Contributions to the Analysis and Forecasting of Oil Prices

by

Pedro Moreno Alonso

supervised by

Prof. Dr. Antonio Muñoz San Roque

ICAI School of Engineering

Comillas Pontifical University

Madrid, 2024

CONSTANCIA REGISTRAL DEL TRIBUNAL DEL ACTO DE LA DEFENSA DE TESIS DOCTORAL

TÍTULO: Contributions to the Analysis and Forecasting of Oil Prices

AUTOR: D. Pedro Moreno Alonso

DIRECTOR: Dr. D. Antonio Muñoz San Roque

TUTOR-PONENTE: No aplica

DEPARTAMENTO:

FACULTAD O ESCUELA: Escuela Técnica Superior de Ingeniería (ICAI)
Miembros del tribunal Calificador:

PRESIDENTE:	**Firma:**
VOCAL:	**Firma:**
VOCAL:	**Firma:**
VOCAL:	**Firma:**
SECRETARIO:	**Firma:**

Fecha de lectura:

Calificación:

Acknowledgments

I want to express my sincere gratitude to Antonio, my thesis supervisor, for their expert guidance, constant support, and valuable contributions throughout the research process. I would also like to mention Isabel, who has been a great source of encouragement during these last few years of research.

I deeply appreciate Noelia's unwavering support, understanding, and patience during this academic journey. I am also grateful to Aitana for being my daily inspiration and motivation.

To my parents Aurelio y Victoria, my brother Pablo, and his family Verónica, Victor and Laura, I thank them for their unconditional love, encouragement, and sacrifices to provide me with the educational opportunities that have led me to this moment.

I also extend my gratitude to my family and friends for their unwavering encouragement, understanding, and relentless support along this challenging path. Special mention to Oscar and David, and my ICAI friends for encouraging and accompanying me through his own experience.

Your support has been crucial to me and has made the achievement of this academic milestone possible. I will be forever grateful for your generosity and love.

Resumen

Siendo una de las materias primas más relevantes, la predicción de los precios del petróleo tiene un gran impacto en diversos sectores, incluidos energía, finanzas y geopolítica. Esta tesis tiene como objetivo proporcionar nuevas herramientas al proponer nuevas metodologías para modelar los precios del petróleo.

Los recientes desarrollos en los mercados energéticos han destacado la naturaleza dinámica e incierta del mercado del petróleo crudo, lo que ha dado lugar a importantes fluctuaciones de precios en las últimas dos décadas. Esta volatilidad implica un desafío considerable para predecir los precios del petróleo, ya que la literatura existente ofrece una amplia gama de metodologías de predicción, pero carece de un consenso sobre el enfoque metodológico más apropiado.

Uno de los principales hallazgos de este estudio concierne a la relación entre los costes de producción de petróleo y los precios del petróleo. Contrariamente a la creencia predominante de que los costes de producción impulsan principalmente las fluctuaciones de los precios del petróleo, nuestra investigación indica que los cambios en los precios del petróleo preceden a los ajustes en otras variables, incluidos los costes de producción. A través de un análisis riguroso que emplea la definición de causalidad de Granger y la metodología de Toda-Yamamoto, el estudio revela un patrón consistente donde las fluctuaciones de los precios del petróleo influyen en los cambios posteriores en los costes de producción, dando forma a la dinámica de la industria. Además, el análisis se extiende a contratos de futuros más largos, reafirmando la relación duradera entre los precios del petróleo y los costes de producción con el tiempo. En consecuencia, el estudio desafía la sabiduría convencional y enfatiza el papel fundamental de los precios del petróleo como el principal determinante de los cambios en los costes de producción, con implicaciones críticas para comprender la dinámica del mercado del petróleo y tomar decisiones informadas en el sector energético.

Por otro lado, esta tesis presenta un marco de modelado híbrido que combina el modelo de regresión clásico con enfoques de aprendizaje automático, utilizando específicamente el método GAM y la Función de Transferencia con el enfoque de ruido ARIMA. Al incorporar capacidades no lineales flexibles, el método propuesto captura no linealidades y permite la interpretación de variables de entrada a través de coeficientes de regresión estimados. El modelo identifica dos principales impulsores que explican los precios del petróleo: la variable Fundamental, que mide el equilibrio físico del mercado, y la variable financiera, que captura el interés especulativo de los inversionistas en petróleo crudo. El análisis de sensibilidad confirma la influencia significativa de la variable Fundamental en los precios del petróleo, seguida por la variable Financiera, el dólar y las variables de volatilidad.

Además, el modelo propuesto demuestra una capacidad de pronóstico superior en comparación con otras referencias, incluidos los precios de futuros y las predicciones de analistas de Bloomberg. También es altamente adecuado para el análisis de escenarios, cuantificando el riesgo asociado con escenarios hipotéticos alternativos sobre la demanda futura de petróleo y las condiciones de oferta. Esto incluye escenarios como condiciones de ajuste del mercado por recortes en la producción o tensiones geopolíticas en el Medio Oriente. En general, el estudio subraya la relevancia de los fundamentos de la oferta y la demanda en la determinación de los precios del petróleo y ofrece ideas valiosas para la gestión del riesgo en corporaciones e instituciones energéticas ante múltiples fuentes de incertidumbre en los mercados energéticos.

Palabras clave: previsión precio petróleo, modelo GAM, modelo Función de Transferencia, Análisis de Escenarios, Futuros Brent

Abstract

As one of the most critical commodities, the accurate prediction of oil prices holds significant implications for various sectors, including energy, finance, and geopolitics. This thesis aims to provide new tools by proposing new methodologies for modeling oil prices.

Recent developments in energy markets have highlighted the dynamic and uncertain nature of the crude oil market, resulting in significant price fluctuations over the past two decades. This volatility implies a considerable challenge for forecasting oil prices, with existing literature offering a wide array of forecasting frameworks but lacking a consensus on the most appropriate methodological approach.

One key finding from this study concerns the relationship between oil production costs and oil prices. Contrary to the prevailing belief that production costs primarily drive oil price fluctuations, the research indicates that changes in oil prices precede adjustments in other variables, including production costs. Through rigorous analysis employing Granger's causality definition and the Toda-Yamamoto methodology, the study reveals a consistent pattern where oil price fluctuations influence subsequent changes in production costs, shaping industry dynamics. Additionally, the analysis extends to longer futures contracts, reaffirming the enduring relationship between oil prices and production costs over time. Consequently, the study challenges conventional wisdom and emphasizes the pivotal role of oil prices as the primary determinant of changes in production costs, with critical implications for understanding the oil market dynamics and making informed decisions in the energy sector.

This thesis introduces a hybrid modeling framework that combines the classical regression model with machine learning approaches, specifically utilizing the GAM method and the Transfer Function with the ARIMA noise approach. By incorporating flexible non-linear capabilities, the proposed method captures non-linearities and allows input variable interpretation through estimated regression coefficients. The model identifies two main drivers explaining oil prices: The Fundamental variable, measuring the physical market balance, and the financial variable, capturing crude oil investors' speculative interest. Sensitivity analysis confirms the significant influence of the Fundamental variable on crude oil prices, followed by the Financial variable, dollar, and volatility variables.

Furthermore, the proposed model demonstrates superior forecasting ability compared to benchmark techniques, including futures prices and Bloomberg analysts' predictions. It is also highly suitable for scenario analysis, quantifying the risk associated with alternative hypothetical scenarios about future oil demand and supply conditions. This includes scenarios such as market tightening conditions from production cuts or geopolitical tensions in the Middle East. Overall, the study underscores the relevance of supply and demand fundamentals in determining oil prices and offers valuable insights for risk management in energy corporations and institutions amidst multiple sources of uncertainty in energy markets.

Keywords: Oil prices forecasting; Brent futures, GAM model; Transfer Function models; Scenarios Analysis

Contents

List of Figures

List of Tables

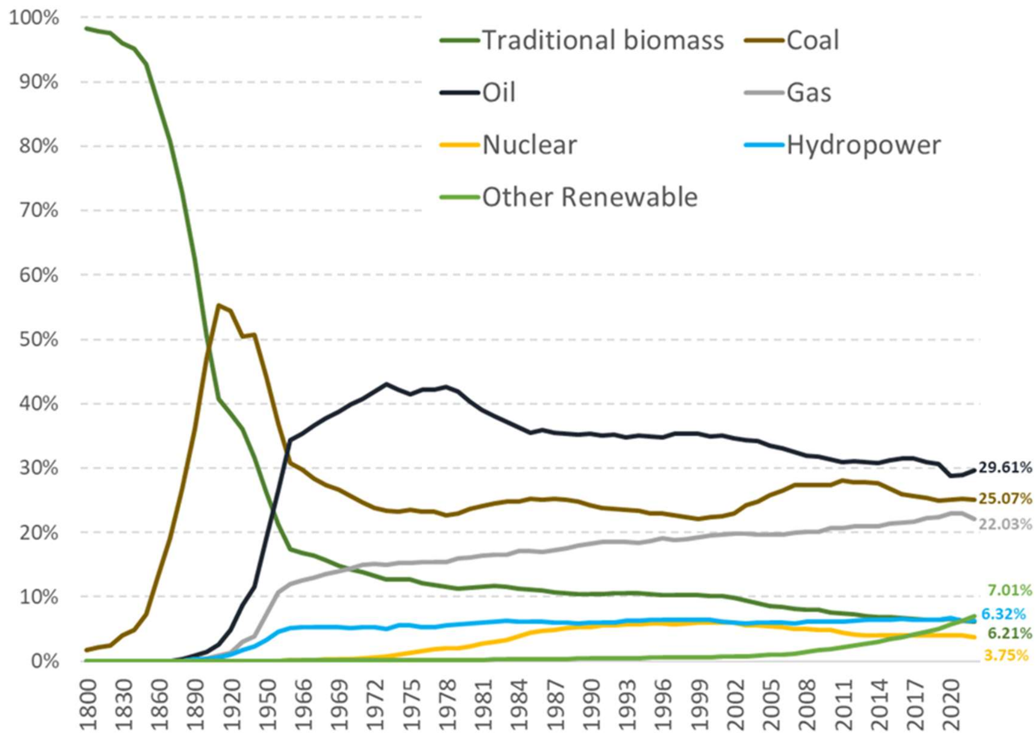
Chapter 1

Introduction, motivation, and objectives

1.1 Introduction

One of the most important discoveries for early hominids was the mastery of fire. Since then, mankind has been tied to different types of fuels. The introduction of the steam engine in the Industrial Revolution intensified the use of energy sources. For centuries, wood and coal were the main energy references (Figure 1), used for heating, lighting, cooking food, and as fuel for the first steam engines from the Industrial Revolution onwards. When the internal combustion engine appeared, oil quickly became a strategic commodity of vital importance. There are two main advantages of oil over other fossil fuels. Oil is much easier to extract than coal, and oil's calorific value per volume is much higher than natural gas, directly impacting its transport use. In addition, some of its derivatives can even lubricate the engine. For this reason, oil was the perfect fuel for a new transport revolution in the 20th century (the age of the automobile). Today, as a result of its economic and social relevance, oil is considered the benchmark for energy prices.

Figure 1: Market share of Primary energy sources along years



Note: Source: Energy Institute - Statistical Review of World Energy (2023); Smil (2017) – with major processing by Our World in Data. "Primary energy from other renewables". All data before 1965 is sourced from Smil (2017). All data from 1965 onwards, except for traditional biomass, was sourced from the Energy Institute Statistical Review of World Energy. Smil's estimates of traditional biomass are only available until 2015. From 2016 onwards, traditional biomass consumption has assumed a similar level. This is approximately in line with recent trends in traditional biomass based on Smil's data.

1.2 Motivation: The importance of Forecasting Crude Oil Prices

Crude oil, often called "black gold," plays a fundamental role in the global economic framework. Its prices affect almost every aspect of our lives, from the cost of fuel and manufacturing processes to the prices of goods and services. Consequently, forecasting crude oil prices is crucial for governments, businesses, and investors. However, given the complex and often unpredictable factors that influence oil markets, this task is fraught with challenges.

1.2.1 Economic Planning and Policy Making

Throughout the last century, the price of oil has suffered numerous shocks usually associated with international conflicts, showing the geopolitical value of the reserves of this fuel. The first oil shock after the Second World War was provoked by the Yom Kippur War in 1973 as the Arabic countries penalized the Western countries that supported Israel with an oil embargo, the "Arab embargo." The market reaction was that oil prices quadrupled, triggering an economic crisis and the concern of more efficient use and interest in finding new energy sources. From that moment, oil became a powerful political and economic weapon. Barsky and Kilian (2004) explain the impact of oil price increases on economic phenomena such as recessions, productivity declines, and reduced economic growth.

Furthermore, due to the intensive use of energy in the production chain of any other consumer product, these spikes impact the real economy by increasing inflation. This distorts economic activity, so it is one of the most essential variables that monetary authorities constantly track. Governments need accurate crude oil price forecasts to draft economic policies, set budgets, and plan for future energy needs. Predicting fluctuations in oil prices can help mitigate the impacts of sudden economic shocks and ensure stability in national economies. The direct correlation between oil prices and economic indicators necessitates vigilant monitoring and flexible policy frameworks. By understanding these dynamics, policymakers can better navigate the uncertainties of the global oil market, making informed decisions that safeguard economic stability and promote sustainable growth.

A) Impact on Inflation

As an extended idea, crude oil prices should directly and profoundly impact inflation. As the cost of oil rises, it increases the cost of fuel and transportation, which, in turn, raises the price of goods and services throughout the economy. The relationship between oil prices and inflation is theoretically positive due to the significant role of oil as an input in economies. Numerous studies have explored the impact of oil prices on inflation. For instance, Burbidge and Harrison (1984) examined several economies and found varying effects of oil price changes on inflation rates. Gisser and Goodwin (1986) observed a stronger inflationary effect of oil price increases before 1973, which diminished over time. Cunado and De Gracia (2005) identified a significant and asymmetric relationship between oil prices and inflation in Asian countries. Research indicates a mixed but generally positive correlation between oil prices and inflation rates (Huntington (2005), Alvarez et al., (2011) and Bouchouev (2021)). In high-dependency oil countries, oil prices indirectly affect inflation through increased production costs and real exchange rate fluctuations, while in low-dependency countries, direct effects are observed due to their status as oil exporters (Sek et al., 2015). Studies also highlight differences in the inflationary effects of oil shocks between oil-exporting and importing countries (Cologni and Manera (2008) and Salisu et al. (2017)).

Furthermore, Choi et al. 2018 revealed asymmetrical pass-through effects of oil price changes on consumer and producer prices in advanced and developing economies. The positive oil price shocks would have a more significant effect than negative ones. Sekine (2020) observed a weakening transition effect from oil prices to inflation in the USA over time in line with Conflitti and Luciani (2019) for the euro area and similar to what Kilian and Zhou (2023) explain about the lower pass-through from gasoline prices to inflation in the recent spike of inflation. Syzdykova et al. (2022) explored asymmetric relationships between oil prices and inflation in BRIC countries, finding diverse causal relationships between oil price shocks and inflation shocks across these nations.

While the relationship between oil prices and inflation varies across countries and time periods, evidence suggests a generally positive correlation, particularly in the short term, with significant economic policy and stability implications. This inflationary pressure can erode purchasing power and consumer confidence, leading to slower economic growth. Policymakers must carefully monitor oil price trends to implement timely adjustments in monetary policy Kilian and Lewis (2011), such as altering interest rates, to curb inflationary pressures without stifling economic growth.

B) Trade Balances

Net importers countries of oil are particularly vulnerable to rising crude oil prices, as these can exacerbate trade deficits. Higher oil prices can drain foreign currency reserves and put downward pressure on the national currency. A relatively small number of studies focus mainly on short-run dynamics. Kilian et al. (2009) analyzed the impact of demand and supply shocks in the global crude oil market on various aspects of oil exporters' and importers' external balances, indicating that the effect of these shocks on trade balances and current accounts hinges significantly on the response of non-oil trade balances, highlighting an intermediate level of international financial integration. This scenario requires careful economic policy adjustments, including measures to improve energy efficiency, diversify energy sources, and possibly adjust trade policies to mitigate the impact of higher oil prices on the trade balance.

C) Fiscal Planning

Oil export countries' governments often rely on oil production revenues to fund their budgets, as indicated by Sturn et al. (2009). Following the Arab Spring revolution, several Arab countries implemented various measures to address social and political unrest. These measures included increased government spending on social welfare programs, infrastructure projects, and subsidies to alleviate economic hardships and appease public discontent. However, such expansionary fiscal policies strained government budgets, heavily reliant on oil revenues. Fluctuating oil prices can lead to revenue volatility, complicate fiscal planning, and potentially lead to budget deficits when prices fall. Furthermore, as Mirzoev et al. (2020) explained, the oil market is experiencing a significant transformation due to new technologies increasing oil supply from various sources and growing environmental concerns driving a gradual shift away from oil dependency. This presents a considerable challenge for oil-exporting nations, which are implementing reforms to diversify their economies and fiscal revenues.

1.2.2 Investment and Financial Markets

Understanding potential changes in crude oil prices is crucial for companies, investors, and financial institutions as it enables them to effectively hedge risks, devise investment strategies, and enhance returns. The values of crude oil futures and derivatives are directly influenced by oil price forecasts, underscoring the importance of accurate predictions in guiding decision-making processes and optimizing financial outcomes for stakeholders in the market.

A) Energy Sector Investments and intensive oil consumer sectors

Crude oil prices directly affect the valuation and profitability of companies in the energy sector, influencing decisions on exploration, production, and capital expenditure. Higher prices may encourage investments in new projects and technologies, while lower prices can lead to reduced spending and project delays or cancellations. On the other hand, industries with high oil consumption are highly dependent on oil prices to define and control their operational costs, profit margins, and strategic planning. Sectors such as transportation, manufacturing, and power production heavily rely on oil as a primary input, making them particularly sensitive to fluctuations in oil prices. Accurate price forecasts enable these industries to anticipate and mitigate potential financial risks, adjust production levels, optimize supply chain management, and develop effective pricing strategies.

B) Portfolio Management and Commodity Trading

Including commodities in asset allocation is crucial for building diversified investment portfolios and managing risk effectively. Commodities, such as oil, gold, agricultural products, and metals, often exhibit low correlation with traditional asset classes like stocks and bonds, providing valuable diversification benefits, as explained by Jensen et al. (2002) or, more recently, Gao and Nardari (2018). Different factors can influence their performance, including supply and demand dynamics, geopolitical events, inflation, and currency fluctuations, making them resilient to market downturns and economic uncertainties. By incorporating commodities into asset allocation, investors can hedge against inflationary pressures, mitigate portfolio volatility, and enhance overall risk-adjusted returns. Furthermore, commodities offer unique opportunities for capital appreciation and portfolio diversification, especially during periods of market turbulence or when traditional asset classes underperform. Therefore, integrating commodities into asset allocation strategies is essential for optimizing portfolio performance and achieving long-term investment objectives.

Traders in commodity markets use futures, options, and other derivatives to speculate on price movements or hedge against them. Volatility in oil prices, driven by geopolitical, economic, and environmental factors, can lead to significant trading opportunities and risks.

C) Oil relationship with other asset classes

Sudden changes in crude oil price forecasts can contribute to financial market volatility as unexpected shifts can affect investor confidence and lead to broader market fluctuations. In the case of equity markets, there is no consensus about the relationship with oil prices. Kilian and Park (2009) demonstrate that the impact of oil prices on stock returns can vary depending on the type of shock experienced. Specifically, they find that demand shocks stem from uncertainties surrounding future oil supply shortages and tend to produce a negative correlation between oil prices and stock returns. Conversely, when oil prices rise due to an unexpected global economic expansion, they observe a positive effect on stock returns. They suggest that during the business cycle's early stages, a positive correlation exists between oil prices and stock returns, as robust demand for industrial commodities propels both oil prices and stock returns upward. The relationship between oil prices and fixed-income investment comes through inflation. As exposed previously, higher oil prices can lead to increased costs for goods and services, prompting central banks to adjust monetary policies, including interest rates. The currencies are also affected by the oil trade. Although it is not easy to isolate the effect in more complex economies, as Habib and Kalanova (2008) explained, countries that are major oil exporters or importers can see their currencies fluctuate significantly based on changes in oil prices.

1.2.3 Energy Policy and Sustainability

Oil prices play a crucial role in strategic planning and decision-making across multiple sectors, particularly concerning sustainability and ecological transition. Oil prices directly impact the economic viability of renewable energy projects and clean technologies, as well as the competitiveness of alternative fuels. Accurate oil price forecasting enables investors, businesses, and governments to assess the financial risks of transitioning to a more sustainable, low-carbon economy. Moreover, it helps design adequate energy and environmental policies, fostering the adoption of environmentally friendly practices and reducing reliance on fossil fuels. Mohamued et al. (2021) explain how oil price and volatility significantly influence GHG emissions, with asymmetrical effects observed between oil-exporting and -importing economies. While oil price increases in oil-importing countries decrease emissions, they rise in oil-exporting nations. In this context, oil price forecasting becomes a fundamental tool for driving the transition towards a greener and more sustainable economy.

1.3 Challenges and Limitations in Forecasting Crude Oil Prices

Forecasting crude oil prices presents several challenges and limitations due to the complex and dynamic nature of the oil market. These challenges stem from various factors, including the influence of geopolitical events, the interplay between supply and demand dynamics, data limitations, and model complexity.

Geopolitical events are one of the main concerns for oil price forecasting. Political instability in major oil-producing regions, such as the Middle East, threatens supply chain disruptions and leads to sudden spikes in oil prices. Despite this clear relationship, the connection between geopolitical developments and oil prices is complex and not always straightforward. Historically, there has been no clear correlation between oil prices and geopolitical events, such as tensions between countries or terrorist attacks. For instance, following the 9/11 attacks, Brent prices initially rose by 5% but dropped by around 20% within two weeks due to concerns about the economic impact, which could weaken oil demand.

Similarly, when Russia invaded Ukraine in February 2022, Brent prices surged by almost 30% within two weeks but then returned to pre-invasion levels after about several weeks. More recently, after the terrorist attacks in Israel on October 7, 2023, Brent prices increased by about 4% before stabilizing. Geopolitical shocks can cause two different impacts on oil prices. Initially, it involves that financial markets price higher risks to future oil supply, increasing the cash value of holding oil contracts and putting upward pressure on prices, also known as convenience yield. At the same time, heightened geopolitical tensions can act as a negative global demand shock, increasing uncertainty and reducing consumption, investment, and international trade. This leads to a contraction in global economic activity, reducing oil demand and prices. El-Gamal and Jaffe (2018) explore the different military conflicts, regime changes, and political tensions, concluding that only when the conflicts destroy production facilities or disrupt transportation networks could lead to sustained, long-term prices.

Moreover, the balance between supply and demand dynamics adds another layer of complexity to oil price forecasting. Changes in global economic growth, technological advancements, and shifts in energy consumption patterns can all impact oil demand, while factors such as OPEC production decisions, investment decisions of oil companies, or natural disasters can affect oil supply. Forecasting oil prices requires accurately assessing these supply and demand dynamics, which can be inherently uncertain and subject to change.

In addition to these challenges, limitations in data availability and the inherent uncertainty of future events contribute to the difficulty of accurately forecasting crude oil prices. Data availability and quality can pose challenges for oil price forecasting. Some relevant data, such as production figures or geopolitical events, may be challenging to obtain or unreliable, leading to forecast inaccuracies. Even with the effort of cooperation of the international agencies related to crude oil (Organization of the Petroleum Exporting Countries OPEC, International Energy Agency IEA, and Energy Information Administration), there is no unique figure for world production or world consumption of crude oil. Besides, the frequency of data is a factor that restricts the availability of the methodology to forecast oil prices. While advanced modeling techniques and data analytics tools can help mitigate some of these challenges, forecasting oil prices remains an inherently challenging task due to the multifaceted nature of the oil market and the multitude of factors that can influence price movements.

1.4 Objectives and Scope of the thesis

The ultimate goal of this thesis is to create a model that can explain the dynamics observed in oil prices to predict future values, improving upon currently proposed models. The forecast horizon is 12 months, considering that monthly data are used. It could be longer, but it would depend on the availability of forecasts on explicative explanatory variables.

To achieve this, we begin with a state-of-the-art review to identify models that best fit the characteristics of the oil price series. These will also serve as benchmark models our proposal should aim to surpass.

Once the benchmark models are defined, the variables influencing price formation will be analyzed. The work in this part will focus on identifying possible causal relationships among all variables associated with the oil market and their explanatory and predictive abilities. This facilitates their use in adjusting the model to be developed, although the intention is to utilize those that make the most sense statistically and fundamentally. Thus, we aim to capture the main factors. This is one of the main contributions to Oil Price Analysis and Forecasting. Generally, published models use few variables and do not typically mix them before price modeling. This thesis proposes combining a fundamental variable that attempts to capture the physical balance of oil with another financial variable that reflects market sentiment. These two variables will be created from other observable variables in the market. As we achieve the goal of creating a model, they would justify their use in future work.

Comparison with predictions from the U.S. Department of Energy, futures contracts in financial markets, and Bloomberg's survey of oil industry professionals regarding price forecasts will gauge the model's commercial usability in financial markets.

1.5 Thesis outline and contents

Apart from this introductory chapter, this thesis comprises five additional chapters addressing the objectives mentioned above.

The second chapter presents a state-of-the-art literature review on oil price forecasting modeling. A primary distinction is made between the two types of methodologies. The first one is related to classical statistics and econometrics and is commonly used by regulators. The second has emerged in the last 15 years, thanks to advancements in computing, enabling very different algorithms, from Machine Learning to Fuzzy Logic. Additionally, models that utilize more than one method and then combine them to make the final forecast will be included.

The third chapter presents all the variables considered in this study and the studies related to variable selection. A description of each variable will be provided, including how frequently they are updated and where they can be sourced. A preliminary classification will be made between fundamental variables related to the physical oil market, financial variables, and risk measurement variables. Besides, the causality study conducted on production cost variables will be introduced, resulting in the published article: "Oil Costs and Prices: An Empirical Causality Analysis." Finally, the Feature Engineering methodology for selecting variables for the proposed model will be introduced, as well as the construction of the chosen variables.

The fourth chapter explains the methodology developed for the proposed model, which combines a Generalized Additive Model (GAM) with a Linear Transfer Function

(LTF) model. This combination offers multiple benefits for modeling oil prices, as it will show.

The fifth chapter will present the results of the proposed model, published in "Forecasting oil prices with non-linear dynamic regression modeling". A second version will also be introduced with some variable modifications that allow for the inclusion of geopolitical risk premium as an explanatory variable.

The sixth chapter, the thesis's concluding section, brings the work's culmination. Firstly, it explores the conclusions drawn from the research. Subsequently, it highlights the significant contributions made by the thesis. Finally, it outlines potential avenues for future research in modeling oil prices.

Chapter 2

State of the art

2.1 Introduction

In recent years, there has been significant volatility in oil prices, with sharp increases and drops occurring associated with different factors. Given the strategic importance of oil and its profound impact on the global economy, understanding the determinants/drivers of oil price fluctuations has been a key focus for energy researchers and economists. Consequently, achieving reliable and highly accurate forecasts of crude oil prices has been vastly researched. To this end, various techniques have been employed to forecast movements, fluctuations, or volatility in crude oil prices. While econometric methods have traditionally been widespread, computational approaches, such as artificial neural networks, have gained traction in financial markets due to their flexibility and potential for increased accuracy. However, there remains no consensus on which methods are most reliable. Furthermore, comparing them is challenging due to variations in the frequency of data used, forecast horizons, and other factors to consider.

2.2 Econometric Models

We have opted to classify the models related to conventional approaches into three main categories: Time series, financial, and Structural models.

2.2.1 Time Series Models

Time series models can take various forms, ranging from univariate models that directly utilize oil prices as the explanatory variable to multivariate models incorporating additional factors such as future curves or volatility. These models rely on historical data to generate forecasts, offering considerable flexibility in terms of data frequency, with applications ranging from daily to monthly observations, typically focusing on short-term forecast horizons. Various techniques have been extensively tested across different domains, with the choice of model depending on the underlying pattern of the time series data. For instance, an ARIMA model alone may suffice if the data exhibits homoscedasticity. However, given the non-constant variance often observed in oil prices (heteroscedasticity), combining ARIMA with GARCH models to capture residual variability becomes necessary.

The ARIMA methodology, pioneered by Box-Jenkins et al. (1994), has gathered widespread adoption due to its versatility, serving both as a primary modeling approach and as a benchmark for evaluating the performance of alternative prediction methods. For instance, studies such as Lamm (2013) and Akpanta and Okorie (2014) have employed ARIMA as a foundational technique for modeling purposes, while others, including Chinn et al. (2005) and Alquist et al. (2012), have utilized ARIMA as a reference point to assess the efficacy of various forecasting methodologies. In these analyses, the variable of interest is often represented as returns or log returns of prices, allowing for a comprehensive examination of predictive accuracy and model performance.

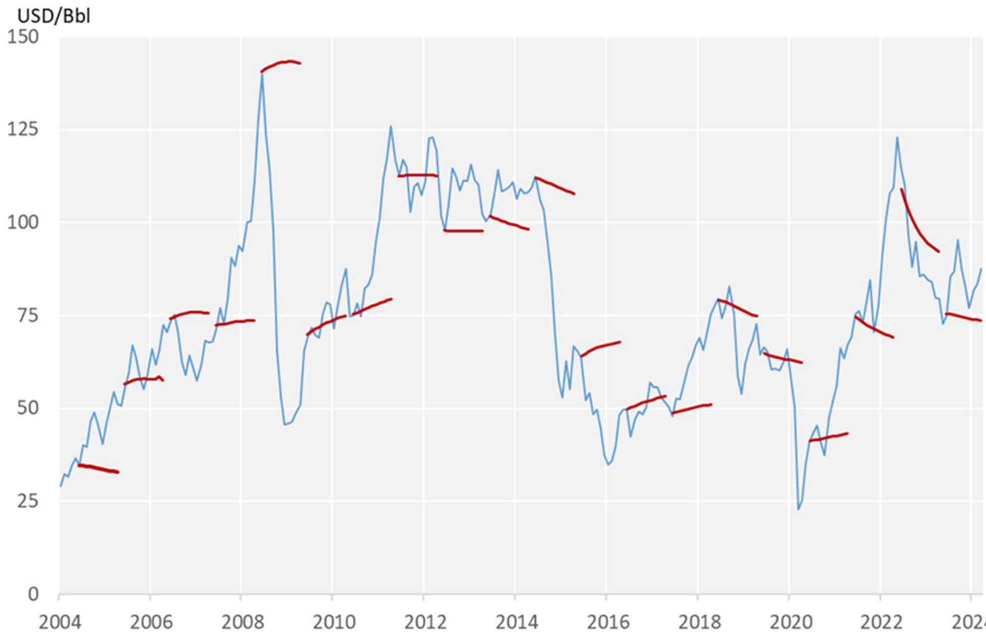
Given the heteroscedastic nature of oil prices, the modeling framework is extended by incorporating a (G)ARCH component to capture the evolving dynamics of error terms within an ARIMA context. The generalized autoregressive conditional heteroscedastic (GARCH) model, introduced by Kristjanpoller and Minutolo (2016), offers a method for forecasting oil price volatility by integrating artificial neural networks with GARCH modeling. Furthermore, researchers such as Sadorsy (2006) commonly utilize GARCH models to directly analyze and forecast crude oil price volatility, exploring various GARCH types tailored to specific assets, as observed in Narayan and Narayan (2007). However, empirical tests by Wei et al. (2010) reveal that no single linear or nonlinear GARCH-class model consistently outperforms others across all scenarios. Moreover, due to structural breaks in oil prices, Aroui et al. (2012) advocate for applying non-parametric GARCH models, such as FIGARCH, which have demonstrated superior forecasting accuracy under such conditions. Lastly, Bildirici and Ersin (2015) propose innovative nonlinear models by combining GARCH with STAR models, further advancing the predictive capabilities of volatility forecasting methodologies.

Until the early 2000s, an alternative method explored for modeling commodities was the mean reversion process, although its prominence diminished in the face of the aggressive movements and prolonged consolidation at historically high levels witnessed until 2014. Schwartz and Smith (2000) introduce a two-factor model characterized by an equilibrium level and a short-term departure from equilibrium, both reverting to a zero mean. Similarly, Skorodumov (2008) demonstrates the presence of mixed mean reversion in spot prices across various commodities based on data from 1990 to 2008.

2.2.2 Financial Models

Financial models' forecasting relies on the futures curve and price adjustments. There is a widespread belief among market participants that futures prices capture all pertinent information, rendering them the most dependable forecast of spot prices. Consequently, institutions like central banks and the International Monetary Fund routinely lean on oil futures prices as an indicator of market expectations. However, an examination of the predictive efficacy of futures prices in forecasting spot prices is depicted in Figure 2. The forward curve (in red) at different temporal points overlaps the historical Brent spot (in blue). Future contracts may not be a reliable predictor of oil prices.

Figure 2: Brent spot compared to future curves, 2004-2003



Note: Futures curves do not seem to be a reliable predictor of oil prices. Source: Bloomberg.

The inconsistency in the predictive power of futures has spurred extensive investigation by numerous scholars, yielding varied conclusions over the years. Moosa and Al-Loughami (1994) discovered that future prices fail to serve as unbiased or efficient forecasters of spot prices. Conversely, Gulen (1998) revisited the issue, extending the study's timeframe and accommodating structural breaks in the data, concluding that future prices efficiently predict spot prices. This ongoing discourse has seen conflicting findings, with some studies advocating for futures' reliability in spot price prediction, as evidenced by Kawamoto and Hamori (2009), Chinn et al. (2005), and Reeve and Vogfusson (2011). However, dissenting views persist, as articulated by Morana (2001), Stevens and Lamirande (2014), and Alquist and Killian (2010). Investigating the use of futures prices for modeling stochastic behavior in commodity prices, Schwartz (1997) and Fileccia and Sgarra (2015) offer notable examples. In such cases, forecasts must align with futures prices to forestall arbitrage opportunities.

2.2.3 Structural Models

The core objective of econometric models in oil pricing revolves around uncovering the intricate relationship between oil prices and fundamental variables. Typically, these variables revolve around physical oil's current and anticipated availability, encompassing factors such as reserves, production, consumption, and geopolitical tensions. However, econometric models also consider broader economic indicators, including equity indices, Eurodollar exchange rates, and other macroeconomic variables, which could influence oil prices. By discerning these relationships, econometric models strive to provide insights into the complex interplay between oil markets and broader economic conditions, facilitating more informed decision-making for market participants and policymakers alike.

Deeply entrenched in traditional econometrics and macroeconomic theory, these models explore medium- to long-term dynamics, rendering them valuable tools for guiding monetary policy decisions. Typically employing monthly or quarterly data, their horizon forecast extends up to two years ahead.

Regression models have been widely used among the various approaches because the relationships obtained among variables are easily understood. The number of explaining variables is limited because fundamental data are typically registered in monthly or quarterly data. For instance, Zamani (2004) developed a quarterly forecasting model for West Texas Intermediate (WTI) oil prices incorporating factors such as OECD stocks, non-OECD demand, OPEC supply, and a dummy variable (to distinguish the Iraq war period). Additionally, nonlinear relationships have been explored by researchers like Yen et al. (2006), who forecasted oil prices in the short-term (1-3 months) using OECD stocks and variables derived from stock levels, finding that low-inventory variables are more significant than the high-inventory variables. OPEC supply and OECD inventories are the most recurrent variables in the literature, but there are other approaches to consider. King et al. (2011) identified political events as the primary catalysts for significant price movements in 2007 and 2008, alongside notable impacts from OPEC decisions and surprises in EIA inventories.

Vector Autoregression (VAR) has emerged as the model of choice for medium-term oil price forecasting, offering a systematic approach to analyze multiple variables simultaneously. In a VAR model, each variable in the system is regressed on its own lagged values and the lagged values of all other variables in the system. This means that the current value of each variable is predicted based on its past values and the past values of all other variables in the system. The main idea behind VAR models is to capture the dynamic interactions and feedback loops between variables. By estimating the coefficients of these lagged variables, VAR models can capture how changes in one variable affect changes in other variables over time. Kilian (2006), (2008), and (2009) pioneered the VAR model to analyze oil prices by decomposing them, trying to identify the underlying demand and supply shocks (using just these two variables) and their different effects on price (persistence). The model incorporates variables such as the percentage change in global oil production, an index of cyclical variation in global real economic activity, and the logarithm of the real oil price. A storage variable is included in Kilian and Murphy (2014), that allows oil price expectations to affect the market. Variations of the VAR model have been explored in the literature, with Antolín-Diaz and Rubio-Ramirez (2018) adjusting sign restrictions to coefficients for improved performance. Another approach of the VAR model is Zagaglia (2009), which extends the search for variables to 230 series (global macroeconomic indicators, financial market indices, quantities and prices of energy products), extracts common factors (principal component analysis) and uses a Factor Augmented VAR (FAVAR). These latent factors, combined with returns, improve forecasting performance. Baumeister and Kilian (2012) developed a real-time forecasting model (VAR) that recreates data available at every moment (revisions of crude oil supply and demand are typical). They use VAR not only to analyse past price fluctuations but also to create conditional forecasts that consider potential future market conditions. Beckers and Beidas-Strom (2015) compare several VAR with different variables (Global oil supply, global I.P., OECD Inventories, CPI, F.X., Short-term Interest Rate, Interest Rate Spread, I.P. advanced, I.P. emerging…). Despite the outperformance of this methodology versus different benchmarks such as Random Walk, Futures-based forecast, and ARMA… they warn about some instability of the model and propose to combine it with other approaches, including futures and random walk.

When variables are integrated into the same order (at least order 1), it is possible to use a VAR model if the series are cointegrated (the error term is stationary). This model is called the Vector Error Correction Model (VECM), and there have been several approaches. Coppola (2008) used it, focusing on deviations from the long-run equilibrium between spot and futures prices as the equilibrium error. Using the crack spreads to build the model, Murat and Tokat (2009) identified significant causal impacts in the long and short terms, improving the random walk model's forecasting abilities. In Merino and Albacete (2010), OECD Inventories, non-commercial position, and OPEC spare capacity are included as variables providing a deeper understanding of the oil market.

The information from futures can be enhanced with other fundamental variables to improve the oil price predictions. Pagano and Pisani (2009) include a real-time U.S. business cycle indicator (degree of utilized capacity in manufacturing), producing significantly better forecasts, particularly at horizons above six months. Figuerola et al. (2021) demonstrate that oil future prices outperform analyst forecasts at the aggregate level.

The Dynamic Stochastic General Equilibrium (DSGE) seeks to describe a whole economic system, so consumers, producers, and investments will also be modeled. These models are based on a set of equations that represent the behavior of different economic agents, such as households, firms, and policymakers, within a unified framework. Balke et al. (2010) created a system with two manufacturing countries (one of them the U.S.) and modeled the evolution of oil reserves to explain the contradiction between the oil price movement and the U.S. economy in the 1970s, 1980s, and 2000s. Nakov and Nuño (2013) address the model from the point of view of the dominant supplier (Saudi Arabia). The system comprises three regions: one oil-importing and two oil-exporting (one of them Saudi Arabia). The model focuses on Saudi Arabia's production policy but includes an oil price forecast.

2.3 Machine Learning Models

The growing computational capacity allows us to face complex problems with new approaches and techniques, such as Artificial Intelligence. Forecasting chaotic financial time series has been really challenging, so with the development of computational intelligence, many proposals have been made to address the price of oil. Every new technique that has appeared has tried to improve previous results. The following classification comprises most recurrent techniques (not all) but just some studies to illustrate the variety of options attempted until now.

The Wavelet Transform, parallel to the Fourier Transform, decomposes the oil price time series into various subseries, each representing different scales of variation. This decomposition allows for a more detailed analysis of the underlying patterns within the data. Additionally, by extending each subseries, one can reconstruct the original signal, facilitating forecasting over various time horizons, as demonstrated by Yousefi et al. (2005). Moreover, the decomposed subseries can serve as input for further modeling using various techniques. For instance, one may employ multiple linear regression for daily crude oil price forecasting, as demonstrated by Shabri and Samsudin (2014). Alternatively, other quantitative methods like Neural Networks, as explored by Jammazi and Aloui (2012), can be utilized to model and predict the behavior of the decomposed series. This approach offers a versatile framework for understanding and forecasting oil price movements across different time scales.

Artificial Neural Networks (ANN) Haykin (2009) are one of the families of alternative methods that have experienced fantastic development. They are inspired by the human neural system. ANN are interconnected, creating a system with at least three different layers (input, one hidden, and output). Each connection has numerical weights, and each node corresponds to a function. They have robust classification and pattern recognition capabilities and are used for diverse tasks in different fields, such as robotics, data processing, classification, control, or function approximation. Many different algorithms depend on the internal structure and how the ANN is trained. The most common ANNs are Feed Forward Neural Networks (FFNNs) with a back-propagation learning algorithm. This methodology has been used for different purposes, using only the endogenous variable or adding exogenous ones. Kulkarni and Haidar (2009) used it to forecast crude oil prices in the short-term (up to three days), Wang and Yang (2010) to explore the efficiency of intraday futures markets, and Ou and Wang (2011) to solve the nonlinear relation in a GARCH model. Another Neural Network is the Radial Basis Function Neural Network (RBFNN). Qunli et al. (2011) propose an RBFNN to solve a wavelet decomposition and forecast future oil prices. Including exogenous variables, Chandar et al. (2015) consider gold prices, Standard & Poor's 500 stock index, and foreign exchange rates as inputs to feed an RBFNN. Other increasing complexity techniques, such as Recurrent Neural Networks (RNN), have been explored. RNN presents a structure with recurrent inputs at the hidden layer connected to the intermediate layer. Using gold as an input variable, Mingming and Jinliang (2012) apply this methodology to predict world oil prices and claim that other commodities can be used as input variables. Abdullah and Zeng (2010) utilized neural networks to incorporate 22 quantitative input variables, including factors related to demand, supply, economy, inventory, and population, alongside qualitative data sourced from expert opinions and news. Their aim was to forecast oil prices for both short and long-term periods.

ANNs often suffer from the problem of local minima during training or overfitting. When training in a Support Vector Machine (SVM), the solution is unique, avoiding these problems. Xie et al. (2006) compare the performance of SVM with ARIMA and FFNN (SVM outperforms both methods). Khashman and Nwulu (2011) use eight different variables with an innovative formulation because SVM performance and efficiency depend on input features (years, seasonal demand, the average price of the previous week, the total number of weeks, yearly number of weeks, world events impact factor, global demand, and future prices). Bao et al. (2019) conducted a comparative analysis of recursive and direct approaches for multi-step ahead prediction of WTI and Brent crude oil spot prices using support vector regression for long-time horizons.

Deep learning models are used to take advantage of complex and rich features from data. Cen and Wang (2019) create a bidirectional long short-term memory (BI-LSTM) that exploits the information in both directions (forward and backward). As this model is computationally intensive, Wang et al. (2020) proposed an optimizer to integrate the forecasting intervals obtained in this kind of model, capturing the uncertainty in the oil price.

Fuzzy Logic is a logic system representing variables with values ranging between 0 and 1, reflecting degrees of truth or membership. This flexibility enables handling imprecise or vague data, making it particularly useful for decision-making in scenarios where conventional binary logic may fall short. For example, it addresses linguistic or qualitative problems where statements may not have clear-cut true or false values. By leveraging fuzzy sets and membership functions, Fuzzy Logic can effectively model uncertainty and ambiguity, offering a robust framework for decision-making in complex and uncertain environments. Zhang et al. (2010) propose a straightforward model that performs very well in short-term WTI crude oil price forecasting. Combining an ANN with fuzzy logic principles (learning capability and inference system of fuzzy if-then rules), Panella et al. (2011) got the best results when several modeling techniques, including other ANN, were tested.

There are many different approaches, each with its advantages and limitations. The idea of creating hybrid models is to complement methodologies trying to mitigate limitations and achieve synergies. Hybrid systems are the combination of two or more techniques to solve problems. On the one hand, this combination could be achieved by using successively different methods to handle different steps of the forecasting process, as in Zhang et al. (2015). As with many models, it starts with a decomposition of crude oil prices. The nonlinear component will be treated through a least square support vector machine together with the particle swarm optimization (PSO) and the time-varying component through a GARCH. On the other hand, combining many different predictions from different models and combining them to create the final forecast is possible. A specific relationship rules this combination. Gabralla et al. (2013) developed three models (SVM, Instance-Based Learning, and K Star), and the combination outperforms every single one. Since the combined methodologies are typically used for short-term forecasts, hybrid models focus on forecasting one day or week ahead. Huang and Wang (2018) combined random wavelet N.N.s with a random time-effective function that effectively exploits historical crude oil time series data. Their empirical results confirmed the advantages of the proposed model over traditional shallow N.N.s and SVMs.

2.4 Combination of models

Many forecasting models have been presented in this study, reflecting that there are many different approaches to forecasting oil prices using similar data and information. However, no dominant methodology has yet been identified. In other words, the forecast error of each model will fluctuate over time, so it will be possible to find better and worse predictions in any given case. Ideally, it would be possible to select the best model at any given time.

That is the idea inherent in model combination. In Timmermann (2006), the reasons for using forecast combination are explained: a) it is a form of diversification that leads to improved individual forecasts, b) it helps to protect against structural breaks, c) it helps to protect against misspecification bias, and d) it helps to alleviate the bias that arises when the underlying forecasts may be based on different loss functions (asymmetric sensitivity). There are two proposed model combinations for predicting the oil price. Manescu and Van Robays (2014) evaluate the performance of ten models with different approaches: the no-change, random walk, futures, risk-adjusted futures, non-oil commodity index, a DSGE model based on, and four alternatives from (VAR models) and no dominant forecasting method is found over futures or random walk. However, a combination of four equally weighted models (futures, risk-adjusted futures, a VAR, and DSGE) is created that more accurately predicts Brent oil prices. In Baumeister (2014), a combined model is proposed that uses the futures curve "model", a world oil market model (oil production, economic activity index, and inventories), an industrial commodity model, and a refined product spreads model (crack spreads). However, we find an essential limitation to this approach. It is not helpful for exploring hypothetical assumptions about future economic conditions, so a structural model complements it.

Chapter 3

Feature engineering for oil price forecasting

3.1 Introduction

This section describes the variables analysed and the process of selecting and building the variables that will be included in our model. The oil market boasts immense data, comprising various factors ranging from geopolitical events and economic indicators to supply chain dynamics and weather patterns. This wealth of data is paramount in constructing robust and accurate models for forecasting oil prices. With access to such extensive datasets, analysts can discern intricate patterns and correlations, enabling them to develop sophisticated predictive models. Moreover, the sheer scale of available data allows for comprehensive analysis, facilitating a deeper understanding of the complex interplay of factors influencing oil prices.

Consequently, leveraging this abundance of data is essential for enhancing the precision and reliability of oil price forecasting models, thereby aiding stakeholders in making informed decisions and mitigating risks in the volatile oil market. Choosing the correct variables is critical in model construction as it significantly impacts the model's performance and interpretability. The selection process involves identifying factors relevant to the problem and has a meaningful impact on the predicted outcome. While correlation between variables can provide insights into their relationships, it is essential to distinguish between correlation and causation. Correlation indicates that two variables are associated with each other, but it does not imply causation, meaning one variable causes changes in the other. Failing to discern between the two can lead to erroneous conclusions and ineffective models. Therefore, a thorough understanding of the underlying mechanisms and context is necessary to establish causal relationships accurately. By incorporating variables based on their causal significance rather than solely on correlation, models can better capture the proper drivers of the phenomenon being studied, resulting in more accurate predictions and actionable insights.

3.2 Description of oil price-related variables

In the dynamic landscape of oil markets, understanding the multitude of variables that impact oil prices is principal for stakeholders across various sectors. We will start to meticulously examine three distinct types of data: Fundamental, Financial, and Risk Data. Each type plays a crucial role in shaping the dynamics of oil prices, from underlying supply and demand fundamentals to intricate financial market dynamics and geopolitical risks. The reader can gain valuable insights into the complex interplay of factors influencing oil price movements by providing a comprehensive overview of these variables as well as some combination of some of them. A brief explanation of price references will also be included.

3.2.1 Fundamental variables

Fundamental variables refer to any factor that directly influences the supply, demand, and overall market dynamics of crude oil. These variables are critical in determining the price movements of crude oil and are used by analysts, traders, and policymakers to forecast future price trends and make informed decisions

A) Supply

A1) World Supply

Looking ahead to collect oil-related variables, we find several data sources regarding production and demand. This can make it challenging to achieve conclusions of the relations obtained. Although the production levels that are issued are very similar, differences can be found in the variations, even opposite signs. There are up to 3 international agencies that could provide this data: The International Energy Agency

(IEA), the U.S. Department of Energy (DoE), and the Organization of Petroleum Exporting Countries (OPEC).

A2) OPEC Supply

There are two reasons why OPEC is highly relevant in the oil world. On the one hand, it is responsible for 35% of world production; on the other hand, it has even more market dominance in proven oil reserves, as the countries of this organization have about 80%. These facts result in an increasing dependency on OPEC. In order to achieve price stability, OPEC has self-imposed production quotas to avoid oversupplying the market. Nevertheless, it is worth saying that most of these countries do not comply with their assigned quota. OPEC supply data can be obtained from the three agencies monthly. On top of that, DoE makes forecasts monthly for up to two years and IEA quarterly (demand is treated similarly).

A3) Spare OPEC Supply

In addition to tracking production levels, data on spare capacity are also meticulously recorded for this region. Spare capacity is defined as the amount of oil that can be extracted and brought to market in less than a month. This metric holds similar significance to inventory levels because it serves as a buffer to address quickly and correct imbalances between supply and demand. In sudden demand spikes or supply disruptions, spare capacity can be rapidly utilized to stabilize the market. For those interested in monitoring this crucial aspect, spare OPEC supply data are accessible through reputable sources such as the International Energy Agency (IEA) and the U.S. Department of Energy (DOE). These organizations provide detailed reports and analyses that help stakeholders understand spare capacity's availability and potential impact on the global oil market.

A4) Baker Hughes Rig Count

Baker Hughes has been offering weekly count operating rigs (rotary drilling equipment) in the U.S. since 1944 and worldwide every month since 1975. Changes in the number of operating drilling should reflect oil production with some delay. This delay is not a fixed value; five months is a valid estimation.

A5) CAPEX, new proved reserves bring in today's dollars, Cash Cost, Full Cycle Cost, and Accounting Breakeven

Oil-producing companies report quarterly data on their financial statements and cash flow. We can use these reports to calculate these ratios only for part of the oil industry since we are missing data from all the national companies in OPEC. These ratios can help determine future production, the breakeven price to modify the production in the short term, and the costs of exploration to get oil in 5 to 7 years.

A6) Production Cost IHS, Production Cost Labor

Considering cost calculation, some other indexes can help create industry proxies. The company IHS CERA has created several indexes to measure inflation in costs related to the different parts of the oil industry that are reported every quarter. Besides, the Bureau of Labor Statistics indexes prices for various activities and related industry oil extraction products monthly.

B) Inventories OECD, World, USA

Supply and demand balance changes in the oil market should be directly reflected in the inventories. Therefore, inventories are variables highly relevant in the oil market analysis. It is worth mentioning that the volume of the inventories is helpful to handle unexpected changes in supply. This is aimed at satisfying better the oil demand. OECD inventories are the global benchmark reference. Both IEA and DOE release this data every month. Additionally, DOE publishes data about U.S. inventories for oil, gasoline, and other products every week. PJK International (a private company for oil market analysis) provides aggregated oil inventories of independent storage companies in the ARA region in Europe every week.

C) World Demand

Likewise, the supply variables have the same agencies that provide them. Regarding data periodicity, it is possible to obtain monthly consumption data from every country. On top of that, it is essential to analyze the factors that affect consumption patterns.

D) World economic growth

The growth of oil demand has been closely linked to the economic development of countries. GDP data can be found in the IMF quarterly (not monthly). Other monthly parameters, such as industrial production or manufacturing indices, can be used as proxies for some regions are available monthly.

E) Weather

The weather impacts the oil demand, so the number of days of heating or cooling is registered in addition to the oil demand. The DoE reports U.S. monthly data, but it is difficult to make an aggregate globally.

F) Energy Intensity

Technological improvements also influence the oil demand. The amount of energy needed to produce consumer goods is decreasing due to the evolution of technology. Nevertheless, measuring that impact is problematic given that the surge of new technologies, like Artificial Intelligence, requires enormous energy. The relationship between GDP growth and oil demand growth is used as an indicator.

G) Crack Spreads

Mathematically, it is just the differential between refined products and crude oil prices, but that means the profit margin for an oil refinery. There are different configurations for refineries, so there are some options. The most followed reference is 321 (buy 3 barrels of oil and sell 2 barrels of gasoline and 1 of distillate fuel). It can be computed directly with the futures prices for every product (daily).

3.2.2 Financial Data

The vision of commodities as an investment asset has been widespread since the early 2000s, mainly due to their excellent diversification properties. This interest in the financial world may have impacted the oil market by intensifying oil price movements or even creating bubbles. Undoubtedly, the financial demand must be considered when studying the oil price. Since the financial demand is part of the equation, we need variables that can give us more details about the state of investors. The U.S. Commodity Futures Trading Commission (CFTC) is an independent agency of the U.S. government that regulates futures and options markets. This agency generates the Commitments of Traders report every week. This report shows a breakdown of commodity contracts operating under U.S. exchanges. Note that the values are from Tuesday, and it is released on Friday. The following subchapters describe the most relevant figures of the CFTC report according to the CFTC website.

A) Open Interest (CFTC)

Open interest is the total of all futures and/or option contracts entered into and not yet offset by a transaction, delivery, or exercise. The aggregate of all long open interest is equal to that of all short open interest.

B) Long-Short non-commercial positions (CFTC)

When an individual reportable trader is identified to the Commission, the trader is classified as "commercial" or "non-commercial." A trader's reported futures positions in a commodity are classified as commercial if the trader uses futures contracts in that particular commodity for hedging as defined in CFTC Regulation 1.3(z), 17 CFR 1.3(z). A trading entity generally gets classified as a "commercial" trader by filing a statement with the Commission, on CFTC Form 40: Statement of Reporting Trader, that it is commercially "...engaged in business activities hedged by the use of the futures or options markets." To ensure that traders are classified accurately and consistently, Commission staff may exercise judgment in re-classifying a trader if it has additional information about the trader's use of the markets. A trader may be classified as a commercial trader in some commodities and a non-commercial trader in others. A single trading entity cannot be classified as a commercial and non-commercial trader in the same commodity. Nonetheless, a multi-functional organization that has more than one trading entity may have each trading entity classified separately in a commodity. For example, a financial organization trading in financial futures may have a banking entity whose positions are classified as commercial and a separate money-management entity whose positions are classified as non-commercial. A long (short) non-commercial positions refer to the futures contracts held by traders who are not involved in the actual production, processing, or merchandising of the underlying commodity. These traders, often referred to as speculators, take long (short) positions when they anticipate that the commodity's price will rise (decline). A long position means the trader has purchased a futures contract to sell it later at a higher price.

C) Disaggregated Commitments of Traders Report (CFTC)

The Disaggregated COT report, only available on 22 major physical commodity markets, increases transparency from the legacy COT reports by separating traders into the following four categories of traders:

- Producer/Merchant/Processor/User: an entity that predominantly engages in the production, processing, packing, or handling of a physical commodity and uses the futures markets to manage or hedge associated risks.
- Swap Dealers: an entity that deals primarily in swaps for a commodity and uses the futures markets to manage or hedge the risk associated with those swaps transactions. The swap dealer's counterparties may be speculative traders, like hedge funds, or traditional commercial clients managing risk arising from their dealings in the physical commodity.
- Money Manager is a registered commodity trading advisor (CTA), a registered commodity pool operator (CPO), or an unregistered fund identified by the CFTC. These traders manage and conduct organized futures trading on behalf of clients.
- Other Reportables: every other reportable trader is not included in one of the other three categories.

D) Traders in Financial Futures Report (CFTC)

The TFF report divides the financial futures market participants into the "sell-side" and "buy-side." This traditional functional division of financial market participants focuses on their respective roles in the broader marketplace, not whether they are buyers or sellers of futures/options contracts. For instance, the "dealer/intermediary" category represents sell-side participants. Typically, dealers and intermediaries earn commissions on selling financial products, capturing bid/offer spreads, and accommodating clients. The remaining three categories ("asset manager/institutional," "leveraged funds," and "other reportables") represent the buy-side participants. These are essentially clients of the sell-side participants who use the markets to invest, hedge, manage risk, speculate, or change the term structure or duration of their assets.

E) Futures curve and convenience yield

Today's oil price is traded in the market, but so are future oil prices. The futures curve is built with the prices for the following months until five years with significant volumes.

Futures prices of commodities differ from other financial assets. These prices include the cost of capital and the cost of capital storage (custody), referred to as the spot price, which prevents arbitrage. Sometimes, a prime is linked to the benefit of physical commodity ownership. This prime is known as convenience yield, and its value can be obtained from spot prices, future prices, capital costs, and storage costs. It is possible to get daily data on all these variables except for the storage cost (which is pretty stable).

F) USD, Equity, and Baltic Dry Index

The data of these variables are available with high frequency, daily at least.

- U.S. Dollar: This is the oil currency since most oil contracts are quoted in U.S. dollars. There is exposure to the dollar's exchange value for suppliers and producers outside the U.S. DXY index, which is used to remove the effect of the dollar on oil prices.
- Equity Market: Although the influence of the oil price is higher in the Equity market than vice versa. The equity market is a factor to consider in the oil price analysis. For this reason, global indexes such as MSCI or regional indexes Standard&Poors or Eurostoxx should be considered.
- Baltic Dry Index (BDI): An economic indicator the London-based Baltic Exchange issued daily. Not restricted to Baltic Sea countries, the index provides "an assessment of the price of moving the major raw materials by sea." The index covers Handysize, Supramax, Panamax, and Capesize dry bulk carriers carrying a range of commodities, including coal, iron ore, and grain. BDI is seen as an economic indicator that predicts future economic activity based on commodities related to industrial activity.

3.2.3 Risk Data

A) Volatility

Oil volatility has been typically estimated using Garch models. It is commonly used in the calculation of Risk management and Options value. If we talk about realized volatility, it is the standard deviation of oil returns and an essential parameter in the oil price modeling. Implied volatility quoted in the derivative market could also help to forecast oil prices.

B) Risk Reversal

It is a measurement based on the volatility skew of derivatives. It shows where the market expects the higher risk of changing prices. For example, the difference in the price of puts and calls and their volumes can reflect the market fear depending on the evolution of the oil price.

C) Geopolitical Risk

Measurement of geopolitical risk is not a simple task. Many agents use their property metrics as strategic and secret tools. However, Caldara and Iacoviello (2022) develop a metric for adverse geopolitical occurrences and their associated risks by analyzing a compilation of newspaper articles that report on geopolitical tensions. They investigate the progression of this geopolitical risk (GPR) index and its economic ramifications dating back to 1900. The index exhibits significant spikes during pivotal events such as the two world wars, the onset of the Korean War, the Cuban Missile Crisis, and the post-9/11 periods. Elevated geopolitical risk tends to forecast declines in investment, stock prices, and employment levels. Moreover, it correlates with heightened probabilities of economic crises and poses substantial downside risks to the global economy.

D) St Louis Fed Financial Stress Index

The fourth version of the St Louis Fed Financial Stress Index (STLFSI4) measures the financial stress in the markets and is constructed from 18 weekly data series: seven interest rate series, six yield spreads, and five other indicators. Each of these variables captures some aspect of financial stress. Accordingly, as the level of financial stress in the economy changes, the data series are likely to move together. It was first published in early 2010, with data going back to 1993, to measure levels of financial stress.

3.2.4 Combined Data

In addition to explaining the previous variables, we have also considered combining them to create indexes that better explain the behavior of the oil market.

The most relevant is the oil inventories' days of supply. Ideally, we should have data on all inventories worldwide. In practice, this value is only based on the OCDE inventories. This measurement contains the oil market balance and trend and the capability to satisfy the oil demand during supply shocks.

Other indexes can be created by using financial ratios. For instance, the ratio between the long and short non-commercial positions differs, and the number of open interests is also affected.

This ratio makes capturing more relevant information in the models with fewer variables possible.

3.2.5 Price references, real vs. nominal price

When defining a reference price for oil, there are different alternatives. In 1974, the market started using the refiner acquisition costs in the USA. Even though it is irrelevant, the DoE still monitors this value. Other price references are WTI (American) and Brent (European). Both of these were equally considered oil benchmarks until five years ago. Currently, Brent has remained the most reliable benchmark. WTI has experienced some distortions due to anomalies in the local market in recent years.

Regarding periodicity, daily data are available for Brent and WTI, while the refining acquisition cost values in the USA are provided weekly.

Apart from choosing the price oil reference to model, it is critical to determine if Inflation will be considered. Economists are more interested in real or inflation-adjusted prices. Mathematicians usually take directly nominal prices.

In the remainder of the document, we will refer to the oil price as Brent, which is currently the undisputed international benchmark.

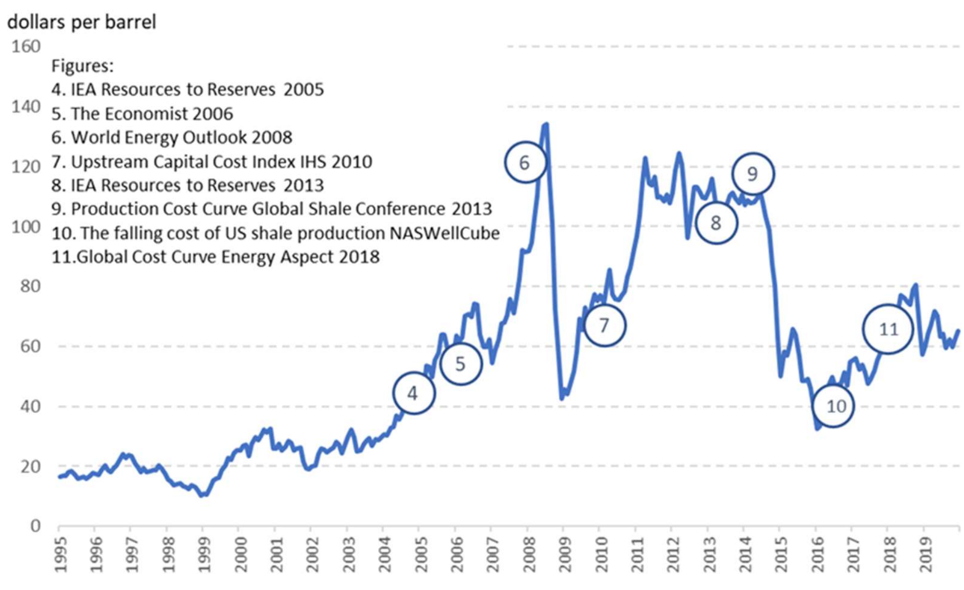
3.3 Variable Selection

3.3.1 Causality versus Correlation

A) Motivation

In an efficient market, the marginal cost of production of a particular product should establish a clear benchmark for its price. According to the theory of Exhaustible Resources in Hotelling (1931), for a resource like oil, non-renewable and limited, the price should exceed the marginal cost before the possibility of future scarcity by introducing a premium. Supply and demand shocks may produce price deviations over time, which should be initially addressed with inventory variations. Considering the market structure, competition among producers should return the price to the path of marginal cost, adding supply or eliminating it according to market needs. On the demand side, there should also be competition that tends to stabilize the market. This breakeven is not so simple to calculate and is more dynamic than anyone could imagine, as claimed by Kleinberg et al. (2017), but it should be the reference for the hedging strategy of the companies. Since the beginning of the 21st century, oil prices have been in a roller-coaster market with extreme fluctuations. Baumeister (2016) explained that this volatility makes investment planning extremely difficult for companies, having severe implications on economic decisions (e.g., inflation, salaries, and available resources)

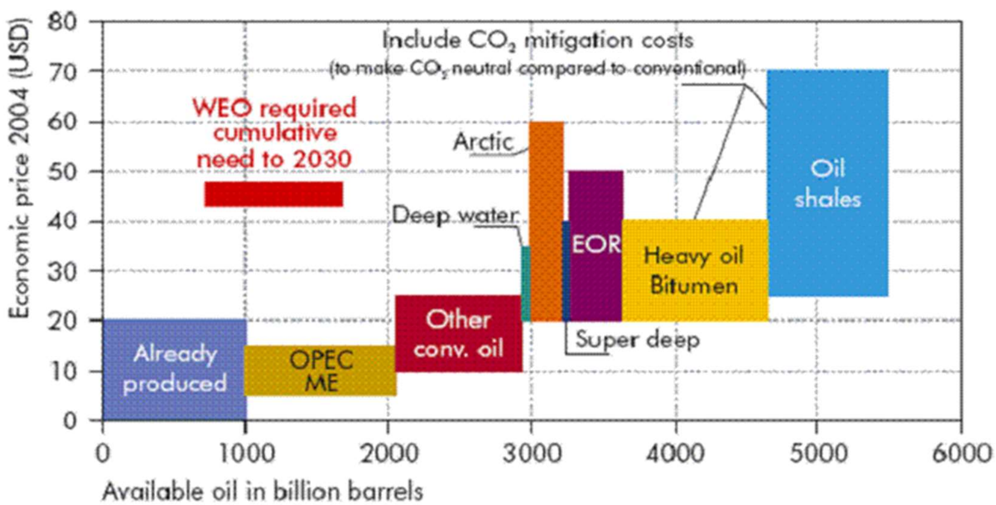
Figure 3: Chronology of cited figures along with Monthly Average Brent price



Note: Chronological identification of reports about oil costs with different oil price levels.

In the late '90s, the petroleum reached historical minimums at $11/bbl. Then, the new century began with a recovery in oil prices, leaving behind the crisis of the Asian economies. However, it relapsed again with the puncture of the Dot-com bubble and the post-September-11 instability. After that, the second Gulf War and steady growth of the world economy, led by emerging countries, caused historically high oil records in nominal and real prices. Between 2003 and 2008, the price went from $25 to $138/bbl, with the inability of the supply to cope with the rapid increase in demand (2% per year). Traditionally, this dramatic price increase would have caused highly negative consequences in the global economy. However, in this case, the economy's strength was the main reason for skyrocketing prices, according to Kilian (2009). Given the need to increase the oil offer, new deposits that had not been used until now were explored. In "Resources to Reserves 2005" (OECD Publishing, International Energy Agency 2005), the cost curve for different technology deposits (Figure 4) is included for the first time in IEA research. With oil peaking above $60/bbl, large volumes of oil not yet exploited were beginning to be competitive to face rising prices.

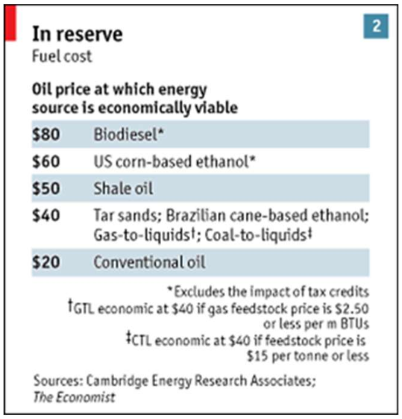
Figure 4: Production Cost by kind of oilfield in 2004



Note: The x-axis represents cumulative accessible oil. The y-axis represents the price at which each type of resource becomes economical. Source: IEA, Resources to Reserves - Oil and Gas Technologies for the Energy Markets of the Future, 2005

Just one year later, in April 2006, with Brent trading above $70/bbl, "The Economist" in The Oil Industry: Steady as she goes (2006) published new price ranges at which unconventional or unexploited oilfields would be viable (Figure 5). Shale oil seems to have a breakeven of $50/bbl, and biofuels start at $60. These figures almost doubled the numbers published by the IEA in the previous year. However, these new sources did not seem mature enough to contribute to the supply in the short term, and the price rally continued for a couple more years. Shale Oil production did not take off until 2010 and did not acquire relevance until 2014, with sustained production of 3.5Mbbl/d and higher costs than estimated previously.

Figure 5: Fuel Cost in 2006



In reserve

Fuel cost

Oil price at which energy source is economically viable

$80	Biodiesel*
$60	US corn-based ethanol*
$50	Shale oil
$40	Tar sands; Brazilian cane-based ethanol; Gas-to-liquids†; Coal-to-liquids‡
$20	Conventional oil

*Excludes the impact of tax credits

†GTL economic at $40 if gas feedstock price is $2.50 or less per m BTUs

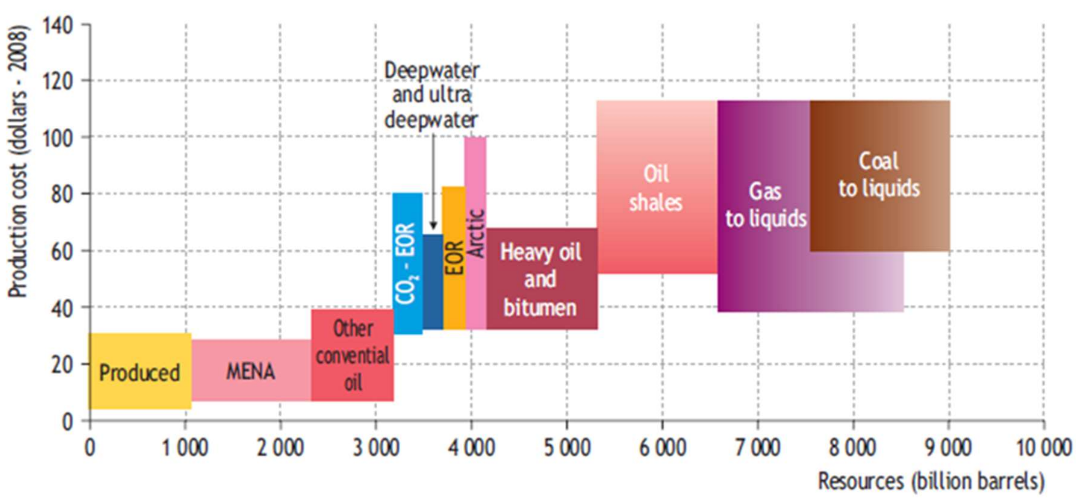
‡CTL economic at $40 if feedstock price is $15 per tonne or less

Sources: Cambridge Energy Research Associates; *The Economist*

Note: Source: Cambridge Energy Research Associates; The Economist

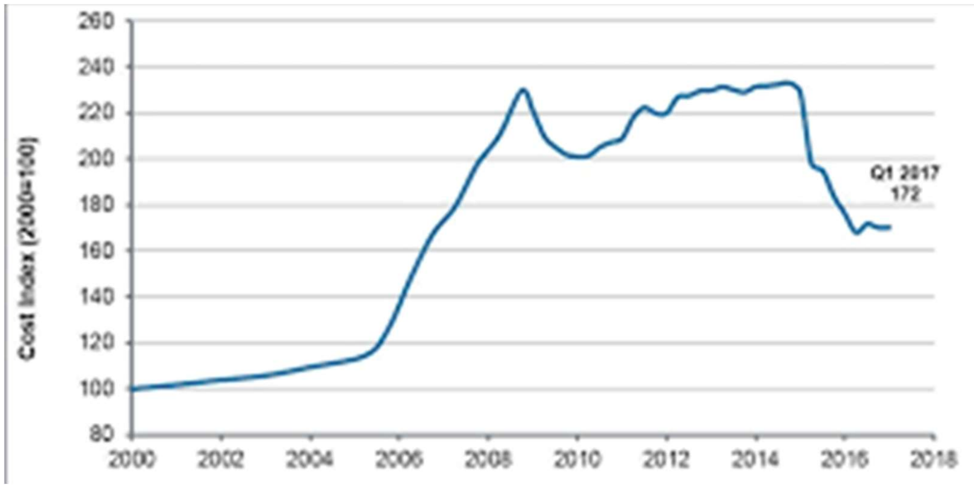
The World Energy Outlook of 2008 (International Energy Agency 2008) devotes an entire chapter to the cost increase of new production capacity, and the graph of the cost curve is updated (Figure 6). It shows the concern for both the delay in capital expenditures by companies and an environment of rising costs, which reduces the investment impact. This fact makes us consider that the breakeven calculated with oil prices at 2006 levels will no longer be valid with the oil at historically high values. An increase in costs was one of the main arguments to justify the highs reached in July 2008. On the 25th of June 2008, Daniel Yergin (Chairman of Cambridge Energy Research Associates) explained to the U.S. Congress Joint Economic Committee the causes of the oil price, pointing out the increase in production costs (Yergin, 2008). He argues that costs doubled between 2004 and 2008, according to the indexes calculated by his organization (Figure 7).

Figure 6 Production Cost by kind of oilfield 2008



Note: The x-axis represents cumulative accessible oil. The y-axis represents the price at which each type of resource becomes economical. Source: World Energy Outlook 2008, IEA

Figure 7 Upstream Capital Cost Index



Note: Source IHS

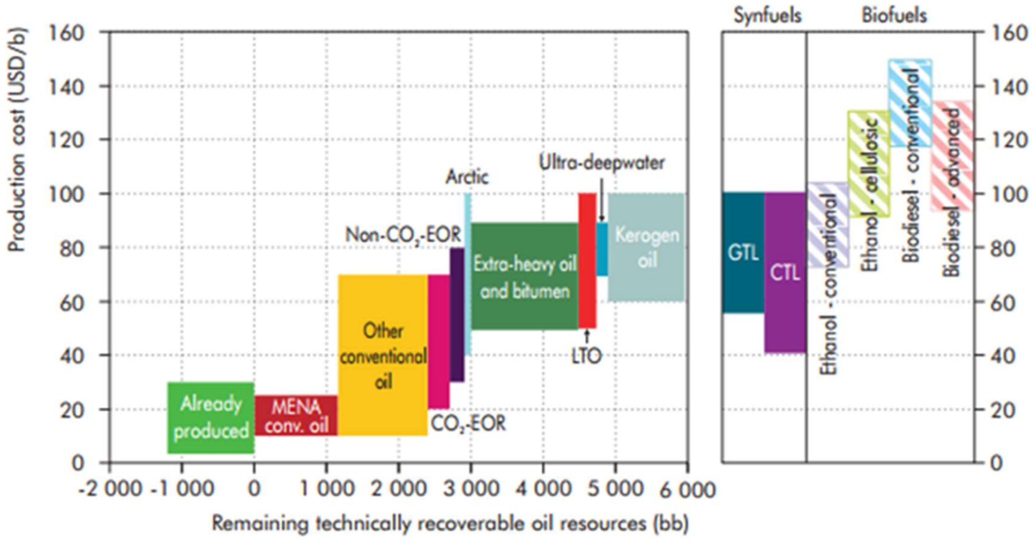
The 2008 financial crisis spread to oil in the second half of the year. Oil prices plummeted to $40/bbl by the end of the year. Later, in 2009, oil almost doubled its price. Reluctantly, the investment in new capacity was recovered after the announcement of budget cuts in public companies at the beginning of the year. In the following years until Q4 2014, oil continued escalating and consolidating the $100/bbl.

During these years, it was accepted that the structure of the oil sector had changed due to its scarcity, as stated by James Hamilton. Hamilton (2014) claims that oil needs to stabilize at a price close to $100/bbl to exploit the new oilfields.

The theory of Hubbert's Peak also attracted supporters in the analyst community. This theory highlighted that since 2005, oil production has been practically stagnant (Chauvet et al. (2012). The increase in output would have been faced by spared capacity, not new projects. Due to the strengthening of demand, the physical balance would be pushing up the price of oil permanently.

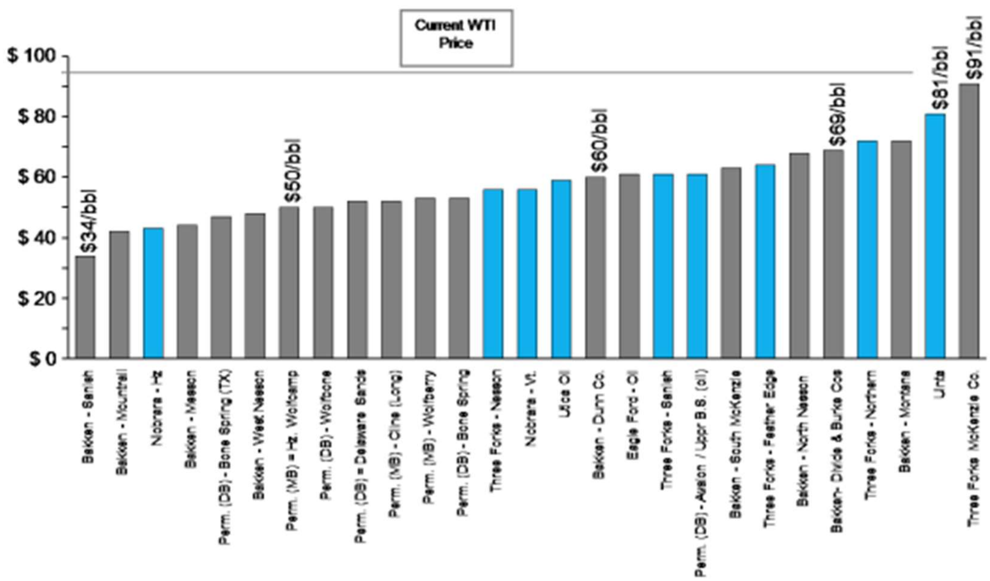
With oil prices consolidated above $100/bbl, IEA published "Resources to Reserves 2013", pointing out that production costs for shale oil (kerogen) were between $60/bbl and $100/bbl dollars (Figure 8). Technology development for this type of oilfield was becoming active, and some cost containment was beginning to be seen simultaneously. In fact, Sandrea (2014) completely reviewed the American Shale gas and oil industry. It showed how most shale oil projects were profitable while shale gas projects were not. The wide variability between projects should be mentioned, ranging from $34/bbl to $91/bbl, as shown in the cost curve (figure 9).

Figure 8: Production Cost by kind of oilfield 2013



Note: The x-axis represents cumulative accessible oil. The y-axis represents the price at which each type of resource becomes economical. Source: Resources to Reserves 2013 IEA

Figure 9: Production Cost Curve



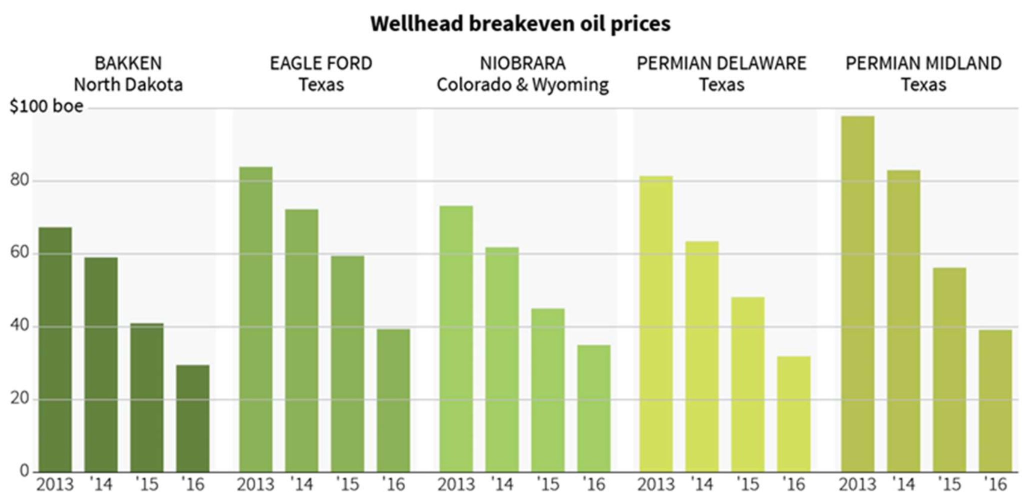
Note: Source: TPH and HPDI (Global Shale Conference, November 21, 2013)

In the last quarter of 2014, there was a sharp drop in the oil price, close to 50%. The oil price will keep lowering until 2016, reaching a value below $30/bbl (the lowest level since 2003). Several studies associate this slump with three principal causes. Baumeister and Kilian (2006) considered that two of them would be predictable and would have accounted for half of the fall in 2014: a slight global slowdown in demand and a positive supply surprise. The remaining factor is the unexpected weakening of the world economy at the end of 2014, leading to a fall in oil price expectations. Given such a price fall, it is argued that supply adjustments are needed, and the most expensive producers should limit their output volumes.

In this market situation, the references used as cost proxies have ceased to be valid. According to Rystad Energy, the production costs of different Shale Oil fields were halved between 2013 and 2016 (figure 10) while keeping their supply during the price drop. Surprisingly, some of the technologies being considered expensive could reduce production costs.

Both Verleger Jr (2016) and Behar and Ritz (2017) affirmed that the justification for the price collapse is a change in the oil industry structure. This change is a consequence of the withdrawal of production restrictions by the OPEC, which ruled for decades until 2014. Without these restrictions, each country is allowed to produce according to its needs and criteria. This lack of quotas would have caused an offer at a much lower cost, leading to a price collapse.

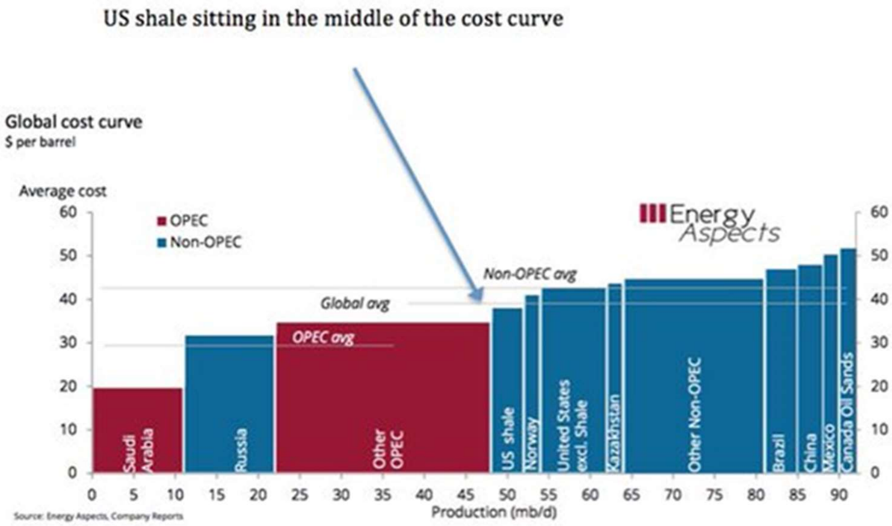
Figure 10: The falling cost of U.S shale production



Note: Source: NASWellCube Rystad Energy

From the '90s until nowadays, oil price analysts have described numerous episodes of dramatic movements. Economists and industry experts have used production costs as a critical variable that drives these movements, which are used to predict, usually unsuccessfully, oil price trends (upwards or downwards). The cost curve seems helpful (figure 11), indicating potential production loss if prices fall. This reduction in supply should underpin prices. However, the energy price is a cost source for some oil production technologies, so its variation will be closely related. Therefore, the connections between oil prices and production costs are unclear.

Figure 11: Global Cost Curve



Note: Source: Energy Aspects

Literature dealing with this relationship is less abundant than newspaper references when significant oil price movements happen. Toews and Naumov (2015) have approached the connection between oil prices and costs in the industry by estimating a structural model. They find that a 1% increase (decrease) in oil price increases (decreases) global drilling activity by 1% and costs of drilling by 0,5% with a lag of a year. However, shocks to the costs of drilling do not have a permanent effect on the oil price. A different approach (Anderson et al. 2014) using a Hotelling model explores drilling activity, prices, and costs for a local production area (Texas). Their main finding is that pre-existing wells do not respond to oil price shocks, while new wells and drilling rig rental rates are strongly co-varying with oil prices. Other relationships among variables (rig counts) have also been studied by Khalifa et al. (2017). They verify that the impact of changes in oil prices on rig counts lags up to one quarter.

B) Causality Analysis

In order to determine the direction of causality between variables, the Granger Causality Test is a widely used and helpful tool (Granger 1969). This test tries to distinguish mere correlation from causality. A universally accepted definition of causation may well not be possible. Still, a reasonable definition to many is the following: "Let Ω_n represent all the information available in the universe at time n. Suppose that at time n, optimum forecasts are made of X_{n+1}, using all of the information in Ω_n and also using all of this information apart from the past and present values Y_{n+j}, j ≥ 0, of the series Y_t. If the first forecast, using all the information, is superior to the second, then the series Y_t has some specific information about X_t, not available elsewhere, and Y_t is said to cause X_t" (Schmalensee et al. 1980). The test is based on linear regression modeling of stochastic processes.

The initial idea is to compare two linear regression models. The first one, the autoregression of Y_t, explains the output variable Y_t from its own lags (restricted model):

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-1} + \cdots + \alpha_j Y_{t-j} + \varepsilon_t \quad (1)$$

The second one is the previous model augmented by including lagged values of X:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \cdots + \alpha_j Y_{t-j} + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_j X_{t-j} + u_t \quad (2)$$

The null hypothesis for Granger causality test is that lagged values of X are not statistically significant, so they do not improve the explanation of the variation in Y. Granger Causality test compares the Sum of Squared Error of the restricted model (SSE_r) with the Sum of Squared Error of the unrestricted model (SSE_u) using an F-test. The F-statistic is given by:

$$F = \frac{(SSE_r - SSE_u)/m}{SSE_u/(n-k)} \quad (3)$$

where m is the number of lagged X values used in the unrestricted regression, n is the number of observations in our sample, and k is the number of parameters estimated in the unrestricted model (constant included).

The time series involved need to follow stationary processes to conduct the test. In the case of integrated processes, as with oil prices, Gujarati (2006) showed that the F-test procedure is invalid, as the test statistics do not have a standard distribution.
To deal with integrated time series, Toda and Yamamoto (1995) propose an interesting and straightforward procedure, estimating an augmented Vector Autoregressive model VAR (order p'), with d extra lags, where d is the order of integration of the variables. This modification guarantees the asymptotic distribution of the Wald statistic since the testing procedure is robust to the integration and cointegration properties of the process. The first step is determining the order of integration of the time series through the Dickey-Fuller test (Dickey, Fuller 1981). An information criterion is used (Akaike Information Criterion (Akaike 1974) or Bayesian Information Criterion (Schwarz 1978)) to determine the optimal lag length of the VAR model. The order of the VAR model could be increased in case there is a serial correlation in the residuals to define the appropriate model:

$$Y_t = C_1 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \sum_{i=1}^{d} \beta_i X_{t-i} + u_{1t} \quad (4)$$

$$X_t = C_1 + \sum_{i=1}^{p} \gamma_i X_{t-i} + \sum_{i=1}^{d} \delta_i Y_{t-i} + u_{2t} \quad (5)$$

The next step is estimating a VAR by adding d lags, so the model contains p plus d lags. The null hypothesis for the Granger non-causality test is that lagged values of X are not statistically significant, so they do not improve the explanation of the variation in Y. Therefore, the null hypothesis is for equation 4:

$$H_0 : \sum_{i=1}^{d} |\beta_i| = 0 \tag{6}$$

While the alternative hypothesis:

$$H_1 : \sum_{i=1}^{d} |\beta_i| \neq 0 \tag{7}$$

It is important to note that coefficients for the extra d lags are not included when performing the Wald tests. They were added to fix up the asymptotic but were not used afterward. The Wald test statistics will be asymptotically chi-square distributed with p degree of freedom. The rejection of the null hypothesis supports the presence of Granger causality.

Therefore, the outcomes obtained for both time series are if X → Y (notation for X Granger causing Y) and also if Y → X (Y Granger causing X). For that reason, four different outcomes are explored:

1) Unidirectional causality X → Y but not Y → X
2) Unidirectional causality Y → X but not X → Y
3) Dual causality where X → Y and Y → X
4) No Granger causality

C) Data and Materials

Development time in the upstream industry could last from some months to years, depending on the characteristics of the project and the sector activity. Therefore, the lead-lag effect between different variables and the oil price will be shown in addition to this methodology. It is a simple method, but studying correlations between oil price and the other variables (leading up to 9 periods and lagging up to 5 periods) will provide more evidence in terms of temporal precedence. Revenue changes could modify company decisions, but the effects between variables are not instantaneous.

The upstream industry involves crude oil exploration, development, and extraction. The time series, whose causality will be investigated, is the price of oil and some production cost indicators. Two data sources were used to provide a broader view of the sector.

The Brent price refers to the benchmark price for crude oil extracted from the North Sea, specifically from the Brent, Forties, Oseberg, and Ekofisk oil fields. It is one of the most widely used benchmarks for oil prices globally. Brent serves as a major reference price for setting the prices of other crude oils worldwide (Bossley 2017). In order to determine if companies could be using longer terms of the oil curve for hedging, four price contracts will be explored (1 month, 12, 24, and 60 months). Descriptive statistics of Brent's different contracts are shown in Table 1.

Table 1: General Statistics

Monthly	1M	12M	24M*	60M**
Average	55,50	55,16	59,35	77,36
SD	32,86	32,74	30,98	17,37
P95	112,36	107,66	105,04	102,79
P05	15,61	15,93	16,12	55,93
Integraged	I(1)	I(1)	I(1)	I(1)
AugD-F Test	-1,545	-1,415	-1,476	-1,497

*data from 31/01/1998
**data from 31/01/2006

Note: Source: Bloomberg

There are some difficulties when calculating the cost in any industry, so there is no single cost index. Two sources of data have been chosen with relevant time series that reflect the costs experienced by the oil industry:

A. **Producer Price Indexes from the U.S. Bureau of Labour Statistics**: Three different Producer Price Indexes published by the U.S. Bureau of Labour Statistics are employed: Drilling Oil and Gas Wells (Drilling), Support Activities for Oil and Gas Operations (Support), and Oil and Gas Field Machinery and Equipment (Machinery). Their higher calculation frequency and greater consistency provide them with excellent suitability for the needs of the study. There are other production cost indices, such as those published by IHS/CERA, or those that can be extracted from the companies' balance sheets but present some difficulties in terms of periodicity. The time considered for this analysis is from January 1995 to December 2019. Monthly data are going to be used. Descriptive statistics of our four variables are shown in Table 2.

Table 2: General Statistics

Monthly	Drilling	Support	Machinery	Production
Average	271,05	164,10	215,15	85,59
SD	111,78	34,96	46,81	9,14
P95	442,36	199,61	269,00	100,31
P05	110,39	110,60	152,10	71,48
Integraged	I(1)	I(1)	I(1)	I(1)
AugD-F Test	-1,4668	-1,5077	-1,5953	-0,5598

Note: Source: U.S. Bureau of Labour Statistics.

B. **Cost built from the Oil and Gas Companies Reports**: Data from 20-F reports and Annual Reports have been analyzed for the 47 largest public oil companies to establish the global cost of production. In order to compute this number, an annual curve cost has been built, starting from finding, development and acquisition (FD&A) costs, lifting and selling, general and administrative (SG&A) costs. FD&A refers to costs incurred when a company purchases, researches, and develops properties to establish oil reserves. FD&A costs are calculated by adding exploration, development, and acquisition costs and dividing them by the oil reported by the company (discoveries, improvements in recovery, and revisions). This term is the most volatile cost because investment in one year could bloom in the following years. Lifting costs (also called production costs) are the costs to operate and maintain wells and related equipment and facilities per barrel of oil equivalent (boe) of oil and gas produced by those facilities after the hydrocarbons have been found, acquired, and developed for production. Lifting costs and selling general and administrative costs are obtained directly from SEC reports.

Table 3: Companies included in Cost Curve 2017

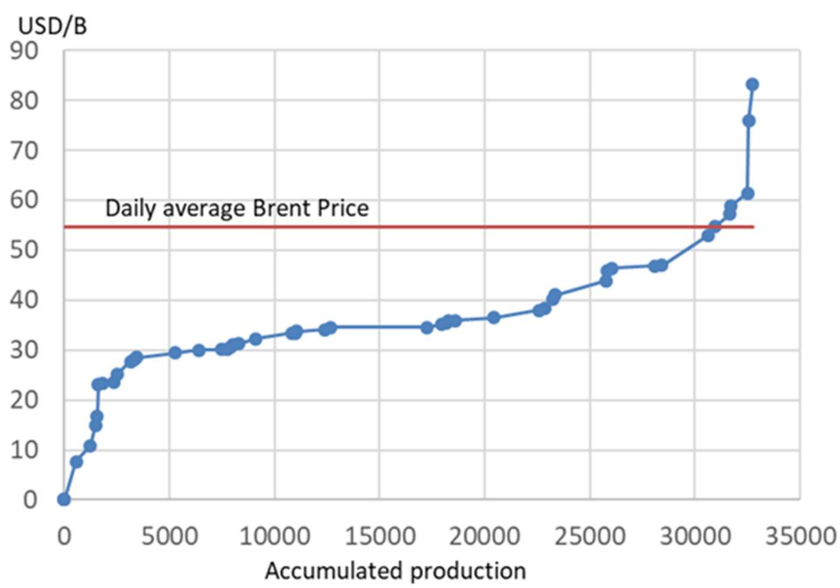
Company	Country	Production Kb/d	% Oil	Cost 2017 USD/Bbl
ROSNEFT OIL CO PJSC	Rusia	5,718	80.05	34.59
PETROCHINA CO LTD-H	China	3,994	60.85	43.93
EXXON MOBIL CORP	USA	3,985	50.89	46.81
ROYAL DUTCH SHELL PLC-A SHS	UK	3,664	49.81	36.47
BP PLC	UK	3,551	60.92	37.99
PETROBRAS - PETROLEO BRAS-PR	Brazil	2,767	79.33	52.97
CHEVRON CORP	USA	2,728	63.16	33.40
TOTAL SA	France	2,566	52.46	34.05
LUKOIL PJSC	Rusia	2,269	79.51	29.38
EQUINOR ASA	Norwey	2,080	54.76	30.03
ENI SPA	Italy	1,816	46.92	32.26
NOVATEK PJSC	Rusia	1,410	16.72	14.87
CONOCOPHILLIPS	USA	1,377	52.36	35.09
CNOOC LTD	China	1,288	82.67	30.11
OIL & NATURAL GAS CORP LTD	India	1,268	53.69	10.83
CHINA PETROLEUM & CHEMICAL-H	China	1,209	67.14	61.40
ECOPETROL SA	Colombia	715	76.22	23.50
REPSOL SA	Spain	695	36.69	38.37
SUNCOR ENERGY INC	Canada	685	100.00	57.17
ANADARKO PETROLEUM CORP	USA	672	52.83	47.01
EOG RESOURCES INC	USA	609	55.30	35.97
OCCIDENTAL PETROLEUM CORP	USA	602	63.29	40.13
CHESAPEAKE ENERGY CORP	USA	548	16.5	33.78
DEVON ENERGY CORP	USA	543	44.94	31.30
CENOVUS ENERGY INC	Canada	471	72.83	54.85
APACHE CORP	USA	457	53.41	34.58
MARATHON OIL CORP	USA	397	51.13	46.35
NOBLE ENERGY INC	USA	381	34.38	28.48
OMV AG	Austria	348	51.65	83.33
IMPERIAL OIL LTD	Canada	335	94.03	30.16
HUSKY ENERGY INC	Canada	323	66.55	23.39
ENCANA CORP	Canada	313	24.36	35.97
HESS CORP	USA	306	57.84	28.01
PIONEER NATURAL RESOURCES CO	USA	272	58.23	30.51
CONTINENTAL RESOURCES INC/OK	USA	243	57.06	25.14
CONCHO RESOURCES INC	USA	193	61.82	33.41
CRESCENT POINT ENERGY CORP	Canada	176	79.54	41.00
SEVEN GENERATIONS ENERGY - A	Canada	175	31.8102	75.93
MURPHY OIL CORP	USA	164	54.5446	31.06
WPX ENERGY INC	USA	110	55.8288	23.11
MOL HUNGARIAN OIL AND GAS PL	Hungary	99	38.06	58.78
GALP ENERGIA SGPS SA	Portual	93	87.3662	16.82
DIAMONDBACK ENERGY INC	USA	79	74.07	45.93

Note: Source: Bloomberg and Companies Reports

Since the output of these companies also includes gas, an exercise of splitting expenses has been completed. It has been considered that costs incurred have been proportional to revenues obtained in producing that oil or gas at the year's average price. For that distribution, it has been considered that a barrel of oil equivalent -boe- of gas is equivalent to 5.8MMBTU. If 1 barrel of oil is around $60 and the price for an MMBTU of gas is about $3, the revenue from one boe of oil is higher than income for one boe of gas ($60 versus $17.4). When companies allocate resources, costs and revenues should remain close. The 2017 production cost curve obtained is shown in Figure 12, and the list of the companies is exhibited in Table 3. Integrated oil companies, like B.P. or Shell, have a diversity of oilfields that provide lower production costs. Producers related to shale oil, like Suncor or Cenovus Energy, have, in general, higher costs than conventional producers, as was expected. In this case, the time considered is from 1998 to 2019 annually. Descriptive statistics of our variables are shown in Table 4.

Two production costs have been considered in the study. The "Cost 100%" considers all companies' marginal costs the highest. The "Cost 90%" is the highest cost, covering the 90 percentile of total production. This last measure seems more robust and meaningful than the marginal cost of oil production. Since the Granger Causality Test requires determining the order of integration of time series, the Dickey-Fuller test (Dickey, Fuller 1981) has been evaluated. A transformation should be done previously to use the Granger Causality Test for the oil, as shown in Tables 1, 2, and 4. In order to reduce observed heteroscedasticity, a logarithmic transformation is applied.

Figure 12: 2017 Oil Cost Built from Publicly Companies Form 10-K



Note: Source: Bloomberg and Companies Reports

Table 4: General Statistics

Yearly	Cost 90%	Cost 100%
Average	44,41	64,57
SD	23,30	28,51
P95	70,34	94,19
P05	12,08	22,65
Integraged	-1,3531558	-1,4483236
AugD-F Test	-1,4668	-1,5077

Note: Source: Bloomberg and Companies Reports

D) Empirical Results

a) **Producer Price Indexes from the U.S. Bureau of Labour Statistics**

Table 5 shows the p-values of the Toda-Yamamoto causality test for the data set related to industry costs. Values below 0.05 will show Granger-causality with a 95% confidence level. The main conclusion is that no factor has been identified to Granger cause the price of Brent. The hypothesis that higher/lower costs result in higher/lower Oil prices does not seem valid. Some bidirectional relationships among the other variables have been found. In addition, there is strong evidence that the price would cause every cost variable to be analyzed.

Table 5: Toda-Yamamoto Causality Test (p-value)

p_values					
Brent 1M	--	**0,00028****	**0,0072****	**0,00043****	0,051
Drilling	0,970	--	**0,000024****	**0,0042****	0,620
Support	0,450	0,720	--	**0,0033****	0,330
Machinery	0,094	0,087	**0,046***	--	**0,0042****
Production	0,820	0,180	**0,011***	**0,028***	--
Granger Cause -->	Brent	Drilling	Support	Machinery	Production

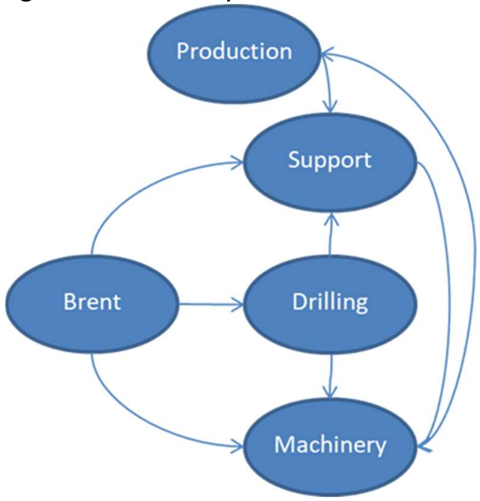
* 95% confidence level

** 99% confidence level

Note: Bold p-values show Granger Causality

In order to visualize the interactions, a diagram relating to causality has been elaborated. Relationships between variables with p-values below 0.05 have been considered. In those cases where both directions are involved, a bold double arrow is used. The dynamic system proposed is summarized in Figure 13.

Figure 13: Causality direction relationship among variables



Note: Brent seems to be the primer to changes in other variables

Since oil companies normally hedge their production some months or even years ahead, the production cost could affect other contracts instead of those closest to maturity. Therefore, the calculus has been repeated with the 12-, 24- and 60-month contracts. In Tables 6 and 7 p-value from Toda-Yamamoto Causality test is shown. The relationships between the different contract prices and the cost indices remain in the same direction.

Table 6: Toda-Yamamoto Causality Test (p-value) Brent Granger-causes

Granger Cause -->	Drilling	Support	Machinery	Production
Brent 1M	**0,00028****	**0,0072****	**0,00043****	0,051
Brent 12M	**0,0000037****	**0,0024****	**0,000024****	0,055
Brent 24M	**0,0000038****	**0,00013****	**0,017***	**0,0015****
Brent 60M	**0,0011****	**0,000058****	0,079	**0,041***

* 95% confidence level

** 99% confidence level

Note: Bold p-values show Granger Causality

Table 7: Toda-Yamamoto Causality Test (p-value) Cost Granger-causes

Granger Cause -->	Brent 1M	Brent 12M	Brent 24M	Brent 60M
Drilling	0,970	0,960	0,770	0,970
Support	0,450	0,810	0,440	0,940
Machinery	0,094	**0,060**	**0,062**	0,170
Production	0,820	0,980	0,580	0,510

* 95% confidence level

** 99% confidence level

Note: Bold p-values show Granger Causality

Even in the maturities dominated by production oil companies, the longest one, it cannot find a relationship that the production cost would be causing the oil price (table

VII). This fact implies that oil cost would not be the main driver in determining the hedging price in companies' strategies. It is evident that production cost is an essential figure, challenging to determine, for an oil producer but with no direct impact on the market oil price.

In order to provide further evidence for this conclusion, the cross-correlation of Brent with the other variables (up to 9 periods) was calculated. When variables within the same industry respond in the same instant to shocks, it is said that they are coincident. In that case, the correlation between variables reaches its highest value when comparing contemporary time series. Therefore, causal relationships could be found in any of the two-way directions (or even dual). In the case of dual causality, both variables could respond to another inceptive factor. However, when a variable leads to another, causality is likely in one direction but not in the other.

We have restricted our study to Brent as the primary variable. As shown in Table 8, in the cases of Drilling, Support Activities, and Machinery costs, the maximum occurs with oil leading between 7 and 9 months. This fact means that changes in oil prices this month would imply changes in the other variables that will take place several months from now. So, speaking in temporal precedence, Brent would cause the other variables' movements.

Table 8: Lead-Lag Cross-Correlations Brent vs. other variables

Log Brent	months	Drilling	Support	Machinery	Production
Lagging	t=-5	0.855	0.875	0.826	0.732
	t=-4	0.864	0.880	0.831	0.734
	t=-3	0.873	0.885	0.836	0.737
	t=-2	0.883	0.890	0.842	0.739
	t=-1	0.893	0.896	0.847	0.741
	t=0	0.903	0.902	0.852	0.744
	t=+1	0.912	0.907	0.857	0.748
	t=+2	0.921	0.913	0.862	0.751
	t=+3	0.928	0.918	0.865	0.750
	t=+4	0.934	0.922	0.869	0.750
Leading	t=+5	0.939	0.925	0.873	0.750
	t=+6	0.943	0.9272	0.876	0.749
	t=+7	0.945	0.9278	0.880	0.748
	t=+8	0.946	0.9271	0.883	0.746
	t=+9	0.946	0.926	0.885	0.746

Note: Correlations increase when Brent leads the other variables

b) **Cost built from the Oil and Gas Companies Reports**

In the case of cost built from the Oil Companies Reports, the conclusion is similar to the previous case. There is no evidence that production costs Granger cause Brent price as shown in the first column of Tables 9 and 10 (values above 0.05). Besides, it can be concluded that the level of Brent price Granger causes the costs reported by the companies, and these relationships also remain along the futures curve. It does not matter what the cost percentile used is.

Table 9 and 10: Toda-Yamamoto Causality Tests (p-value)

Companies	p_values		
Brent 1M	--	**0,00071****	**0,0067****
Perc90	0,390	--	
Perc100	0,810		--
Granger Cause -->	logBrent	Perc90	Perc100

* 95% confidence level
** 99% confidence level

Companies	p_values		
Brent 12M	--	**0,00000****	**0,00039****
Perc90	0,350	--	
Perc100	0,650		--
Granger Cause -->	logBrent	Perc90	Perc100

* 95% confidence level
** 99% confidence level

Note: Bold p-values show Granger Causality

The explanation would be that when companies detect an alteration in the level of their incomes (different prices), they react by correcting the budget in exploration and production to the new situation (changing the industry's cost). In a period of high prices, the increase in Capital Expenditure creates a scarcity of resources. Companies that provide oilfield services will use the opportunity to increase their prices. The opposite seems to happen when the oil prices go down.

When creating the cost curve, FD&A costs are volatile and challenging to apply to an exact period. Furthermore, the alteration of results on exploration from previous expectations on oil recovery could alter figures in both directions. If revisions have increased/reduced the amount of oil found, the costs would be underestimated/overestimated. In order to minimize this volatility, an analogous exercise has been completed using the three-year average of finding costs with similar results.

The marginal costs obtained would have been more accurate if the cost curve had been built from the oilfields instead of the companies. Wood McKenzie and Rystad Energy are some of the companies that are trying to collect these data. Due to budget limitations, the study has not included these time series.

E) Discussion

In order to recap the results, it is essential to highlight that the study seeks to determine the causal relationships (and the direction of these) between the price of oil and production costs (upstream). In that sense, none of the analyzed cases indicate that the production costs would be causing the price of oil. In fact, it seems that it is just the opposite. Brent's inception would affect drilling, machinery, and support costs. These, in turn, would cause changes in the rest of the variables analyzed.

In order to find an explanation for the upstream industry's business, these relationships indicate that a change in the price of oil would be transferred to the producing companies in the form of a change in income. The higher or lesser availability of revenues would motivate these companies to modify their budgets for the future, thus adapting their needs to develop new oilfields. Depending on these labor, machinery, and material needs, the cost indicators would also reflect this change. The connection between production and oil prices is not well defined, but in any case, the causality relationship would be that high prices incentivize to augment production.

The observed delay between the price of oil and the activities related to oil extraction is part of the investment cycle in the industry. Prospecting projects can take a quarter to several years to start operations from the investment decision. This delay may have been identified in our study. The apparent lag production costs take to react to higher oil prices seems to be between 7 and 9 months. It could be considered even a short period for the complexity of the projects in this industry.

Oil production costs have been a recurrent justification for the rise in oil prices during the last decades. No evidence of that justification has been found in this study. New technology and regulation have allowed the exploitation of shale oil fields and other unconventional resources, triggering higher-cost production, but this would not be the reason for the increase in oil prices. In fact, the possibility of exploiting fields with higher costs would be given by the rise in the price of oil.

Based on Granger's definition of causality and the Toda-Yamamoto methodology, this study has analyzed the interactions between price and production costs. According to the results obtained, oil price changes would lead to changes in the rest of the variables. The explanation of this relationship would imply that an increase (decrease) in Brent would cause an increase (decrease) in revenues in oil-producing companies. Considering that companies prefer to maintain dividends steady, the budget for exploration and drilling would increase (reduce), pressing (softening) the prices of professionals, machinery, and raw materials and increasing (decreasing) production costs.

Exploring the Brent future curve has shown that production costs are not the primary driver in determining the heading price in companies' strategies. There is no change in the relationships between oil prices and production costs when conducting the study with longer futures contracts.

Therefore, when trying to explain or predict the movement of oil prices in the future, we should not consider the related costs of the industry (unfortunately often used) as explanatory variables. It is the price of oil that precedes the movements of production costs. The direction of this causality has been well observed in many cases, especially during the correction of 2014-2015.

3.3.2 Feature engineering for oil price forecasting

"Feature engineering could be defined as the process of using domain knowledge to extract features (characteristics, properties, attributes) from raw data (N.G. 2019). The motivation is to use these extra features to improve the quality of results from a machine learning process, instead of supplying only the raw data to the machine learning process". This study aims to utilize previously transformed and combined variables to construct a model. Thanks to those variables, this model will allow a better understanding of the oil market. Feature engineering has been a successful tool in machine learning models, and in the case of oil, price forecasting could also be an advantage. In econometrics models, authors often test multiple raw variables to build oil models, as in Beckers and Beidas-Strom (2015) or Zagaglia (2010). However, there have also been successful attempts to create new variables, such as the well-known Kilian indicator for economic activity in Kilian (2009).

Many factors affect oil markets, and it is possible to find information for most of them. Unfortunately, there are several sources of information, and they do not always match together. Therefore, this study will use the US Energy Information Administration (EIA) as a fundamental data source. The EIA Short Energy Outlook reports data series related to the fundamental balance in the crude oil market. This agency releases data for production, production capacity, consumption, stocks, price, and other variables for different regions. As shown below, these are used to construct the fundamental variable measure and provide input forecasts based on EIA data.

In addition to the fundamental data, it is possible to find several financial data that should help the model capture price movements. For that purpose, data relative to investor positions will be obtained from the Commodity Futures Trading Commission. The CFTC releases weekly data on investor positions to construct the Financial variable. Data on Long and Short Positions of non-commercial agents and open contracts were obtained for the entire sample period. Other position data that were initially considered but not selected as input data are specified in Table 11. Weekly data are transformed into monthly averages for analysis. Front-month Brent Intercontinental Exchange (ICE) crude oil daily data are downloaded from Bloomberg. This is used to calculate the monthly average price. The log of the monthly Brent price is the target variable within the model. However, model forecasts (provided in logs) are transformed into level Brent spot data to allow comparison with benchmark forecasting models. The nominal Brent spot price returns are used to construct the historical (realized) volatility measure. We use daily quotations of the DXY dollar index to calculate a monthly measure of the dollar variable. We also download Daily Brent ICE futures prices for the remaining available maturities (2-12 months) to construct the futures price benchmark as an alternative forecast measure.

Table 11: Raw data description

Raw Variable	Frequency	History	Source	Model Variable
Brent Log(Brent)	Monthly (average daily data)	from Jan 1995	Bloomberg	Fundamental Variable
Total World Production OPEC Production Spare OPEC Production Total World Consumption OECD Consumption China Consumption OECD Commercial Inventory OECD Total Inventory Stocks Consumption Ratio	Monthly	from Jan 1995	U.S. Energy Information Administration	Fundamental Variable
Long non-commercial Futures Short non-commercial Futures Net non-commercial Futures Open Interest Futures	Monthly (average weekly data)	from Jan 1995	Commodity Futures Trading Commission	Financial Variable
Long non-commercial F&O Short non-commercial F&O Net non-commercial F&O Open Interest F&O		from March 1995		
DXY USD/EUR	Monthly (average daily data)	from Jan 1995	Bloomberg	Dollar
Implied Volatility Realized Volatility	Monthly (average daily data)	from Jan 1995	Bloomberg Price	Volatility

Note: This table describes the data used in the initial stage of algorithm implementation. The second to fourth columns provide variable frequency, data history, and data source. The label "Model variable" in the last column describes the category of the given data series within the Fundamental, Financial, dollar, and volatility, according to dimensions for model input variables.

The analysis covers from 1995 to 2023 (358 observations). The highest frequency possible to get all these data is monthly. The data pre-processing step combines different variables to create a new one representing some market feature. The correlation between the oil price and other variables will be used to decide which combination will be chosen and can be explored in Tabla 12. A brief description of the variables used in the model (created using Feature Engineering and direct variables), headlined by the related market driver, will be explained in the following paragraphs.

Table 12: Correlation matrix of primary variables used in the analysis

	Brent	Log(Brent)	Total World Production	OPEC Production	Spare OPEC Production	Total World Consumption	OECD Consumption	China Consumption	OECD Commercial Inventory	OECD Total Inventory	Stocks Consumption Ratio	Fundamental Variable	Long non-commercial Futures	Short non-commercial Futures	Net non-commercial Futures	Open Interest Futures	Long non-commercial F&O	Short non-commercial F&O	Net non-commercial F&O	Open Interest F&O	Financial Variable Futures	Financial Variable F&O	DXY	USD/EUR	Implied Volatility	Realized Volatility
Brent	1.00	0.96	0.63	0.66	-0.15	0.66	-0.15	0.60	0.00	0.36	-0.79	-0.81	0.51	0.53	0.41	0.61	0.53	0.46	0.48	0.73	0.44	0.36	-0.57	0.55	-0.23	-0.24
Log(Brent)	0.96	1.00	0.74	0.74	-0.16	0.77	-0.10	0.70	0.11	0.49	-0.86	**-0.89**	0.60	0.64	0.49	0.71	0.62	0.56	0.56	0.81	**0.51**	0.45	**-0.52**	0.49	-0.18	**-0.21**
Total World Production	0.63	0.74	1.00	0.80	-0.17	0.98	-0.22	0.96	0.58	0.79	-0.79	-0.80	0.90	0.73	0.81	0.92	0.90	0.70	0.85	0.87	0.80	0.78	-0.11	0.08	0.00	0.02
OPEC Production	0.66	0.74	0.80	1.00	-0.56	0.78	0.03	0.65	0.35	0.63	-0.71	-0.72	0.64	0.76	0.49	0.70	0.66	0.73	0.56	0.77	0.53	0.48	-0.38	0.35	-0.08	-0.06
Spare OPEC Production	-0.15	-0.16	-0.17	-0.56	1.00	-0.12	-0.41	0.05	0.20	0.07	0.26	0.29	0.05	-0.13	0.11	0.02	0.04	-0.16	0.09	-0.01	0.08	0.09	0.18	-0.16	0.08	-0.01
Total World Consumption	0.66	0.77	0.98	0.78	-0.12	1.00	-0.13	0.95	0.55	0.77	-0.84	-0.80	0.89	0.74	0.79	0.92	0.89	0.70	0.84	0.87	0.78	0.77	-0.14	0.10	-0.06	-0.06
OECD Consumption	-0.15	-0.10	-0.22	0.03	-0.41	-0.13	1.00	-0.38	-0.45	-0.36	-0.16	-0.01	-0.43	-0.14	-0.47	-0.36	-0.42	-0.18	-0.45	-0.29	-0.50	-0.46	0.06	-0.07	-0.06	-0.13
China Consumption	0.60	0.70	0.96	0.65	0.05	0.95	-0.38	1.00	0.64	0.80	-0.72	-0.71	0.93	0.68	0.87	0.95	0.93	0.65	0.91	0.86	0.85	0.85	-0.09	0.07	-0.01	0.00
OECD Commercial Inventory	0.00	0.11	0.58	0.35	0.20	0.55	-0.45	0.64	1.00	0.88	-0.02	-0.05	0.71	0.57	0.65	0.65	0.70	0.61	0.64	0.54	0.65	0.66	0.16	-0.17	0.15	0.14
OECD Total Inventory	0.36	0.49	0.79	0.63	0.07	0.77	-0.36	0.80	0.88	1.00	-0.38	-0.41	0.85	0.81	0.73	0.85	0.85	0.81	0.76	0.81	0.75	0.74	-0.16	0.13	0.07	0.05
Stocks Consumption Ratio	-0.79	-0.86	-0.79	-0.71	0.26	-0.84	-0.16	-0.72	-0.02	-0.38	1.00	0.93	-0.60	-0.56	-0.51	-0.68	-0.61	-0.47	-0.58	-0.71	-0.51	-0.49	0.27	-0.24	0.13	0.14
Fundamental Variable	-0.81	**-0.89**	-0.80	-0.72	0.29	-0.80	-0.01	-0.71	-0.05	-0.41	0.93	1.00	-0.59	-0.56	-0.50	-0.67	-0.60	-0.48	-0.56	-0.73	-0.50	-0.46	0.28	-0.26	0.02	0.04
Long non-commercial Futures	0.51	0.60	0.90	0.64	0.05	0.89	-0.43	0.93	0.71	0.85	-0.60	-0.59	1.00	0.67	0.96	0.97	1.00	0.66	0.98	0.85	0.94	0.93	-0.13	0.10	-0.10	-0.05
Short non-commercial Futures	0.53	0.64	0.73	0.76	-0.13	0.74	-0.14	0.68	0.57	0.81	-0.56	-0.56	0.67	1.00	0.44	0.73	0.68	0.96	0.52	0.83	0.49	0.47	-0.35	0.32	0.14	0.10
Net non-commercial Futures	0.41	0.49	0.81	0.49	0.11	0.79	-0.47	0.87	0.65	0.73	-0.51	-0.50	0.96	0.44	1.00	0.90	0.95	0.44	0.99	0.72	0.95	0.95	-0.03	0.00	-0.18	-0.10
Open Interest Futures	0.61	0.71	0.92	0.70	0.02	0.92	-0.36	0.95	0.65	0.85	-0.68	-0.67	0.97	0.73	0.90	1.00	0.97	0.70	0.94	0.93	0.86	0.84	-0.21	0.19	-0.09	-0.05
Long non-commercial F&O	0.53	0.62	0.90	0.66	0.04	0.89	-0.42	0.93	0.70	0.85	-0.61	-0.60	1.00	0.68	0.95	0.97	1.00	0.67	0.98	0.86	0.94	0.92	-0.15	0.12	-0.11	-0.06
Short non-commercial F&O	0.46	0.56	0.70	0.73	-0.16	0.70	-0.18	0.65	0.61	0.81	-0.47	-0.48	0.66	0.96	0.44	0.70	0.67	1.00	0.49	0.77	0.51	0.47	-0.30	0.27	0.10	0.09
Net non-commercial F&O	0.48	0.56	0.85	0.56	0.09	0.84	-0.45	0.91	0.64	0.76	-0.58	-0.56	0.98	0.52	0.99	0.94	0.98	0.49	1.00	0.79	0.95	0.95	-0.08	0.06	-0.16	-0.10
Open Interest F&O	0.73	0.81	0.87	0.77	-0.01	0.87	-0.29	0.86	0.54	0.81	-0.71	-0.73	0.85	0.83	0.72	0.93	0.86	0.77	0.79	1.00	0.70	0.65	-0.40	0.38	0.03	0.03
Financial Variable Futures	0.44	**0.51**	0.80	0.53	0.08	0.78	-0.50	0.85	0.65	0.75	-0.51	-0.50	0.94	0.49	0.95	0.86	0.94	0.51	0.95	0.70	1.00	0.97	-0.10	0.06	-0.18	-0.11
Financial Variable F&O	0.36	0.45	0.78	0.48	0.09	0.77	-0.46	0.85	0.66	0.74	-0.49	-0.46	0.93	0.47	0.95	0.84	0.92	0.47	0.95	0.65	0.97	1.00	-0.03	0.00	-0.18	-0.12
DXY	-0.57	**-0.52**	-0.11	-0.38	0.18	-0.14	0.06	-0.09	0.16	-0.16	0.27	0.28	-0.13	-0.35	-0.03	-0.21	-0.15	-0.30	-0.08	-0.40	-0.10	-0.03	1.00	-0.98	0.26	0.26
USD/EUR	0.55	0.49	0.08	0.35	-0.16	0.10	-0.07	0.07	-0.17	0.13	-0.24	-0.26	0.10	0.32	0.00	0.19	0.12	0.27	0.06	0.38	0.06	0.00	-0.98	1.00	-0.23	-0.23
Implied Volatility	-0.23	-0.18	0.00	-0.08	0.08	-0.06	-0.06	-0.01	0.15	0.07	0.13	0.02	-0.10	0.14	-0.18	-0.09	-0.11	0.10	-0.16	0.03	-0.18	-0.18	0.26	-0.23	1.00	0.84
Realized Volatility	-0.24	**-0.21**	0.02	-0.06	-0.01	-0.06	-0.13	0.00	0.14	0.05	0.14	0.04	-0.05	0.10	-0.10	-0.05	-0.06	0.09	-0.10	0.03	-0.11	-0.12	0.26	-0.23	0.84	1.00

Note: This table shows the correlation matrix for every time series initially considered for building the final input (predictive) variables.

a) **Balance in the physical market (FUN):**

In the case of the Fundamental data, many variables have been explored. Some of them, like OPEC's spare capacity, provide relevant information, but the relationship is not clear in time. For instance, a negotiated reduction of OPEC output (looking for higher prices) will increase OPEC spare capacity. This type of policy arises when the market is oversupplied (a signal of lower prices). However, it could take time to comply with quotas, so it is unclear when the physical balance will adjust.

In conclusion, supply, demand, and stock data will be considered to define the Fundamental Variable. The difference between supply and demand should be reflected in the inventories. It is worth mentioning that the stocks help to cope with unexpected supply changes, so oil inventories' days of supply could be the first proxy.

Baumeister and Kilian (2015) include the percentage change in global oil production, the change in global crude oil inventories, and global economic activity, among other factors. Similar ratios as proxies of fundamentals are considered in Figuerola-Ferretti et al. (2015) in their study of bubble characteristics of non-ferrous metals. They define the consumption-supply ratio (CSR) as a measure of market fundamentals. An enhancement of this measure is to incorporate production into inventories, obtaining all the available oil to meet demand. Ideally, we should have data on all the stocks worldwide, including tankers for storage or Chinese oil storage facilities. Actually, this value is based only on the OCDE inventories, the global benchmark reference. We propose to create a ratio in which the numerator will be the sum of 30 days of world oil production (moving average twelve months (MAV)) with OCDE inventories (moving average six months), and the denominator will be 30 days of world oil consumption (moving average twelve months). This fundamental variable (FUN) contains as much essential information as possible, should reflect the physical market balance, and is specified under equation (8). We can see from Table 12 that it exhibits an inversely proportional relationship with Brent crude oil with a correlation equal to -0.81.

$$Fun = \frac{MAV\left(6M(Commercial\ OECD\ Stocks)\right) + 30 \cdot MAV(12\quad(World\ Supply))}{30 \cdot MAV(12M(World\ Demand))} \tag{8}$$

b) **Speculation in the crude oil market (FIN):**

The variable used to capture the speculative activity and investors' sentiment concerning oil prices is constructed with CFTC data. This requires the definition of the following input ratios. Open Interest is the total amount of futures and/or option contracts that remain open overnight (and thus not offset by a transaction, delivery, or exercise). Note that all long open interest aggregate equals short open interest. Secondly, we use "commercial" or "non-commercial" CFTC classifications and define a "net non-commercial ratio" that considers net (long minus short) "non-commercial" positions in the numerator and total Open Interest in the denominator. The objective is to provide a metric gauging the direction of the market sentiment as "non-commercial" positions are defined as trades not designed for hedging purposes. The second measure is the sum of long and short "non-commercial" positions divided by the total open interest. This aims to provide the magnitude or impact of investors (or speculators) taking oil market positions. The proposed financial variable (FIN) is the product of two ratios. Note that this metric is related to Working's T-index, which has been used as a futures speculation proxy by Figuerola et al. (2020) in the crude oil price case by Haase et al. (2019) for multiple commodity markets and for food commodities. See also Figuerola-Ferreti et al. (2015) and Etienne et al. (2015) for the non-ferrous and agricultural market cases, respectively. While the FIN variable correlates with the Working's T index, it better fits the proposed forecasting model and is more closely related to the speculation-related measures used in the crude oil forecasting literature Chai et al. (2018). The underlying presumption is that a high (low) level of speculation will encourage higher (lower) prices, as shown by a correlation coefficient between the FIN variable and the crude oil Brent price, which is reported to be 0.44 in Table 12. The financial variable is therefore defined as:

$$Fin = \frac{Net\ Long\ NonCommercial\ Positions}{Open\ Interest} \cdot \frac{Total\ NonCommercial\ Positions}{Open\ Interest} \tag{9}$$

c) **U.S. Dollar (DXY):**

The U.S. Dollar is the numeraire in most oil contracts quoted in U.S. dollars. We use the DXY index to address the effect of the U.S. Dollar on the oil price. As underlined, changes in the exchange rate can be translated into changes in oil consumption for oil-importing countries and non-US-based investors. The dollar index (as well as the euro-dollar exchange rate) is considered by Chai et al. (2018) in a recent oil forecasting exercise. Table 12 shows that the correlation coefficient of the DXY index and the log of the Brent price is -0.57.

d) **Realized Volatility (VOL):**

We follow and use a metric of uncertainty related to the crude oil market. Specifically, the realized volatility of Brent front-month futures prices is used. Volatility is often related to market risks and, therefore, has a negative impact on the price of oil. As reported in Table 12, the correlation coefficient of realized volatility with the oil price equals -0.24.

3.3.3 Descriptive Statistics

Table 13 summarizes the series selected to construct the final variables, including data sources. Table 14 shows the correlations across the log of the Brent price and the main variables selected by the algorithm. Results show that reported correlations between explanatory variables remain below 0.55, suggesting that the model will not suffer from multicollinearity problems.

Tabla 13: Description of selected input variables

Model Variable (Equation)	Raw Variable	Frequency	History	Source
Fundamental Variable (8)	Total Crude Oil Supply (World) Total Crude Oil Demand (World) Total Commercial OECD Stocks	Monthly	from Jan 1995	U.S. Energy Information Administration
Financial Variable (9)	Non-Commercial Long Futures WTI Non-Commercial Short Futures WTI Open Interest Futures WTI	Monthly (average weekly data)	from Jan 1995	Commodity Futures Trading Commission
Volatility Realized	Price First Brent Contract	Monthly (average daily data)	from Jan 1995	Price
Dollar	DXY Index	Monthly (average daily data)	from Jan 1995	Bloomberg

Note: This table describes the input data for GAM model implementation. The third to fifth columns provide variable frequency, data history, and data source. The label "Model variable" in the first column describes the category of the given data series within the Fundamental, Financial, dollar, and volatility according to dimensions for model input variables. The fundamental and financial variable definitions are linked to definitions specified in equations 8 and 9, respectively.

Tabla 14. Correlation matrix of selected predictive variables and the target variable

	Log(Brent)	Fundamental Variable	Financial Variable	Dollar	Realized Volatility
Log(Brent)	-	-0.89	0.51	-0.52	-0.21
Fundamental Variable	-0.89	-	-0.50	0.28	0.04
Financial Variable	0.51	-0.50	-	-0.10	-0.11
Dollar	-0.52	0.28	-0.10	-	0.26
Realized Volatility	-0.21	0.04	-0.11	0.26	-

Note: This table reports correlation coefficients across the variables selected as final input variables. Correlations with the output variable defined as the log of the Brent spot price are also reported.

Table 15 in the main text reports descriptive statistics of the selected input variables and the output or forecasted variable, which is the log of the Brent spot price labeled as log(Brent). Estimates are based on a sample of monthly data ranging from January 1995 to December 2023 (358 observations). We can see that the Brent spot price level exhibits the highest standard deviation and maximum level.

Table 15. Summary statistics

Variables	n	Mean	Median	Std	Skew	Kurtosis	Min	Max
Brent	348	58.18	56.81	32.28	0.353	-0.972	10.19	133.81
log(Brent)	348	1.68	1.75	0.28	-0.442	-0.947	1.01	2.13
Fundamental	348	2.05	2.03	0.09	0.673	-0.336	1.88	2.28
Financial	348	0.03	0.02	0.03	0.476	-0.869	-0.03	0.11
Volatility	348	0.32	0.30	0.16	2.788	14.452	0.08	1.54
DXY	348	92.29	92.83	10.69	0.366	-0.428	72.08	119.04

Note: This table reports summary statistics of the Brent spot price, the log of the spot Brent price (log(Brent)), and the selected variables used in the forecasting exercise. The table shows mean, median, standard deviation (Std), skew, kurtosis, minimum (Min), and maximum (Max) variable values.

Normality and unit root test results are reported in Table 16. Results of the Jarque-B test and Ljung-Box show that the null hypothesis of normality and white noise errors is rejected for all variables considered. This table also reports results for the Augmented Dicky-Fuller (ADF) (1981), the Phillip-Peron (P.P.) (1988), and Kwiatkow-ski-Phillips-Schmidt-Shin (KPSS) (1992) unit root test. Reported results show that the unit root hypothesis is accepted for all variables except the volatility variable (Vol).

Table 16. Normality and unit root test results.

Variables	Jarque-B	Ljung-Box	ADF	PP	KPSS
Brent	21.06 ***	338.26 ***	-2.85	-2.51	1.04 ***
log(Brent)	24.39 ***	340.21 ***	-2.51	-2.22	1.38 ***
Fun	28.42 ***	338.99 ***	-3.48 **	-2.8	0.89 ***
Fin	24.21 ***	322.99 ***	-3.19 *	-3.99 ***	1.04 ***
Vol	3489.10 ***	129.37 ***	-9.15 ***	-9.07 ***	0.13
DXY	10.49 ***	339.84 ***	-1.77	-1.6	1.09 ***

Note: This table provides normality and unit root test results. ***, **, and * denote rejection of the null hypothesis at the 1, 5, and 10 percent levels, respectively.

Table 17. Bai and Perron structural breaks test results for log(Brent)

Sequential F-statistic determined breaks: **5**

Break test	F-statistics	Scaled F-statistic	Critical Value**	Break dates:	Dates:
0 vs 1 *	1080.393	1080.393	8.58	1	Sep-99
1 vs 2 *	75.975	75.975	10.13	2	Oct-04
2 vs 3 *	55.274	55.274	11.14	3	Aug-10
3 vs 4 *	112.946	112.946	11.83	4	Nov-14
4 vs 5 *	23.992	23.992	12.25	5	Apr-19

Note: This table provides the results of the structural breaks test. The test report results for the null hypothesis H0 of no structural break. The alternative hypothesis H 1 test for k structural breaks). There are five structural breaks. * Significant at the 0.05 level, **Bai-Perron critical values are used.

The Bai and Perron test (2003) for detecting multiple structural changes has also been performed for the logarithm of Brent spot price and the regression with the selected input and explained variables in differences. Results are reported in Tables 17 and 18, respectively. They show that the log of Brent prices exhibits five breakpoints along our sample period. When we run the regression in differences (with log Brent as the explained variable and the changes in the fundamental, financial dollar, and volatility as input variables), the reported results do not show evidence of structural breaks. The fact that structural breaks are no longer reported for the regression in differences shows that the input variables have been appropriately selected.

Table 18. Bai and Perron's structural breaks test for equations in differences

Sequential F-statistic determined breaks: **0**

Break test	F-statistics	Scaled F-statistic	Critical Value**
0 vs 1	2.321	11.605	18.23

Note: This table provides the results for the structural breaks test for an equation that estimates changes in log(Brent) as a function of the differences in the fundamental, financial, volatility, and dollar variables. The test report results for the null hypothesis H0 of no structural break. The alternative hypothesis H 1 tests for k structural breaks). There are no structural breaks. * Significant at the 0.05 level, **Bai-Perron critical values are used.

Once we have developed and defined our variables, the next step involves establishing the methodology and constructing the model. This process entails creating a theoretical framework and carefully selecting appropriate analytical tools and techniques to address our research problem effectively.

Chapter 4

Methodology proposed and Model identification

4.1 Introduction

This section describes the two-step method proposed to model the oil price. The aim of this model is to create monthly forecasts for up to 12 months. The data for building the model covers the Jan 1995 to June 2023 period and aims to forecast the monthly crude oil Brent price series as the current global crude oil price benchmark. The in-sample period runs up to December 2016. This selection makes the in-sample size comparable to the recent literature. Baumeister and Kilian (2015) use an in-sample period ranging from 1997:12 to 2010:6. The variable that will be modeled is the logarithmic monthly average price of Brent's prompt contract. As non-linear relationships between the explanatory variables and oil price have been detected, the Generalized Additive Model is the approach selected. In order to guarantee uncorrelated residuals and no cross-correlation between the residuals and the regressors, the LTF method with ARIMA noise is applied. Once we have decided on the methodology to apply, we will proceed with model identification using actual data and refine the proposed final version. During this process, we must ensure that the model accurately and comprehensively captures the relationships between variables and Brent.

4.2 Modeling approach for oil price forecasting

Generalized Additive Models (GAM) offer a general framework for extending a standard linear model by allowing non-linear functions of each variable while maintaining additivity. They offer a natural way to extend the multiple regression model to allow for non-linear relationships between each explanatory variable (feature) and the explained variable (response variable). The smooth functions are used as a replacement for the alternative detailed parametric relationship on the covariates. Moreover, this methodology is appropriate for the monthly data required in this study due to the low-frequency availability of oil fundamental data. The GAM methodology supersedes competing machine learning algorithms, such as neural networks, when large volumes of data are unavailable. It is also a preferred method because it allows a straightforward interpretation of results. This method calculates the sensibilities of the forecasted variable concerning changes in input values, allowing a deeper understanding of underlying relationships than under competing Machine Learning Models.

In essence, a Generalized Additive Model (GAM) is a Generalized Linear Model (GLM) in which the linear predictor is given by a sum of smooth non-linear functions of at least some (or possibly all) covariates, as explained in Wood et al. (2015). The family of smooth functions is defined as the basis functions. The logarithmic function or a polynomial (cubic spline) are good examples of this specification class. Each basis function transforms the vector of explanatory variables x in terms of the type of basis considered.

The GAM can be formally expressed as:

$$y_t = \beta_0 + \sum_{i=1}^{n} f_i(x_{i,t}) + \varepsilon_t \tag{10}$$

Where i=1, …, n, x_i are the n independent input variables, f_i are unknown non-parametric smooth functions of x_i, and ε_t is a i.i.d random error. This structure captures the non-linear relationships while providing a flexible framework for understanding the (linear or non-linear impact) of every variable considered.

We impose restrictions on the number of smooth functions allowed in the framework to prevent problems related to overfitting. For this reason, the specified models are usually fit by penalized likelihood maximization, and each penalty is multiplied by an associated smoothing parameter to control the balance between over and underfitting. The MGCV implementation of GAM in R is applied. This module characterizes the smooth functions using penalized regression splines with smoothing parameters selected by Restricted Maximum Likelihood (REML).

In order to make the reported method robust to the existence of residual autocorrelation and dynamic causal effects, we consider a Linear Transfer Function (LTF) with ARIMA noise Box et al. (1994) for the variables transformed by the GAM model $x'_{i,t} = f(x_{i,t})$.

Assume the series, y_t, and $x_{1,t}$, …, $x_{n,t}$ are stationary variables. The classical multiple linear regression model is given by:

$$y_t = c + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_n x_{n,t} + \varepsilon_t \tag{11}$$

assumes that the system's noise ε_t is white noise and uncorrelated with the explanatory variables. In order to guarantee uncorrelated residuals and no cross-correlation between the residuals and the regressors, the LTF method with ARIMA noise, introduced by Liu and Hanssens (1982), is applied. The dependent variable is modeled as a function of its past values and lagged values of the explanatory variables. The following specification is used for this purpose:

$$y_t = c + \frac{\omega(L)}{\delta(L)} x'_{i,t-b} + v_t \tag{12}$$

$$\omega(L) = (\omega_0 - \omega_1 L - \omega_2 L^2 - \cdots - \omega_s L^s) \tag{13}$$

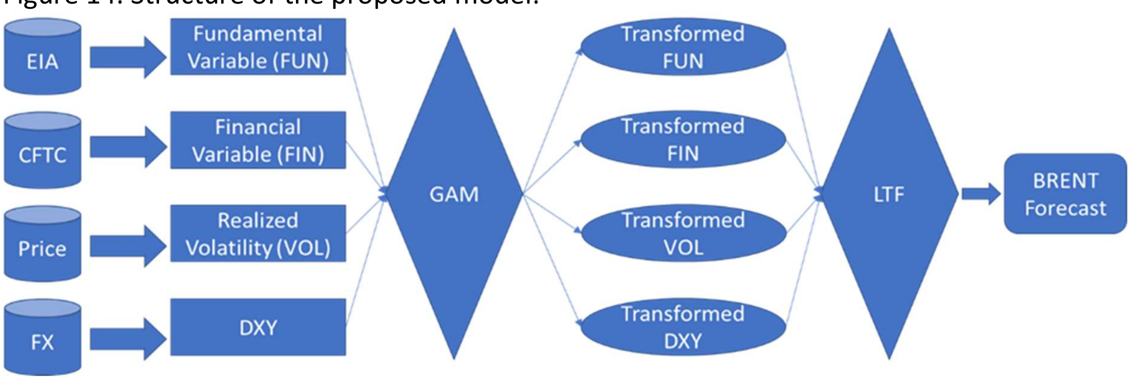
$$\delta(L) = (1 - \delta_1 L - \delta_2 L^2 - \cdots - \omega_s L^r) \tag{14}$$

$$v_t = \frac{(1-\theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)}{(1-\varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p)(1-L)^d} \varepsilon_t \tag{15}$$

where y_t is the dependent output variable at time t, $x_{i,t}$ represents the i-th independent or explanatory input variables, v_t: is an autocorrelated ARIMA(p,d,q) noise, r, s, b constant integers, and ω(L) and δ(L) are lagged polynomials and εt is white noise and $x'_{i,t} = f(x_{i,t})$ are the input variables transformed by the GAM model.

Figure 14 illustrates the complete process of the proposed methodology to forecast oil prices. The starting point is the data obtained from multiple sources, such as the EIA or the Commodity Futures Exchange Commission (CFTC). The data is then used to build four variables (FUN, FIN, VOL, and DXY), which are transformed through a GAM model into the final input variables used by the Linear Transfer Function Model.

Figure 14. Structure of the proposed model:



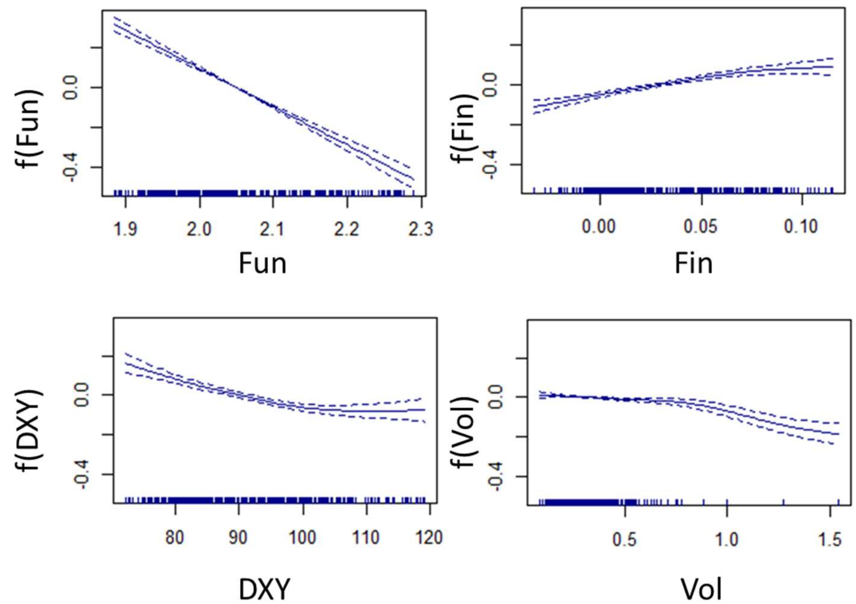
Note: This figure exhibits the structure of the proposed modeling approach: Step 1 involves raw data extractions; step 2 requires the creation of featured variables; step 3 transforms through a GAM model into the final variables used by the Linear Transfer Function model in step 4 to create the Brent Forecast.

4.3 The Proposed Model

We construct the model by fitting a Generalized Additive Model (GAM) using the four variables defined in the previous chapter: Fundamental, Financial, Volatility, and Dollar. This initial step establishes a foundation for the subsequent modeling process, providing insights into the relationships among these key variables and the oil price. The tool used is the MGCV implementation of GAM in R (Wood et al. 2015), which characterizes the smooth functions using penalized regression splines with smoothing parameters selected by Restricted Maximum Likelihood (REML).

Figure 15 illustrates the partial effects obtained with the GAM model of the transformed variables on the oil price. For instance, the top left-hand side (LHS) panel of Figure 15 illustrates the fundamental variable on the x-axis and the transformed variable on the y-axis, indicating the effect of the fundamental on the oil price metric. The dotted lines illustrate 5% confidence intervals. Model results show non-linearities in every variable considered except for the fundamental metric. This is corroborated by the Effective Degrees of Freedom (EDF) reported in column 2 of Table 19, which measures the degree of non-linearities within a given curve. Note that when reported EDFs are equal to one, as is the case for the fundamental variable, this implies that the curve can be accepted as linear. The volatility variable depicted in the bottom right-hand side (RHS) panel exhibits the highest level of non-linearity, followed by the dollar in the bottom LHS panel and the financial metric in the top RHS panel.

Figure 15. Partial Effects illustration under the GAM model:



Note: This fig. illustrates the non-linear relationship between the variables considered and the oil price under the GAM estimation. The row input variables (represented by dots on the horizontal axis) are transformed using the basis functions (denoted by f()). The transformed variables are introduced in the LTF model at a later stage. The effects of the fundamental and financial variables are illustrated under the top left-hand side (LHS) and right-hand side (LHS) panels. The dollar and volatility variables are depicted under the bottom LHS and RHS panels. 95% confidence intervals are depicted as dotted lines.

Table 19. Summary of estimated Coefficients under a GAM Model specification

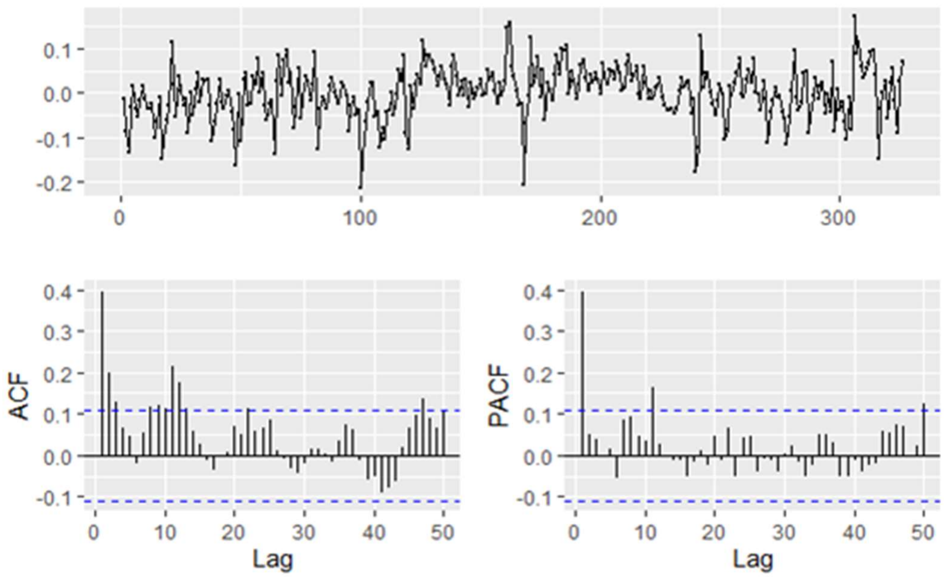
Approximate significance of smooth terms:					
Variable:	**Edf**	**Ref Edf**	**F**	**p-value**	
Fundamental	1.00	1.00	293.16	<2e-16	***
Financial	3.015	3.985	17.20	7.5e-13	***
Volatility	4.385	5.471	13.34	1.57e-12	***
Dollar	3.713	4.900	14.34	2.11e-12	***
Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1'' 1					
R-sq- (adj) = 0.883			Deviance explained = 88.8%		
fREML = -637.32			Scale est. = 0.0035994		
Box-Pierce test = 294.28, df = 1, p-value < 2.2e-16					

Note: This table reports estimates of the GAM model specification. EDF: reflects the degree of non-linearity of a curve. An EDF equal to 1 is equivalent to a linear relationship. P-values represent calculated p-values from the Wald test (significance of each parametric and smooth term of the model)

In what follows, we interpret the plots in Fig 15, illustrating the partial effects for every explanatory variable. Note that the four signs of the function slopes are aligned with the correlation coefficient calculated with the oil prices, reported in Table 14. The top LHS panel in Fig 15 shows that the Fundamental Variable, which takes a low value under fundamental shortages, exhibits a well-fitted negative linear relationship with the oil Brent price, showing that excess supply market conditions lead to lower oil prices. While the effect of the Financial Variable on the oil price is almost linear, we can see that the financial variable presents some non-linear features. Specifically, we can see that the slope is slightly smoothed when the market sentiment becomes bullish so that positive investor bets outnumber the negative counterparts. The bottom LHS panel in Fig 15 shows a non-linear inverse relationship between the dollar index and oil benchmark that significantly smoothed when the index exceeds 105. The negative influence of the Dollar value on oil prices has been widely documented in the literature by Beckmann et al. (2020) and Miao et al. (2017). The relationship estimated implies that Brent prices increase under low dollar conditions. A lower dollar leads to higher demand and higher prices as producers try to protect the dollar-adjusted value of their revenues. Oil becomes relatively cheap for foreign investors, and that increases demand. However, results illustrated in the LHS panel of Fig 15 suggest that the dollar's impact on crude oil prices is lower when the dollar is under stronger conditions. Results depicted in the bottom RHS panel in Fig 15 show the effect of volatility, which is highly significant under high volatility regimes and negatively affects prices. Episodes of extreme volatility (such as that seen during the 2014 oil price shock) are expected to decrease the oil price while the volatility effect is reduced under normal market conditions. In fact, we can see that when the volatility is below 40%, it exhibits a reduced impact on oil prices. The existence of volatility-driven regime changes has been considered in the forecasting literature by Miao et al. (2017), who document a "volatility upward regime" via the TVIP-MRS model and forecast the crude oil price.

The preliminary estimation results reported in Table 19 show that the adjusted R2 and the deviance explained demonstrate that the model fits the data correctly. The Box-Pierce test suggests that there is residual autocorrelation. Details can be found in Figure 16.

Figure 16. PACF and ACF structure of the error term under a GAM Model



Note: This figure shows the PACF and ACF for the initial GAM model error. A clear autocorrelation can be observed in the regression errors since some lags exceed confidence limits.

In order to correct this autocorrelation, we apply a Linear Transfer Function model with ARIMA noise in a second step, using the four GAM-transformed variables as inputs. We estimate the LTF specification using the identification, estimation, and diagnosis procedure proposed in Pankratz (1991), following a similar approach to constructing the univariate Box-Jenkins ARIMA model explained in Box et al. (1994). The identification requires fitting a multiple regression model, adding as many lags of the regressors as required, and a low-order Autoregressive model for the error term to capture most of the autocorrelation and be able to estimate the impulse response. If regression errors are not stationary, variables are differentiated. The next stage is identifying the transfer function and selecting the appropriate values for b, r, and s. We can identify the orders (b, r, s) by visually comparing the estimated impulse response function with some standard theoretical functions. Then, the ARMA model for the regression errors must be determined to fit the complete model. Finally, several diagnostic tests are applied to determine the model selection model based on resulting cross-correlation and autocorrelation tests.

Explanatory variables are determined using the previously estimated GAM process. The final model identification suggests an ARIMA (1,1,0) for the residuals. Estimation results are reported in Table 20. We can see that the four independent variables are statistically significant, and the residuals do not exhibit serial correlation, with Box-Pierce failing to reject that residuals are independently distributed. The Partial autocorrelation function (PACF) and Auto-correlation function (ACF) confirm the absence of serial correlation (see Fig 17). Note that results reported for the regression in differences under the Bai and Perron test (see Table 18) show that we failed to reject the null hypothesis of no structural breaks. This confirms that the LTF model can be applied to the residuals.

Table 20. Summary of estimated Coefficients under final specification:

Approximate significance of smooth terms:					
Variable:	**Estimate**	**Std Error**	**Z Value**	**p-value**	
ar	0.268	0.057	4.741	2.12e-06	***
f(Fundamental)	0.51	0.110	4.654	<2e-16	***
f(Financial)	1.125	0.106	10.622	<2.11e-12	***
f(Volatility)	0.882	0.097	9.088	<7.5e-13	***
f(Dollar)	0.510	0.137	3.710	<1.57e-12	***
Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1''					
Box-Pierce test = 0.000040065, df = 1, p-value = 0.984					

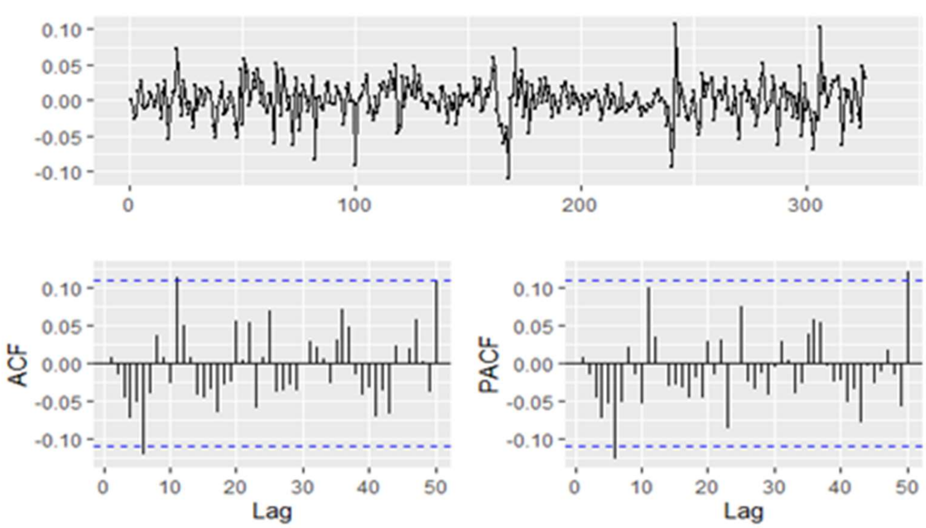
Note: This table reports estimated coefficient estimates corresponding to specifications (3) and (4). Columns 2-6 report estimated coefficients, standard errors, z statistics, and p-values.

The final equation of the complete model will be:

$$y_t = \omega_{1,0}x'_{1,t} + \omega_{2,0}x'_{2,t} + \omega_{3,0}x'_{3,t} + \omega_{4,0}x'_{4,t} + \frac{\varepsilon_t}{(1-\varphi L)(1-L)} \quad (16)$$

where $x'_{1,t} = f_1(x_{1,t})$; $x'_{2,t} = f_2(x_{2,t})$; $x'_{3,t} = f_3(x_{3,t})$; $x'_{4,t} = f_4(x_{4,t})$ are the variables transformed by the GAM model (see Figure 2) and ε_t is a white noise.

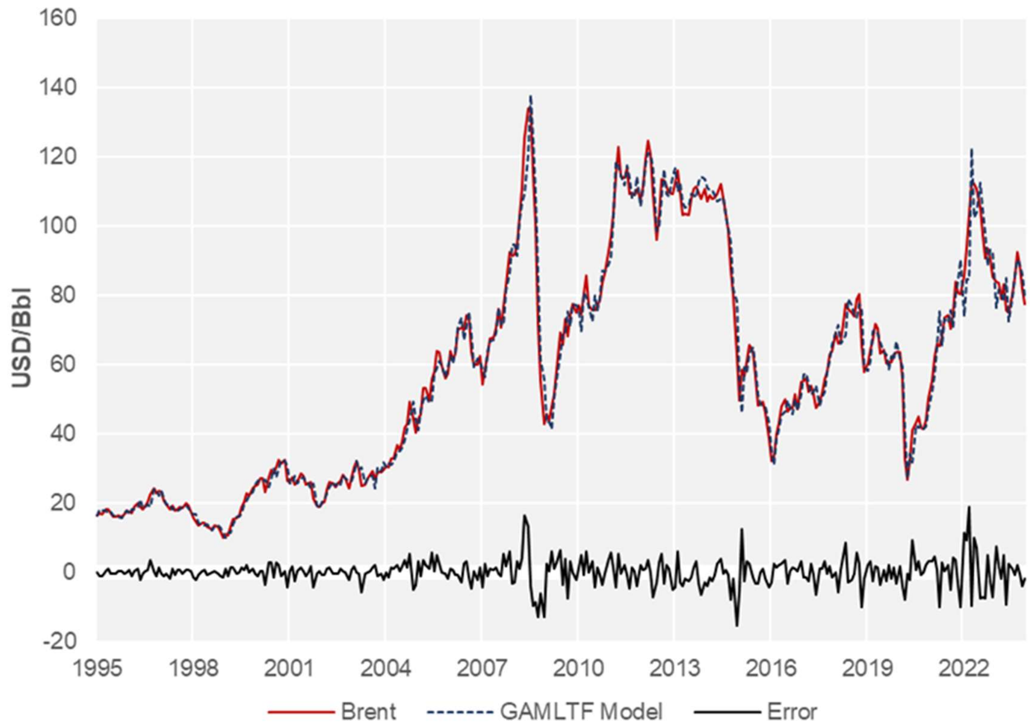
Figure 17. PACF and ACF structure of the error term under GAM with LTF Specification



Note: This figure shows the PACF and ACF for the proposed model error. No problem of autocorrelation can be appreciated.

Figure 18 depicts the one-month-ahead forecast of the Brent crude oil price under the proposed model versus the observed Brent price, and the error is defined as the difference between estimated and observed values. A closer look at the figure shows that the goodness of fit is high but deteriorates in times of increased uncertainty, such as during the 2008 crisis, the 2014 crude oil price collapse, or the 2020 COVID crisis.

Figure 18. Evolution of the Brent, the forecasted Brent, and the Model error:



Note: This figure illustrates the time series evolution of the observed price (Brent), the estimated Brent spot price (model), and the forecast error. The GAM model is estimated using the selected input values.

Chapter 5

Forecasting Results and Sensitivity Analysis

5.1 Introduction

This section describes the proposed model's forecasting results and sensitivity analysis. A extended model that includes the geopolitical risk index and financial stress index instead of volatility has also been created. The model is tested for the 2017-2023 out-of-sample period. Note that this constitutes seven years of monthly data leading to 82 observations. While this out-of-sample window may be considered short compared to other benchmark analyses Baumeister and Kilian (2015) and Garratt et al. (2018), recent research in the literature considering the existence of non-linearities Miao et al. (2017) has used shorter out-of-sample periods. Specifically, they evaluate the out-of-sample forecast performance for the 2009:M5 to 2016:M11 period. We, therefore, follow the recent literature that addresses the sources of non-linearities and use shorter out-of-sample periods in our forecasting exercise. The sources of recent non-linearities include the collapse of the 2014-2016 crude oil price, the 2020 COVID-19 pandemic shock, the ongoing war in Ukraine, and a shift to green energy. Forecasting performance is measured in terms of MAPE values and the absolute ratios of MAPE concerning the no change. RMSE is also computed in the principal analysis as a means of robustness. The same out-of-sample period is considered for the sensitivity analysis.

5.2 Forecasting Results

In what follows, we quantify the predictive performance of the proposed model specification. The analysis takes the 2017Q1 to 2023Q4 time frame for the out-of-sample test (21% of data). A forecast for different quarters within a window of 12 months (4 quarters) is made at the beginning of every quarter. Data for the last seven years of the sample have been used to compare model performance with four forecasting methods. This implies that there are 25 quarterly forecasting periods. The average monthly forecast for each quarter is considered, and the Mean Absolute Percentage Error (MAPE) for each method considered is reported in Table 5. Note that this period represents the recovery from the 2014-2016 price slump and the COVID-induced crude oil price collapse. As discussed in the introduction, crude oil prices have experienced many different price swings over the forecasting period. Therefore, we believe it is essential to provide an appropriate testing framework to account for the observed non-linearities in the data.

We benchmark the proposed model against the no-change or spot reference price. This uses the last available monthly spot price observation. The no-change forecast is set as the spot price under the previous month of the forecast during the whole forecast period. Next, we consider the forecasting performance of the Intercontinental Exchange (ICE) Brent futures prices. This price aggregates expectations for future price delivery across market participants. The benchmark built based on futures prices takes the average of the 1st, 2nd, and 3rd month generic future contracts (Brent) for the first quarter forecast and the average of the 4th, 5th, and 6th month contracts for the second quarter forecast. The same method is applied to forecast prices in subsequent quarters. The benchmark is constructed the day before the forecast period begins, and as previously specified, the data source for the price of the futures prices is Bloomberg.

As an alternative analysts' forecast benchmark, we first use the monthly forecast of the Department of Energy of the U.S. (EIA or DoE) released under the Short-Term Energy Outlook every month. This report calculates monthly Brent price forecasts for maturities ranging between 1, 12, or 24 months. We construct quarterly forecasts by calculating three-month averages using the last report before the start of the forecast period.

The second benchmark source of analysts' forecasts is the prediction provided by the Bloomberg survey with crude oil analyst forecasts (BBG). This offers industry experts price forecasts for different maturities. The median forecast for each quarter reported in this survey is taken as forecasts the day before the forecast period starts. See Figuerola-Ferretti et al. (2020) for a detailed description of the Bloomberg analysts' forecast survey.

We report forecasts for the GAMLTF with forecasted EIA fundamental inputs as well as from actual input data. To measure the contribution of the GAM framework, we report a forecast for the LTF with no GAM. We select the Mean Absolute Percentage Error (MAPE) as a metric for evaluating the performance of the forecasting methods is defined as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t-\tilde{y}_t}{y_t}\right| \tag{17}$$

The choice of the MAPE metric is motivated by the high oil price variability during the sample period considered. Oil prices range between \$30 and \$140, implying that absolute differences in high-price states will be difficult to compare with absolute differences in low-price states. However, the RMSE metric is also included in the main forecasting analysis as a means of robustness.

The forecasting performance of a model with exogenous variables will depend on the forecast accuracy of the future values of the selected regressors. For that reason, we also test under two explanatory variables' predictions. In the Real Data Model, the observed values of the future selected explanatory variables are used for forecasting purposes. In the Forecast Data Model, every explanatory variable is forecasted. In this sense, we use forecasts of the Fundamental and U.S. Dollar variables from the EIA, available in its Short-term Energy Outlook, providing information for World Production, World Demand, and OECD inventories. Therefore, we incorporate forward-looking information (based on EIA predictions) into our forecast framework. ARMA models are estimated for the Financial and Volatility variables.

Table 21. MAPE error measures for different forecasting methods

	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2	2018Q3	2018Q4	2019Q1
Constant	6.19%	10.09%	24.62%	20.06%	10.14%	7.30%	10.47%	20.80%	9.96%
Futures	9.84%	10.29%	20.53%	19.21%	9.45%	5.91%	11.75%	23.31%	14.37%
BBG Analysts Median	6.50%	8.77%	12.08%	20.77%	19.48%	8.73%	3.96%	13.30%	9.48%
Department of Energy EIA	5.72%	9.81%	20.50%	22.63%	16.15%	12.05%	4.66%	16.99%	5.68%
GAMLTF Forecasted Inputs	**5.78%**	**9.75%**	**18.72%**	**15.58%**	**14.05%**	**7.10%**	**8.49%**	**21.33%**	**6.41%**
GAMLTF Actual Inputs	**16.11%**	**6.15%**	**8.86%**	**4.80%**	**5.78%**	**11.45%**	**3.86%**	**11.30%**	**7.72%**
LTF Actual Inputs no GAM	17.65%	6.32%	8.13%	9.07%	9.27%	22.79%	3.65%	4.13%	7.09%

	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2
Constant	9.83%	29.17%	34.49%	54.64%	24.77%	22.29%	30.57%	28.53%	15.94%
Futures	10.87%	29.24%	29.81%	48.13%	20.86%	19.84%	26.89%	27.36%	22.64%
BBG Analysts Median	16.04%	39.54%	35.67%	42.36%	10.02%	25.24%	26.14%	33.24%	23.26%
Department of Energy EIA	7.38%	37.00%	32.62%	43.62%	32.28%	16.61%	25.21%	24.85%	20.14%
GAMLTF Forecasted Inputs	**10.39%**	**30.62%**	**42.76%**	**42.15%**	**29.97%**	**23.36%**	**10.80%**	**7.59%**	**13.10%**
GAMLTF Actual Inputs	**5.86%**	**9.83%**	**16.25%**	**12.47%**	**19.46%**	**17.67%**	**12.96%**	**5.95%**	**3.79%**
LTF Actual Inputs no GAM	14.35%	63.13%	93.00%	94.80%	27.15%	27.26%	17.22%	8.85%	13.11%

	2021Q3	2021Q4	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1	**TOTAL**
Constant	16.66%	21.17%	22.66%	20.27%	36.84%	8.48%	3.10%	**19.96%**
Futures	19.15%	21.30%	23.45%	6.74%	13.33%	5.29%	4.55%	**18.16%**
BBG Analysts Median	23.92%	26.33%	22.15%	5.12%	16.60%	14.79%	10.73%	**18.97%**
Department of Energy EIA	19.95%	20.19%	23.10%	10.64%	13.21%	12.29%	3.86%	**18.29%**
GAMLTF Forecasted Inputs	**11.98%**	**9.50%**	**13.59%**	**32.05%**	**32.49%**	**8.98%**	**5.86%**	**17.30%**
GAMLTF Actual Inputs	**11.72%**	**14.91%**	**21.22%**	**26.12%**	**20.87%**	**9.29%**	**4.12%**	**11.54%**
LTF Actual Inputs no GAM	8.31%	8.27%	9.96%	43.39%	42.51%	11.08%	5.30%	**23.03%**

Note: This table reports the forecasting performance in terms of the MAPE measure of the proposed framework for forecasted and actual input data, as well as alternative benchmarks, including the LTF framework with no GAM. The in-sample period is 1995-2016, and the out-of-sample or forecasting period is 2017-2023. Forecasting is performed for the next four quarters. The following forecasting methods are considered:
No-change: Forecasts are the average price of the previous month for the whole forecast period.
Futures: Forecasts are the average of Brent 1st, 2nd and 3rd month contracts for the first quarter, 4th, 5th and 6th month contracts for the second quarter… the day before beginning the period of forecast
BBG: Bloomberg quarterly surveys are taken as forecasts the day before beginning the period of forecast
EIA: Average monthly forecasts to create quarterly forecasts are taken from the last EIA report before beginning the period of forecast
GAMLFT with Forecasted Inputs: proposed new model fed by forecasted inputs
GAMLFT with Actual Inputs: proposed new model fed by actual inputs
LFT with Actual Inputs no GAM: Linear Function Transfer model fed by actual inputs

The results reported in Table 21 show that the performance of each model varies over time. Average MAPE errors indicate that the best model is GAM-LTF with actual inputs followed by the GAM-LTF. However, a close look at the table shows that the no change and the futures forecast outperform in periods of high volatility, such as 2020Q3 and 2022Q2. Bloomberg analysts' forecasts perform worse on average than futures prices, consistent with previous results reported in the literature by Figuerola-Ferretti et al. (2020). However, it outperforms all the benchmarks considered during 2017Q3 and 2020Q1. Given that the best results at the average level are achieved when we know the variable data (GAMLTF Actual Inputs), we propose using the model for scenario analysis, as reported results suggest that it accurately captures the relationships between variables. This analysis is performed in the next chapter.

Table 1A in the appendix provides the same forecasting results under the RMSE measure. Reported figures are qualitatively similar to those reported in Table 21, suggesting that relative forecasting ability is not dependent on the forecast performance measure selected for the analysis.

Table 22. MAPE for different forecasting methods and horizons:

	No-change	Futures	BBG	Department of Energy EIA	**GAMLTF with Forecasted Inputs**	**GAMLTF with Actual Inputs**	LTF with Actual Inputs no GAM
1Q Forecast	8.1%	11.6%	10.2%	9.5%	7.7%	6.0%	8.2%
2Q Forecast	19.8%	21.3%	18.2%	19.5%	17.7%	11.0%	23.1%
3Q Forecast	24.8%	25.0%	22.6%	22.2%	21.1%	12.0%	31.3%
4Q Forecast	30.4%	28.8%	26.9%	27.2%	25.1%	12.5%	39.4%

Note: This table illustrates the model accuracy in terms of the MAPE measure with different fore-casting horizons ranging from Q1 to Q4

In order to provide a deeper analysis of the results we report, Table 22 provides forecast accuracy in terms of MAPE metrics for four maturities of the different models analyzed. The average forecast for each quarter is reported. For instance, if the forecast maturity is one quarter, in Q1 of 2016, the forecast for Q2 2016 is performed for each of the models considered and is used to calculate the average forecast for the Q2 Forecast period. Similarly, in Q2 of 2016, the forecast for Q3 2016 is performed for each of the models considered for the reported average for the Q3 Forecast. The same procedure is followed to calculate the forecast for longer horizons.

Our main findings can be summarized as follows: i) in line with the previous literature by Figuerola-Ferretti et al. (2020), forecast accuracy decreases with maturity. (ii) The best forecasting performance for all horizons considered is reported for the proposed model with actual values of input variables. Furthermore, the second-best performance is observed for the proposed model with forecasted input. This confirms that the proposed model can be used as an optimal tool for scenario analysis purposes (details will be provided in section 5). (iii) The introduction of the GAM specification in the model, considering the non-linearities in the input/output relationships between the explanatory variables and the oil price, is important to improving forecasting results, as can be seen by comparing the last column with columns 5 and 6. The forecast provided by the LTF approach with no GAM is less accurate than that provided under the GAMLTF with actual and forecasted inputs. (iv) The Model with forecasted variables (Forecast Data Model) improves the forecasting performance compared to other benchmarks for all quarters considered.

Table 23 reports the forecasting performance of the different models in terms of the MAPE metric in relation to the No Change forecast. The prediction horizon ranges from 1Q in the first panel to Q4 in the fourth. In this case, a moving window of six quarters is used to calculate the MAPE metric. The purpose is to quantify the evolution of predictive ability and robustness for the different models considered. Note that this requires changing the in-sample and out-of-sample periods for every calculation. For instance, the forecast estimates corresponding to 2019Q4 include 2019Q4, 2020Q1, 2020Q2, 2020Q3, 2020Q4, and 2021Q1. Therefore, the in-sample period covers the Jan 1995 to December 2019Q3 range. However, the forecast estimate corresponding to 2020Q1 calculates the average prediction for 2020Q1, 2020Q2, 2020Q3, 2020Q4, 2021Q1, and 2021Q2 and uses the 1995Q1-2019Q4 as an in-sample period. Unlike the results reported in Table 21, we provide forecasting results for every period of the out-of-sample window under each different quarter to analyse the persistence of the relative performance of the different methodologies considered. This is relevant given the high performance of regime-changing events seen during the 2017-2023 window, including the COVID crisis, the war in Ukraine, and the higher-than-expected recovery with high inflation and interest rate rises. Under this reporting format, the ratio takes a value of 1 if a given model performs equally well as the naïve (no change model). A close look at Table 23 shows that every model's forecasting performance varies across time. The calculated results confirm the findings reported in Table 21. The proposed Model with Actual Inputs performs best for almost all subsamples considered. The only exceptions are documented in 2018, a period dominated by the Fed tightening monetary policy. Results also demonstrate that the model with forecasted inputs is, on average, the second best when the horizon ranges from one quarter to two quarters. The model with forecasted inputs does not exhibit a clear outperformance for higher horizons. Since this specification is run based on predicted data, performance depends on the forecast accuracy of the different (EIA forecasted) inputs. We see that the longer the forecast horizon, the lower the forecast accuracy. Reported results confirm that the proposed model can be used to consider different (twelve-month maturity) scenarios underlying the selected explanatory variables.

The forecasting ability of futures prices and the Bloomberg analyst survey can be compared with the results reported by Figuerola-Ferretti et al. (2020), which demonstrate that futures prices outperform (at the aggregate level) analyst forecasts when considering forecasts performed yearly. The current analysis makes it unclear whether future prices will beat Bloom-berg analysts' forecasts. This may be explained by the different periods and prediction horizons considered in the forecasting exercise. While Figuerola-Ferretti et al. (2020) consider the average forecast for a given year with Chicago Mercantile Exchange (CME) WTI futures prices for a sample ending in

December 2019, the analysis in this study uses ICE Brent futures prices and a six-quarter rolling window and includes forecasts ending the last quarter of 2023.

Table 23. Performance Evolution versus no-change forecast (a six-quarter window)

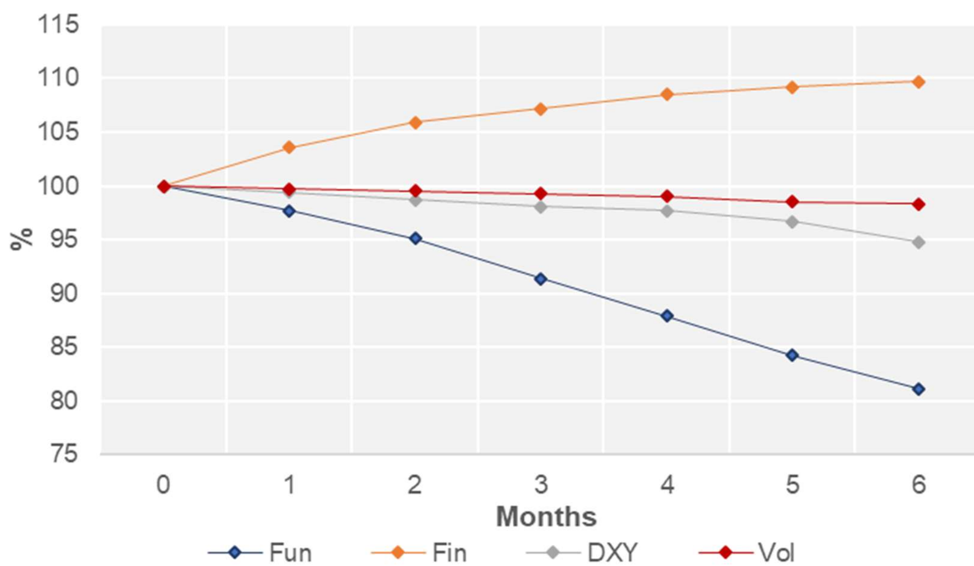
1st Quarter Forecast																				
6 Quarter Rolling from	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2	2018Q3	2018Q4	2019Q1	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4
to	2018Q2	2018Q3	2018Q4	2019Q1	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1
Futures	1.28	1.167	1.317	1.447	1.35	1.447	1.568	1.595	1.712	1.589	1.629	1.434	1.135	1.272	0.927	0.937	1.115	0.930	0.892	0.943
BBG	1.511	1.62	1.204	1.283	1.158	1.17	1.029	0.749	0.89	1.152	1.181	1.150	1.346	2.095	1.914	1.371	1.818	1.507	1.550	1.571
Department of Energy EIA	1.281	1.295	1.231	1.316	1.196	1.191	1.145	0.836	1.309	1.386	1.448	1.123	1.134	1.605	0.655	0.645	0.682	0.609	0.630	0.643
GAMLTF Forecasted Inputs	1.024	1.011	0.999	1.013	0.993	0.893	0.774	0.736	0.854	1.138	1.061	0.956	1.224	1.810	1.256	0.848	1.198	1.386	1.127	1.018
GAMLTF Actual Inputs	0.878	0.746	0.651	0.661	0.77	0.745	0.796	0.567	0.649	0.726	0.709	0.707	0.595	0.821	0.658	0.665	0.799	0.763	0.787	0.838
LTF Actual Imputs no GAM	1.423	1.187	0.899	0.931	0.897	0.765	0.474	0.72	0.993	1.139	1.104	1.048	1.255	1.637	1.388	0.919	1.518	1.916	1.951	1.857
2nd Quarter Forecast																				
6 Quarter Rolling from	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2	2018Q3	2018Q4	2019Q1	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4
to	2018Q2	2018Q3	2018Q4	2019Q1	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1
Futures	1.155	1.178	1.143	1.204	1.276	1.316	1.457	1.072	1.03	0.995	0.963	0.968	0.935	0.964	1.011	1.027	0.930	0.769	0.654	0.645
BBG	1.093	0.951	0.816	0.832	0.905	0.89	0.891	0.918	0.864	0.955	0.899	0.889	0.909	1.050	1.243	1.165	1.083	0.864	0.884	0.906
Department of Energy EIA	1.216	1.173	1.051	1.064	1.063	1.048	0.908	0.898	1.005	0.996	0.961	0.914	0.948	1.071	0.917	0.964	0.914	0.762	0.755	0.760
GAMLTF Forecasted Inputs	0.873	0.935	0.939	0.931	1.006	0.966	1.104	0.992	1.016	1.11	0.997	0.883	0.889	0.826	0.649	0.562	0.749	0.915	0.802	0.843
GAMLTF Actual Inputs	0.795	0.583	0.512	0.57	0.616	0.693	0.714	0.449	0.446	0.475	0.521	0.456	0.388	0.535	0.538	0.555	0.649	0.740	0.792	0.869
LTF Actual Imputs no GAM	1.078	0.876	0.663	0.696	0.572	0.631	0.921	1.522	1.629	1.77	1.625	1.413	1.298	0.777	0.617	0.469	0.684	1.030	1.026	1.038
3rd Quarter Forecast																				
6 Quarter Rolling from	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2	2018Q3	2018Q4	2019Q1	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4
to	2018Q2	2018Q3	2018Q4	2019Q1	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1
Futures	1.089	1.026	1.047	1.152	1.214	1.425	1.097	1.05	0.983	0.956	0.943	0.877	0.906	0.958	1.024	1.051	0.949	0.790	0.646	0.578
BBG	0.878	0.757	0.784	0.996	1.118	1.075	1.05	1.044	0.963	0.947	0.9	0.888	0.867	0.962	1.109	1.115	0.980	0.800	0.787	0.742
Department of Energy EIA	1.075	0.995	0.973	0.943	0.824	0.879	0.828	0.861	0.904	0.904	0.898	0.868	0.906	0.972	0.985	1.038	1.002	0.844	0.819	0.750
GAMLTF Forecasted Inputs	0.861	0.876	0.904	0.922	1.012	0.978	1.094	1.009	1.043	1.049	0.928	0.817	0.662	0.620	0.521	0.443	0.645	0.780	0.831	0.909
GAMLTF Actual Inputs	0.616	0.402	0.445	0.597	0.806	0.763	0.503	0.394	0.449	0.494	0.454	0.406	0.376	0.479	0.474	0.462	0.625	0.703	0.828	0.835
LTF Actual Imputs no GAM	0.715	0.435	0.445	0.63	0.858	1.27	1.815	2.027	2.053	1.965	1.73	1.540	1.065	0.672	0.548	0.410	0.654	0.824	0.796	0.958
4th Quarter Forecast																				
6 Quarter Rolling from	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2	2018Q3	2018Q4	2019Q1	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4
to	2018Q2	2018Q3	2018Q4	2019Q1	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2	2021Q3	2021Q4	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1
Futures	0.871	0.885	0.901	0.95	1.197	1.095	1.065	1.012	0.976	0.957	0.875	0.863	0.912	0.970	1.028	1.068	0.954	0.786	0.750	0.665
BBG	0.781	0.728	0.745	0.876	1.034	1.051	1.063	1.051	0.981	0.95	0.889	0.805	0.818	0.881	1.028	1.083	0.925	0.779	0.710	0.641
Department of Energy EIA	0.998	0.889	0.859	0.76	0.687	0.89	0.924	0.955	0.938	0.93	0.92	0.853	0.852	0.894	0.972	1.058	1.003	0.848	0.832	0.749
GAMLTF Forecasted Inputs	0.925	0.832	0.858	0.873	0.954	0.975	1.042	0.985	0.979	0.937	0.835	0.675	0.503	0.514	0.413	0.371	0.576	0.700	0.851	0.929
GAMLTF Actual Inputs	0.578	0.434	0.449	0.488	0.568	0.395	0.312	0.321	0.385	0.394	0.367	0.367	0.418	0.484	0.458	0.503	0.682	0.762	0.861	0.882
LTF Actual Imputs no GAM	0.849	0.714	0.59	0.674	1.236	1.76	1.944	1.963	1.985	1.864	1.679	1.253	0.825	0.615	0.506	0.426	0.610	0.708	0.824	0.916

Note: This table reports the forecasting performance in terms of the ratio of MAPE of the selected method and the no-change method. (forecasting horizon is a six-quarter average window ahead). The performance of the proposed GAMLTF framework for forecasted and actual input data and alternative benchmarks, including the LTF framework with no GAM. The in-sample period is 1995-2016, and the out-of-sample or forecasting period is 2017-2023.

5.3 Sensitivity Analysis

The next step is to provide a sensitivity analysis, developed to show the future evolution of the crude oil price, given a one standard deviation shock to some of the explanatory variables over a six-month horizon, keeping the remaining variables constant. Results are reported in Figure 19.

Figure 19. Sensitivity Analysis:



Note: This figure illustrates the effect of a one standard deviation shock in each explanatory variable on the Brent crude oil price over a 6-month horizon. The Jan1995-Dec2016 in-sample period is considered for this purpose. The sensitivity analysis was performed for the first two quarters of 2017.

We assume that variables were shocked in December 2016 and evaluated over the next six months.

They show that the variable with the most significant influence on crude oil is the Fundamental variable, which decreases the crude oil price by 20% for a given one standard deviation shock. The second most important variable in terms of price impact is the Financial variable, which has a positive 10% effect on Brent prices for a given one-standard-deviation shock. The same shock applied to the dollar and volatility variables exert a negative impact of 5% and 2%, respectively. Our findings are consistent with the literature supporting supply and demand fundamentals as the main drivers of crude oil prices, as in Baumeister and Kilian (2012), Baumeister and Kilian (2015), Figuerola et al. (2020) and Kaufmann and Ullman (2009). The market fundamental variable is the most important factor explaining the time series evolution of crude oil prices, with shocks remaining important after six months. Speculators are informed investors who enter the market to exploit fundamental trends, as explained by Kaufmann and Ullman

(2009). Indeed, table 14 in the appendix shows that the Fundamental and Financial variables exhibit a significant negative correlation of -0.56, implying that they share common characteristics. When fundamentals are tight, the market has a more significant inflow of speculative activity.

5.4 Extended Model

In the current times, it may be surprising to propose a fundamental oil price model that does not include geopolitical risk. That is because the relationship between geopolitical risk and Brent crude oil prices is complex and dynamic. Geopolitical tensions, such as conflicts in oil-producing regions or disruptions to major oil supply routes, can increase uncertainty and volatility in the oil market, causing Brent prices to fluctuate. Additionally, geopolitical events that threaten stability in key oil-producing countries or regions can prompt concerns about future oil supply, exerting upward pressure on Brent prices. The complexity arises when the increase in oil prices translates into inflation and may ultimately harm economic growth and, consequently, oil demand. Overall, geopolitical risk plays a significant role in shaping market sentiment and influencing Brent crude oil prices, highlighting the importance of monitoring global geopolitical developments for oil market participants and analysts.

In our opinion, increasing geopolitical risk enhances expectations of future tight fundamentals or restricted supply. We have illustrated in Figure 20 the relationship between our fundamental variable and the oil price. The figure illustrates the historical fitted regression between the oil price (on the y-axis) and the value of the fundamental variable on the x-axis. The depicted observations show that during times of high geopolitical risk, such as during the 2022 initiation of the war between Russia and Ukraine or the increased tensions in the Middle East (starting in 2023), the price deviates from its expected value. In fact, geopolitical risk drives prices to a level associated with tighter fundamentals. Therefore, we argue that Geopolitical Risk could be reflected in the Proposed Model's explanatory variables as a proxy. In the case of a war involving oil-producing countries, it will reduce the oil supply and, therefore, decrease the value of our fundamental variable (the ratio between supply, stocks, and demand).

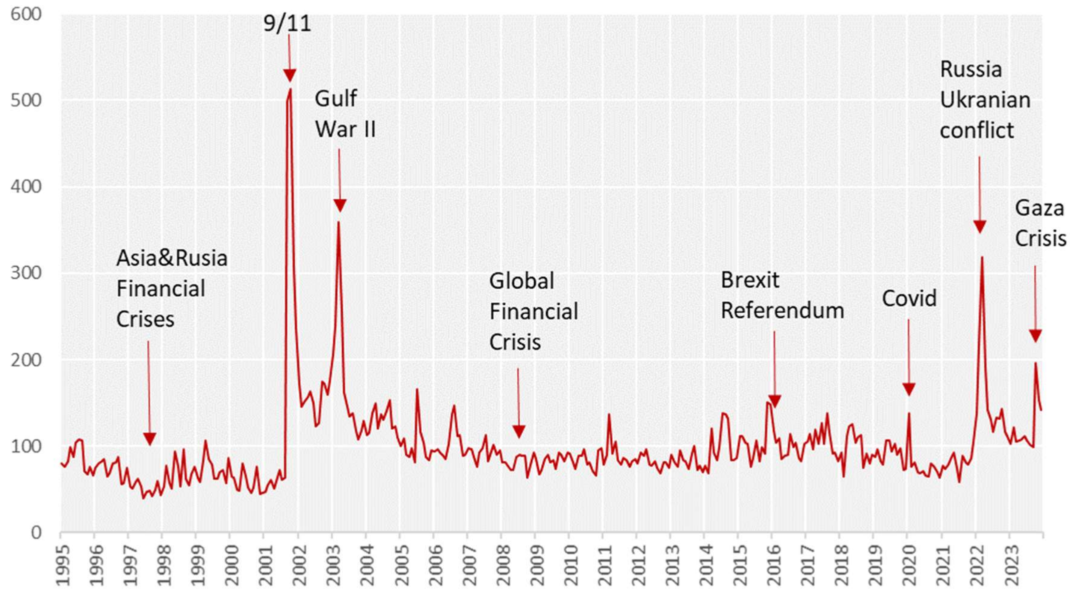
Figure 20. Fundamental Variable and Geopolitical Risk

Note: In the scatter between Fundamental Variable and Brent, there are some points far from the regression line, which in many cases are associated with geopolitical tensions

However, due to the conflict between Ukraine and Russia, we have succumbed to the temptation of attempting to include a variable of geopolitical risk in the oil model. Traditionally, tensions in the Middle East have been a focal point for oil price spikes, although these often manifest in supply and demand variables with a delay. Therefore, we have included this variable while removing the historical volatility variable. The geopolitical risk variable is defined by Caldara and Iacoviello (2022), who conduct an automatic keyword search of archives from 10 newspapers (Chicago Tribune, the Daily Telegraph, Financial Times, The Globe and Mail, The Guardian, the Los Angeles Times, The New York Times, USA Today, The Wall Street Journal, and The Washington Post). The index is calculated by counting the number of articles related to adverse geopolitical events in each newspaper each month, considering the total number of news articles. Figure 21 displays the indicator alongside the most significant geopolitical risk-related events of recent years.

The inclusion of this variable makes the model more reactive to specific moments of an increase in the geopolitical risk premium, although there is a part of the response that will continue to be reflected by the expectations of the fundamental variable.

Figure 21 Geopolitical Risk Index

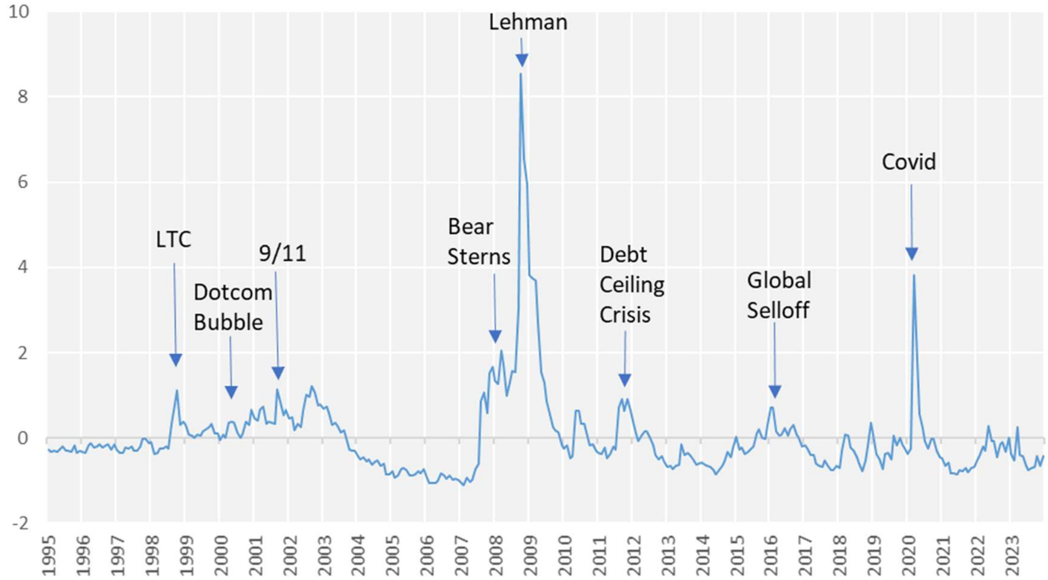


Note: Geopolitical Risk Index according to methodology Caldara and Iacoviello (2022), and events related

Incorporating geopolitical risk into Brent crude oil price forecasting models offers market participants and analysts valuable insights. Geopolitical factors significantly impact oil market dynamics, influencing supply disruptions, investor sentiment, and price volatility. Analysts can better capture the relationships between geopolitical events and Brent prices by integrating geopolitical risk indicators into forecasting models. This approach enables more accurate predictions and improved risk management strategies, helping market participants navigate uncertainties and capitalize on market opportunities. Overall, leveraging geopolitical risk data enhances the robustness and reliability of Brent price forecasts, making it a valuable tool for informed decision-making in the oil market.

The geopolitical variable provides us with a measure of risk, although it does not capture everything that volatility contributed to the model. Therefore, we need another variable that measures risks more closely associated with financial markets. For that reason, the financial stress index developed by the St Louis Fed has been included. Incorporating a financial stress variable into Brent crude oil price forecasting models can provide valuable insights into market dynamics. Financial stress indicators capture broader economic conditions and market sentiment, which can influence oil prices. This approach enhances the robustness of Brent price forecasts and enables market participants to make more informed decisions, particularly during economic uncertainty. Overall, integrating financial stress variables into Brent price models improves their accuracy and reliability, making them essential tools for navigating the complexities of the oil market.

Figure 22 Financial Stress Index



Note: Financial Stress Index according to methodology Federal Reserve Bank of St. Louis, and periods related

The financial conditions variable is the fourth version of the indicator constructed by the Federal Reserve Bank of St. Louis from 18 different data series: seven of interest rates, five of interest rate differentials, and other variables such as the three-month implied volatility of the Standard & Poor's (S&P), inflation, or the S&P financial index. A value of zero represents normal financial conditions. When it is below zero, it indicates relaxed financial conditions; when it is above, it indicates stressed conditions. Figure 22 displays the indicator alongside the most significant financial stress-related events of recent years.

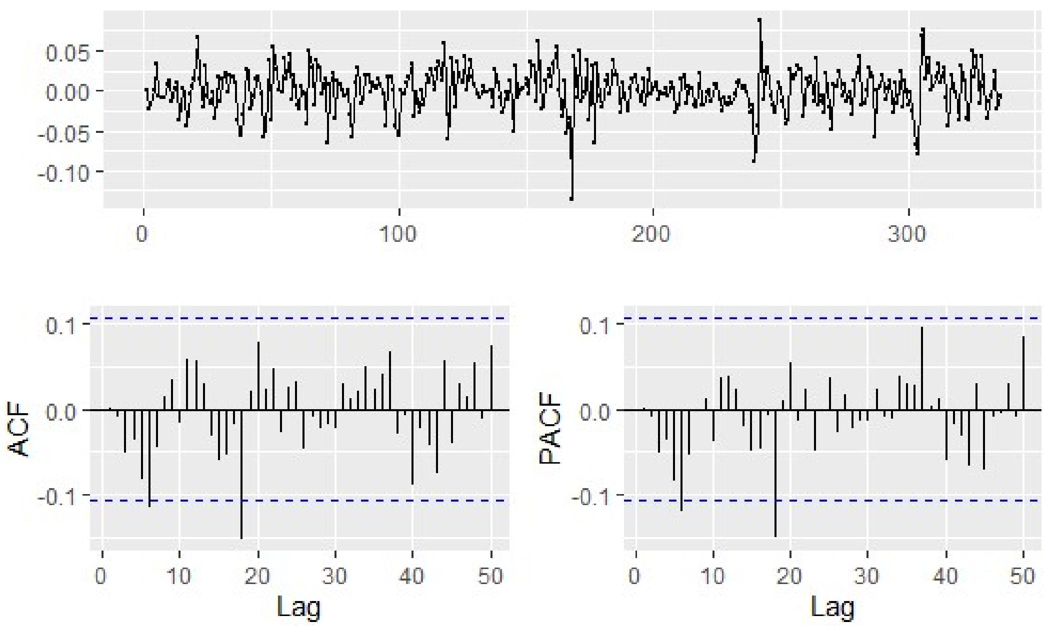
Apart from replacing volatility with geopolitical risk and financial stress variables, the modeling approach will be similar to the previous model. In this case, the error structure has been modeled according to an ARIMA (0,1,1), and the coefficients obtained are all significant, as shown in Table 24. The error analysis is depicted in Figure 23, and with the extended model, we have addressed the issues with the error structure.

Table 24. Summary of estimated Coefficients under final specification:

Approximate significance of smooth terms:					
Variable:	**Estimate**	**Std Error**	**Z Value**	**p-value**	
ma	0.288	0.053	5.477	4.32e-08	***
f(Fundamental)	0.638	0.101	6.301	2.95e-10	***
f(Stress)	1.066	0.129	8.270	<2.20e-16	***
f(Geopol)	1.281	0.384	3.342	8.32e-4	***
f(Dollar)	0.508	0.138	3.681	2.32e-4	***
f(Financial)	1.148	0.109	10.573	<2.20e-16	***
Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1''					
Box-Pierce test = 0.000456, df = 1, p-value = 0.983					

Note: This table reports estimates of the final model specification with the coefficient of the regression calculated.

Figure 23. PACF and ACF structure of the error term under GAM with LTF Specification



Note: This figure shows the PACF and ACF for the proposed model error. No problem of autocorrelation can be appreciated**.**

The same performance metric (MAPE) has been calculated for this model, using the actual values for the variables, and has been compared with those of the previous model, as shown in Table 25. As observed, the overall performance of this second model is slightly worse, considering the entire out-of-sample period. However, comparing individual periods in isolation, there are more instances where it does exhibit better performance.

Table 25. MAPE error measures for different models with actual variable data

	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2	2018Q3	2018Q4	2019Q1
Proposed M.	**16.11%**	6.15%	8.86%	**4.80%**	**5.78%**	**11.45%**	3.86%	11.30%	7.72%
Extended M.	16.20%	**5.87%**	**7.76%**	6.26%	9.12%	13.88%	**3.27%**	**8.91%**	**3.26%**

	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2
Proposed M.	5.86%	**9.83%**	**16.25%**	**12.47%**	**19.46%**	**17.67%**	**12.96%**	5.95%	**3.79%**
Extended M.	**4.99%**	15.53%	20.29%	17.81%	24.65%	28.64%	16.43%	**5.43%**	8.55%

	2021Q3	2021Q4	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1	**TOTAL**
Proposed M.	11.72%	14.91%	**21.22%**	26.12%	20.87%	9.29%	4.12%	**11.54%**
Extended M.	**10.21%**	**14.21%**	29.79%	**4.37%**	**12.82%**	**3.51%**	**3.81%**	**11.82%**

Note: This table reports the forecasting performance in terms of the MAPE measure of the proposed framework for the Proposed Model and the Extended Model.

Another issue with this model is that we have another variable for which we would need to make its own forecast to calculate the forecast for oil prices. However, as mentioned earlier, the model's ability to explain oil prices allows us to use it for scenario analysis. This model has an advantage as it isolates the geopolitical risk premium and the financial stress premium. For instance, it allows us to analyze which part should be attributed to a temporary surge in Ukraine's conflict or issues in economies facing widespread interest rate hikes. In this way, by setting similar scenarios for the shared variables, we can incorporate these new factors.

Chapter 6

Model application: Generation of oil price scenarios

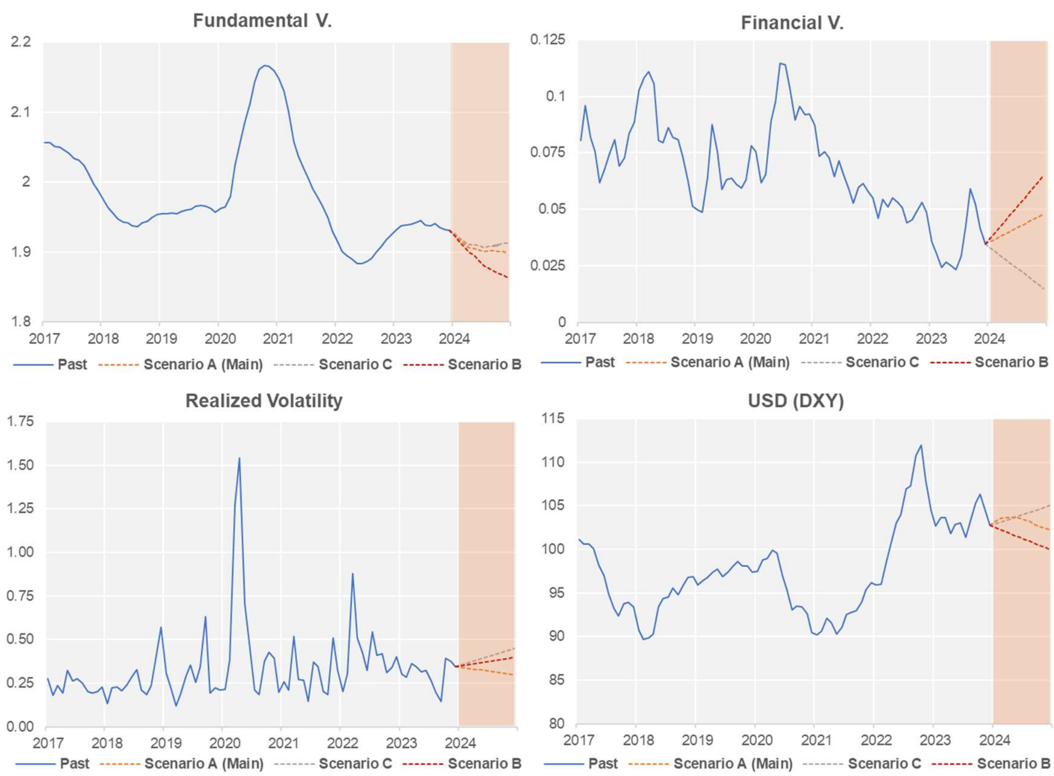
6.1 Introduction

This chapter demonstrates the application of the created model. The models developed will be used to generate various scenarios regarding oil prices based on explanatory variables. Having tools for scenario analysis is invaluable for decision-makers in the energy sector and beyond. These models not only provide insights into potential future price movements but also allow stakeholders to assess the impact of various factors and events on oil prices. By simulating different scenarios, decision-makers can better understand the potential risks and opportunities, formulate strategies to mitigate risks and capitalize on favorable conditions. Additionally, scenario analysis helps in contingency planning and policy development, enabling organizations to adapt more effectively to changing market conditions and uncertainties. Overall, having robust scenario analysis tools enhances decision-making processes and facilitates more informed and strategic actions in response to dynamic market environments.

6.2 Initial Proposed Model

We have seen in the previous section that the proposed model can explain and forecast very accurately when the observed (and not forecasted) values of the explanatory variables are used in the forecasting process. This tool can help understand the interaction of factors that determine the past oil price evolution and the future paths under different scenarios, quantifying the risk associated with a particular scenario compared to an alternative baseline forecast (selected as the EIA forecast). The proposed model identifies key variables driving upside and downside risks in the oil price forecast. For expository purposes, three scenarios involving hypothetical future oil market conditions are explored, starting in the first quarter of 2024. These primary raw data correspond to world production, world demand, OECD stocks, non-commercial long and short positions, open interest, historical volatility, and the U.S. dollar, which help us build our variables. Figure 24 illustrates the twelve-month forecasts for the four variables in the three scenarios defined in December 2023 for 2024. The illustrative scenarios are focused on the implications of shocks arising from the supply relative to demand conditions.

Figure 24. Forecasts of Input variables under different scenarios:



Note: This figure illustrates the evolution of the input variables in different scenarios. The main scenario: Scenario A uses EIA forecast input data. Scenario B analyzes the possibility of physical tightening. Scenario C addresses the case of low OPEC.

Scenario A: Main benchmark scenario with EIA forecast

The Main scenario (Scenario A) uses the U.S. Department of Energy forecast of the fundamental variable for the next 12 months, performed in December 2023. This includes the concern expressed by the DoE regarding the weakening global economic situation, which leads to lower expectations for global oil demand growth. An increase in demand of 1.3mb/d is thus considered under this scenario. These views about the economy can potentially offset the upward pressure on prices stemming from lower short-term oil supply due to OPEC's and Russia's supply cuts in crude oil production. Oil production cuts were first announced in October 2022 for a cut of 2 mb/d and were enhanced in April 2023 to 3.5 mb/d.

Furthermore, in June 2023, OPEC and Russia decided to extend cuts to December 2024. In July, Saudi Arabia additionally announced voluntary cuts (details can be found at https://www.reuters.com/business/energy/saudi-arabia-expected-extend-voluntary-oil-cut-september-analysts-say-2023-07-28/). Full compliance (-3.5mb/d from the level registered in August 2022) is not expected, but the agency forecasted in December 2023 that production will increase by 0.6mb/d, representing a slowdown when compared to growth levels reported of 1.6 mb/d in 2023. The Fundamental variable is predicted to stay near last year's lows. With the tightening of the physical market, investors will increase their positions. The U.S. dollar stabilizes around 102 as monetary policies are becoming more aligned in the U.S. and Europe. Crude oil price volatility returns to normal conditions, considering the Ukrainian crisis causes no other uncertainty-related spikes.

Scenario B: Physical market tightening

This represents the case of full compliance with OPEC's quota supported by increased tensions in the Middle East (particularly in the Red Sea) and a robust economy that sustains oil consumption with a rebound in consumption driven by the airline sector, as forecasted by data from S&P Global. In this case, the Fundamental Variable will fall to the lowest value registered over our sample period. Under this scenario, investors will be attracted to exploit the upward price trend. The U.S. Dollar will be weaker than in the Main scenario, and volatility will rebound mildly because of increasing geopolitical pressures. Note that the OPEC plus group has announced an extension of 3 months of their voluntary cuts, making this scenario less likely. See the Financial Times article "Opec+ members extend production cuts to boost oil price," 4 March 2024.

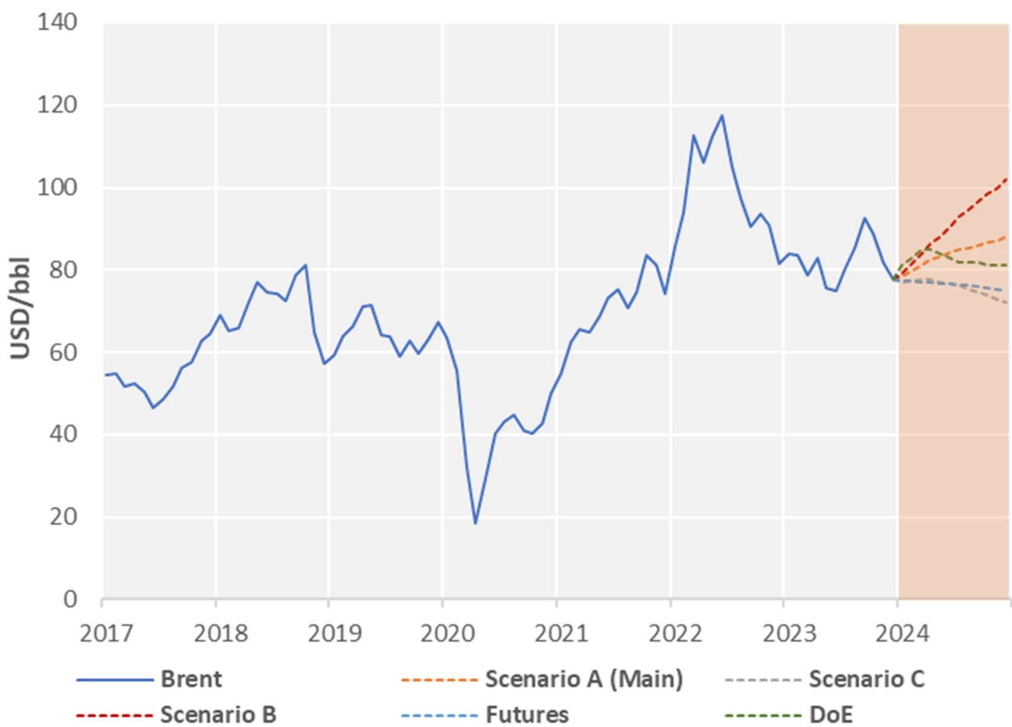
Scenario C: Low OPEC compliance and delay on the end of monetary tightening

OPEC compliance is less stringent than the main scenario, implying that production stands at 1mb/d during 2023Q3-2024Q4. Oil demand growth moderates because of the delay in the monetary tightening. Under this scenario, investors will reduce their oil exposure in their portfolios, volatility will pick up, and the dollar will appreciate slightly.

Our model also allows us to do reverse engineering. This feature implies that we can calculate the values of the underlying variables implied by futures prices. In order to match quoted futures prices observed in December 2023, our framework shows that there should be low compliance with the announced OPEC cuts in the first half of 2024. Prices are similar to Scenario C.

Forecasts under the different scenarios, including the Main and EIA forecasts, are illustrated in Figure 25. First, our main benchmark scenario for the next 12 months is slightly more bullish than that reported by the EIA. Under the supply-stressed scenario (B), oil prices are expected to be higher than \$100, given the context of deteriorated levels in the physical balance. The increase in geopolitical risk, partly driven by the recent moves by Saudi Arabia and Russia to extend their voluntary supply cuts, drives fears of future inflation and new periods of prolonged periods of low investment in new capacity. This fact is fundamental under the currently announced OPEC's production cuts. While OPEC compliance has always been hesitant, the possibility of future supply cuts remains a primary concern for Western governments already struggling to contain inflation. We could see prices stabilizing around the $72/bbl level in the case of low OPEC compliance. The term structure of futures prices is in December 2023 in mild backwardation, with the December 2024 futures price trading at $75/bbl. This implies improved fundamentals compared to the term structure seen in November 2023.

Figure 25. Oil price forecasts in different scenarios:



Note: The figure illustrates forecasting prices under the different scenarios considered, including the Main (using EIA forecast), the EIA forecast (labeled DoE), the forecast implied by Futures (labeled as Futures), scenario B (tight fundamentals), and scenario C (low OPEC compliance).

Note that the set of scenarios envisaged for the explanatory variables allows the simulation of different geopolitical situations. Given that increased geopolitical tensions influence the price of oil, the proposed tool can be used to consider changes in the explanatory variables (and the corresponding crude oil price forecast) affected by increased geopolitical uncertainty. For instance, we expect that there will be supply disruptions under the surge of an armed conflict. These disruptions will reduce the value of the fundamental variable and, therefore, lead to a scenario similar to that described in scenario B.

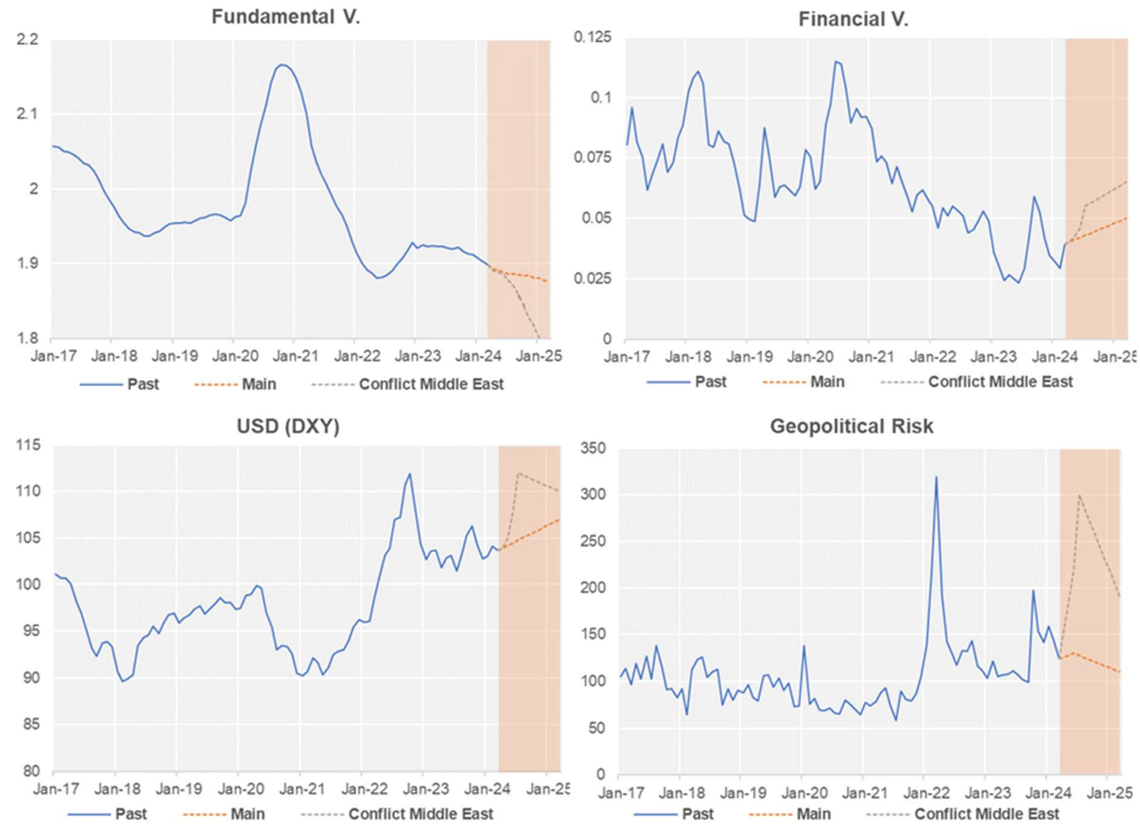
6.3 Extended Model (geopolitical risk and financial stress)

The current instability in the Middle East, particularly the ongoing conflict between Israel and Iran, has significant implications for the price of oil. Given its vast oil reserves and strategic importance in global energy markets, the region has long been a focal point for geopolitical tensions. The escalation of hostilities between these two nations raises concerns about potential disruptions to the region's oil production and supply routes. Any military confrontation between these two countries could disrupt oil shipments through strategic waterways such as the Strait of Hormuz, a critical chokepoint for global oil trade.

For this reason, the extended model can help contextualize possible movements in oil prices. The prospect of supply disruptions in the Middle East, combined with increased geopolitical risk, tends to exert upward pressure on oil prices. Investors and traders closely monitor developments in the region, reacting swiftly to any signs of escalation or instability. Even the perception of heightened risk can lead to speculative buying and price spikes in the oil market.

In this case, two scenarios involving hypothetical future oil market conditions are explored, starting in the second quarter of 2024. These main variables and estimated parameters correspond to world production, world demand, OECD stocks, non-commercial long and short positions, open interest, geopolitical risk, and the U.S. dollar. We will keep the financial stress variable constant as it should not have too much influence on the concerned scenarios. Figure 26 illustrates the twelve-month forecasts for the four variables in the two scenarios defined in March 2024 for the next twelve months.

Figure 26. Forecasts of Input variables under different scenarios (conflict Middle East):



Note: This figure illustrates the evolution of the input variables in different scenarios. The main scenario: Scenario A uses EIA forecast input data. The second scenario is considering an armed conflict in the Middle East.

Scenario A: Main scenario with EIA forecast

The Main scenario uses the U.S. Department of Energy forecast of the fundamental variable for the next 12 months, performed in March 2024. In these first months of the year, there is less concern about the weakening global economic situation, and IEA has revised the global oil demand growth. They expect oil inventories to not increase until 2025, when the OPEC+ supply cuts expire. With this tight market, investor flow may return, and the dollar will continue its appreciation of the last year in a relatively moderated geopolitical risk environment.

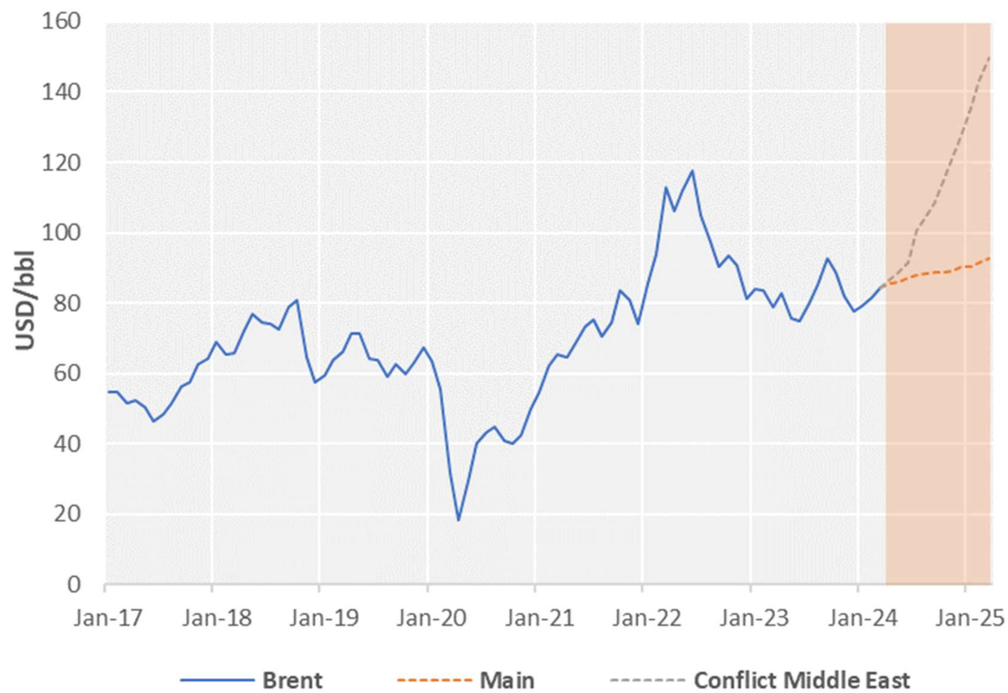
Scenario Conflict in the Middle East

We imagine that the current situation escalates into an armed conflict after a couple of months, potentially affecting oil production or transportation. In this regard, we have reduced global production by 2 million barrels starting in June, impacting inventories and, thus, the fundamental variable. Geopolitical Risk premiums would increase, along with the dollar as a safe-haven asset amid such conflicts. We also consider it possible that speculative investor positions may increase.

Figure 27 depicts the oil price forecasts under the extended model and new scenarios. The central scenario shows a mild bullish trend, consolidating above $90, while the Middle East conflict scenario would push oil prices close to the 2008 historical highs. While it is true that oil's reaction to such events is typically more pronounced, the model values illustrate the magnitude that the movement could potentially reach.

If we look at the forecast made in December 2023 with the previous model, we can see that oil prices in the first quarter of 2024 have behaved quite similarly to what was expected by scenarios A, B, and the DoE. All of them predicted a tightening of the market, which has ultimately occurred. If we look ahead to the remainder of the year, our Main scenario (Scenario A) continues to point to a gradual climb above $90.

Figure 27. Oil price forecasts in different scenarios (Middle East conflict):



Note: The figure illustrates forecasting prices under the different scenarios considered, including the Main (using EIA forecast) and a scenario with a conflict in the Middle East

Chapter 7

Conclusions, Contributions, and Future Research

7.1 Introduction

This final chapter of the thesis marks the culmination of the work presented and contains the main conclusions and findings drawn from the research. Comprising three parts, it begins with a summary of the studies and conclusion presented within the thesis. Secondly, it elaborates on the contributions of forecasting methodologies and models presented in the thesis. Finally, it identifies and proposes several lines for future research from the results and developments achieved in this thesis.

7.2 Summary and Conclusions

Recent developments in energy markets have shown that the crude oil market is exposed to time-changing uncertainty. As a result, crude oil prices have been subject to significant fluctuations over the past two decades. This fact makes oil price prediction a very challenging task. While the forecasting frameworks developed in the literature are wide and varied, there is no consensus about the appropriate methodological framework to apply.

The first highlighted conclusion comes from studying the relationship between production costs and oil prices. The results shed light on the prevailing belief that oil production costs are the predominant factor driving fluctuations in oil prices. Contrary to this assumption, the research findings indicate that changes in oil prices precede adjustments in other variables, including production costs. Employing Granger's definition of causality and the Toda-Yamamoto methodology, the study rigorously examines the interactions between oil prices and production costs. It uncovers a consistent pattern where fluctuations in oil prices influence subsequent changes in production costs rather than the reverse. This suggests that oil prices significantly influence production costs, shaping industry dynamics.

Moreover, the analysis extends to explore longer futures contracts, revealing that the relationship between oil prices and production costs remains consistent over time. As a result, the study challenges the conventional wisdom and advises against solely considering industry-related costs as explanatory variables when explaining or predicting future oil price movements. Instead, it underscores the pivotal role of oil prices as the primary determinant of changes in production costs. This insight holds critical implications for understanding the dynamics of the oil market and making informed decisions in the energy sector.

This thesis combines the classical regression model with machine learning approaches in a hybrid framework, selecting the GAM method across the feature engineering literature jointly with the Transfer Function with the ARIMA noise approach. Machine learning methods help to incorporate flexible non-linear capability in the modeling process.

Compared to competing machine learning approaches, the advantage of the proposed method is that it captures non-linearities under the analysis of partial effects. This feature allows input variable interpretation through estimated regression coefficients. The method identifies two main drivers explaining oil prices. The first and most important variable is the Fundamental variable, which measures the physical market balance. The second is the financial variable quantifying and capturing crude oil investors' speculative interest. The volatility and the dollar variables contribute to a lower impact on oil price movements. Results show that the non-linear effects are remarkably significant in the dollar and volatility variables. The impact of the dollar index is significant only under weak dollar conditions, while volatility is essential for forecasting purposes under high volatility states.

We show that the algorithm may be applied using U.S. EIA forecasts of the fundamental variable and dollar and forecasting the other input variables. The forecasting ability of the proposed framework outperforms benchmark techniques, including the futures prices and analysts' crude oil price predictions provided by Bloomberg and the EIA.

Sensitivity analysis confirms that the variable with the most significant influence on crude oil prices is the Fundamental Variable. One standard deviation increase in this variable results in an oil price reduction of 20%. The financial variable is the second most important, exerting an impact of 10% for a one standard deviation increase. The impact of the one standard deviation change in the dollar and volatility variables are 5% and 2% price change, respectively.

The proposed model is also highly suitable for scenario analysis. The algorithm demonstrates the ability to quantify the risk associated with a benchmark forecast based on an extensive analysis of how this forecast changes under alternative hypothetical scenarios about future oil demand and oil supply conditions. The main scenario (December 2023) predicts a rebound in oil prices towards USD88/bbl, delivering higher prices than the EIA. Two additional situations are proposed for 2024, with the market balance variable acting as the main driving force. Under market tightening conditions arising from compliance with OPEC's and Russia's production cuts, prices are pushed to new highs (above USD 100/bbl). Under a lower OPEC compliance scenario and lower consumption due to higher-than-expected interest rates, we could see a moderation in prices towards $72/bbl.

In light of the resurge of hostilities in the Middle East during the first quarter of 2024, the extended model has been used to simulate scenarios where we can directly vary geopolitical risk. If the conflict escalates into an armed conflict between some Middle Eastern countries and there is a surge in geopolitical risk premium, we could see oil approaching the historical highs of 2008.

Our results show the relevance of supply and demand fundamentals in the price determination process and confirm that events that disrupt global oil supply are expected to increase oil prices, while events that suppress oil consumption growth will generally decrease oil spot prices. The proposed hybrid model can be applied to risk management systems of energy corporations and institutions. It can also provide a quantitative assessment of the impact of a range of hypothetical events on the crude oil price. This is crucial in times of multiple sources of uncertainty arising from factors such as geopolitical tensions, interest rate risk, and energy transition-related shocks.

7.3 Contributions and Publications

The contributions of the thesis to the existing literature on modeling and forecasting oil prices are significant and multifaceted. Firstly, when forecasting future oil price movements, it is essential to focus on the price of oil itself rather than industry-related costs as explanatory variables. This observation holds significant implications for understanding the dynamics of the oil market and making informed decisions in the energy sector.

Secondly, we propose the creation of new variables to enhance our understanding of oil movements, bringing both physical and financial perspectives. By incorporating diverse variables, the model provides a more comprehensive view of the factors influencing oil prices, allowing for a deeper analysis of market dynamics.

Thirdly, the thesis proposes a novel modeling approach capable of capturing nonlinear relationships while maintaining a straightforward interpretation of variable behavior. This innovative approach represents a departure from traditional linear models, enabling a deeper understanding of the oil price dynamics. By incorporating nonlinear elements, the model can better account for the relationships between various factors influencing oil prices, resulting in more accurate forecasts. In addition, the additive nature of the model facilitates its explainability.

Furthermore, the thesis demonstrates a notable improvement in forecasting accuracy compared to existing benchmarks in the field. The proposed model achieves superior results in predicting oil prices through practical application, offering valuable insights for market participants and decision-makers.

In summary, the thesis contributes significantly to advancing modeling and forecasting techniques for oil prices. The thesis offers valuable insights and tools for understanding and predicting oil price movements in today's complex and dynamic market environment by introducing new variables, proposing an innovative modeling approach, and demonstrating practical improvements in forecasting accuracy.

Articles published in peer-reviewed academic journals:

- P. Moreno Alonso, A. Muñoz San Roque, Oil Costs and Prices: An Empirical Causality Analysis. International Journal of Energy Economics and Policy. Vol. II, nº 3, págs. 546-554, Diciembre 2020-Enero de 2021. ISSN: 2146-4553. Repositorio: http://hdl.handle.net/11531/55372.
- Moreno, P.; Figuerola-Ferretti, I.; Muñoz, A. Forecasting Oil Prices with Non-Linear Dynamic Regression Modeling. Energies 2024, 17, 2182. https://doi.org/10.3390/en17092182

Currently, we are awaiting the review of the article 'Enhancing Natural Gas Price Forecasts: A Hybrid Model Integrating GAM and Transfer Function' by the journal International Journal of Oil, Gas and Coal Technology.

7.4 Future Research

First and foremost, the variables created could be helpful across various methodologies. On one hand, in the case of structural models, their inclusion could be assessed. Given their monthly frequency, no transformation would be necessary. However, when considering integration into other machine learning methodologies, efforts should be made to increase frequency. As the financial variable is reported weekly, no adjustment would be required. However, some form of proxy would need to be devised for the fundamental variable. Concerning inventories, an approximation could be made using the weekly data reported in the USA and Europe. Nevertheless, supply and production present a more significant challenge due to the absence of relevant weekly frequency data.

The primary constraint of the model for forecasting oil prices lies in the uncertainty surrounding the future values of explanatory variables at the time of forecasting. Therefore, a second line of research would involve investigating deeper into modeling these variables. Certain variables, such as the fundamental one, exhibit significant short-term correlations with the convenience yield or futures slope. In fact, the difference between the 12th and the 3rd future oil contracts appears to anticipate the fundamental variable by several periods, suggesting it could serve as a promising starting point for further investigation.

Hybrid models are a powerful tool in predicting oil prices, combining the strengths of two or more modeling approaches to achieve more accurate and reliable results. By merging different techniques such as statistical analysis, machine learning, and artificial intelligence, these models can capture a wide range of patterns and relationships in data, making them more robust against the complexity and volatility inherent in the oil market. There are some successful examples of this methodology forecasting oil prices, so one of the lines of future work could include mixing this model with other models.

Bibliography

Abdullah, S.N. and Zeng, X. (2010) “Machine learning approach for crude oil price prediction with artificial neural networks-quantitative (ann-q) model”. In: The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1–8

Akaike, H., (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), pp. 716-723.

Akpanta, A.C. and Okorie, I.E., (2014) “Application of Box-Jenkins Techniques in Modeling and Forecasting Nigeria Crude Oil Prices”. Journal of Statistics and Applications, 4(6), pp. 283-291,

Alquist, R., and Kilian, L., (2010) “What Do We Learn from the Price of Crude Oil Futures?”. Journal of Applied Econometrics, 25, 539-573.

Alquist, Ron & Kilian, Lutz and Vigfusson, Robert. (2011). Forecasting the Price of Oil. SSRN Electronic Journal. 10.2139/ssrn.1911194.

Alvarez, L., Hurtado, S., Sánchez, I. and Thomas, C. (2011). The Impact of Oil Price Changes on Spanish and Euro Area Consumer Price Inflation, Economic Modelling, [e-journal] vol. 28, no. 1, pp.422–431

Anderson, S.T., Kellogg, R., and Salant, S.W., (2014). Hotelling Under Pressure. w20280.

Antolín-Díaz and Rubio-Ramírez J.F., (2018). Narrative Sign Restrictions for SVARs. The American economic review, 108(10), pp. 2802-2829.

Arouri,M.E.H., Lahiani, A., Lévy, A., and Nguyen,D.K., (2012). “Forecasting the conditional volatility of oil spot and futures prices with structural breaks and long memory models”. Energy Economics. 34, pp. 283–293.

Bai, J. and Perron, P., (2003) Computation and analysis of multiple structural change models. Journal of Applied Econometrics, 18, 1–22.

Balke, N. S., Brown, S.P.A., and Yücel M. K., (2010) "Oil Price Shocks and U.S. Economic Activity: An International Perspective", mimeo, Federal Reserve Bank of Dallas.

Bao, Y., Yang, Y., Xiong, and T., Zhang, J. (2011) A comparative study of multi-step-ahead prediction for crude oil price with support vector regression. In: 2011 Fourth International Joint Conference on Computational Sciences and Optimization (CSO), pp. 598–602

Barsky, R. B. and Kilian, L. (2004). Oil and the Macroeconomy Since the 1970s, Journal of Economic Perspectives, [e-journal] vol. 18, no. 4, pp.115–134

Baumeister, C. and Kilian, L. (2015) Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach, Journal of Business and Economic Statistics, 33(3) pp 338-351

Baumeister, C. and Kilian, L., (2012) "Real-Time Forecasts of the Real Price of Oil", Journal of Business and Economic Statistics, vol 30, 2, pp326-336.

Baumeister, C. Recent Developments in (Modeling) Energy Market Dynamics, keynote at the workshop on "Energy and Climate: Macroeconomic Implications," B.I. Norwegian Business School, Oslo, 18 August 2022

Baumeister, C., (2014) "The Art and Science of Forecasting the Real Price of Oil", Bank of Canada Review, Spring.

Baumeister, C., 2016. Forty years of oil price fluctuations. The journal of economic perspectives, 30(1), pp. 139-160.

Baumeister, C., and Killian, L., (2016). Understanding the Decline in the Price of Oil since June 2014. Journal of the Association of Environmental and Resource Economists, 3(1), pp. 131-158.

Beckers, B. and Beidas-Strom, S., (2015) "Forecasting the Nominal Brent Oil Price with VARs—One Model Fits All?", International Monetary Fund, WP/15/251.

Beckmann, J., Czudaj, R. and Arora V. (2020) The Relationship between Oil Prices and Exchange Rates: revisiting theory and evi-dence. Energy Economics, Volume 88, Article 104772

Behar, A. and Ritz, R.A. (2017) OPEC vs US shale: Analyzing the shift to a market-share strategy,

Bildirici, M., and Ersin, Ö.Ö. (2015) "Forecasting volatility in oil prices with a class of nonlinear volatility models: smooth transition RBF and MLP neural networks augmented GARCH approach". Petroleum Science. 12 pp 534–552.

Bossley, L., 2017. Saving the Brent benchmark. The Journal of World Energy Law & Business, 10(4), pp. 259-273.

Bouchouev, I. (2021). Is the Oil Price-Inflation Relationship Transitory? Oxford Institute for Energy Studies

Box, G., Jenkins, G., and Reinsel, G., (1994) Time Series Analysis, Forecasting and Control. 3rd ed. Englwood Cliffs Prentice Hall

Bruno, V., Büyüksahin, B. and Robe, M. A. (2017) The Financialization of Food? American Journal of Agricultural Economics, 99 (1): 243-264

Burbidge, J., and Harrison, A. (1984). Testing for the effect of oil price rises using vector autoregressions. International Economic Review, 25(2), 459–484.

Caldara, D. and Iacoviello, M. (2022), "Measuring Geopolitical Risk," American Economic Review, April, 112(4), pp.1194-1225.

Cen, Z., and Wang, J. (2019). Crude oil price prediction model with long short-term memory deep learning based on prior knowledge data transfer. Energy, 169, 160–171.

Chai, J., Xing, L. M., Zhou, X. Y., Zhang, Z. G., and Li, J. X. (2018) Forecasting the WTI crude oil price by a hybrid-refined method. Energy Economics, 71, 114-127.

Chandar, K., Sumathi, M., and Sivanandam, N., (2015) "Crude Oil Prediction using a Hybrid Radial Basis Function Network", Journal of Theoretical and Applied Information Technology", 72.

Chauvet, M., Selody, J.G., Laxton, D., Kumhof, M., Benes, J., Kamenik, O., and Mursula, S., (2012). The Future of Oil: Geology Versus Technology. IMF Working Papers, 12(109), pp. 1.

Chinn, M. D., LeBlanch, M. and Coibion, O., (2005) "The Predictive Content of Energy Futures: An Update on Petroleum, Natural Gas, Heating Oil and Gasoline". NBER Working Paper 11033

Choi, S., Furceri, D., Loungani, P., Mishra, S., and Poplawski-Ribeiro, M. (2018), Oil prices and inflation dynamics: Evidence from advanced and developing economies. Journal of International Money and Finance, 82, 71-96

Cologni, A. and Manera, M. (2008). Oil Prices, Inflation and Interest Rates in a Structural Cointegrated VAR Model for the G-7 Countries, Energy Economics, [e-journal] vol. 30, no. 3, pp.856–888

Conflitti, C., and Luciani, M., (2019). Oil price pass-through into core inflation. Energy J. 40, 221–247.

Coppola, A. (2008), Forecasting oil price movements: Exploiting the information in the futures market. J. Fut. Mark., 28: 34-56.

Cunado, J., and Gracia, F. P. (2003). Do oil price shocks matter? Evidence for some European countries. Energy Economics, 25(2), 137–154.

Dickey DA and Fuller WA. (1981) Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica. 49(4):1057-72.

Dickey, D.A. and Fuller, W.A. (1981) Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica, 49, 1057–1072.

El-Gamal, M. A., and Jaffe, A. M. (2018). The Coupled Cycles of Geopolitics and Oil Prices. Economics of Energy & Environmental Policy, 7(2), 1–14.

Energy Economics, Volume 63, Pages 185-198

Energy Institute - Statistical Review of World Energy (2023)

Etienne, X. L., Irwin, S. H., and Garcia, P. (2015). Price explosiveness, speculation, and grain futures prices. American Journal of Agricultural Economics, 97(1), 65-87

Figuerola-Ferretti, I., Gilbert, C. L., and McCrorie, J. R. (2015) Testing for mild explosivity and bubbles in LME non-ferrous metals prices. Journal of Time Series Analysis, 36(5), 763-782.

Figuerola-Ferretti, I., McCrorie, J. R., and Paraskevopoulos, I. (2020) Mild explosivity in recent crude oil prices. Energy Economics, 87, 104387.

Figuerola-Ferretti, I., Rodriguez, A., and Schwartz, E. 2021. Oil price analysts' forecasts. The Journal of Futures Markets, 41 (9), pp. 1351-1374

Fileccia, G., and Sgarra, C., (2015) "Historical and Risk-Neutral Estimation in Commodity Market with two factor stochastic volatility model". International Journal of Economic and Econometrics, 5,4.

Gabralla, L. A., Jammazi R. and Abraham, A. (2013) "Oil price prediction using ensemble machine learning," INTERNATIONAL CONFERENCE ON COMPUTING, ELECTRICAL AND ELECTRONIC ENGINEERING (ICCEEE), Khartoum, Sudan pp. 674-679.

Gao, X. and Nardari, F. (2018) Do Commodities Add Economic Value in Asset Allocation? New Evidence from Time-Varying Moments. Journal of Financial and Quantitative Analysis, Volume 53, Issue 1, February 2018, pp. 365 – 393

Garratt, A., Vahey, S.P. and Zhang, Y., (2018) Real-Time Forecast Combinations for the Oil Price, CAMA Working Paper No. 38.

Gisser, M., and Goodwin, T. H. (1986). Crude oil and the macroeconomy: Tests of some popular notions. Journal of Money, Credit and Banking, 18(1), 95–103.

Granger CWJ. (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica. 37(3):424-38.

Gujarati, D.N. (2006), Basic Econometrics. 4th ed. Rio de Janerio: Elsevier.

Gulen, S.G. (1998) “Efficiency in the Crude Oil Futures Markets”. Journal of Energy Finance and Development, 3, pp. 13-21.

Haase, M., Seiler, Y. and Zimmermann, H. (2019) Permanent and transitory price shocks in commodity futures markets and their relation to speculation. Empirical Economics. 56. 10.1007/s00181-017-1387-2.

Habib, M.M., and Kalamova, M.M. (2007). Are there oil currencies? the real exchange rate of oilexporting countries. European Central Bank Working Paper 839

Hamilton, J.D., (2014). The Changing Face of World Oil Markets. w20355.

Haykin, S. (1994) Neural Networks: A Comprehensive Foundation. Prentice Hall PTR, Upper Saddle River

Hotelling, H., (1931). The Economics of Exhaustible Resources. Journal of Political Economy, 39(2). London: SAGE Publications Ltd, pp. 119.

Huang, L., and Wang, J. (2018). Global crude oil price prediction and synchronization-based accuracy evaluation using random wavelet neural network. Energy, 151, 875–888.

Huntington, H. G. (2005). The economic consequences of higher crude oil prices. Energy Modeling Special Report, 9. Energy Modeling Forum workshop (February 8, 2005), at Marymount University, Ballston Campus, Arlington, Virginia.

INTERNATIONAL ENERGY AGENCY, (2008). World Energy Outlook 2008. Paris and Washington, D.C.: Organization for Economic Co-operation and Development, 2008, pp. 569. pp. 569.

Jammazi, R., and Aloui, C., (2012) “Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling”. Energy Economics, 34, pp 828–841.

Jensen, G.R., Johnson, R.R. and Mercer, J.M., (2002). Tactical asset allocation and commodity futures. Journal of Portfolio Management, 28(4), p.100.

Kashman, A. and Nwulu, N.I., (2011) “Intelligent Prediction of Crude Oil Price Using Support Vector Machines”, 9th International Symposium on Applied Machine Intelligence and Informatics.

Kaufmann, R.K. and Ullman, B., (2009) Oil prices, speculation, and fundamentals: Interpreting causal relations among spot and futures prices. Energy economics, 31(4), pp. 550-558

Kawamoto, K. and Hamori, S. (2011) “Market Efficiency among Futures with different maturities: evidence from the Crude Oil Futures Market”. The Journal of Futures Markets, Vol 31, No.5, pp. 487-501.

Khalifa, A., Caporin, M., and Hammoudeh, S., (2017). The Relationship between Oil Prices and Rig Counts: The Importance of Lags. Energy Economics, 63, pp. 213-226.

Kilian, K. and Park, C. (2009) The impact of oil price shocks on the US stock market. International Economic Review, 50 (4) pp. 1267-1287

Kilian, L. and Lewis, L.T. (2011) Does the Fed Respond to Oil Price Shocks? The Economic Journal. vol 121, 555, 1047-1072

Kilian, L. and Murphy, D.P., (2014). The role of inventories and speculative trading in the global market for crude oil. Journal of applied econometrics (Chichester, England), 29(3), pp. 454-478.

Kilian, L. and Zhou X. (2023) A broader perspective on the inflationary effects of energy price shocks, Energy Economics, Volume 125, 106893, ISSN 0140-9883

Kilian, L., (2006). Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. American Economic Review, 99(3), pp. 1053-1069.

Kilian, L., (2008). Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy? Review of Economics and Statistics 90 (2): 216–40

Kilian, L., (2009). Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market. American Economic Review, 99(3), pp. 1053-1069.

Kilian, L., Rebucci, A. and Spatafora, N. (2009) Oil shocks and external balances J. Int. Econ., 77 (2), pp. 181-194

King, K., Deng, A. and Metz, D., (2011) "An Econometric Analysis of Oil Price Movements: The Role of Political Events and Economic News, Financial Trading and Market Fundamentals", Bates White Economic Consulting.

Kleinberg, R.L., Paltsev, S., Ebinger, C.K.E., Hobbs, D.A., and Boersma, T., 2017. Tight oil market dynamics: Benchmarks, breakeven points, and inelasticities. Energy Economics, 70, pp. 70-83.

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

Kulkarni, S., and Haidar, I., (2009) "Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices". International Journal of Computer Science and Information Security, 2,1, June.

Kwiatkowski, D., Phillips, P.C., Schmidt, P. and 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? J. Econ., 54, 159–178.

Lamm, D., (2013) “Time Series Modeling of Monthly WTI Crude Oil Returns” Lincoln College, University of Oxford.

Liu, L. and Hanssens, D.M., (1982) Identification of multiple-input transfer function models. Communications in statistics - Theory and methods, 11(3), pp. 297-314.

Manescu, C. and Van Robays, I., (2014) “Forecasting the Brent Oil Price Addressing Time-Variation in Forecast Performance”, European Central Bank, NO 1735/September

Merino, A. and Albacete, R., (2010) “Econometric modeling for short-term oil price forecasting”, OPEC Energy Review, March.

Miao, H., Ramchander, S., Wang, T., and Yang, D. (2017) Influential factors in crude oil price forecasting. Energy Economics, 68, 77-88.

Mingming, T., and Jinliang, Z., (2012) “A multiple adaptive wavelet recurrent neural network model to analyze crude oil prices”, Journal of Economics and Business, vol 64, pp 275-286.

Mirzoev, M.T.N., Zhu, L., Yang, Y., Zhang, M.T., Roos, M.E. and Pescatori, M.A.; and Matsumoto, M.A. (2020) The Future of Oil and Fiscal Sustainability in the GCC Region; International Monetary Fund: Washington, DC, USA.

Mohamued, E.A., Ahmed, M., Pypłacz, P., Liczmańska-Kopcewicz, K. and Khan, M.A. (2021) Global Oil Price and Innovation for Sustainability: The Impact of R&D Spending, Oil Price and Oil Price Volatility on GHG Emissions. Energies. 14(6):1757

Moosa, I.A. and N.E. Al-Loughani. (1994) “Unbiasedness and Time Varying Risk Premia in the Crude Oil Future Market”. Energy Economics, 16, pp 99-105.

Morana, C., (2001) “A Semiparametric Approach to Short-term Oil Price Forecasting”, Energy Economics 23, pp. 325–338.

Murat, A. and Tokat, E. (2009) Forecasting oil price movements with crack spread futures, Energy Economics, Volume 31, Issue 1, Pages 85-90.

Nakov, A., and Nuño, G., (2013) “Saudi Arabia and the Oil Market”, The Economic Journal, 123, pp 1333-1362.

Narayan, P.K., and Narayan, S. (2007) “Modelling oil price volatility”. Energy Policy 35, pp 6549–6553.

Ng, A. (2019), Machine Learning and AI via Brain simulations, Stanford University

OECD PUBLISHING and INTERNATIONAL ENERGY AGENCY, 2005. Resources to Reserves: Oil and Gas Technologies for the Energy Markets of the Future. Paris: International Energy Agency.

Ou, P., and Wang, H., (2011) “Application of Neural Networks in modeling and forecasting volatility of crude oil markets: Evidences from US and China”, Advanced Materials Research, 230-232, pp953-957.

Pagano, P. and Pisani, (2009) M.,” Risk-Adjusted Forecasts of Oil Prices”, The B.E. Journal of Macroeconomics, vol 9, 1, 24.

Panella, M., Barcellona, F., Santucci, V., and D’Ecclesia, L.D., (2011) “Neural Networks to Model Energy Commodity Price Dynamics”.

Pankratz, A., (1991) Forecasting with Dynamic Regression Models. Wiley-Interscience

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

Qunli, W., Ge, H., and Xiaodong, C., (2011) “Crude Oil Price Forecasting with an Improved Model Based on Wavelet Transform and RBF Neural Network”, International Forum on Information Technology Applications.

Reeve T.A. and R.J. Vigfusson. (2011) "Evaluating the Forecasting Performance of Commodity Futures Prices". International Finance Discussion Papers 1025, Board of Governors of the Federal Reserve System (U.S.)

Sadorsky, P. (2006) “Modeling and forecasting petroleum futures volatility”. Energy Econ. 28, 467–488.

Salisu, Afees A. and Isah, Kazeem O. & Oyewole, Oluwatomisin J. & Akanni, Lateef O., 2017. "Modelling oil price-inflation nexus: The role of asymmetries," Energy, Elsevier, vol. 125(C), pages 97-106.

Sandrea, I., (2014). US shale gas and tight oil industry performance: challenges and opportunities. The Oxford Institute for Energy Studies.

Schmalensee, R., Ashley, R., and Granger, C.W.J., (1980). Advertising and Aggregate Consumption: An Analysis of Causality. Econometrica, 48(5), pp. 1149-67.

Schwartz, E., and Smith, J. E. (2000) “Short-term Variations and Long-term Dynamics in Commodity Prices”, Management Science 46, 893–911.

Schwartz, E.S., (1997) “The stochastic behavior of commodity prices: Implications for valuation and hedging”, Journal of Finance 52, Part 3, pp. 923-973.

Schwarz, G., (1978). Estimating the Dimension of a Model. The Annals of Statistics. Baltimore, MD: Institute of Mathematical Statistics, pp. 461-464.

Sek, S. K., Teo, X. Q., and Wong, Y. N. (2015). A comparative study on the effects of oil price changes on inflation. “4th World Conference on Business, Economics and Management, WCBEM”. Procedia Economics and Finance, 26(2015), 630–636.

Sekine, A. (2020). "Oil price pass-through to consumer prices and the inflationary environment: a STAR approach," Applied Economics Letters, Taylor & Francis Journals, vol. 27(6), pages 484-488, March.

Shabri, A., and Samsudin, R., (2014) “Crude Oil Price Forecasting Based on Hybridizing Wavelet Multiple Linear Regression Model, Particle Swarm Optimization Techniques, and Principal Component Analysis”. The Scientific World Journal.

Skododumov, B., (2008) “Estimation of mean reversion in Oil and Gas markets”, Mitsui Technical Report.

Smil, V. (2017). Energy and Civilization: A History (REV-Revised, 2). The MIT Press. http://www.jstor.org/stable/j.ctt1pwt6jj

Stevens, J., and de Lamirande, P. (2014) “Testing the efficiency of the futures market for crude oil in the presence of a structural break”. Applied Economics, 46:33, pp 4053-4059.

Sturm, M., Gurtner, F. and González Alegre, J. (2009). "Fiscal policy challenges in oil-exporting countries: a review of key issues," Occasional Paper Series 104, European Central Bank.

Syzdykova, A., Abubakirova, A., Kudabayeva, L., Zhantayeva, A., and Omarova, A (2022). Asymmetric causality relationship between oil prices and inflation in BRIC Countries. In: International Journal of Energy Economics and Policy 12 (3), S. 184 - 191.

The oil industry: Steady as she goes. (2006). The Economist, 379, (8474).

Timmermann, A., “Forecast Combinations”, Handbook of economic forecasting, vol 1, pp 135-196.

Toda, H.Y, and Yamamoto, T., (1995). Statistical inference in vector autoregressive with possibly integrated processes. Journal of Econometrics, 66(1), pp. 225-250.

Toews, G., and Naumov, A., (2015). The Relationship Between Oil Price and Costs in the Oil Industry. The Energy Journal, 36(1), pp. 1.

Verleger Jr, P.K. (2016) Structure matters: oil markets enter the Adelman era Energy J., 36 (2016), pp. 129-157

Wang, J., Niu, T., Du, P., and Yang, W. (2020). Ensemble probabilistic prediction approach for modeling uncertainty in crude oil price. Applied Soft Computing, 95, 106509

Wang, T., and Yang, J., (2010) “Nonlinearity and intraday efficiency tests on energy futures markets”. Energy Economics, 32, pp 496-503.

Wei, Y., Wang, Y.D., and Huang, D.S. (2010) “Forecasting crude oil market volatility: further evidence using GARCH-class models”. Energy Economics. 32, pp.1477–1484.

Wood, S.N., Goude, Y. and Shaw S. (2015) Generalized additive models for large datasets. Journal of the Royal Statistical Society, Series C 64(1): 139-155.

Xie, W., Yu, L., Xu, S., and Wang, S., (2006) “A New Method for Crude Oil Price Forecasting Based on Support Vector Machines”, Computational Science, 3994, pp 444-451.

Ye, M., Zyren, J. and Shore, J., (2006) “Forecasting short-run crude oil price using high- and low-inventory variables”, Energy Policy, 34, pp2736-2743

Yergin, D., 2008. Oil at the "Break Point". Special Report, Cambridge Energy Research Associates.

Yousefi, S., Weinreich, I., and Reinarz, D., (2005) “Wavelet-based prediction of oil prices”. Chaos, Solitions and Fractals, 25, pp 265–275.

Zagaglia, P., (2010) “Macroeconomic Factors and Oil Futures Prices: A Data-Rich Model”, Energy Economics, vol 32, 2, pp 409-417.

Zamani, M., “An Econometrics Forecasting Model of Short-Term Oil Spot Price”, (2004) 6th IAEE European Conference

Zhang, J-L., Zhang, Y-J., and Zhang, L. (2015) “A novel hybrid method for crude oil price forecasting”, Energy Economics, 49, pp 649-659.

Zhang, X., Wu, Q., and Zhang, J., (2010) “Crude Oil Price Forecasting Using Fuzzy Time Series”, 3rd International Symposium on Knowledge Acquisition and Modeling.

Appendix

Table 1A. RMSE error measures for different forecasting methods

	2017Q1	2017Q2	2017Q3	2017Q4	2018Q1	2018Q2	2018Q3	2018Q4	2019Q1
Constant	4.144	8.880	18.581	15.593	8.447	6.380	8.179	13.759	6.934
Futures	5.636	7.773	15.878	15.226	8.143	5.323	8.619	15.330	9.668
BBG Analysts Median	3.579	5.715	9.903	15.761	14.872	7.939	3.620	8.961	6.447
Department of Energy EIA	4.122	7.213	16.305	17.359	12.688	9.786	3.722	11.440	5.120
GAMLTF Forecasted Inputs	**5.233**	**7.560**	**14.174**	**12.368**	**10.994**	**6.288**	**6.422**	**14.046**	**4.808**
GAMLTF Actual Inputs	**9.409**	**4.215**	**6.703**	**3.854**	**4.197**	**8.550**	**3.273**	**7.585**	**5.452**
LTF Actual Inputs no GAM	10.088	5.606	6.505	7.905	6.736	16.029	2.616	2.757	4.840

	2019Q2	2019Q3	2019Q4	2020Q1	2020Q2	2020Q3	2020Q4	2021Q1	2021Q2
Constant	7.253	16.155	16.047	22.900	16.258	17.497	23.071	22.156	16.696
Futures	7.938	15.825	13.880	20.262	12.529	15.967	20.520	21.594	21.916
BBG Analysts Median	11.110	19.716	16.965	17.981	6.908	16.904	19.816	24.502	21.498
Department of Energy EIA	6.009	18.994	15.179	18.653	15.237	12.696	18.851	20.654	20.239
GAMLTF Forecasted Inputs	**7.748**	**17.138**	**20.165**	**17.989**	**16.629**	**14.133**	**8.360**	**6.372**	**11.279**
GAMLTF Actual Inputs	**4.120**	**5.043**	**7.977**	**5.408**	**12.576**	**11.860**	**9.243**	**4.998**	**5.232**
LTF Actual Inputs no GAM	12.689	35.927	44.212	40.622	16.969	18.266	13.300	7.160	10.398

	2021Q3	2021Q4	2022Q1	2022Q2	2022Q3	2022Q4	2023Q1	**TOTAL**
Constant	22.182	23.929	24.366	20.987	31.913	7.979	3.004	17.040
Futures	25.044	24.891	24.864	7.277	11.495	5.855	3.954	15.323
BBG Analysts Median	27.313	28.984	23.360	10.227	14.378	13.291	9.052	16.031
Department of Energy EIA	26.232	25.283	24.485	10.465	11.659	10.816	3.961	15.400
GAMLTF Forecasted Inputs	**15.823**	**12.759**	**15.853**	**30.541**	**27.898**	**8.644**	**5.754**	**14.341**
GAMLTF Actual Inputs	**16.721**	**17.960**	**22.357**	**25.347**	**18.111**	**8.847**	**3.981**	**11.123**
LTF Actual Inputs no GAM	10.190	9.670	11.454	40.561	37.043	9.519	6.019	20.065

Note: This table reports the forecasting performance in terms of the RMSE measure of the proposed framework for forecasted and actual input data as well as alternative benchmarks, including the LTF framework with no GAM. The in-sample period is 1995-2016, and the out-of-sample or forecasting period is 2017-2023. Forecasts are performed for the next four quarters. The following forecasting methods are considered:
No-change: Forecasts are the average price of the previous month for the whole forecast period.
Futures: Forecasts are the average of Brent 1st, 2nd and 3rd month contracts for the first quarter, 4th, 5th and 6th month contracts for the second quarter the day before beginning the period of forecast.
BBG: Bloomberg quarterly surveys are taken as forecasts the day before beginning the period of forecast.
EIA: Average monthly forecasts to create quarterly forecasts are taken from the last EIA report before be-ginning the period of forecast.
GAMLFT with Forecasted Inputs: proposed new model fed by forecasted inputs. Highlighted in bold.
GAMLFT with Actual Inputs: proposed new model fed by actual inputs. Highlighted in bold.
LFT with Actual Inputs no GAM: Linear Function Transfer model fed by actual inputs.