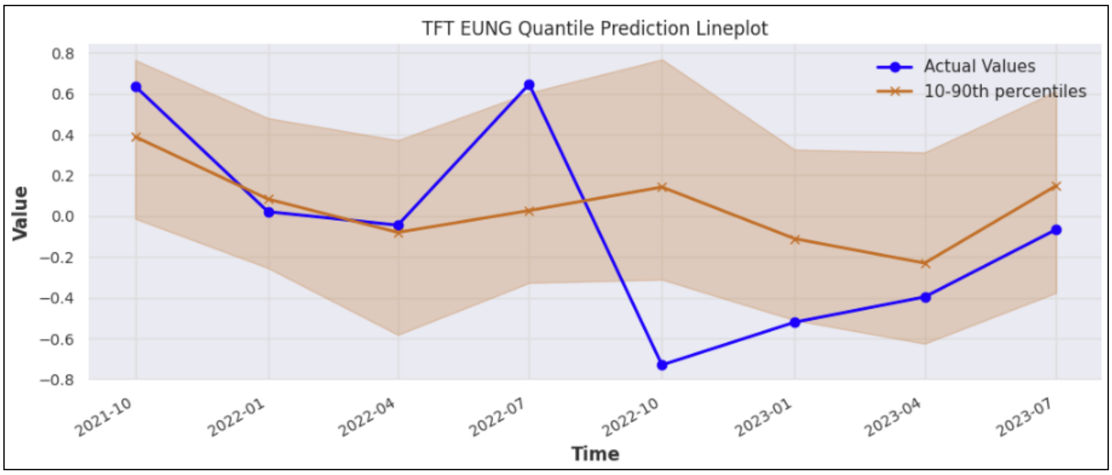


Figure 39: TFT Forecasting for Natural Gas in the European Market for Quarterly Data

Even though challenged by this crucial problem, the model performed admirably and was the best or the top performing in multiple test cases. In general, TFT showed better results in quarterly data than in monthly data when compared to the competition. As an example, Figure 41 represents the worst fit of the model, where underperformed all but two methods.

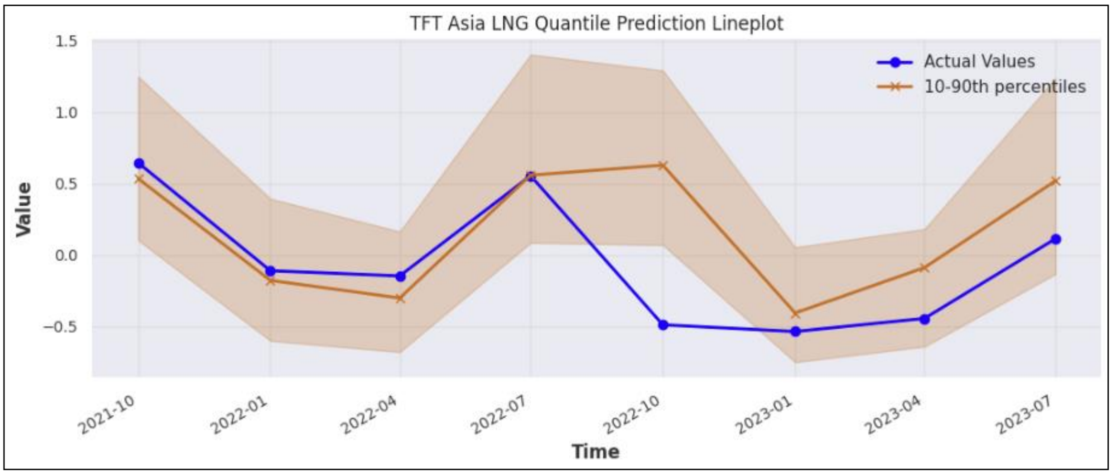


Figure 40: TFT Forecasting for Natural Gas in the Asian Market for Quarterly Data

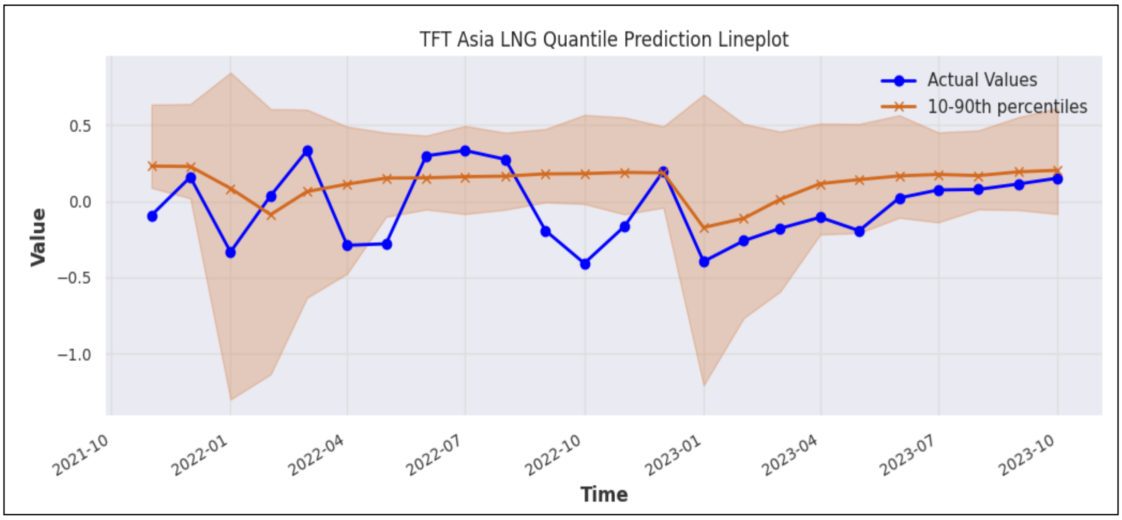


Figure 41: TFT Forecasting for Natural Gas in the Asian Market for Monthly Data

TFT was always able to follow the time series trend but not always with the desired definition. As seen in the most overt example, Figure 32 sometimes feels like it lacks degrees of freedom, and we have reasons to believe that it is more of a tuning problem than an architectural problem as the model is highly complex and more than capable of this desired flexibility.

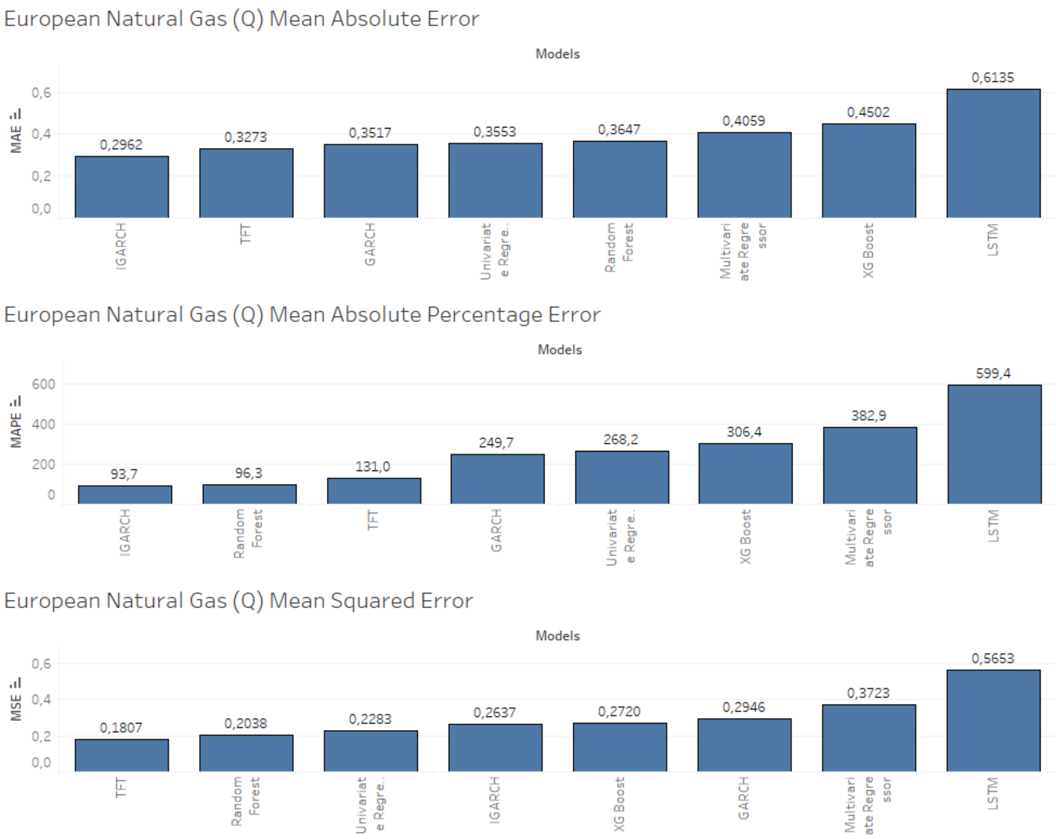


Figure 42: Quarterly European Natural Gas Accuracy Metrics

9.3 Model Comparison

As already mentioned, the model results rarely show the same behavior in all three metrics. So, the model comparison is analyzed into two parts, quarterly and monthly time series, and upon those two timescales, we would attempt to draw conclusions.

Also, it is important to mention that both EGARCH and GJR-GARCH are absent from the Quarterly graphs, the reason being their impossibility to show a robust fit. So, even if their results are somewhat worse from simple GARCH, it would be dishonest to incorporate them for the final analysis. EGARCH is only analyzed in the Monthly analysis where it was possible to form robust fits for the model, but GJR-GARCH, since it achieved a good fit only for one out of six cases, will be omitted entirely.

9.3.1 Quarterly Data

Figure 43: Mean Absolute Error Model Results

Starting with the mean absolute error as our judge, the two best-performing models are the Temporal Fusions Transformers and the IGARCH, with the simple GARCH following close behind. On the other side of the spectrum are the LSTM and the Multiple Regressor. In general, we see that the machine learning models, with one bright exception, are not only less accurate but also less consistent in different markets.

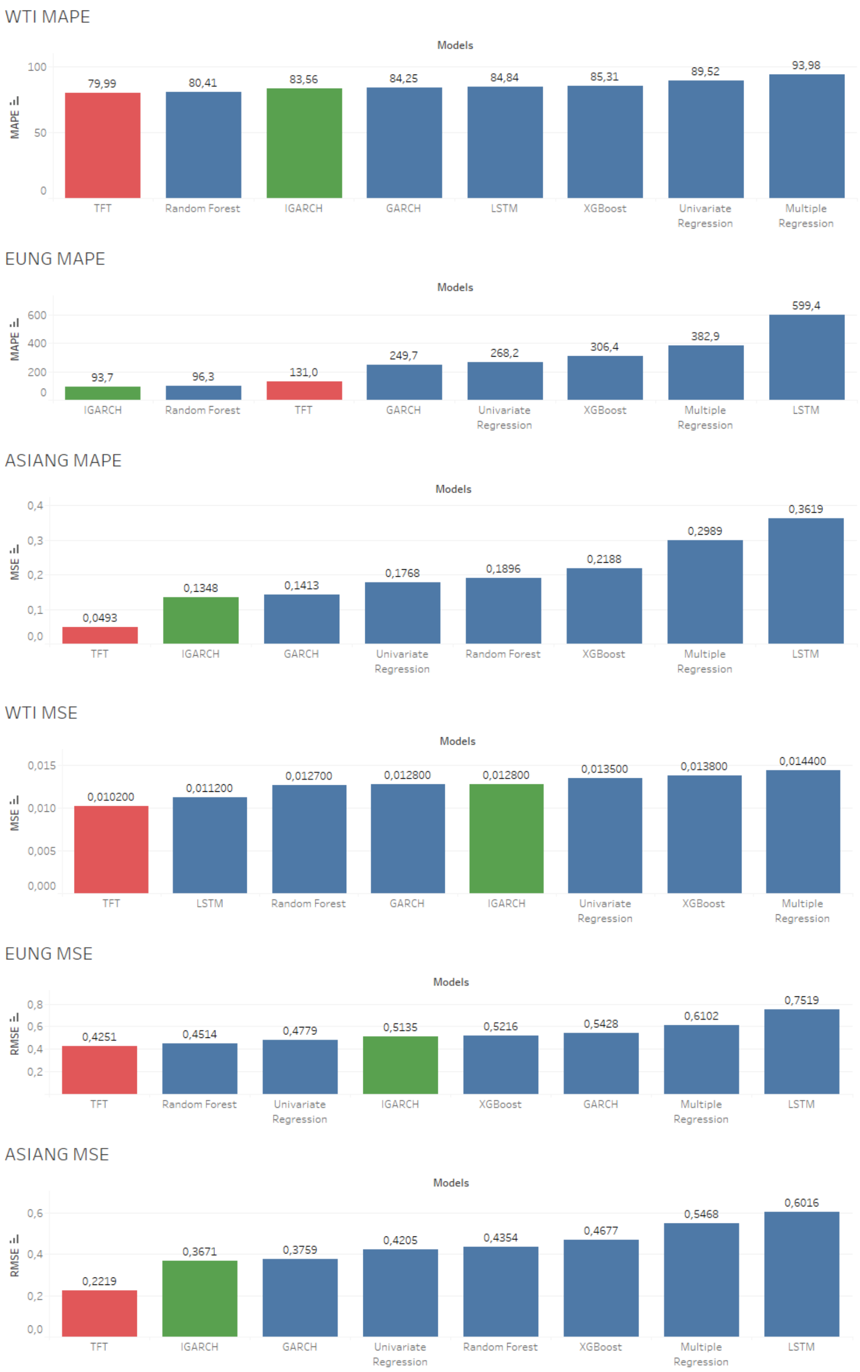


Figure 44: Mean Absolute Percentage Error and Mean Squared Error Model Results

Considering the mean absolute percentage error, there is not much change from Figure 33, except from the Random Forest, which is now considered more accurate. This trend remains for the mean absolute squared Error. Also, for the metrics in Figure 44, IGARCH seems to perform worse than the TFT, something not clearly observed in Figure 43. As for the models for the quarterly prediction, the TFT was shown to be the superior option. Concerning the asset, TFT clearly overperforms for the American crude less profoundly for the ASIA Natural Gas, and it is unclear who is the winner for the European Natural GAS, but it seems to be barely in favor of the Transformers.

9.3.2 Monthly Data

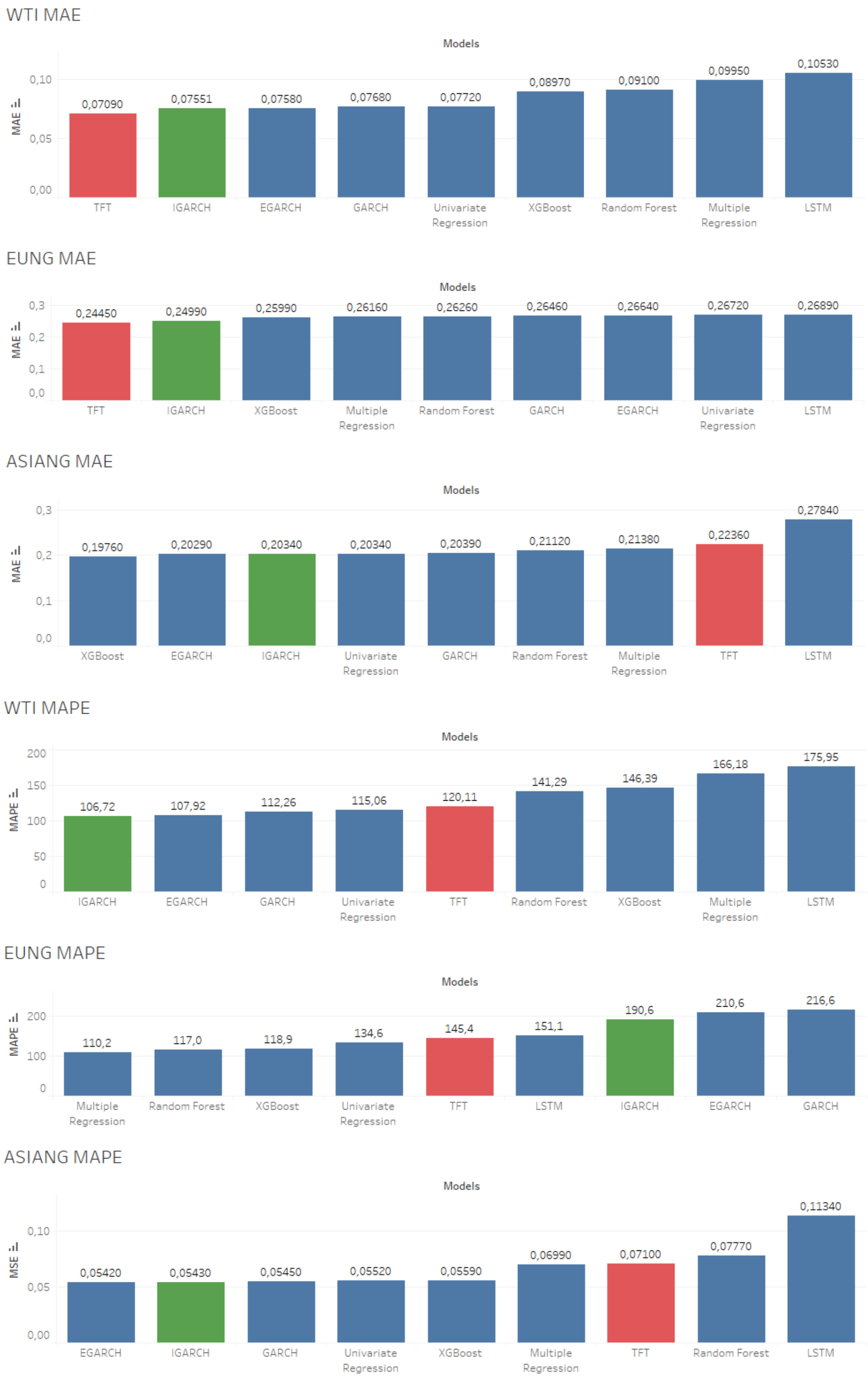


Figure 45: Mean Absolute and Mean Absolute Percentage Error Model Results

For the quarterly data, the Transformers were ahead, but the same does not repeat for the monthly data. Starting at the American Crude, IGARCH is the better model, while TFTs show mediocre results for the MAPE. Accordingly, we must address the performance of EGARCH, which also showed good results like all the GARCH family models.

In contrast with all previous examples, here both the XGBoost and Random Forest have very competitive performances, while the two main contestants fall behind, especially IGARCH. In general, concerning MAPE, the GARCH model performs the worst in this case. As for the Asian Natural Gas, the XGBoost continues to be competitive, and the GARCH models improve, with EGARCH being the best, but this is the example where the TFT appears to have its worst performance. In fact, TFT is consistently in the three worst models. Fact the might indeed show that is truly a problem of bad tuning.

Figure 46: Mean Absolute Percentage Error and Mean Squared Error Model Results

9.3.3 General Findings

The results provide a long list of important conclusions. First of all, the use of LSTMs independently for time series forecasting is a debatable choice, and as seen in this work, it is completely outperformed by every other model and not by a small margin. Out of the six cases, only in one showed competitive results. This finding justifies why, when used in time series, volatility prediction, for example, is often used in a complementary way alongside some other model or in a hybrid model along with another methodology.

While Random Forest and XGBoost are not inherently designed for time series forecasting, with thoughtful feature engineering to capture time-dependent structures and careful consideration of time series characteristics, they can be powerful tools in this domain. In this example, even after an extensive grid search, the hyperparameter tuning was not sufficient for XGBoost to show promise in 2 out of six examples, and all were in the monthly data. Meanwhile, the simpler and more flexible Random Forest performed better, especially in the quarterly timeframes, with four cases of above-average results. In this set of test cases, bagging proved more powerful than boosting.

Some probable explanations are the small size of the train data, the underfitting of XGBoost due to shallow trees selected by the grid search, and the abundance of outliers due to the high volatility of the asset. The Random Forest was the best-performing model outside of the top three.

The true gatekeeper and baseline model that showed very good results was the simple autoregressive moving average generalized autoregressive conditional heteroskedasticity model, which is capable of solving the majority of problems caused by the assumption violation of linear regression. As one of the methodologies designed for time series analysis problems, it has the capacity to capture volatility clustering, leverage effects, and be more transparent in its interpretation. Also, it requires fewer data points to produce reliable forecasts. That gives us a model that behaves very reliably and with only two cases where GARCH has a below-average performance.

That leaves us with the two optimal methodologies that pass our acceptance bar already highly set by the simple, simple autoregressive moving average generalized autoregressive conditional heteroskedasticity model. IGRACH takes the results of GARCH and takes it to the next level. It assumes persistence in shocks to volatility, and also considers that the impact of a shock does not decay over time but has a permanent effect and benefits by the present conditions on the global economy that seem to be a constant stream of different crises (Covid19 lockdowns, supply chain disruptions, Russian-Ukraine war sanctions, pipeline destructions, and trade blockades). Writing this in February, it is unfortunate that our data ended at the start of the previous autumn, so we have no image of the impact of the trade interdiction on the Bab-el-Mandeb straight, one of the most important energy routes of oil and gas from the Arab countries to Europe. While, in theory, IGARCH is prone to shock overestimation, the nature of the problem by itself invalidates that drawback. IGARCH is as effective as the TFT overall, with certain small differences. Said differences become more apparent if we analyze the results by commodity or by timeframe, as it performed better on the monthly data and the Asian Market.

Temporal Fusion Transformers is a brand-new method based on the somewhat new idea of self-attention and is by far the most complex of all models, as described in the literature review. Based on previous implementations, this method was very promising, and indeed, it proved at least as good as the best ARIMA GARCH model, that being IGARCH. It showed better accuracy on the quarterly data from IGARCH, and if not for its abysmal predictions on the monthly data of Asian Natural Gas, the TFT could have been the best performer overall. This outlier of a performance signifies that the main reason was the suboptimal tuning rather than the weakness of the architecture. Taking into consideration that the model was used in its default state, with only minimal tuning, and without extensive random or grid searches, we concluded that the TFT, allowed the necessary resources, is certainly capable of significantly outperforming each of the models used in this thesis. It is a powerful tool for time series modeling and forecasting that every practitioner and researcher should have as a possible solution for more unusual and complex problems where more traditional methods fall short.

10. Conclusions

The purpose of this study was to model and predict energy market series effectively using econometric time series models along with other machine learning methodologies. The comparison of those different models on the predictive ability performance in an out-of-sample horizon is the most valuable product of this work, along with the confirmed impact of different exogenous factors on the price changes of Crude Oil and Natural Gas.

In terms of models, we proved the inability of simple models, such as linear regression, to forecast time series as complex as the returns of an energy commodity. Beyond that, we showed that XGBoost was also ineffective, while the Random Forest performed with higher accuracy comparable to the simple autoregressive moving average generalized autoregressive conditional heteroskedasticity model. While GARCH(1,1) offered acceptable results, not all members of the GARCH family of models, such as EGARCH and GJR-GARCH, were as effective, being from difficult to impossible to produce robust solutions or high accuracy. Where EGARCH and GJR-GARCH failed, IGARCH proved to be the best model to predict energy markets due to its peculiarity, which is that the volatility is considered to be persistent over time. The LSTM was deemed ineffective because of the nature of the problem, having the worst performance. In contrast, the new and highly complex Temporal Fusion Transformer had excellent results for the quarterly predictions, better than IGARCH for quarterly and mixed for monthly data, but still considered the best overall.

The predictive variables had little impact on monthly predictions but were medium to significant for quarterly. Macroeconomic variables such as the Infectious Disease Index, GDP, and lagged observations had a significant impact. However, the most interesting variables were those used to annotate important events. Our experiments make it plain how events completely independent of the assets studied can and do change the market equilibrium—making predictions with great accuracy extremely challenging.

Overall, traditional statistical-econometric methods were very effective when employed properly, but the upcoming TFT, even though severely handicapped, achieved at points better and, in general, very similar accuracy results. There is always room for improvement, and we estimate that deep learning models, and especially self-attention architectures, are not far from providing us with a very strong alternative for modeling and predicting the future.

References

Afees A. Salisu , Ismail O. Fasanya (October 2012) Modelling oil price volatility with structural breaks

Alizadeh, A.H., Nomikos, N.K., Pouliasis, P.K., (2008). A Markov regime switching approach for hedging energy commodities. J. Bank. Finance 32, 1970–1983

Andersen, T. G., Bollerslev, T.(1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review, 39:885–907, 1998

Andersen T., Bollerslev T., Diebold F., Labys P. (2003) “Modeling and forecasting realized volatility”. In: Econometrica 71.2 (2003), pp. 579–625

Arouri, Jouini J., Nguyen D.K (2011) Volatility spillovers between oil prices and stock sector returns: implications for portfolio management J. Int. Money Finance, 30 (7) (2011), pp. 1387-1405

Audrino, F. and Barone-Adesi (2006). Average conditional correlation and tree structures for multivariate GARCH models. Journal of Forecasting 25(8), 579–600

Audrino, F., Buhlmann P. (2001). Tree-structured GARCH models. Journal of the Royal Statistical Society, Series B 63, 727–44

Baillie, R.T., Bollerslev, T., Mikkelsen, H.O., (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. J. Econ. 73, 3–20

Bauwens L., Laurent S. and Rombouts J. V. K. (2006). Multivariate GARCH models: A survey. Journal of Applied Econometrics 21, 79–109

Bekaert, G., Wu, G., (2000). Asymmetric volatility and risk in equity markets. Rev. Finance. Stud. 13, 1–42

Bildirici, M.; Ersin,(2009) O. Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in Istanbul Stock Exchange.

Black, F., Scholes, M.(1973) The pricing of options and corporate liabilities. Journal of political Economy, 81: 637 – 654, 1973

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–27

Bradley T. Ewing A., Farooq M., (March 2017) Modelling asymmetric volatility in oil prices under structural breaks

Breiman L. (2001) Random Forests. Machine Learning 45, 5–32 (2001)

B.T. Ewing, F. Malik (2017) Energy Economics 63 (2017) 227–233

Campbell, J.Y., Hentschel, L., (1992). No news is good news: an asymmetric model of changing volatility in stock returns. J. Finance. Econ. 31, 281–318

Chkili, W., Hammoudeh, S., Nguyen, D.K., (2014). Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. Energy Econ. 41, 1–18

Chen, T., & Guestrin, C.,(2016) XGBoost: A Scalable Tree Boosting System, arXiv:1603.02754 (2016)

Cho K., Merrienboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. (2014) “Learning phrase representations using RNN encoder-decoder for statistical machine translation”. In: arXiv preprint arXiv:1406.1078 (2014)

Christie A., (1982). The stochastic behavior of common stock variances: value, leverage and interest rate effects. J. Finance. Econ. 10, 407–432

Clements, A. and Fuller, J., (2012) Forecasting increases in the VIX: A time-varying long volatility hedge for equities, NCER Working Paper Series 88, National Centre for Econometric Research, 2012

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le V, Q., and Salakhutdinov, R. (2019). Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: 2978–2988

Dellaportas P., Vrontos I. (2007) Modelling volatility asymmetries: a Bayesian analysis of a class of tree structured multivariate GARCH models Econometrics Journal (2007), volume 10, pp. 503–520. doi: 10.1111/j.1368-423X.2007.00219.x

Deng, H., Zhou, Y., Wang, L. et al. Ensemble learning for the early prediction of neonatal jaundice with genetic features. BMC Med Inform Decis Mak 21, 338 (2021)

Dickey, D.A., Fuller, W.A., (1979). Distribution of the estimators for autoregressive time series with a unit root. J. Am. Stat. Assoc. 74 (366a), 427–431

Ding Z., Engle R.F., Granger C.W.J. (1993), A long memory property of stock market return and a new model, Journal of Empirical Finance, Vol. 1(1): 83-10

Elder J. and Serletis A. (2010), Oil price uncertainty J. Money Credit Bank., 42 (6) (2010), pp. 1137-1159

Engle, R.F. (1982) Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation. Econometrica, 50, 1-50

Engle, R.F., Lee, G.G., (1999). A permanent and transitory component model of stock return volatility.

Engle, R.F. (2001) What good is a volatility model?. Quantitative Finance, 1: 237 – 245, 2001

Engle, R.F., Kroner, K.F. (1995), Multivariate Simultaneous Generalized Arch. Econometric Theory Vol. 11, No. 1 (Mar., 1995), pp. 122-150

Engle, R.F., White, H. (Eds.),(2002) Cointegration, Casuality and Forecasting: A Festschrift in Honour of Clive W. J. Granger. Oxford University Press, Oxford, pp. 475–497

Ewing, B.T., Malik, F., (2010). Estimating volatility persistence in oil prices under structural breaks. Financ. Rev. 45, 1011–1023

Elliott, G., Rothenberg, T.J., Stock, J.H., (1996). Efficient tests for an autoregressive unit root. Econometrica 64 (4), 813–836

Farsani, R. M. and Pazouki, E. (2021). A Transformer Self-Attention Model for Time Series Forecasting. Journal of Electrical and Computer Engineering Innovations, 9(1): 1–10

Friedman. J., (2001) "Greedy function approximation: A gradient boosting machine.." Ann. Statist. 29 (5) 1189 - 1232, October 2001

Giamouridis, D., Vrontos, I., (2007). Hedge fund portfolio construction: A comparison of static and dynamic approaches. Journal of Banking and Finance 31, 199–217

Haas, M (2004). A New Approach to Markov-Switching GARCH Models. The Journal of Financial Econometrics, 2004, vol. 2, issue 4, 493-530

Hamilton J., (2020) Time series analysis. Princeton university press, 2020

Harris, R., Sollis, R., (2005). Applied Time Series Modeling and Forecasting, second ed. John Wiley and Sons, London

Hochreiter S., Schmidhuber J. (1997), Long Short-Term Memory. Neural Computation. Vol. 9(8): 1735–1780

Hosker, J., Djurdjevic, S., Nguyen, H. and Slater, R.,(2008) Improving VIX futures forecasts using machine learning methods. SMU Data Sci. Rev., 2018, 1(4), Article 6

Hopfield JJ (1982), Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences, Vol. 79(8): 2554–2558

Hu Y., Ni J., and Wen L., (2020) "A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction". In: Physica A: Statistical Mechanics and its Applications 557 (2020), p. 124907

IEF (2023) Upstream oil and gas investment outlook February 2023

Jo, S. (2014). The effects of oil price uncertainty on global real economic activity. Journal of Money, Credit and Banking, 46(6), 1113–1135

Johannes, F. (2023). Forecasting realized volatility in turbulent times using temporal fusion transformers. FAU Discussion Papers in Economics, No. 03/2023, Friedrich-Alexander-Universität Erlangen-Nürnberg, Institute for Economics, Nürnberg

Jones, D. W., Leiby, P. N., Paik, I. K. (2004). Oil price shocks and the macroeconomy: What has been learned since 1996. Energy Journal, 25, 1–32

Jeong, Y., & Lee, S. (2019). Recurrent neural network-adapted non-linear ARMA-GARCH model with application to S&P 500 index data. Journal of the Korean Data and Information Science Society, 30(5),

Kilian, L. (2008). Exogenous oil supply shocks: how big are they and how much do they matter for the US economy? Review of Economics and Statistics, 90(2), 216–240

Kilian, L., Vigfusson, R. J. (2011). Are the responses of the US economy asymmetric in energy price increases and decreases? Quantitative Economics, 2, 419–453

Kilian, L., Vega, C. (2011). Do energy prices respond to US macroeconomic news? A test of the hypothesis of predetermined energy prices. Review of Economics and Statistics, 93(2), 660-671

Kristjanpoller, W., Minutolo, M. C. (2015). Gold price volatility: A forecasting approach using the Artificial Neural Network–GARCH model. Expert Systems with Applications, 42, 7245-7251

Kristjanpoller, W., Minutolo, M. C. (2016). Forecasting volatility of oil price using an artificial neural network-GARCH model. Expert Systems with Applications, 65, 233-241

Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. Expert Systems with Applications, 103, 25–37

Kwiatkowski, D., Phillips, P. C., Schmidt, P., Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics, 54(1–3), 159–178

Lahmiri S., (2017) Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices, IEEE Syst. J., 11 (2017), 1907–1910

Lamoureux, C.G., Lastrapes, W.D., (1990). Persistence in variance, structural change and the GARCH model. J. Bus. Econ. Stat. 8, 225–234

Lawrence R. Glosten, Ravi J., David E. R., (1993) On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks

Lee, C., Lee, J., (2009). Energy prices, multiple structural breaks, and efficient market hypothesis. Applied Energy 86, 466–479

Lee, Y.-H., Hu, H.-N., Chiou, J.-S., (2010). Jump dynamics with structural breaks for crude oil prices. Energy Economics 32, 343–350

Liu L., Geng Q., Zhang Y., Wang Y. (September 2022) Investors' perspective on forecasting crude oil return volatility: Where do we stand today? Journal of Management Science and Engineering Volume 7, Issue 3, September 2022, Pages 423-438

Liu W.K., & So M.K.P. (2020), A GARCH model with artificial neural networks, Information, Vol. 11(10): 489

Li S. R., Ge Y. L., (2013), Crude oil price prediction based on a dynamic correcting support vector regression machine, Abstr. Appl. Anal., 2013, 666–686

Lim B., Arık S., Loeff N., Pfister T. (2023): Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting

Luo, R. Zhang W, Xu X., Wang J. (2018) "A neural stochastic volatility model". In: proceedings of the AAAI conference on artificial intelligence. Vol. 32. 1. 2018

Lim, B., Arık, S. O., Loeff, N., and Pfister, T. (2021). Temporal Fusion Transformers for ¨ interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4). PII: S0169207021000637: 1748–1764

Leybourne, S., Taylor, R., Kim, T.-H., (2007). CUSUM of squares-based tests for a change in persistence. J. Time Ser. Anal. 28 (3), 408–433

Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. Int. J. Forecast., 36, pp. 54–74

Masih R., Peters S., De Mello L. (2011). Oil price volatility and stock price fluctuations in an emerging market: evidence from South Korea. Energy Econ., 33 (5), pp. 975-986

Meligkotsidou L. and Vrontos I.D. (2008). Detecting Structural Breaks and Identifying Risk factors in Hedge Fund returns: A Bayesian approach. Journal of Banking and Finance, 32, pp. 2471-2481

Michańków, J., Kwiatkowski, L., Morajda, J. (2023). Combining Deep Learning and GARCH Models for Financial Volatility and Risk Forecasting (October 2, 2023)

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. Review of Economics and Statistics, 51, pp. 247 – 257, 1969

Mikosch, T., Starica, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. Rev. Econ. Stat., 86, pp. 378–390

Mertons, R. C. (1964). Lifetime portfolio selection under uncertainty: The continuous-time case. Review of Economics and Statistics, 51, pp. 247 – 257, 1969

Monge, M., Gil-Alana, L. A., Fernando, P. D. G. (2017). Crude oil price behaviour before and after military conflicts and geopolitical events. Energy, 120 (2017), pp. 79–91

Narayan, P., Popp, S., (2010). A new unit root test with two structural breaks in level and slope at unknown time. J. Appl. Stat. 37 (9), 1425–1438.

Narayan, P.K., Narayan, S., (2014). Psychological oil price barrier and firm returns. J. Behav. Financ. 15 (4), 318–333.

Narayan, P.K., Sharma, S., (2011). New evidence on oil price and firm returns. J. Bank. Finance. 35 (12), 3253–3262.

Narayan, P.K., Sharma, S., (2014). Firm return volatility and economic gains: the role of oil prices. Econ. Model. 38, 142–151.

Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 1991, 59, 347–70.

Nomikos, N.K., Pouliasis, P.K., (2011). Forecasting petroleum futures markets volatility: the role of regimes and market conditions. Energy Econ. 33, 321–337.

Oberndorfer, U., (2009). Energy prices, volatility and the stock market: evidence from the Eurozone. Energy Policy 37, 5787–5795

OECD (2022), Competition in Energy Markets, OECD Competition Policy Roundtable

Ozdemir, A.C.; Bulu¸s, K.; Zor, K.(2022) Medium-to long-term nickel price forecasting using LSTM and GRU networks. Resour. Policy 2022, 78, 102906.

Phan D, Sharma, S., Narayan P., (2015). Oil price and stock returns of consumers and producers of crude oil. J. Int. Finance. Mark. Inst. Money 34, 245–262.

Phillips, P.C., Perron, P., (1988). Testing for a unit root in time series regression. Biometrika 75 (2), 335–346.

Ramos-Pérez, E.; Alonso-González, P.J.; Núñez-Velázquez, J.J. (2021) Multi-Transformer: A New Neural Network-Based Architecture for Forecasting S&P Volatility

Sadorsky, P., (2006). Modelling and forecasting petroleum futures volatility. Energy Economics 28, 467–488.

Salisu, A., Fasanya, I., (2013). Modelling oil price volatility with structural breaks. Energy Policy 52, 554–562

Schuster M. and Paliwal K., (1997) “Bidirectional recurrent neural networks”. In: IEEE transactions on Signal Processing 45.11 (1997), pp. 2673–2681.

Shephard N., (2005) Stochastic volatility: selected readings. OUP Oxford, 2005.

Sharma, S., Narayan, P., (2012). Investment and oil price volatility. Econ. Bull. 32, 1428–1433

Swarup, S.; Kushwaha, G.S. (2023) Nickel and Cobalt Price Volatility Forecasting Using a Self-Attention-Based Transformer Model

Taylor S. (1986), Modelling Financial Time Series. Wiley

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention Is All You Need. 31st Conference on Neural Information Processing Systems: 6000–6010

Verma, S.(2016) Forecasting volatility of crude oil futures using a GARCH–RNN hybrid approach. Intell. Syst. Accounting, Financ. Manag. 2021.

Vrontos, S., Galakis, J., Vrontos, I. (2021): Implied volatility directional forecasting: a machine learning approach, Quantitative Finance, DOI: 10.1080/14697688.2021.1905869

Vrontos, S., Vrontos, I., Giamouridis, D., (2008). Hedge fund pricing and model uncertainty. Journal of Banking and Finance 32,

Wang X., Wu C. (2018) Asymmetric volatility spillovers between crude oil and international financial markets. Energy Econ., 74 (2018), pp. 592-604

Wennström A., (2014) Volatility Forecasting Performance: Evaluation of GARCH type volatility models on Nordic equity indices

Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., and Sun, L. (2022). Transformers in Time Series: A Survey. arXiv:2202.07125v3

Williams R.J., Hinton G.E., Rumelhart DE (1986), Learning representations by back-propagating errors, Nature, Vol. 323 (6088): 533–536

Wilson, B., Aggarwal, R., Inclan, C., (1996). Detecting volatility changes across the oil sector. J. Futur. Mark. 16, 313–340

Wu G., Zhang Y. Z. (2014), Does China factor matter? An econometric analysis of international crude oil prices, Energy Policy, 72 (2014), 78–86

Wu, N., Green, B., Ben, X., and O'Banion, S. (2020). Deep Transformer Models for Time Series Forecasting: The Influenza Prevalence Case

Yang, C., Hwang, M., Huang, B., (2002). An analysis of factors affecting price volatility of the US oil market. Energy Economics 24, 107–119

Yu L., Dai W., Tang L., Wu J., (2016) A hybrid grid-GA-based LSSVR learning paradigm for crude oil price forecasting, Neural Comput. Appl., 27 (2016), 2193–2215

Yu L., Zhao Y., Tang L., (2017) Ensemble forecasting for complex time series using sparse representation and neural networks, J. Forecast., 36 (2017)

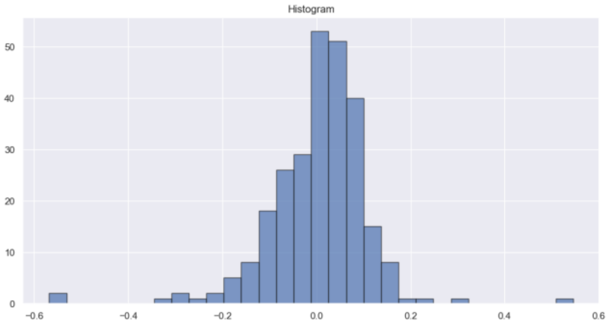
Zakoian J.M. (1994), Threshold heteroscedasticity models. Journal of Economic Dynamics and Control, Vol. 18(5): 931-955

Zhang, L.; Zhu, K.; Ling, S. (2018) The ZD-GARCH model: A new way to study heteroscedasticity. J. Econom. 2018, 202, 1–17

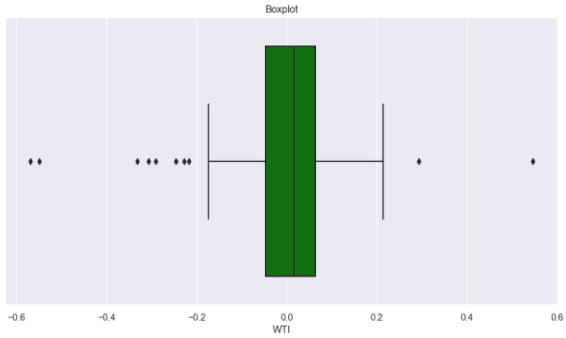
Zhao Y., Li J. P., Yu L.,(2017) A deep learning ensemble approach for crude oil price forecasting, Energy Econ., 66 (2017), 9–16

Appendix I

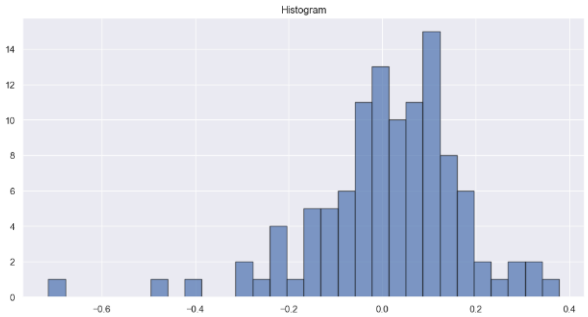
Histogram of Monthly WTI Returns



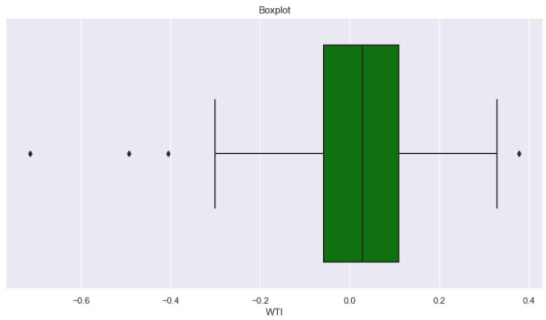
Boxplot of Monthly WTI Returns



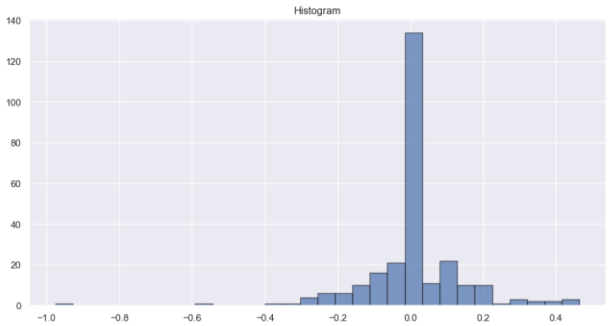
Histogram of Quarterly WTI Returns



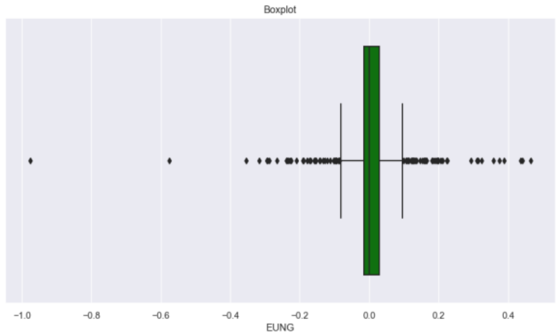
Boxplot of Quarterly WTI Returns



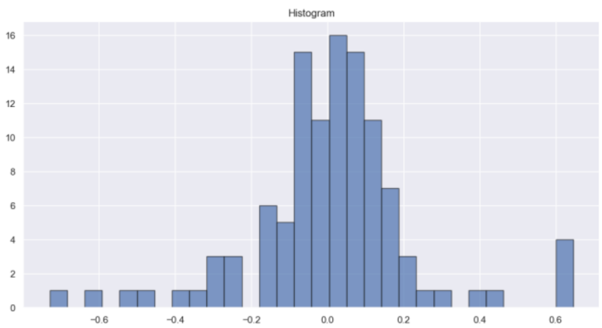
Histogram of Monthly EUNG Returns



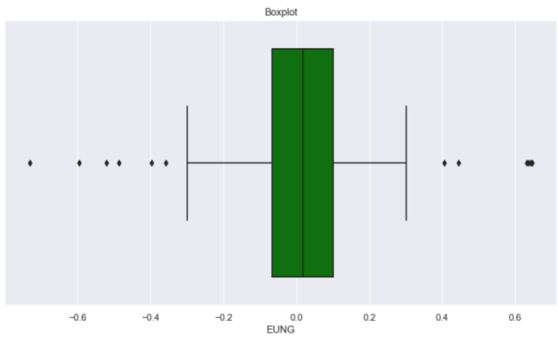
Boxplot of Monthly EUNG Returns

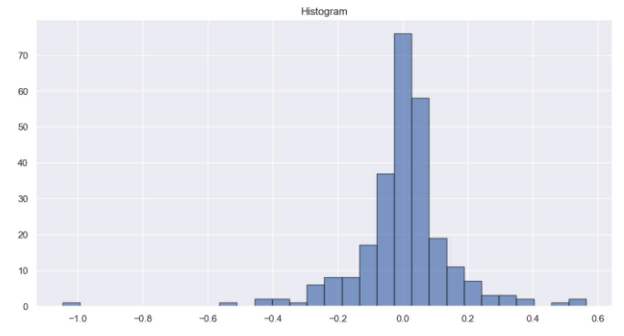
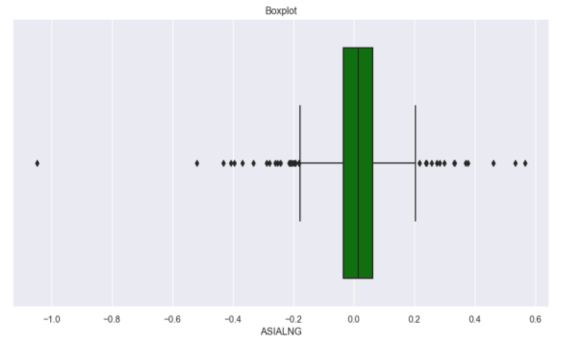
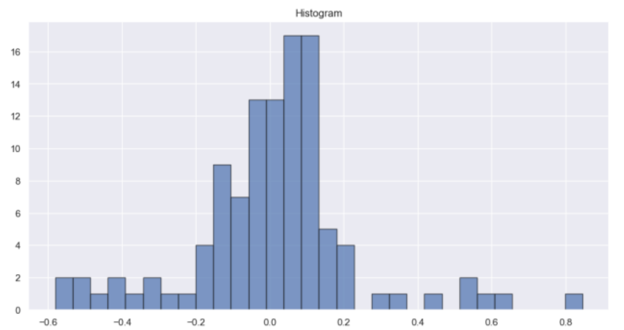
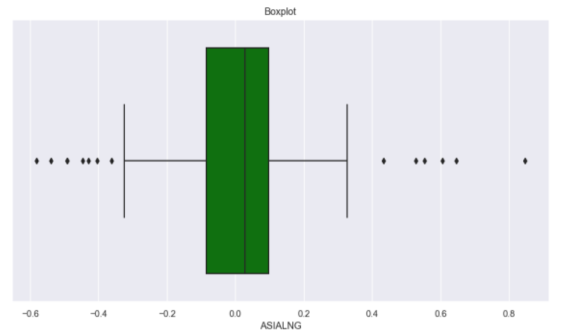


Histogram of Quarterly EUNG Returns



Boxplot of Quarterly EUNG Returns



Histogram of Monthly ASIALNG Returns	Boxplot of Monthly ASIALNG Returns
	
Histogram of Quarterly ASIALNG Returns	Boxplot of Quarterly ASIALNG Returns
	

Appendix II

	Monthly Data								
	WTI			EUNG			ASIANG		
	MAE	**MAPE**	**MSE**	**MAE**	**MAPE**	**MSE**	**MAE**	**MAPE**	**MSE**
EGARCH	0,0758	107,9212	0,0077	0,2664	210,586	0,1337	0,2029	113,0773	0,0542
IGARCH	0,07551	106,7164	0,0075	0,2499	190,6194	0,1206	0,2034	109,2223	0,0543
GARCH	0,0768	112,2574	0,0074	0,2646	216,5708	0,1324	0,2039	114,7778	0,0545
LSTM	0,1053	175,9459	0,0199	0,2689	151,1213	0,1268	0,2784	162,6608	0,1134
Multiple Regression	0,0995	166,1762	0,0146	0,2616	110,2334	0,1266	0,2138	109,7583	0,0699
Random Forest	0,091	141,2878	0,0113	0,2626	116,9904	0,1199	0,2112	136,3037	0,0777
TFT	0,0709	120,1106	0,0076	0,2445	145,3657	0,1147	0,2236	228,4629	0,071
Simple Linear Regression	0,0772	115,0591	0,0076	0,2672	134,554	0,1226	0,2034	109,464	0,0552
XGBoost	0,0897	146,3879	0,0106	0,2599	118,8522	0,1153	0,1976	155,5576	0,0559
	Quarterly Data								
	WTI			**EUNG**			**ASIANG**		
	MAE	**MAPE**	**MSE**	**MAE**	**MAPE**	**MSE**	**MAE**	**MAPE**	**MSE**
IGARCH	0,0983	83,5644	0,0128	0,2962	93,6643	0,2637	0,281	80,1109	0,1348
GARCH	0,0976	84,2451	0,0128	0,3517	249,736	0,2946	0,2898	91,93	0,1413
LSTM	0,0944	84,8424	0,0112	0,6135	599,3878	0,5653	0,4314	127,178	0,3619
Multiple Regression	0,1102	93,977	0,0144	0,4059	382,8954	0,3723	0,3714	118,8345	0,2989
Random Forest	0,0955	80,4109	0,0127	0,3647	96,2809	0,2038	0,3755	97,6876	0,1896
TFT	0,0732	79,9946	0,0102	0,3273	130,9709	0,1807	0,1869	112,7493	0,0493
Simple Linear Regression	0,1051	89,5214	0,0135	0,3553	268,2239	0,2283	0,3857	127,4703	0,1768
XGBoost	0,1036	85,3149	0,0138	0,4502	306,3551	0,272	0,4258	139,8449	0,2188

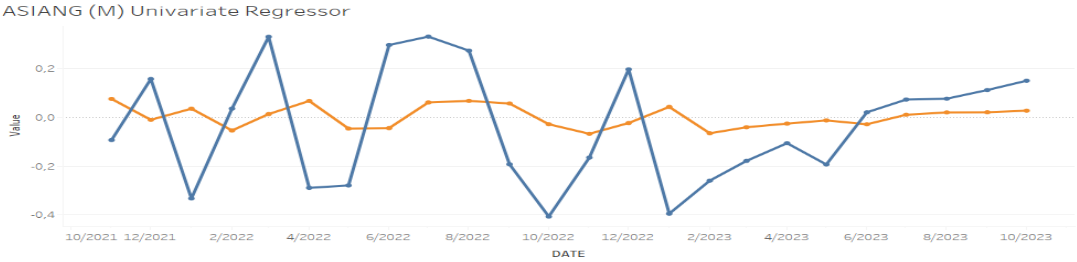
WTI (M) Univariate Regression

WTI (M) Multivariate Regression

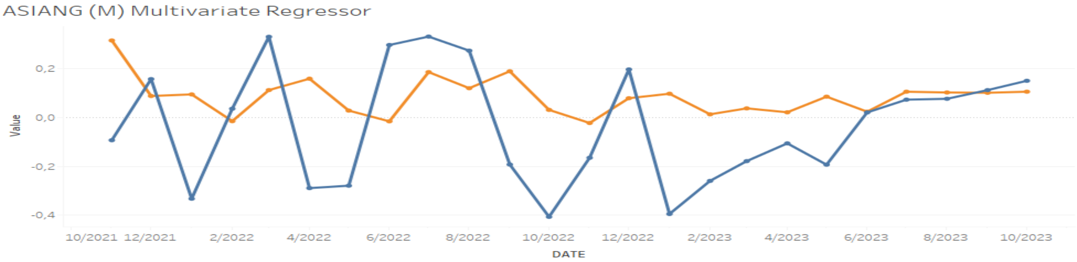
EUNG (M) Univariate Regression

EUNG (M) Multivariate Regression

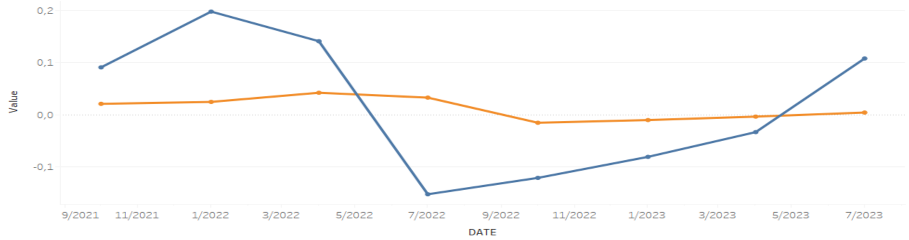
ASIANG (M) Univariate Regressor



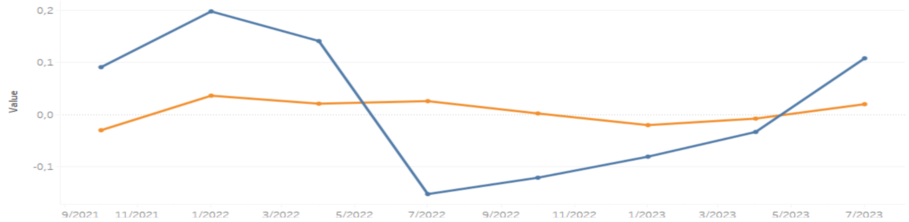
ASIANG (M) Multivariate Regressor



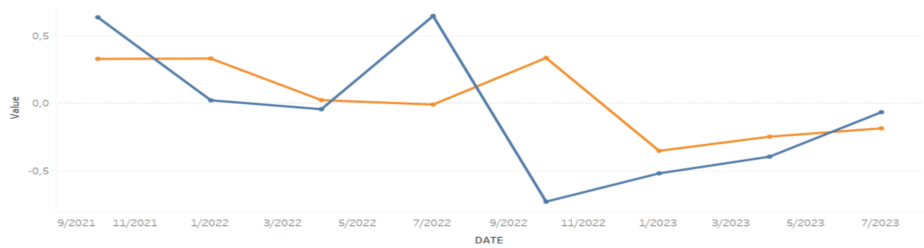
WTI (Q) Univariate Regression



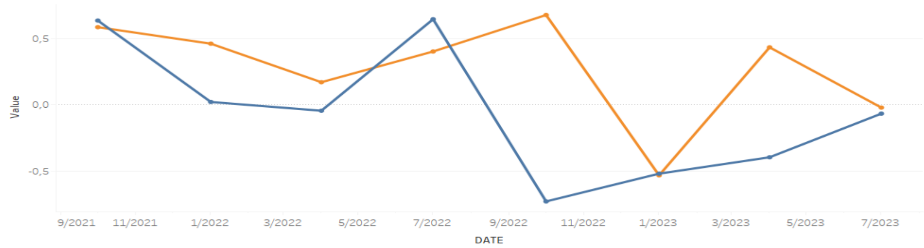
WTI (Q) Multivariate Regression



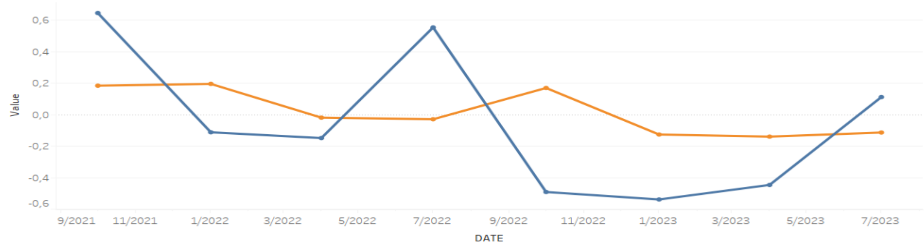
EUNG (Q) Univariate Regression



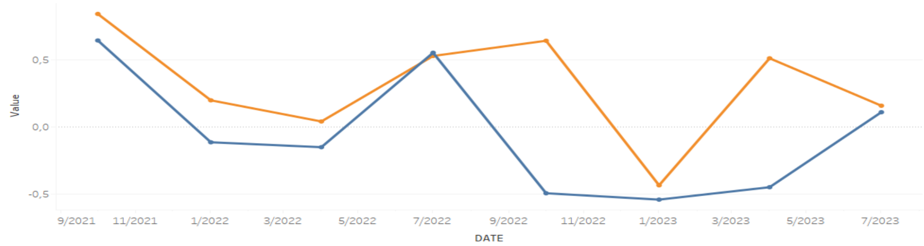
EUNG (Q) Multivariate Regression



ASIANG (Q) Univariate Regression

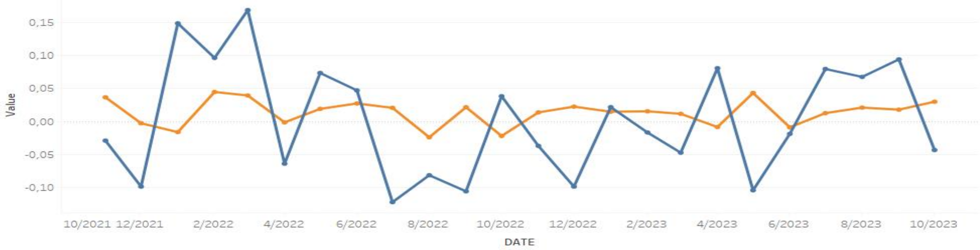


ASIANG (Q) Multivariate Regression

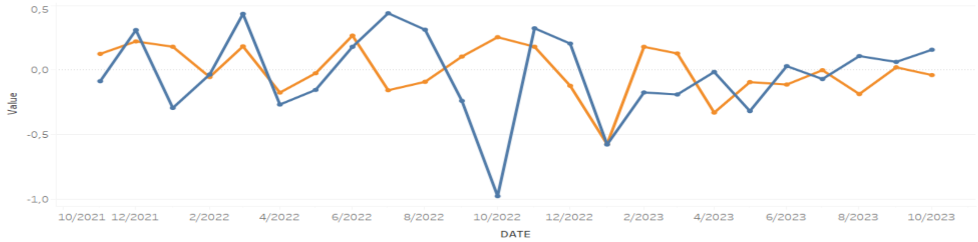


WTI (M) ARIMA GARCH

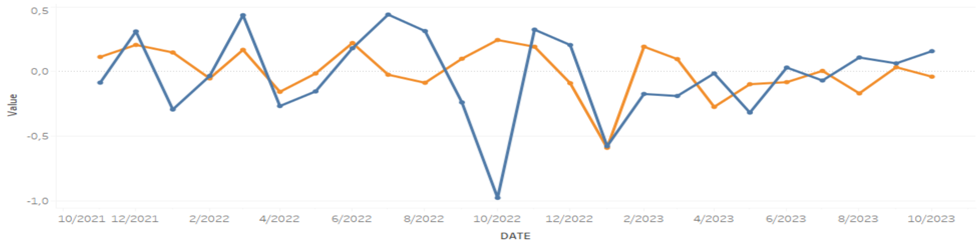
WTI (M) ARIMA IGARCH



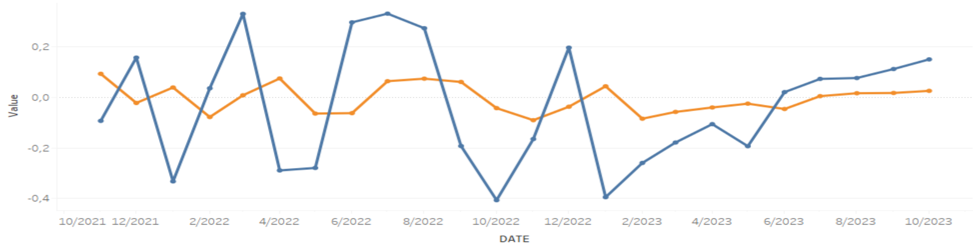
EUNG (M) ARIMA GARCH



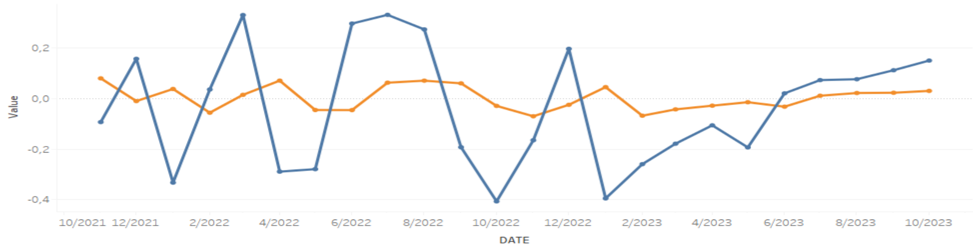
EUNG (M) ARIMA IGARCH



ASIANG (M) ARIMA GARCH



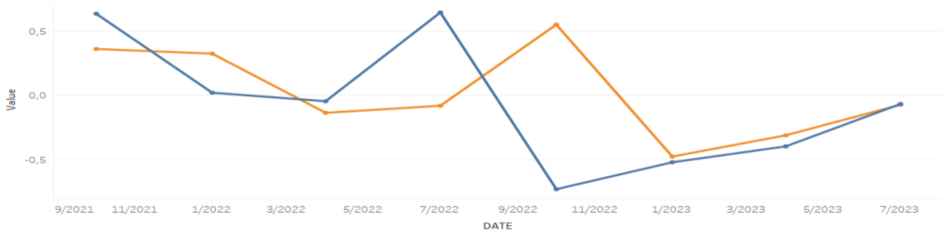
ASIANG (M) ARIMA IGARCH



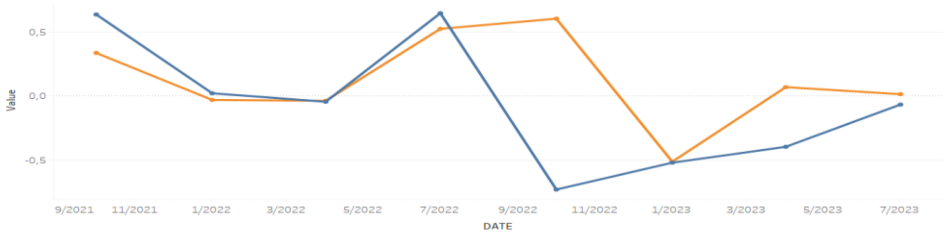
WTI (Q) ARIMA GARCH

WTI (Q) ARIMA GARCH

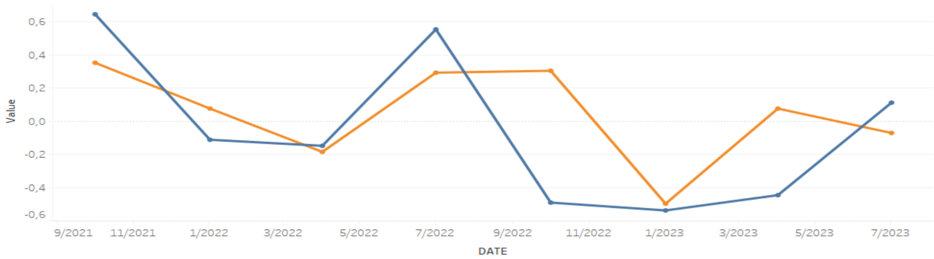
EUNG (Q) ARIMA GARCH



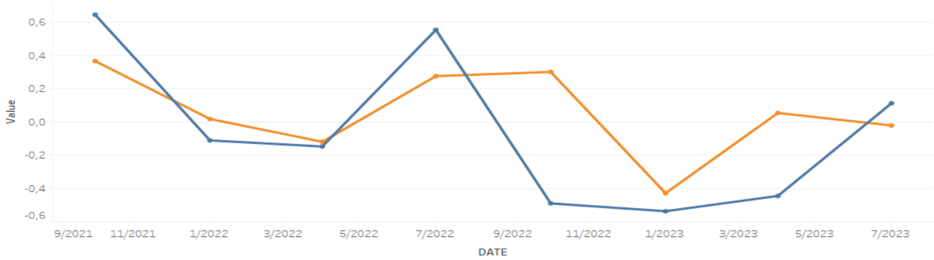
EUNG (Q) ARIMA IGARCH



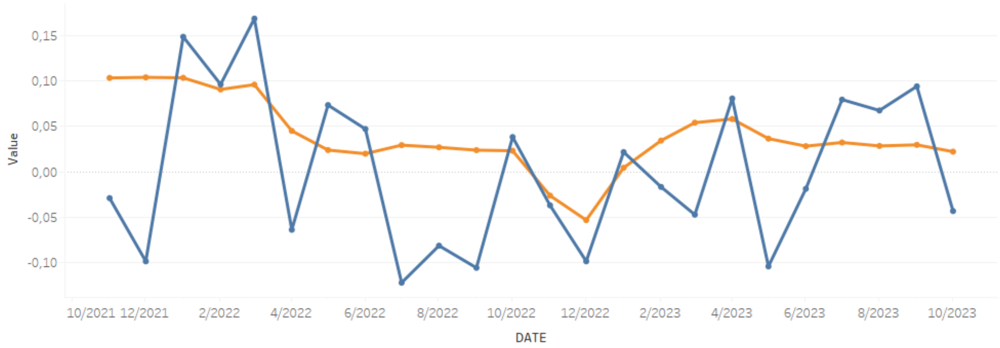
ASIANG (Q) ARIMA GARCH



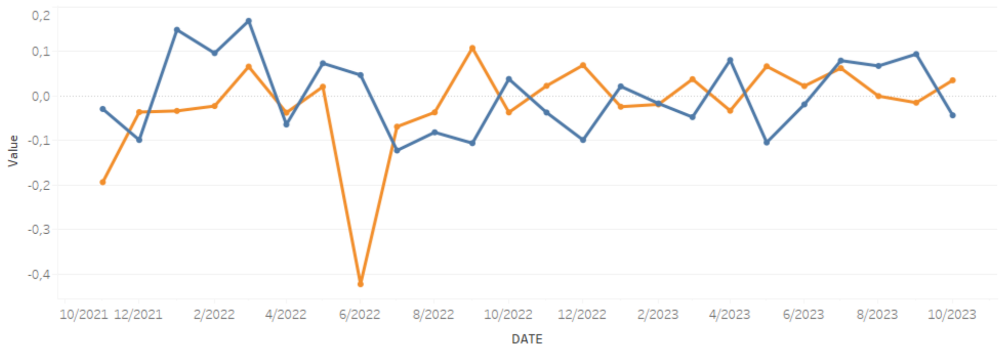
ASIANG (Q) ARIMA IGARCH



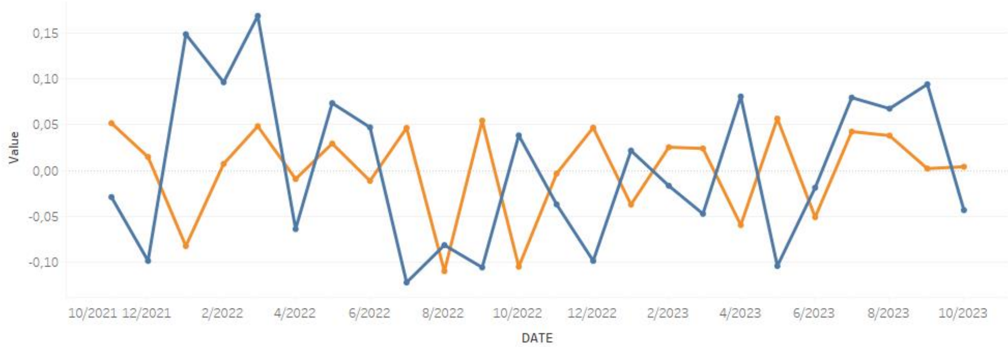
WTI (M) Temporal Fusion Transformers



WTI (M) LSTM

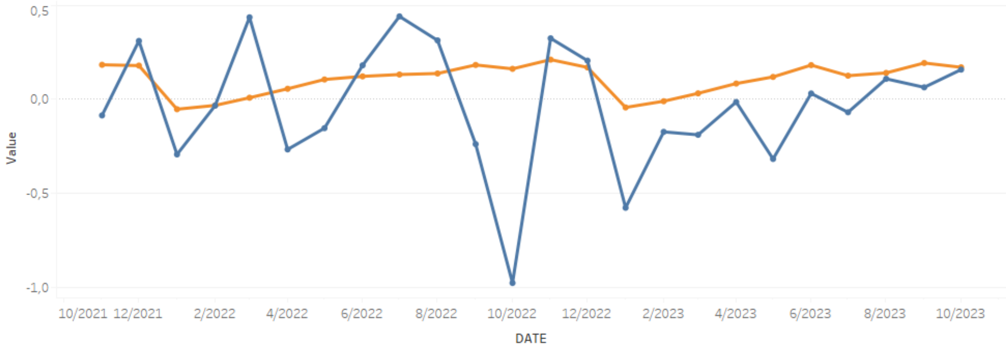


WTI (M) Random Forest

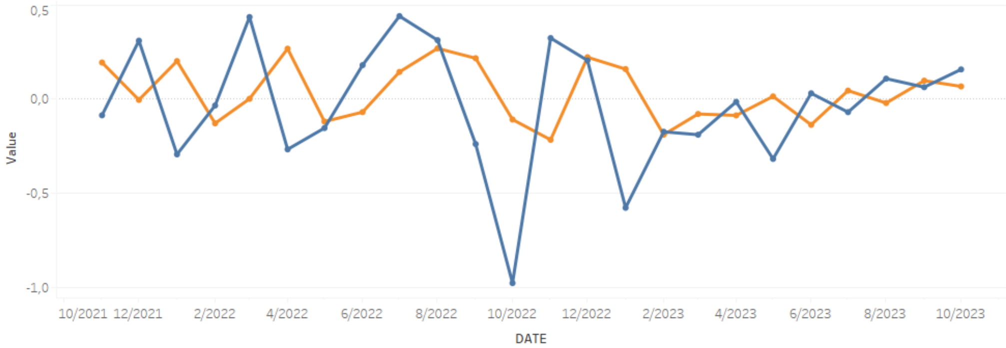


WTI (M) XGBoost

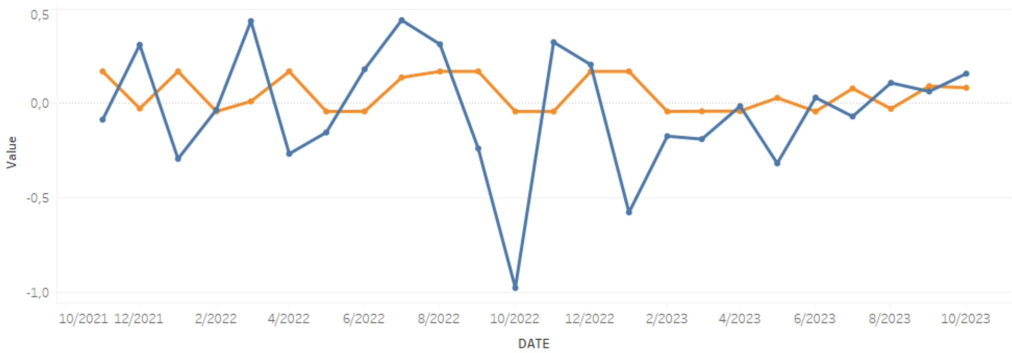
EUNG (M) Temporal Fusion Tranformers



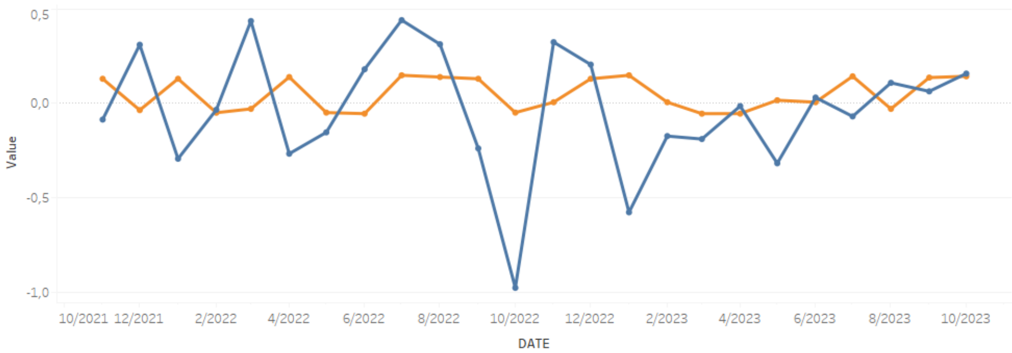
EUNG (M) LSTM

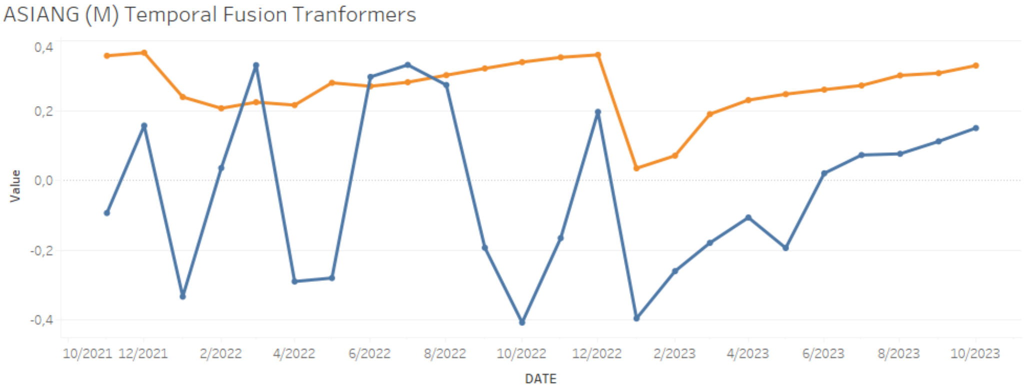


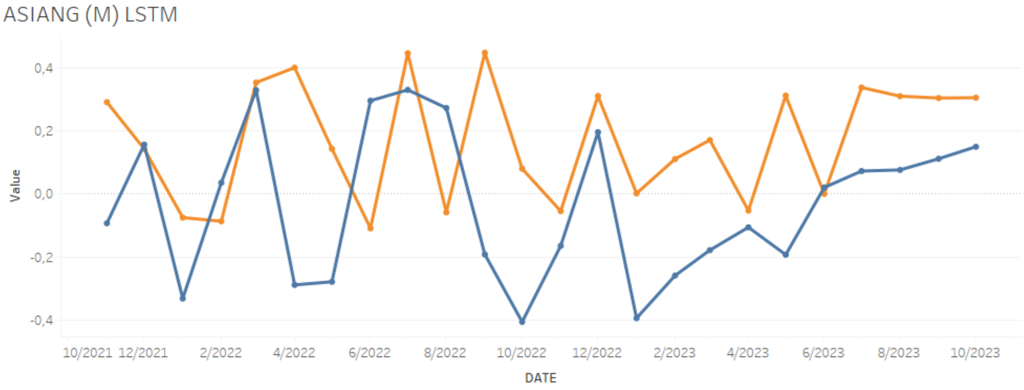
EUNG (M) Random Forest

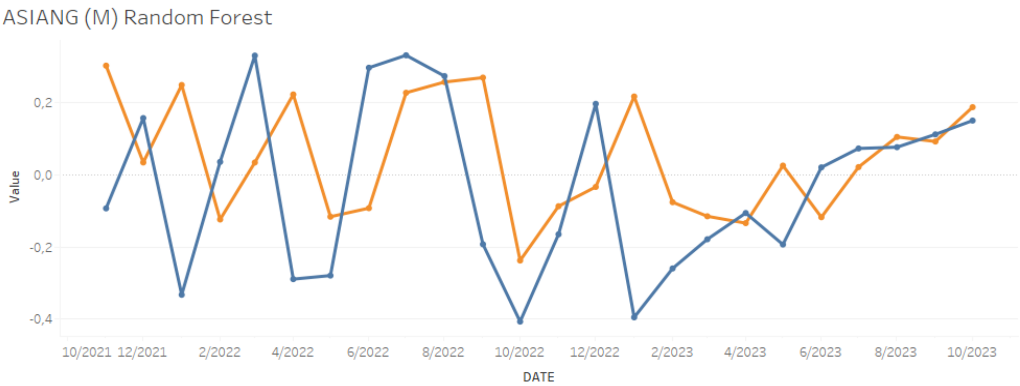


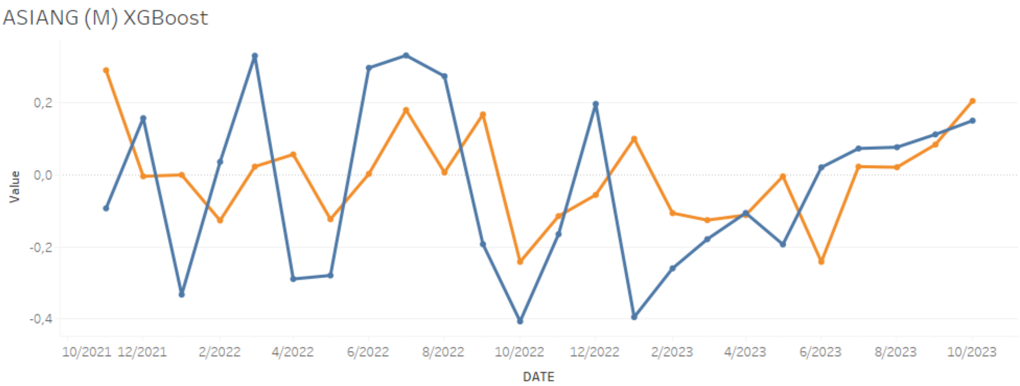
EUNG (M) XGBoost



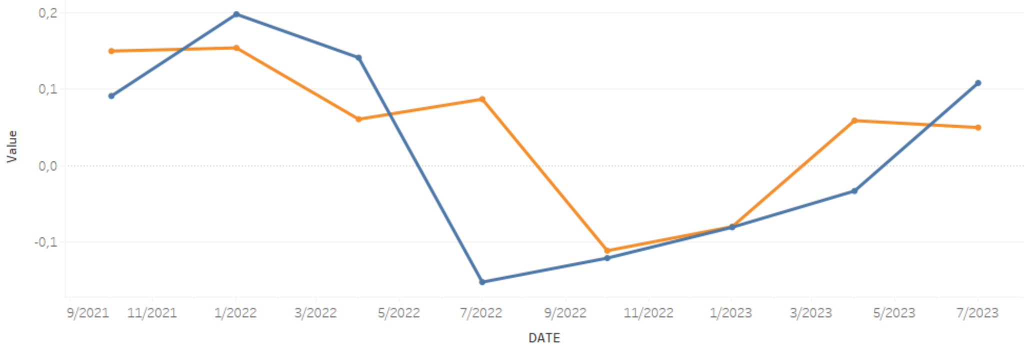

ASIANG (M) Temporal Fusion Tranformers
0,4
0,2
0,0
-0,2
-0,4
Value
10/2021 12/2021 2/2022 4/2022 6/2022 8/2022 10/2022 12/2022 2/2023 4/2023 6/2023 8/2023 10/2023
DATE


ASIANG (M) LSTM
0,4
0,2
0,0
-0,2
-0,4
Value
10/2021 12/2021 2/2022 4/2022 6/2022 8/2022 10/2022 12/2022 2/2023 4/2023 6/2023 8/2023 10/2023
DATE

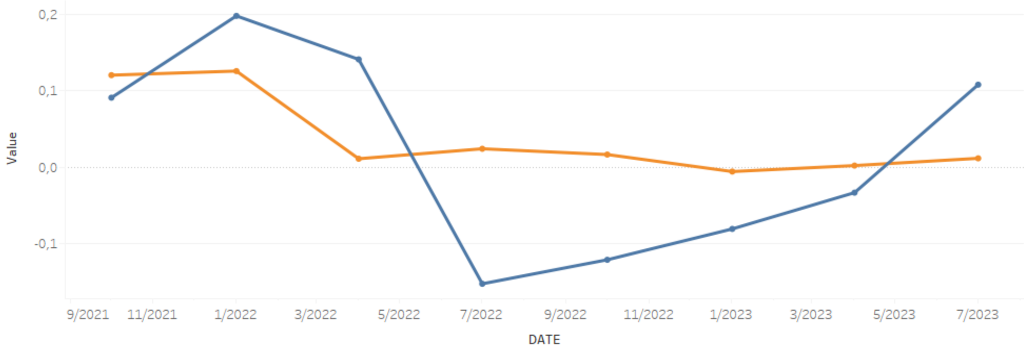

ASIANG (M) Random Forest
0,2
0,0
-0,2
-0,4
Value
10/2021 12/2021 2/2022 4/2022 6/2022 8/2022 10/2022 12/2022 2/2023 4/2023 6/2023 8/2023 10/2023
DATE


ASIANG (M) XGBoost
0,2
0,0
-0,2
-0,4
Value
10/2021 12/2021 2/2022 4/2022 6/2022 8/2022 10/2022 12/2022 2/2023 4/2023 6/2023 8/2023 10/2023
DATE

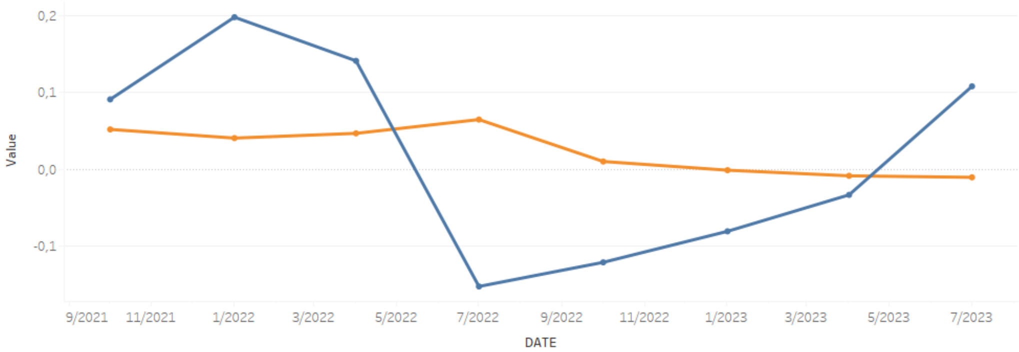
WTI (Q) Temporal Fusion Transformers



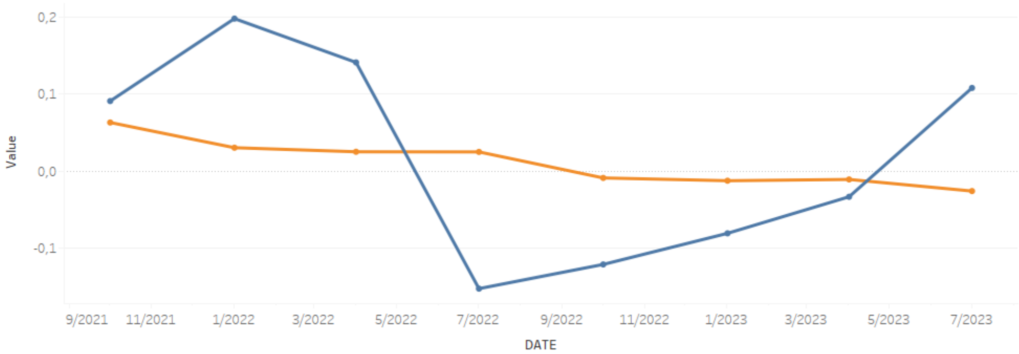
WTI (Q) LSTM



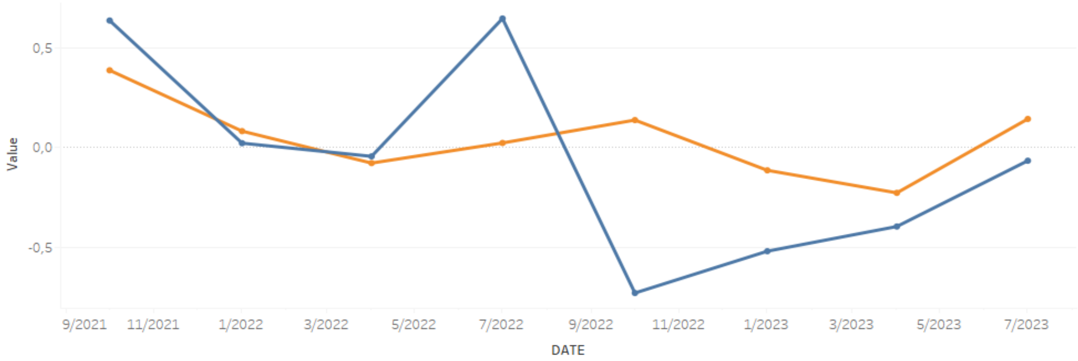
WTI (Q) Random Forest



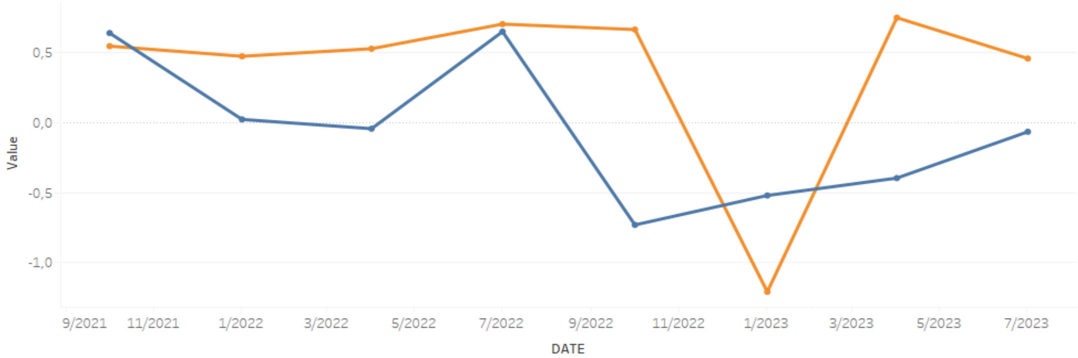
WTI (Q) XGBoost



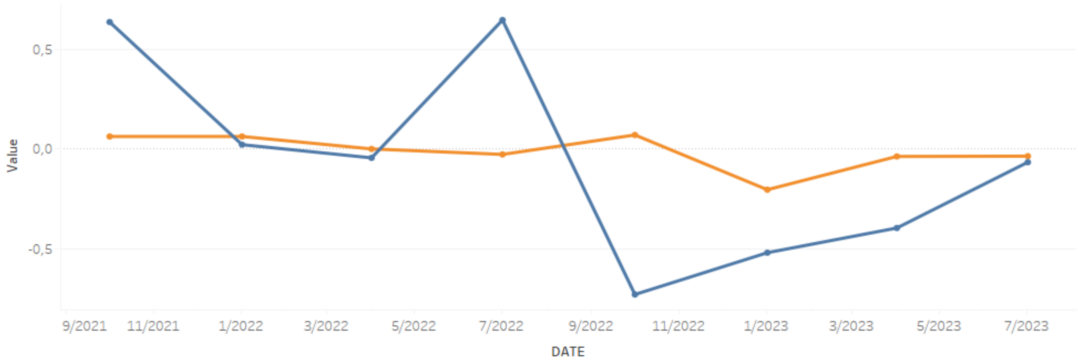
EUNG (Q) Temporal Fusion Transformers



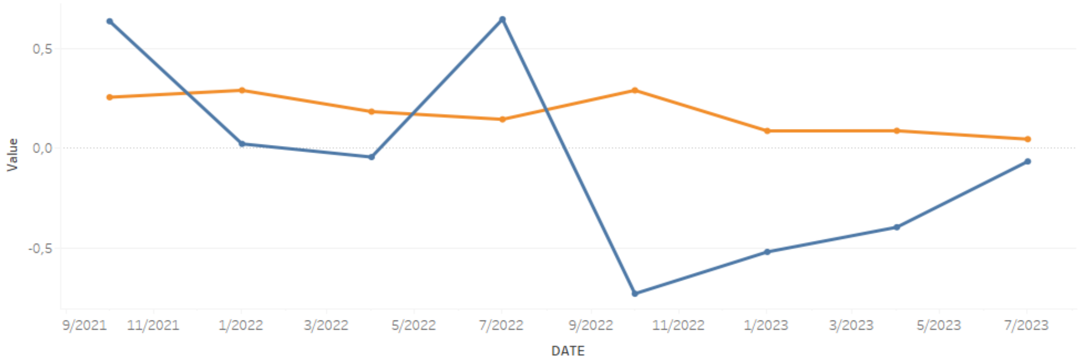
EUNG (Q) LSTM



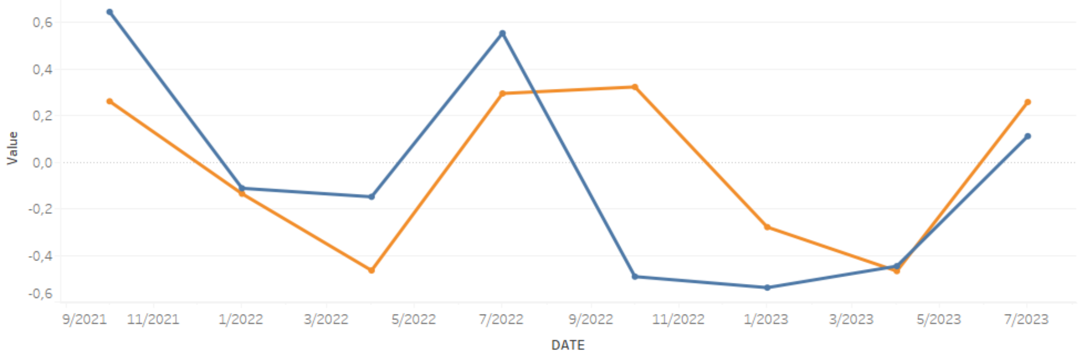
EUNG (Q) Random Forest



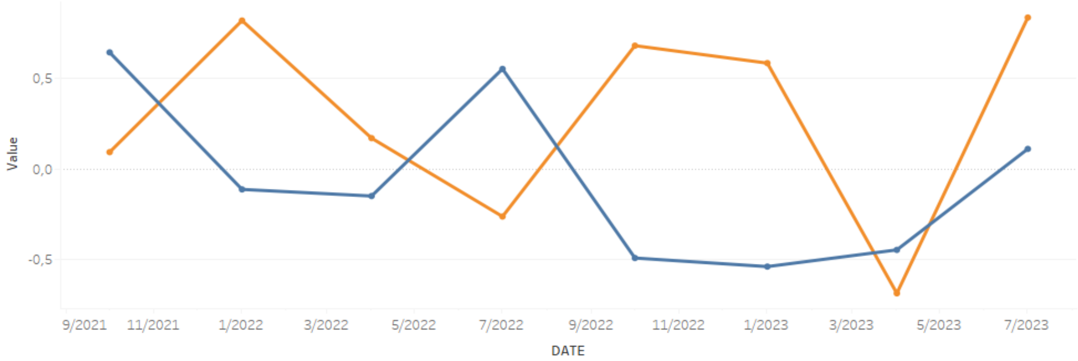
EUNG (Q) XGBoost



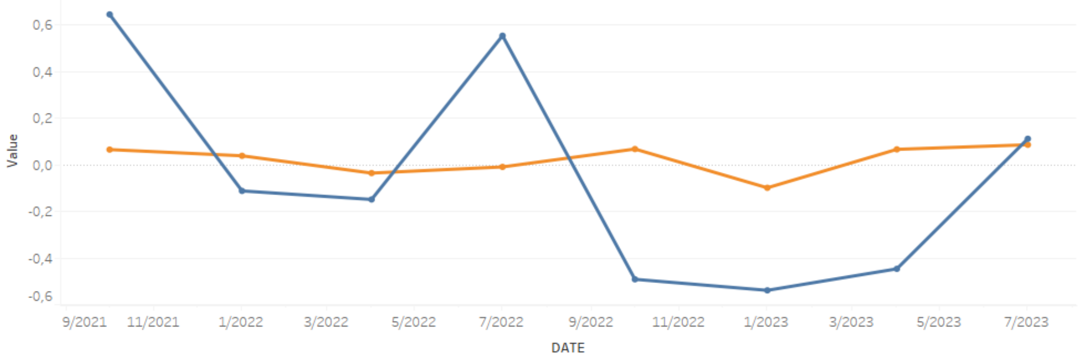
ASIANG (Q) Temporal Fusion Trasforms



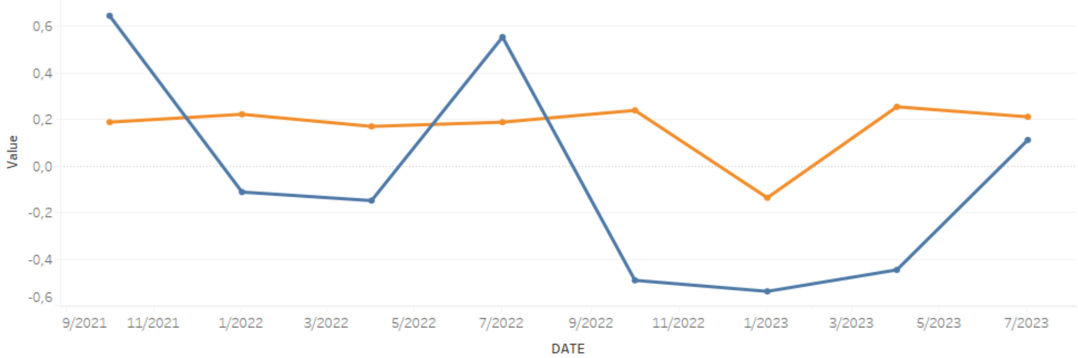
ASIANG (Q) LSTM



ASIANG (Q) Random Forest



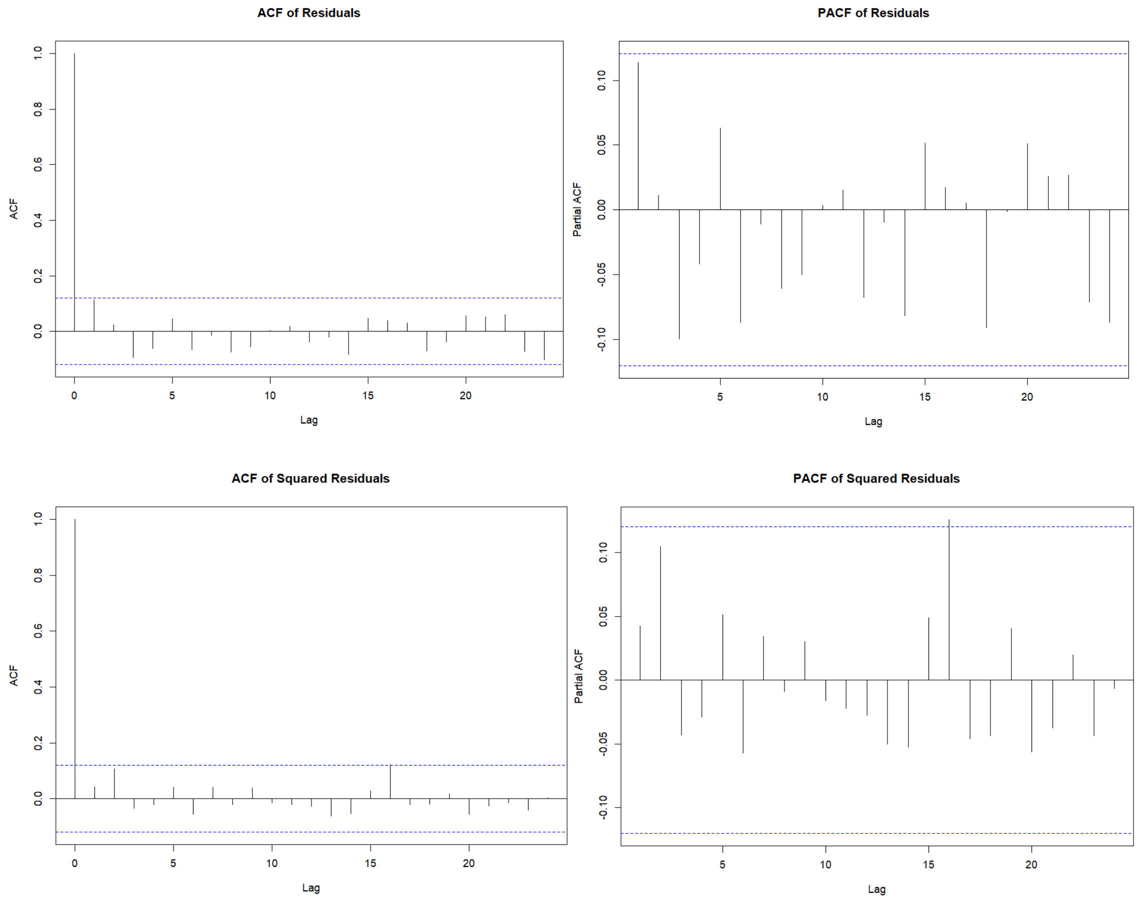
ASIANG (Q) XGBoost



Appendix III

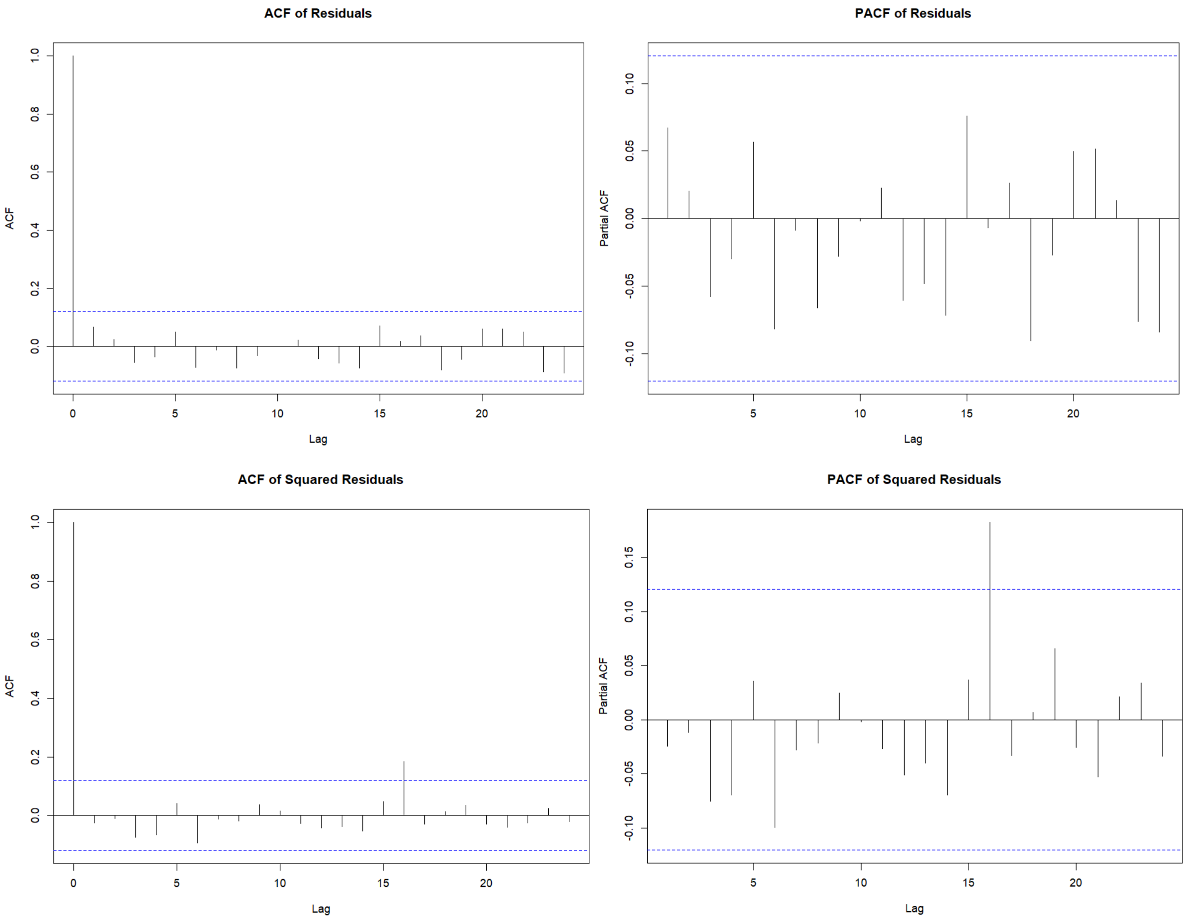
GARCH WTI (M)

```
Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.014029    0.004902   2.8622 0.004208
mxreg1  0.138513    0.061198   2.2633 0.023615
omega   0.005283    0.000675   7.8301 0.000000
alpha1  0.357630    0.136798   2.6143 0.008941
shape   8.997008    3.918326   2.2961 0.021668

LogLikelihood : 274.8747

Information Criteria
------------------------------------

Akaike       -2.0368
Bayes        -1.9692
Shibata      -2.0375
Hannan-Quinn -2.0097
```



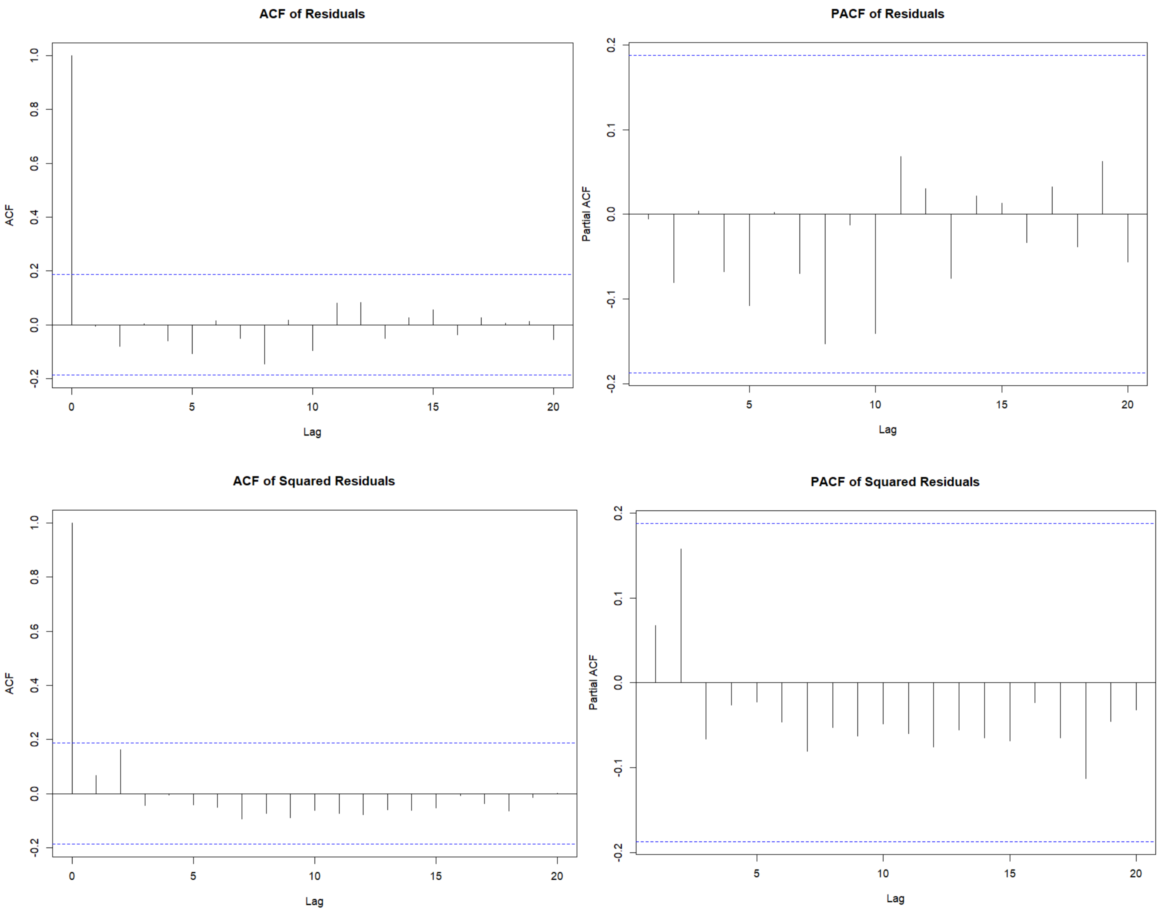
IGARCH WTI (M)

```
Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.012018    0.004953   2.4265 0.015247
mxreg1  0.165213    0.059075   2.7967 0.005163
mxreg2 -0.000240    0.000101  -2.3705 0.017766
omega   0.002125    0.001728   1.2300 0.218705
alpha1  0.664776    0.316071   2.1032 0.035444
beta1   0.335224          NA       NA       NA

LogLikelihood : 267.9415

Information Criteria
------------------------------------

Akaike       -1.9845
Bayes        -1.9169
Shibata      -1.9852
Hannan-Quinn -1.9573
```



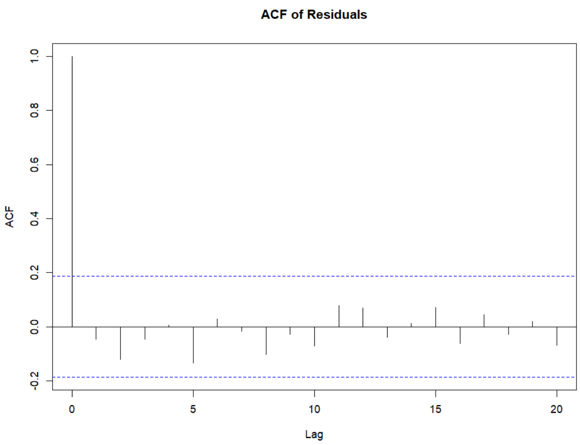
GARCH WTI (Q)

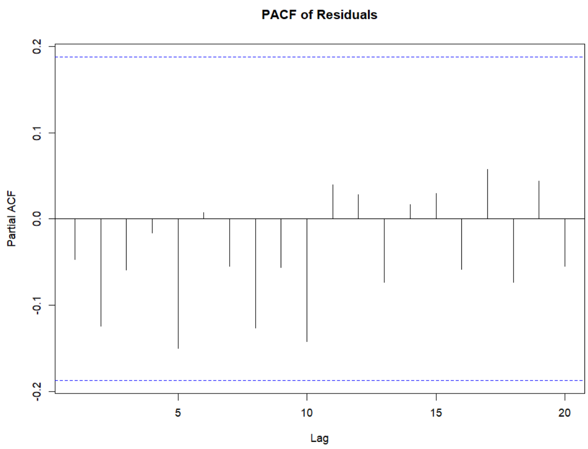
```
Robust Standard Errors:
        Estimate  Std. Error   t value Pr(>|t|)
mu      0.033866    0.022440   1.50917 0.131256
ar1    -0.617614    0.102953  -5.99901 0.000000
ma1     1.291489    0.072794  17.74181 0.000000
ma2     0.532672    0.044569  11.95161 0.000000
mxreg1 -0.327820    0.126549  -2.59046 0.009585
mxreg2 -2.150049    0.729496  -2.94731 0.003206
omega   0.000160    0.000271   0.58975 0.555359
beta1   0.999000    0.008545 116.90435 0.000000
skew    0.712518    0.060597  11.75828 0.000000
shape   3.302350    0.668054   4.94324 0.000001

LogLikelihood : 60.72468

Information Criteria
------------------------------------

Akaike       -0.93073
Bayes        -0.68382
Shibata      -0.94575
Hannan-Quinn -0.83060
```

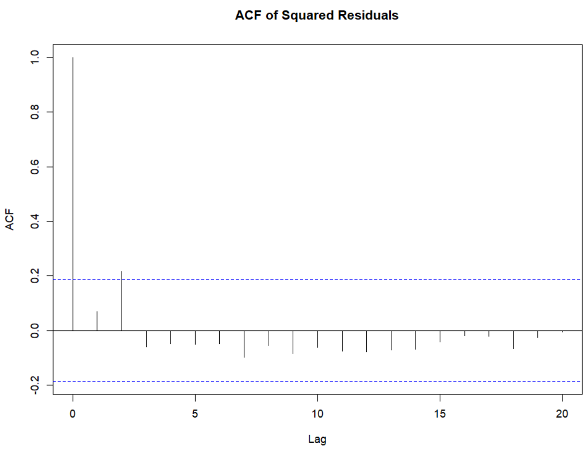


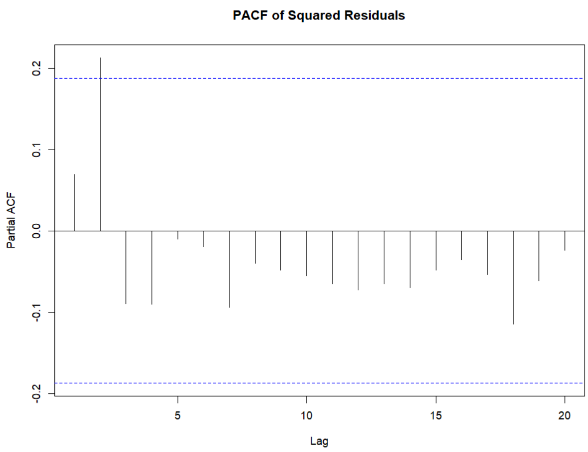
IGARCH WTI (Q)

```
Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.019111    0.012794  1.49369 0.135256
mxreg1  0.281622    0.124503  2.26198 0.023699
omega   0.000253    0.000303  0.83523 0.403586
beta1   1.000000          NA       NA       NA
shape   2.706832    0.532423  5.08399 0.000000

LogLikelihood : 56.2516

Information Criteria
------------------------------------

Akaike       -0.95874
Bayes        -0.85998
Shibata      -0.96131
Hannan-Quinn -0.91869
```



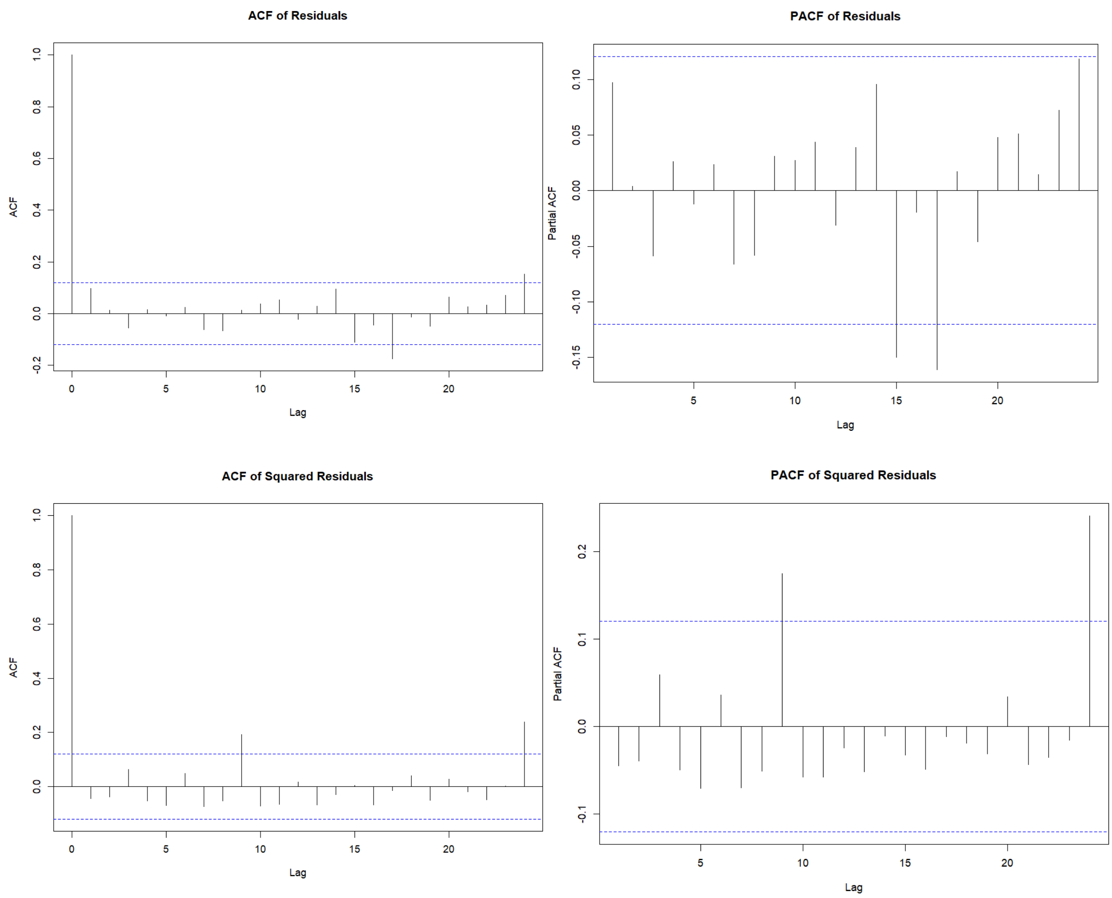






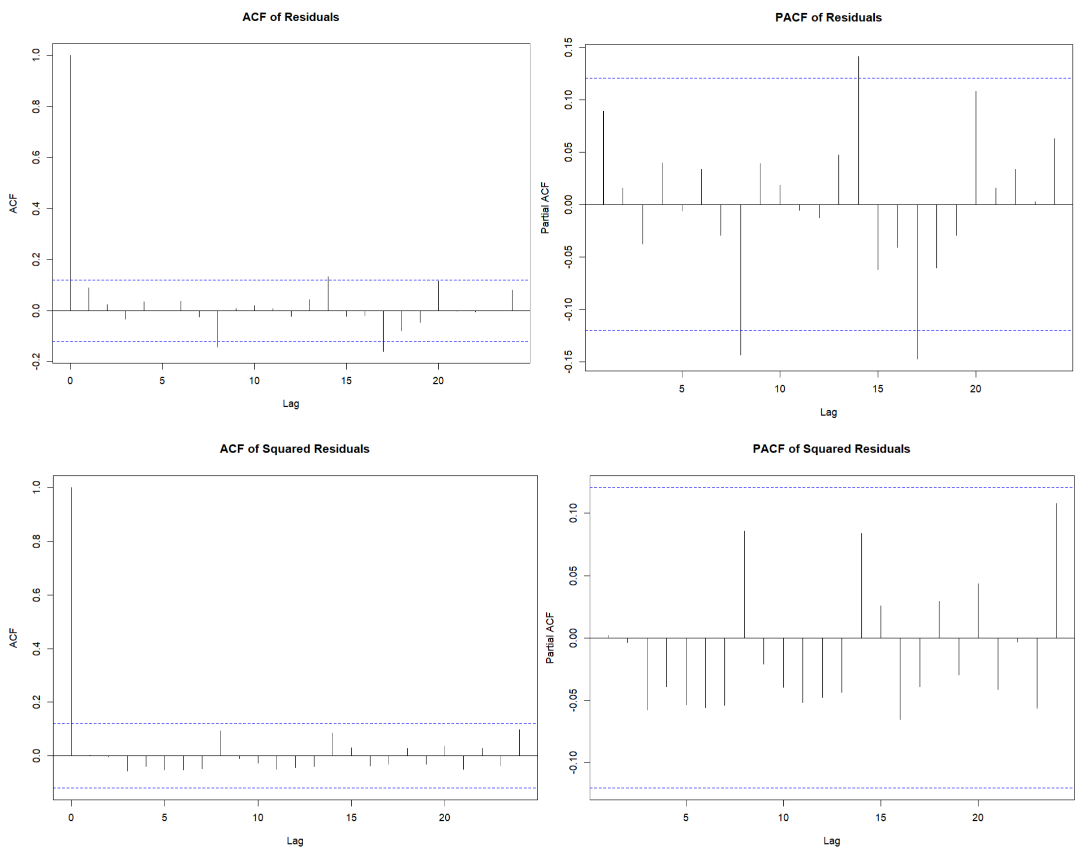
GARCH EUNG (M)

```
Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000485    0.000646  0.75200 0.452051
ar1    -0.803063    0.113982 -7.04550 0.000000
ar2    -0.673975    0.159436 -4.22724 0.000024
mxreg1  0.815178    0.084213  9.67995 0.000000
omega   0.000051    0.000114  0.45043 0.652399
alpha1  0.338731    0.054026  6.26976 0.000000
beta1   0.095286    0.087952  1.08339 0.278635
beta2   0.564983    0.089251  6.33031 0.000000
shape   3.024066    0.382071  7.91494 0.000000

LogLikelihood : 326.1572

Information Criteria
------------------------------------

Akaike       -2.3936
Bayes        -2.2721
Shibata      -2.3958
Hannan-Quinn -2.3448
```



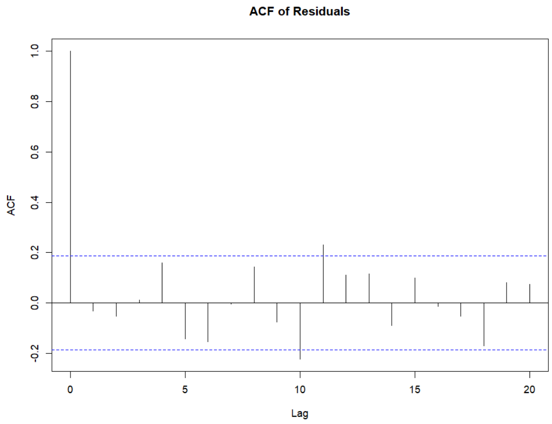
IGARCH EUNG (M)

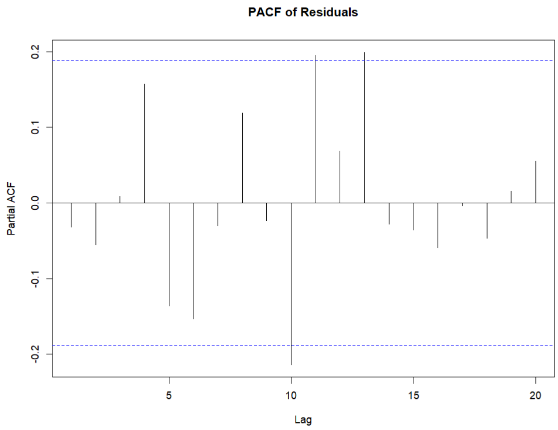
```
Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000289    0.000010  29.5540 0.000000
ar1    -0.012452    0.001746  -7.1335 0.000000
ar2    -0.007941    0.003916  -2.0277 0.042589
ar3     0.478076    0.024470  19.5376 0.000000
mxreg1  0.065803    0.011452   5.7460 0.000000
omega   0.000052    0.000006   8.0012 0.000000
alpha1  0.093594    0.001789  52.3087 0.000000
alpha2  0.009386    0.001085   8.6509 0.000000
alpha3  0.906185    0.135854   6.6703 0.000000
beta1  -0.009165          NA       NA       NA
shape   3.199069    0.646171   4.9508 0.000001

LogLikelihood : 369.2697

Information Criteria
------------------------------------

Akaike       -2.7115
Bayes        -2.5764
Shibata      -2.7142
Hannan-Quinn -2.6572
```

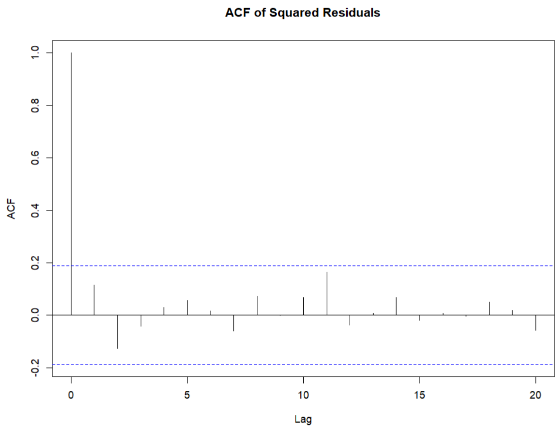


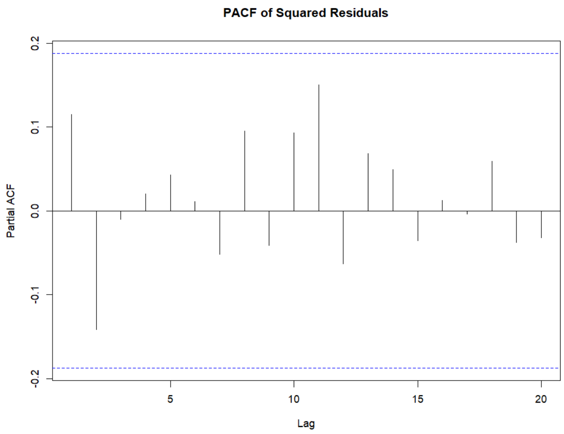
GARCH EUNG (Q)

```
Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.004537    0.007236  0.62705 0.530628
ar1    -0.315865    0.133247 -2.37052 0.017763
mxreg1  0.993242    0.168889  5.88104 0.000000
mxreg2 -0.427169    0.170122 -2.51097 0.012040
mxreg3  0.045497    0.013652  3.33264 0.000860
omega   0.001253    0.000775  1.61553 0.106195
alpha1  0.638190    0.100720  6.33627 0.000000
beta1   0.360810    0.143643  2.51185 0.012010
shape   9.593404    8.125839  1.18060 0.237760

LogLikelihood : 77.91376

Information Criteria
------------------------------------

Akaike       -1.2645
Bayes        -1.0423
Shibata      -1.2768
Hannan-Quinn -1.1744
```



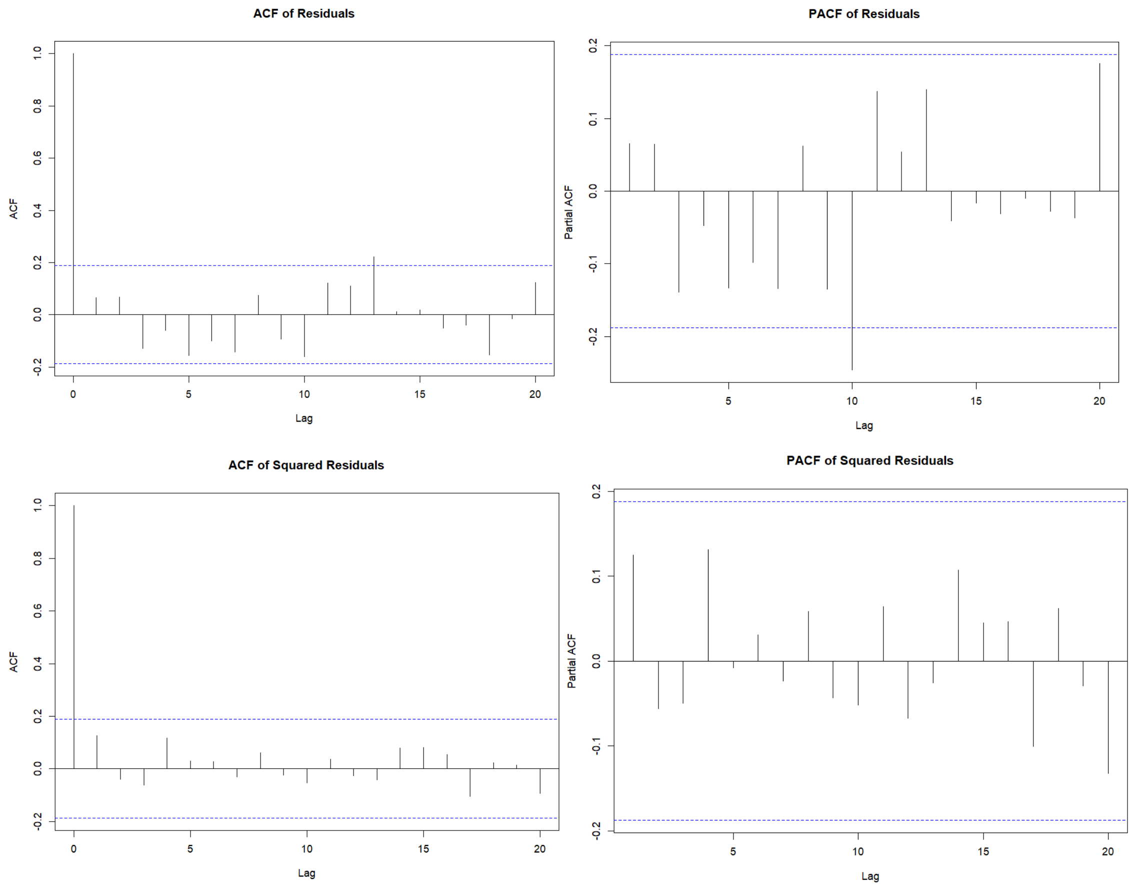






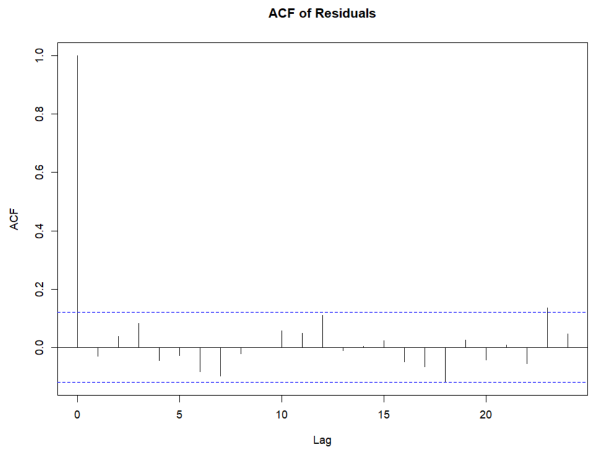
IGARCH EUNG (Q)

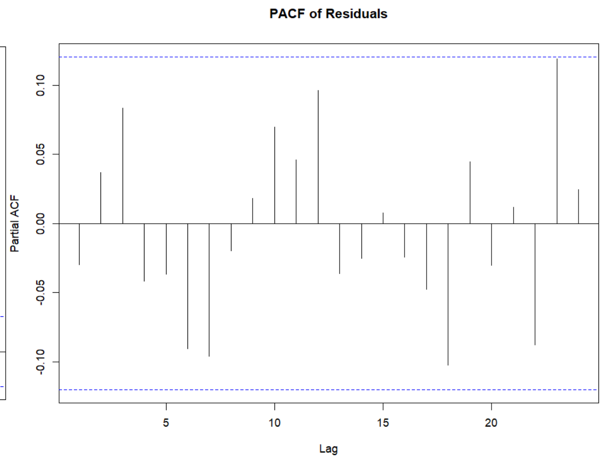
```
Robust Standard Errors:
        Estimate  Std. Error    t value Pr(>|t|)
mu     -0.005079    0.000035 -145.23802 0.000000
ar1    -0.270080    0.002255 -119.76035 0.000000
ar2    -1.165850    0.004721 -246.96938 0.000000
ar3    -0.064657    0.001245  -51.91332 0.000000
ma1     0.192975    0.003208   60.15064 0.000000
ma2     1.210538    0.004824  250.92562 0.000000
mxreg1  0.613641    0.021218   28.92081 0.000000
mxreg2  0.047577    0.003111   15.29440 0.000000
omega   0.001761    0.005938    0.29660 0.766770
alpha1  0.837132    0.211709    3.95417 0.000077
beta1   0.162868          NA         NA       NA
shape   2.008758    5.456869    0.36812 0.712787

LogLikelihood : 87.79648

Information Criteria
------------------------------------

Akaike       -1.4091
Bayes        -1.1375
Shibata      -1.4271
Hannan-Quinn -1.2990
```

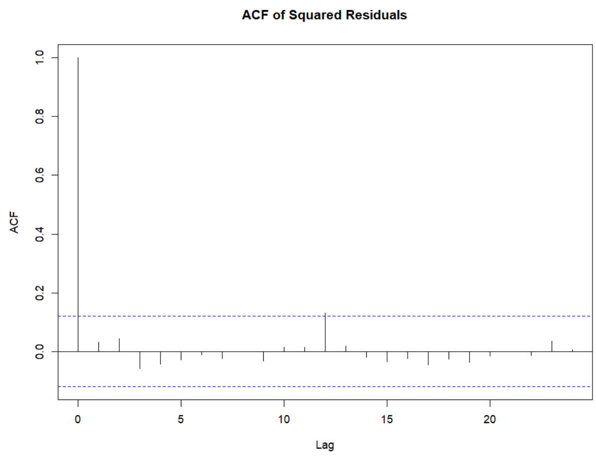


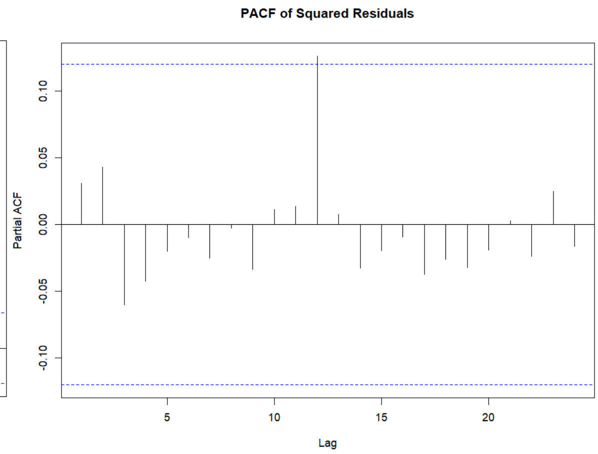
GARCH ASIANG (M)

```
Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu     -0.000041    0.000015  -2.7072 0.006785
mxreg1  0.236550    0.066071   3.5802 0.000343
omega   0.000369    0.000124   2.9827 0.002858
alpha1  0.220705    0.017483  12.6239 0.000000
beta1   0.778289    0.024555  31.6956 0.000000
skew    0.820422    0.046083  17.8031 0.000000
shape   1.092358    0.120162   9.0907 0.000000

LogLikelihood : 249.8756

Information Criteria
------------------------------------

Akaike       -1.8330
Bayes        -1.7385
Shibata      -1.8344
Hannan-Quinn -1.7950
```



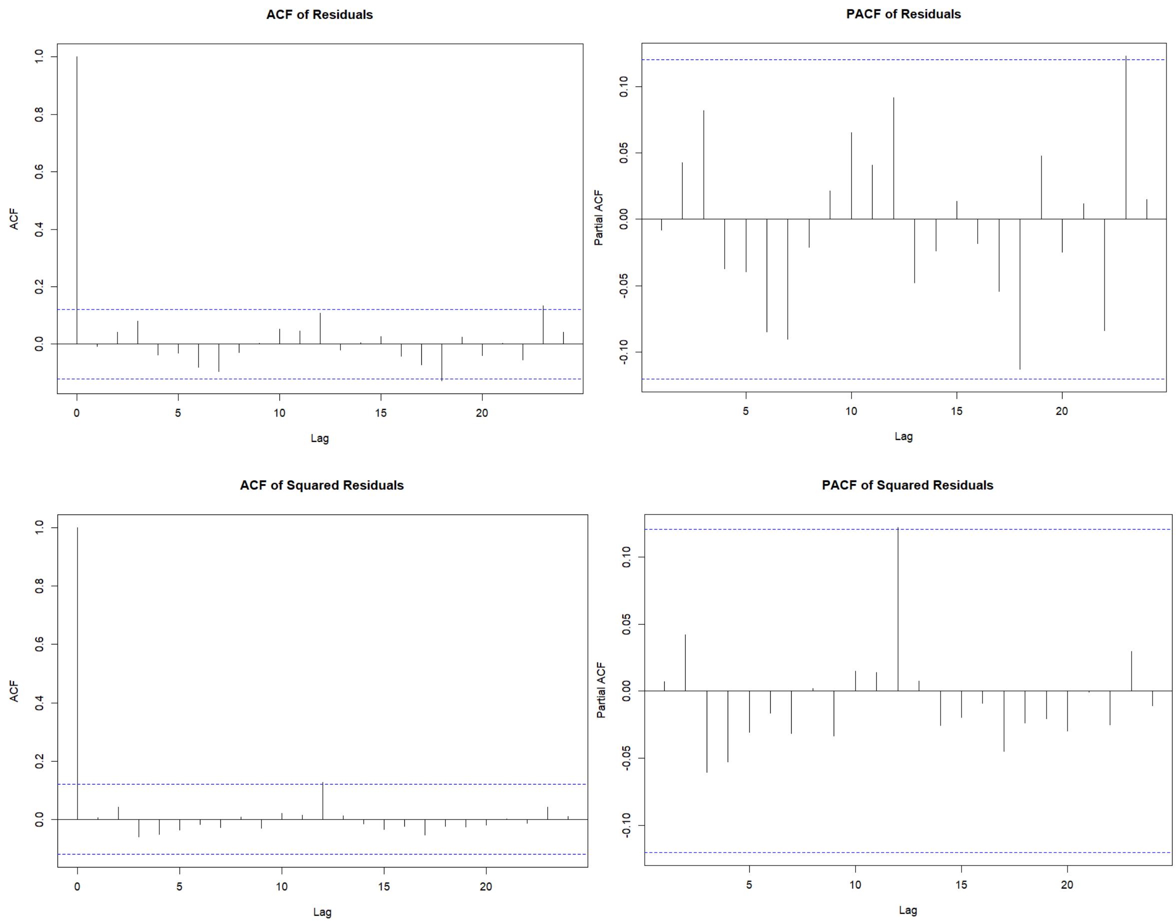






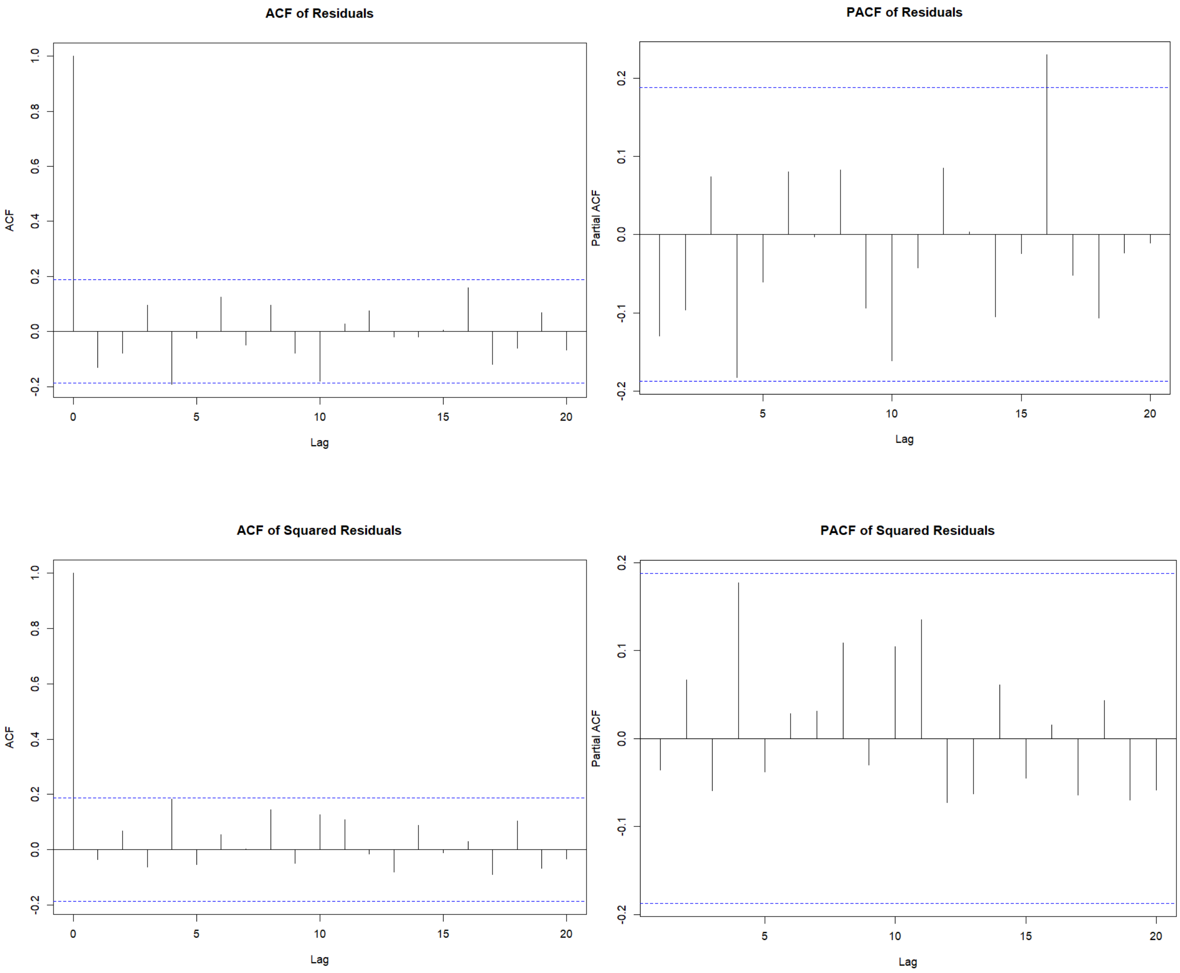
IGARCH ASIANG (M)

```
Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.007563    0.000598  12.6480 0.000000
mxreg1  0.205733    0.015723  13.0846 0.000000
omega   0.000519    0.000326   1.5904 0.111739
alpha1  0.309707    0.100128   3.0931 0.001981
beta1   0.690293          NA       NA       NA
shape   1.138153    0.142556   7.9839 0.000000

LogLikelihood : 245.2403

Information Criteria
------------------------------------

Akaike       -1.8131
Bayes        -1.7456
Shibata      -1.8138
Hannan-Quinn -1.7860
```



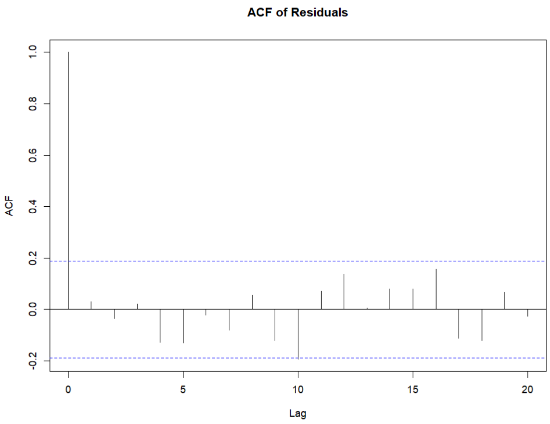
GARCH ASIANG (Q)

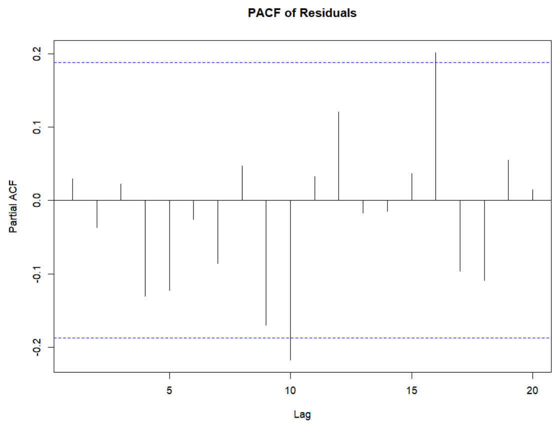
```
Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.002892    0.007168  0.40347 0.686603
ar1    -0.155820    0.097367 -1.60033 0.109525
ar2    -0.761114    0.111683 -6.81496 0.000000
mxreg1  0.710818    0.075012  9.47600 0.000000
mxreg2  0.386085    0.156397  2.46862 0.013564
mxreg3 -0.443265    0.150725 -2.94088 0.003273
omega   0.011937    0.005409  2.20690 0.027321
alpha1  0.937145    0.368364  2.54408 0.010957
shape   4.523802    4.270013  1.05943 0.289402

LogLikelihood : 51.89039

Information Criteria
------------------------------------

Akaike       -0.78698
Bayes        -0.56476
Shibata      -0.79928
Hannan-Quinn -0.69686
```

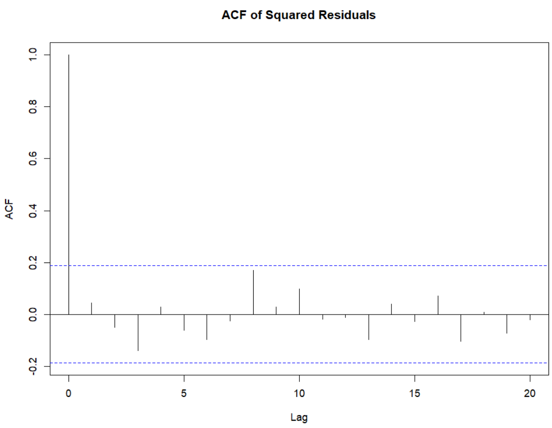


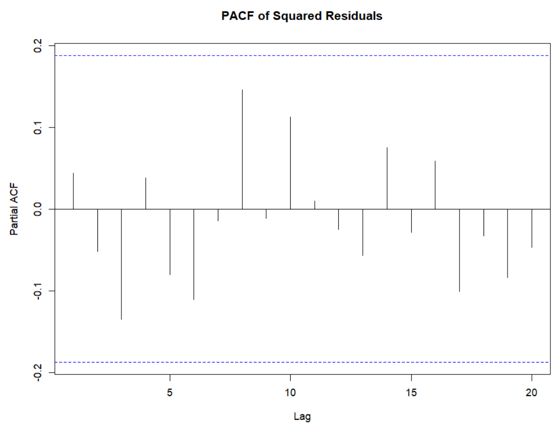
IGARCH ASIANG (Q)

```
Robust Standard Errors:
        Estimate  Std. Error   t value Pr(>|t|)
mu     -0.012675    0.011051  -1.14702 0.251375
ar1    -0.053896    0.071856  -0.75006 0.453216
ar2     0.359268    0.054169   6.63232 0.000000
mxreg1  0.422030    0.089164   4.73319 0.000002
mxreg2 -0.581113    0.038603 -15.05360 0.000000
mxreg3  0.405923    0.070096   5.79099 0.000000
omega   0.002959    0.003018   0.98038 0.326897
alpha1  0.446644    0.202323   2.20758 0.027274
beta1   0.553356          NA        NA       NA
skew    0.757567    0.043118  17.56966 0.000000
shape   1.185365    0.223865   5.29499 0.000000

LogLikelihood : 56.38202

Information Criteria
------------------------------------

Akaike       -0.85105
Bayes        -0.60413
Shibata      -0.86607
Hannan-Quinn -0.75091
```









Appendix IV

RANDOM FOREST HYPERPARAMETERS

	Monthly Data			Quarterly Data		
	WTI	EUNG	ASIALNG	WTI	EUNG	ASIALNG
NUMBER OF ESTIMATORS	200	400	200	300	500	200
THE MINIMUM NUMBER OF SAMPLES REQUIRED TO SPLIT AN INTERNAL NODE	2	2	2	12	12	2
THE MINIMUM NUMBER OF SAMPLES REQUIRED TO BE AT A LEAF NODE	6	6	1	4	6	6
THE MAXIMUM DEPTH OF THE TREE	10	30	10	10	10	10
THE CALCULATION METHOD OF THE NUMBER OF FEATURES TO CONSIDER WHEN LOOKING FOR THE BEST SPLIT	sqrt	sqrt	sqrt	sqrt	auto	sqrt
USE OF BOOTSTRAP	YES	YES	YES	YES	YES	YES

XG BOOST HYPERPARAMETERS

	Monthly Data			Quarterly Data		
	WTI	EUNG	ASIALNG	WTI	EUNG	ASIALNG
NUMBER OF ESTIMATORS	100	500	100	100	250	100
SUBSAMPLE RATIO OF COLUMNS WHEN CONSTRUCTING EACH TREE	0.7	0.7	0.9	0.7	0.7	0.9
LEARNING RATE	0.1	0.3	0.1	0.3	0.2	0.3
THE MAXIMUM DEPTH OF THE TREE	3	3	3	3	2	3
L1 REGULARIZATION TERM	1	1	0.1	1	1	1
L2 REGULARIZATION TERM	1.5	1.5	2	1.5	1	1
FRACTION OF OBSERVATIONS TO BE RANDOM SAMPLES FOR EACH TREE	0.7	0.9	1	0.7	0.5	0.9

LSTM HYPERPARAMETERS & ARCHITECTURE

	Monthly Data			Quarterly Data		
	WTI	EUNG	ASIALNG	WTI	EUNG	ASIALNG
NUMBER OF LAYERS	5	6	5	6	6	6
NUMBER OF NEURONS FIRST LAYER	3	3	4	5	5	5
NUMBER OF NEURONS SECOND LAYER	30	90	40	40	50	50
NUMBER OF NEURONS THIRD LAYER	270	270	200	400	200	200
NUMBER OF NEURONS FORTH LAYER	27	270	40	100	100	100
NUMBER OF NEURONS FIFTH LAYER	1	20	1	40	25	25
NUMBER OF NEURONS SIXTH LAYER	-	1	-	1	1	1
NUMBER OF DROPOUT LAYERS	3	4	3	3	4	4
FIRST DROPOUT LAYER DROPOUT RATE	0.3	0.3	0.3	0.4	0.1	0.1
SECOND DROPOUT LAYER DROPOUT RATE	0.5	0.6	0.5	0.2	0.25	0.25
THIRD DROPOUT LAYER DROPOUT RATE	0.1	0.4	0.1	0.1	0.25	0.2
FOURTH DROPOUT LAYER DROPOUT RATE	-	0.1	-	-	0.1	0.1
NUMBER OF EPOCHS	150	300	200	150	350	200
BATCH SIZE	12	6	12	48	24	

TFT HYPERPARAMETERS

	Monthly Data			Quarterly Data		
	WTI	EUNG	ASIALNG	WTI	EUNG	ASIALNG
NUMBER OF TIME STEPS IN THE PAST TO TAKE AS A MODEL INPUT	24	24	24	24	16	24
NUMBER OF TIME STEPS PREDICTED AT ONCE	1	1	1	1	1	1
HIDDEN STATE SIZE OF THE TFT	128	256	512	128	256	128
NUMBER OF LAYERS FOR LSTM ENCODER AND DECODER	3	1	1	2	2	4
NUMBER OF ATTENTION HEADS	8	8	4	8	14	10
DROPOUT RATE	0.15	0.2	0.25	0.15	0.2	0.2
NUMBER OF TIME SERIES USED IN EACH TRAINING PASS	4	24	6	4	4	4
NUMBER OF EPOCHS	60	60	40	90	25	75