Three essays on crude oil markets

Three Essays on Crude Oil Markets

Carlo Piccari

A dissertation submitted in fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

Birkbeck, University of London.

School of Computing and Mathematical Sciences

Birkbeck, University of London

November 5, 2024

I, Carlo Piccari, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

The oil markets have been at the centre of attention of researchers over the past 50 years. Remarkably, the last 10 years on which the three chapters are focused, brought a large shares of events: starting with the shock collapse of prices in 2014 due to the massive arrival of shale oil in the US output, followed by price recovery and another, steeper, crash in 2020 due to the dramatic effect of COVID-19. In the former case supply was the major explanation of the collapse; in the latter demand vanished away because the world had come to a standstill. We firstly analysed the relationship between lagged WTI oil prices and the credit default swaps of 11 oil producers and 7 oil servicing companies of different sizes and examine how it varies through time. Our findings show a significant inverse effect for over 60% of the dataset. Moreover, this relationship grows in magnitude during periods of high volatility. Then, we applied a Variational Mode Decomposition-Neural Network and -Generalised Additive Models ensemble to forecast 5-minute WTI and BRENT prices in 2020 and in 2024. We highlighted the difficulties due to the structural instability caused by the unprecedented drop of WTI prices in 2020, which do not appear when forecasting data observed in other time-periods. Lastly, we proposed a methodology to account for the counter-intuitive features of sentiment analysis of oil-related news articles over a two years period, encompassing the events of the Ukraine war and Gaza conflict. We showed how adopting a scoring system of words co-occurrences and war-related nouns leads to increased predictability power of next day WTI returns, historical and conditional volatility. Topic modelling coupled with our sentiment measures was applied to forecast WTI volatility and returns, displaying higher accuracy over measures obtained via a standard sentiment analyser.

Acknowledgements

First of all, I would like to express my deepest gratitude towards my supervisors, Professor Helyette Geman and Dr. Alessandro Provetti, for their constant support throughout this long journey. They encouraged me to persevere with my work at a time of deep personal struggle and it is thanks to them that I was able to complete this thesis. I would like to extend my sincere thanks for his support to Professor Ron Smith who, in many occasions, was able to clear my doubts and provide great inspiration. I also thank Dr. Ilaria Peri for the encouragement she gave me during the most difficult times. Further, I am also grateful to my friends and colleagues at Birkbeck: Jose Camarena Brenes, Federico Corrado, Charisios Grivas, whom I shared this experience with. Lastly, but most importantly, I am deeply grateful to my parents, who made this dream achievable.

I gratefully acknowledge the funding that made my PhD possible. I was funded by the Birkbeck Business School through a student scholarship.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the last few decades, crude oil has emerged as the most important commodity traded. There are over 200 grades of crude produced worldwide. The American Petroleum Institute (API) sets the density classification standards of the oil grades. “Light” crude, with an API Gravity above 10 is usually preferred, because of the small amount of residues it contains. Similarly, a lower the amount of sulphur makes the oil “sweeter” and preferred to “sour” grades because of the reduced amount of SO2 emissions. Aside of the conventional extractions methodology which allows the release of oil through pressure, new extractions procedures were introduced in the past two decades. The technology that most impacted the oil markets was the process of hydraulic fracturing, or fracking. This new technique, which rose in popularity in the United States around 2015, allows the recovery of a greater quantity of petroleum and natural gas by creating fractures in shale rock formations through the injection of specialised fluids. Called “shale oil”, this grade of crude is light and sweet, hence often seen by Saudi Arabia as a threat to its heavy and sour counterpart.

The Canadian oil sands are another type of crude which rose in production as of 2012. Due to its composition, namely sand, clay, water and bitumen, it is considered a “heavy” quality of crude oil which is typically harder to recover. The new technologies introduced in the early 2010s allow for the extraction of bitumen directly from oil sands deposits found deeper underground, increasing the recovery rate without disrupting the surface land.

The two main oil benchmarks, West Texas Intermediate (WTI) and Brent, are reference prices in the USA and worldwide, respectively. The major players in the oil markets are the United States, OPEC and Russia, among the producers, Europe and

China, among the consumers. The Organisation of the Petroleum Exporting Countries (OPEC) was founded in 1960 by the governments of Saudi Arabia, Iran, Venezuela, Iraq and Kuwait in order to form a cartel regulating the member's oil production and, as a consequence, oil prices. Saudi Arabia holds the major reserves out of the OPEC countries, followed by Iran and Venezuela.

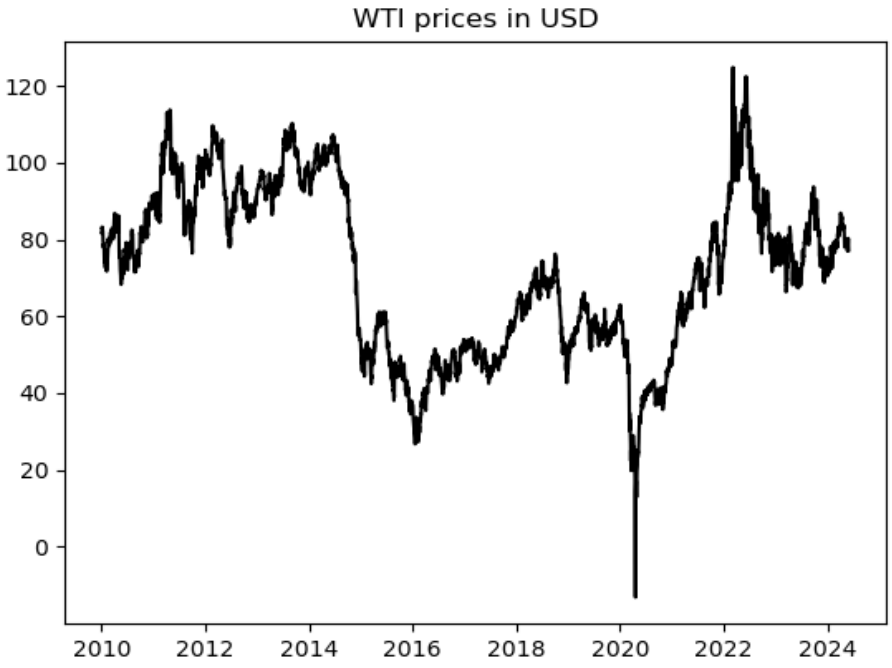


Figure 1.1: WTI price trajectory between January 2010 and June 2024.

A number of factors need to be considered when trying to explain crude oil prices movements. Similarly to all other commodities, the basic principles of the supply and demand also apply to crude oil. Everything else equal, prices are expected to grow if demand increases. On the other hand, prices are expected to drop if supply grows. However, the cost of a barrel of crude is also affected by geopolitical factors like trade tensions between countries, wars and sanctions, among others. Below, we discussed some of the most important geopolitical events that affected crude oil prices in the last decade.

With the Russian's invasion of Crimea (Ukraine), sanctions against Iran and other issues in the middle East contributed to the growth of crude prices to $100 during 2013 and the first half of 2014; the increase of US shale production coupled with a stagnant

global economy and lower oil demand from China and Europe led to the price crash observed in June 2014. By the end of 2014, Brent and WTI reached their lowest prices since 2009. The many attempts of poorer OPEC countries to reduce the cartel's crude output were blocked by Saudi Arabia, who saw the US Congress cancellation of the oil exports' ban in 2015 as a threat to the Kingdom's main source of income. Prices did not recover until 2017, helped by OPEC's production cuts, the introduction of Russia to OPEC+ and the increase of US consumption.

In 2018 the United States became the number one worldwide oil producer, surpassing Saudi Arabia and Russia for the first time in history. The high US production and lift of the sanctions on Iran in 2019 meant that the market was once again oversupplied, ahead of the beginning of the COVID-19 pandemic. In early 2020, the closure of factories around the world and the halt of daily commutes and leisure travels plunged the global consumption to unexpected lows. OPEC did not agree to cut production until April 2020, a decision that was mirrored by the United States. However, despite the acquisition of some 75 million barrels by the American Government, the US storage facilities were reaching their maximum capacity. On April 20th, WTI contracts for delivery in May traded below zero, meaning that for the first time in history traders were paying to sell their barrels of oil. Contracts for later delivery were trading at higher prices as they included a premium for the increasing storage costs, displaying a market feature which was labelled as "super-contango."

2021 was the year of recovery of the crude markets thanks to the advancements in COVID-19 vaccines and the progressive reduction of quarantine measures worldwide. The oil embargo and price cap imposed on Russia's crude as a consequence of its invasion of Ukraine led prices to surpass the $100 mark once again in 2022. WTI prices remained stable over $70 through 2023 and the first quarter of 2024 despite the Israel-Gaza conflict which had no visible impact on oil prices.

Turning to methodology, we applied a combination of classical and modern approaches to argue our theories. Firstly, in Chapter 2 we investigated the links between WTI prices and the risk of default of American crude oil producers and servicing companies through an AutoRegressive Distributed Lag (ARDL) model estimated via Generalised Autoregressive Conditional Heteroskedasticity (GARCH), which, to the best

of our knowledge, has not been analysed in this manner before.

Secondly, in Chapter 3 we applied Variational Mode Decomposition to extract from non-stationary time series of WTI and Brent prices a number of stationary functions, called modes. The modes were then used to train a Neural Network and a Generalised Additive Model to forecast oil prices in various time periods. To overcome limitations of previous works that applied VMD, we presented an innovative way to compute the forecast of the test dataset using exclusively past data, called Recursive Forecast Methodology.

Lastly, in Chapter 4 we analysed the ever-changing relation between crude oil-related news articles and next day WTI prices, historical and conditional volatility. Our innovative framework, which accounts for the counter-intuitive sentiment features found in crude oil-related news articles, achieved higher predictability power measured via Granger Causality and forecast accuracy when coupled with Topic Modelling techniques.

Chapter 2

Exploring the determinants of CDS premia: the case of oil producing and servicing companies

Abstract

This chapter aims to extend the literature on corporate credit spreads in order to *i*) benefit from the liquidity of credit default swaps and analyse their premia; *ii*) focus on the sector of oil producing and servicing companies and its large amount of outstanding debt; *iii*) propose lagged WTI returns as a determinant of CDS premia in a novel way in the credit literature. We found evidence in favour of our theory for eight out of eleven oil producers and four out of seven oil servicing companies. We showed that an increase in oil returns leads to a reduction of the following week's CDS spreads. Furthermore, this effect appeared to be stronger for companies with smaller market capitalisation. Lastly, we observed a negative Pearson correlation between the CDS spreads of over 60% of the companies in analysis and West Texas Intermediate (WTI) returns, with increasing magnitude during periods of high volatility. In parallel, the sensitivity of CDS spreads to oil returns was stronger during the same period for both producers and servicing companies. Interestingly, we found that, for a number of firms, the correlation became positive during periods of steadily growing prices, namely 2017 and 2018.

2.1 Introduction

In the last few decades, crude oil has emerged as the most important commodity traded. With over 200 grades of crude oil extracted all around the world, the prices of WTI and Brent are the main benchmarks in the USA and worldwide, with daily trading volumes reaching 74 million for the former and 267.7 million for the latter in 2023 [1] Remarkably, Brent has recently become the world reference index, with China competing with the US as the first world economy and the geographical position of the UK halfway between the East and the West making it a central market.

Many factors need to be considered when trying to explain crude prices movements. A number of them are geopolitical with key players being the US, Russia, Saudi Arabia and OPEC countries in general playing a central role in the output decisions, leading to changes in supply; while the demand is driven by the world economy and exhibits less abrupt changes (except for the COVID-19 period). The shale revolution that started in early 2015 has contributed to a gigantic increase of oil and gas output from the United States. After cancellation of the oil exports' ban by the US Congress in 2015, the country has become an exporter of natural gas and oil products. Moreover, as of 2018, the United States became the first oil producing country, with an output over 2 million barrels per day greater than Russia and Saudi Arabia [2]

Oil producers can be split in three different categories: upstream, midstream, downstream. The upstream industry explores and produces crude. Firms belonging to the midstream category are responsible for storing crude and shipping it via vessels or pipelines. The downstream sector involves the refinement of the raw commodity, as well as the marketing and distribution of products derived from crude oil.

US crude oil output, according to the Energy Information Administration (EIA) 2023 Annual Energy Outlook [3], should grow from the 12.9 million bbls per day produced in

[1] Source: Intercontinental Exchange (ICE). https://www.ice.com/report/7

[2] Source: the US Energy Information Administration (EIA). https://www.eia.gov/todayinenergy/detail.php?id=61545

[3] Source: the EIA Annual Energy Outlook 2023. https://www.eia.gov/outlooks/aeo/data/browser/#/?id=1-AEO2023&region=0-0&cases=ref2023&start=2021&end=2050&f=A&linechart=ref2023-d020623a.3-1-AEO2023&ctype=linechart&sid=ref2023-d020623a.3-1-AEO2023&sourcekey=0

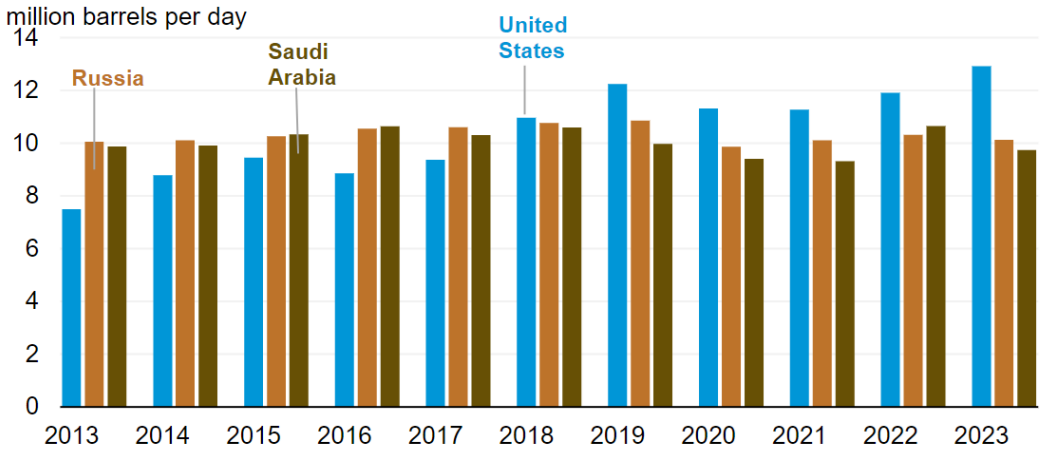


Figure 2.1: Russia, Saudi Arabia and United States crude oil production in million barrels per day between 2013 and 2023. Source: the US Energy Information Administration.

2023 to 14 million bbls per day in 2030 and then plateau until 2050.

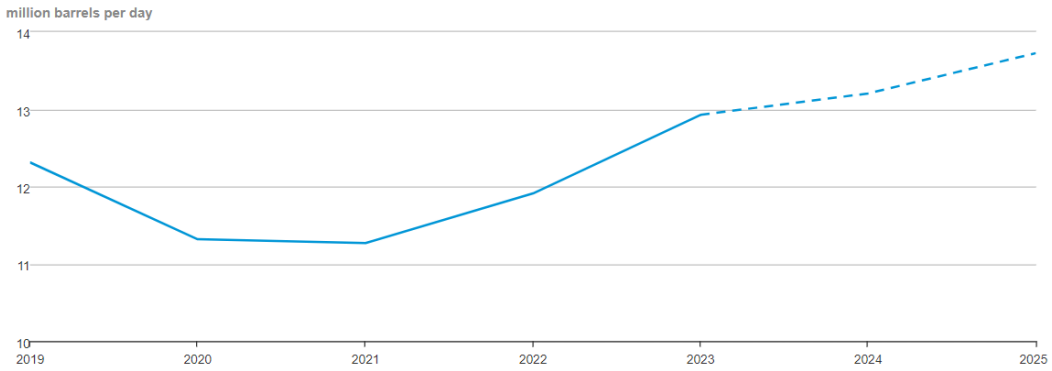
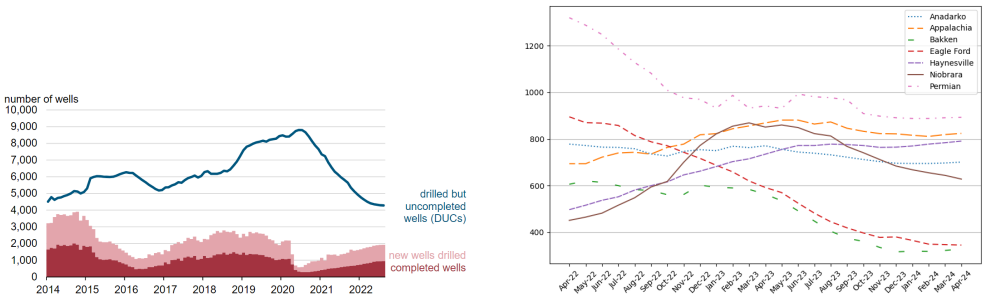


Figure 2.2: US crude oil observed production 2019-2023 and forecast for 2024-2025. Source: the US Energy Information Administration

The number of drilled but uncompleted oil and gas wells in the US halved from the peak reached in 2020, as depicted in Figure 2.3a and 2.3b, signalling a recovery of the sector in the years post COVID-19.

In 2024, the Energy Information Administration adjusted their US crude production forecast from 260.000 barrels per day (bpd) to 280.000 bpd for 2024 and from 460.000 bpd to 510.000 bpd for 2025. Under the new scenario, it is expected that the US oil output will reach 13.21 mn bpd in 2024 and 13.72 mn bpd the following year. This adjustment is due to expected higher prices in 2024 for both WTI and Brent.

On the consumption side, the EIA expects the US to consume 20.4 mn bpd in 2024, up by 200.000 bpd from the previous year, whilst the global consumption is forecasted to



(a) **(b)**

Figure 2.3: a) Number of drilled but uncompleted oil and gas wells in the US from January 2014 to June 2022. Source: the US Energy Information Administration. b) Number of drilled but uncompleted oil and gas wells by region in the US from April 2022 to April 2024. Source: www.statista.com

grow to almost 103 mn bpd. Moreover, the sanctions imposed on Russia at the beginning of 2023 as a consequence of the invasion of Ukraine limited its exports of natural gas and crude oil to Europe and the United States, meaning that the increase in global demand in the mid-term is expected to be met by non-OPEC regions, namely Norther and Southern America [4].

Reserves of a country, like the oil inventory held by an exchange, play a key role in explaining the properties of a given regional market. The Theory of Storage introduced by Kaldor (1939) analyses, among other properties, the relationship between inventory, shape of the forward curve and commodity price volatility - a subject further investigated by Geman & Ohana (2009) in the case of the US oil and natural gas markets. For an oil company, the debt granted by banks is defined by the size of its reserves, according to the principle of "Reserves Based Lending" (RBL). It is clear that the reduction of revenues, in the case of lower oil prices or reserves, threatens the ability of the company to meet its financial obligations and increases the probability of default. Moreover, exploration companies' investments are financed through debt to a great extent: in the situation of a steep growth of the firm's leverage ratio, increasing financing costs and tighter borrowing conditions might force firms to suspend investments in new projects and technologies or even sell their assets to generate new liquidity.

Between the beginning of 2009 and 2015, US oil production grew by more than 70%, with the bulk of the increase coming from shale oil. Obviously, the conjunction of the

[4]Source: the EIA International Energy Outlook 2023. `https://www.eia.gov/outlooks/ieo/pdf/IEO2023_Narrative.pdf`

quantitative easing that followed the financial crisis of 2008 together with the vibrant popularity of shale oil led to very large amounts of debt borrowed after 2008 by oil firms, using oil reserves and revenues as collateral. The debt borne by the oil and gas sector was multiplied by two and a half between 2006 and 2014, from roughly $1 trillion to around $2.5 trillion as oil prices were sharply declining (see Figure 2.4). As a consequence of the price drop in 2014, over 160 oil patches and 170 oilfield services companies filed for bankruptcy between 2015 and August 2018 (Boone, 2020). The number of cases decreased after 2016, whilst the value of E&P debt under Chapter 11 in the first eight months of 2018 was higher than the whole of 2017. Out of the $150 billion total outstanding secured and unsecured debt, over $60 billion debts were converted in equities in 39 E&P filings. The major bankruptcies were worth $8 billion (Seadrill Limited), $5.3 billion (Odebrecht) and over $3 billion (Ocean Rig, CGG Holding and Pacific Drilling). Furthermore, recovery rates were low, reaching an average of 20% of the notional amount, compared to an average of 60% for all defaults before 2015. Moreover, the combined total of bankruptcy debt surpassed $56 billion in 2020 alone, with an unprecedented record high of $1.2 billion average debt per bankruptcy filed (Boone, 2021).

Portfolio managers and bondholders, who invest in oil mining businesses and want to manage their credit risk, are likely to hedge their exposures by purchasing credit derivatives. Credit Default Swaps (CDS) are the most popular credit derivatives because of the simplicity of their design in providing insurance against the risk of default of the reference entity. When a credit event occurs, the seller of the insurance has to purchase the corporate bonds for their face value from the buyer of the CDS. In order to obtain this protection, the buyer of the CDS makes periodic payments to the seller until the end of the life of the CDS or until a credit event occurs. CDS premia are expressed in basis points of the notional principal.

The increase in oil companies CDS premia that occurred in conjunction to the price drop of crude oil that started in mid-2014 has been remarkable. Whilst the importance of crude prices is known to market participants who trade oil miners' shares, no previous research, to the best of our knowledge, has exhibited the important correlation between oil prices and the credit rating of its producers. Accordingly, the aim of our research is to

investigate the role of crude oil prices as a determinant of credit default swaps. Another contribution to the literature is the analysis of a number of state variables that have not been included in previous research, like the Markit CDX Investment Grade index.

Results are in agreement with the theory, displaying a negative correlation between crude oil prices and CDS premia, with an increasing intensity during times of high volatility. There are, however, cases where the correlation became positive when prices were stable around $50 from mid-2016 until Q3 2017. Additionally, default spreads are found positively correlated to leverage ratio, one of the determinants of default risk suggested by the founding paper of Merton (1974); however, we do not find evidence supporting firm-specific equity volatility as a significant determinant for the firms' CDS.

The chapter is organised as follows: Section 2.2 presents a review of the literature as well as our proposed approach and novel state variables to explain the CDS premia. Section 2.4 describes the data used, Section 2.5 presents the methodology applied whilst Section 2.6 reports the result obtained. Section 2.7 concludes the chapter.

2.2 Literature Review

Up to the last decade, the literature on firms' credit features had mainly focused on corporate bond spreads; the interest in credit derivatives increased only after the early 2000's. Previous approaches can be differentiated between "structural" and "reduced-form" models. The reduced-form approach tries to explain credit derivatives' price dynamics by modelling stochastic default probabilities as exogenous and obtaining parameters from market data. These models put default intensity at the centre of the pricing process and have shown a certain degree of versatility in practical approaches as they require a crucial assumption of the recovery rate in the case of default. On the other hand, structural models are based on the founding papers of Black & Scholes (1973) and Merton (1974). The valuation of credit-related instruments depends on the firm's probability of default, namely when the face value of its debt exceeds the value of its assets (the single state variable) at maturity of the debt. The likelihood of a credit event is related to the three main parameters: the firm's leverage ratio, its share volatility and the risk-free interest rate.

In the literature, Longstaff & Schwartz (1995) exhibit a negative correlation between

interest rates and credit spreads. Collin-Dufresne et al. (2001) show that theoretical variables have little power to explain credit spread changes and point out that residuals are highly correlated. Further, they point out that residuals are driven by a single common factor and are highly cross-correlated. Campbell & Taksler (2002) exhibit the importance of firm-specific equity volatility as a determinant for corporate bond yields. Moreover, idiosyncratic firm volatility and credit ratings are found to explain the same level of cross-sectional variation in bond yields. Ericsson et al. (2009) extend this result to CDS premia across various sectors of the economy by analysing the relationship between theoretical determinants of default risk and CDS spreads. They show that 10-year US treasury bonds, firm specific leverage ratios and equity volatility are statistically and economically significant, whilst a principal component analysis confirms that the remaining variation of data not explained by the model has to be attributed to one common factor. Domanski et al. (2015) emphasise the build-up of oil-related debt in their period of analysis and observe that the total debt of the oil and gas sector was standing at 2.5 trillion dollars at the beginning of 2015, more than twice its size of one trillion dollars at the end of 2006. Figuerola Ferretti & Cervera (2018) examine the link between oil prices and credit default swaps as a proxy for credit risk. More specifically, using the "multiple bubble" methodology, they show the existence of two mildly explosive periods in CDS premia, one before the financial crisis and one after the 2014 crude oil collapse.

Another stream of literature that we view as related to our research comprises of the articles that analyse the impact of commodity prices on the share prices of the corresponding extracting industries. Strong (1991) examines this sensitivity in case of the the oil sector, Blose & C.P. Shieh (1995) studied the impact of gold prices on the value of the gold mining companies' stock. Tufano (1998) pursued and extended this analysis by investigating the determinants of stock prices exposure in the gold mining industry. Geman & Vergel Eleuterio (2013) investigated the sensitivities of fertilisers mining companies to fertilisers indexes as well as agricultural commodity prices during a period that contained the wheat and corn spikes of 2008.

We will show that in the case of oil companies, the sensitivity to oil prices depends on the size of the firm and its position in the production chain: for instance, downstream

refiners benefit from the lag between a movement in the input cost and the adjustment of the refined product price while the crackspread remains fairly stable. Downstream firms are also able to lock in their profits through the use of the so-called crackspreads. In contrast, upstream miners are, on average, more susceptible to a drop in the price of a barrel of crude as the production costs might exceed the market price of the final product.

2.3 Candidate determinants of CDS spreads

We propose to use a structural model to explain credit default swaps movements. As observed by Ericsson et al. (2009), the benefits of using CDS instead of corporate bond spreads as a measure of credit risk are numerous. Firstly, by the virtue of being "spreads", Credit Default Swaps avoid the addition of a "noise" created by a wrong choice of risk-free yield curve model or an incorrect procedure to remove the coupon effect in bond prices. Blanco et al. (2005) argue that CDS prices better reflect firm-specific credit risk in the short run, with bond credit spreads achieving a similar explanatory effect only in the long run.

In our approach, firm-specific CDS are explained by lagged oil prices, the Markit CDX Investment Grade Index in addition to the firm's leverage ratio and share price volatility. The motivation for the inclusion of each explanatory variable is presented in the following sections.

2.3.1 *Crude oil returns*

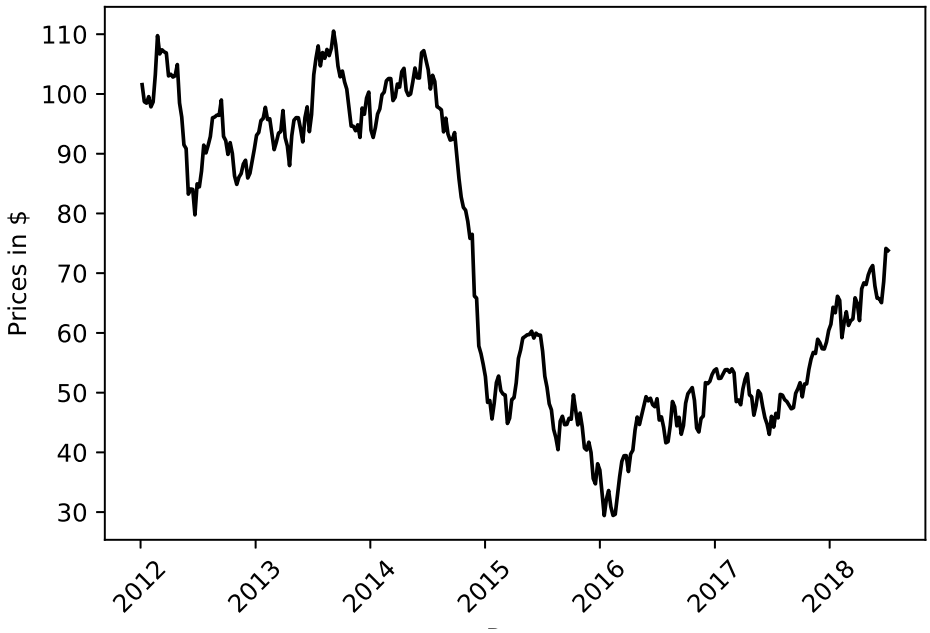


Figure 2.4: WTI price trajectory between 2012 and July 2018.

Selling barrels of crude oil is the main source of revenues for an oil producing company, whose other profits usually come from activities around oil transportation, refining and

financial trading.

Accordingly, when the barrel of crude was traded at over $100 in 2007 and the first half of 2014, oil producing firms were flourishing, generating enough cash flows to pay their creditors, reward their investors and finance new exploration projects. On the contrary, low crude prices meant tighter expenditure budgets and, if the driller's break-even point was not reached, struggle to face financial commitments and shareholders' dividends. Furthermore, banks grant loans to the oil and gas sector according to the rule of Reserved Based Lending, where reserves represent the debt collateral. The lender assesses a loan request on the consideration of the borrower's expected production and ability to generate revenues from it. Hence, the discounted future cash flows depend on the firm's "proven", or to a lesser extent "probable", reserves as well as oil prices. Considering that credit default swaps embed the company's probability of default, it is easy to understand the impact of oil prices on the spread paid by the CDS buyer.

To account for non-stationarity, we computed log returns of the WTI data, as explained in Section 2.5.1. We adopt one lag weekly WTI returns given the markets' delay in updating CDS premia of oil companies subsequent to changes in the cost of a barrel of crude. We expect that an increase (reduction) in WTI returns will reflect in a reduction (increase) of the firms' CDS spreads. As displayed in Table 2.1, the Pearson correlation between the two quantities over the whole time period is highly negative, with values ranging between -0.29 to -0.79 (-0.61 average) for oil producers and between 0.04 and -0.87 (-0.60 average) for oil servicing companies.

2.3.2 *Leverage ratio*

In all sectors of the economy, a highly leveraged firm is considered more likely to default and this is reflected in the premium requested to protect against the event of bankruptcy. Hence, we expect to uncover a positive relation between the two quantities. Our leverage ratio will classically be defined as

$$L = \frac{Debt}{Debt + Equity}$$

Oil Producing Firms	WTI - CDS correlation	Oil Servicing Firms	WTI - CDS correlation
Anadarko Petroleum Corp.	-0.506	Enbridge Inc.	-0.633
Apache Corp.	-0.760	Ensco Plc.	-0.891
Chesapeake Energy Corp.	-0.574	Halliburton Company	-0.531
ConocoPhillips	-0.582	Nabors Inc.	-0.868
Devon Energy Corp.	-0.580	Transocean Inc.	-0.839
Encana Corp.	-0.513	Valero Energy Corp	0.039
Marathon Oil	-0.608	Weatherford Int.	-0.825
Murphy Oil Corp.	-0.771		
Noble Energy Inc.	-0.740		
Pioneer Natural Resources	-0.286		
Whiting Petroleum	-0.788		
Average	-0.610	Average	-0.600

Table 2.1: correlations between U.S. oil prices and the CDS premia of the firms in analysis.

It captures at all times the firm capacity to repay its debt. Market Value of Equity is obtained by multiplying the number of outstanding shares by the corresponding weekly share price. The strong volatility of the latter over our period of analysis caused leverage ratios to vary significantly through time.

2.3.3 *Equity volatility*

As first exhibited in Merton (1974), the probability of default of a firm increases when the share price becomes more volatile. Consequently, the protection buyer will have to pay a higher premium to protect his exposure against a more risky entity. Our expectation in this paper is to uncover a positive relation between firm-specific equity volatility and the firm's credit default swaps. In contrast to Collin-Dufresne et al. (2001), who use implied volatility from VIX data, we choose to focus on firm-specific historical volatility derived from share returns time series. Therefore, we adopt the classic annualised volatility formula for weekly data

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R_i - \overline{R})^2} \cdot \sqrt{52}. \tag{2.1}$$

2.3.4 *Markit CDX Investment Grade Index*

Lastly, in order to represent the market "appetite" for corporate debt and its reward, we introduce the *Markit CDX Investment Grade Index* which covers North America and

exhibits a high liquidity and transparency.

2.4 Data

The sample consists of eleven crude oil producers, nine of which are based in the US, and seven oilfield servicing companies, two of which are located in the US, two in Europe, one in Canada and one has headquarters in Bermuda but operates mainly in the US. We expect that WTI will show a degree of statistical significance for non-US based firms included in the dataset given their world-wide operation network and exposure to the US market. The dataset comprises of weekly observations ranging over the period January 1st, 2012 and July 6th, 2018; the low CDS daily variation justifies the choice of data frequency. The study starts in 2012 to allow the inclusion, as an explanatory variable, of the Markit CDX Investment Grade Index, which has no data available prior to the end of 2011. Noble Energy Inc. and Whiting Petroleum have no data available prior to July 2012 and September 2015, respectively. All the data are obtained from Bloomberg. We use the WTI index, as opposed to other benchmarks like Brent, to represent crude oil prices as 70% of the firms in our dataset are US based. Table 2.2 reports the summary statistics of firm- and non-firm-specific variables averaged across sectors. Figures 2.6 and 2.7 in Section 2.8 show the plot of the firms CDS spreads over the period in analysis.

Table 2.3 reports the country of operation, market capitalisation and the most recent, at time of writing, Moody's short and long-term rating for each firm in both subsets. The market capitalisation of the oil producers ranges from a value of 2.6 billion USD, in the case of Whiting Petroleum, to a maximum of 76.2 billion USD, in the case of ConocoPhillips, averaging at $17.23 billion. On the other hand, the lowest market cap in the oil servicing firms dataset belongs to Weatherford Int. with 0.5 bln USD, whilst the highest is held by Enbridge Inc. with 71.9 billion USD, and a similar average value of $20 billion. The dataset does not include sector giants like BP, Total SA and Royal Dutch Shell among others, as these firms are integrated majors which are engaged on both upstream, midstream and downstream activities. Their size and weight in the market allow these companies to have higher contractual power when entering forward contracts and better sustain downward sloping oil prices. Moody's ratings display another

Oil Producing Firms			
	CDS	Equity Volatility	Leverage Ratio
Number of obervations	340	340	340
Mean	252.02	0.39	0.32
Median	196.39	0.36	0.31
Maximum	1447.89	0.88	0.54
Minimum	66.37	0.17	0.18
Standard deviation	212.38	0.16	0.087
Oil Servicing Firms			
	CDS	Equity Volatility	Leverage Ratio
Number of obervations	340	340	340
Mean	272.61	0.38	0.40
Median	227.25	0.39	0.41
Maximum	924.08	0.72	0.62
Minimum	66.04	0.17	0.24
Standard deviation	179.43	0.13	0.10
Non firm-specific variables			
	Markit CDX Index	WTI	
Number of obervations	340	340	
Mean	75.19	70.95	
Median	70.75	64.3	
Maximum	126	110.53	
Minimum	45.25	29.42	
Standard deviation	16.72	24.19	

Table 2.2: Summary Statistics of the time series of explained and explanatory variables aggregated over each group of firms, and non firm-specific determinants.

common feature of the energy sector outside of the majors, namely low credit ratings. ConocoPhillips is the best rated firm, awarded an A3 long term rating. The remaining firms are rated from Baa1 to B2 and B3, which are described as highly speculative grade rating. Only three oil producing firms display a “positive” outlook, four servicing companies exhibit a “negative” outlook whilst the remaining sample displays a “stable” outlook. Two ratings were “under review” as of January 2019.

Oil Producing Companies	Country of Operation	Market Capitalisation as of 01/2019 (billion USD)	Moody's Rating Short Term	Long Term	Outlook
Anadarko Petroleum Corp.	US	23.59	SGL-2	Ba1	Stable
Apache Corp.	US	11.88	P-3	Baa3	Stable
Chesapeake Energy Corp.	US	3.42	SGL-3	B2	Stable
ConocoPhillips	US	76.23	SGL-1	A3	Stable
Devon Energy Corp.	CA	12.12	SGL-2	Ba1	Stable
Encana Corp.	US	6.51	SGL-1	Ba1	Positive
Marathon Oil Corp.	US	12.98	SGL-1	Ba3	Positive
Murphy Oil Corp.	US	4.80	SGL-1	Ba2	Stable
Noble Energy Inc.	CA	11.00	SGL-2	Ba3	Negative
Pioneer Natural Res.	US	24.41	SGL-3	Baa2	Stable
Whiting Petroleum	US	2.60	SGL-1	B1	Positive
Oil Servicing Companies	**Country of Operation**	**Market Capitalisation as of 01/2019 (billion USD)**	**Moody's Rating Short Term**	**Long Term**	**Outlook**
Enbridge Inc.	CA	71.98	-	Baa3	Under Review
Ensco Plc.	UK	1.92	SGL-1	B2	Under Review
Halliburton Company	US	26.84	P-2	Baa1	Stable
Nabors Industries Ltd.	BM	1.02	SGL-2	Ba3	Negative
Transocean Inc.	CH	5.19	SGL-1	B3	Negative
Valero Energy Corp.	US	33.83	-	Baa2	Stable
Weatherford Int.	US	0.52	SGL-3	B3	Negative

Table 2.3: Description of the two subsets. We report each firms' country of operation, market capitalisation and the most recent short and long-term Moody's rating.

2.4.1 Unit root test

Given the time dimension of the data in analysis, we applied the Augmented Dickey-Fuller (ADF) test to check the presence of unit roots in the data. According to the ADF test, if the Null hypothesis (H_0) fails to be rejected, the time series is said to have a unit root, highlighting its non-stationarity. On the other hand, if the Null hypothesis is rejected, the data is deemed as stationary. Tables 2.7 and 2.8 in Section 2.8 report the results of the ADF test. The results display that all firm's levels data and WTI prices have a unit root (we fail to reject H_0 given that $p-value > 0.05$). On the other hand, we reject the Null hypothesis at the 5% confidence level for all log changes variables and CDXIG, which are considered stationary or integrated of order 1 ($\mathbf{I}(1)$).

2.5 Methodology

In this section we describe the Autoregressive Distributed Lag (ARDL) models, the Generalised AutoRegressive Conditional Heteroskedasticity models (GARCH), the Impulse Response Function (IRF) and the Redundant Variables (RV) test.

2.5.1 Autoregressive Distributed Lag

To account for non-stationarity of the data, we adopted an Autoregressive Distributed Lag model. As discussed in Hendry et al. (1984) and Pesaran & Shin (1995), ARDLs allow to model a time series variable as a function of its lagged values as well as the current and lagged values of a number of exogenous variables. An ARDL model can be described as follows

$$y_t = a_0 + a_1 y_{t-1} + b_0 x_t + b_1 x_{t-1} + u_t, \tag{2.2}$$

$$y_t - y_{t-1} = a_0 + (a_1 - 1) y_{t-1} + b_0 x_t + (b_0 x_{t-1} - b_0 x_{t-1}) + b_1 x_{t-1} + u_t,$$

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 \Delta x_t + \beta_1 x_{t-1} + u_t, \tag{2.3}$$

where x is the vector of endogenous regressors (which in our case include lagged oil returns, CDXIG Index, firm-specific leverage ratio and share price volatility), α_0 is the intercept, $\alpha_1 = a_1 - 1$, b_0, b_1 and β_0 commensurate vectors of regression coefficients, $\beta_1 = b_0 + b_1$, and u_t represents the error term. For each firm, our model can then be written as:

$$\begin{aligned} \Delta CDS_t = {} & \alpha_0 + \alpha_1 CDS_{t-1} + \beta_0^{LEV} \Delta LEV_t + \beta_1^{LEV} LEV_{t-1} + \\ & \beta_0^{VOL} \Delta VOL_t + \beta_1^{VOL} VOL_{t-1} + \beta_0^{WTI_{t-1}} \Delta WTI_{t-1} + \\ & \beta_1^{WTI_{t-2}} WTI_{t-2} + \beta_0^{CDXIG} CDXIG_t + u_t. \end{aligned} \tag{2.4}$$

In the case of WTI prices and credit default swaps $WTI_{t-1} = \ln(WTI_{t-1})$ and $CDS_t = \ln(CDS_t)$, hence it follows that ΔWTI_{t-1} and ΔCDS_t are percentage changes.

On the other hand, ΔLEV_t and ΔVOL_t are first differences. From Equation (2.3) it follows that $y_t \sim I(1), x_t \sim I(1)$. Consequently if $\Delta y_t - \beta_0 x_t \sim I(0)$ then Δy_t and Δx_t are co-integrated and if $\alpha_1 \neq 0, \beta_1 \neq 0$ there exist a long run relationship between the two

variables and we can estimate the Impulse Response Function, presented in Section 2.5.3.

2.5.2 Generalized AutoRegressive Conditional Heteroskedasticity

Given the properties of financial time series, namely fat tails and heteroskedasticity, we estimate Equation (2.4) using a GARCH(1,1) model with Student-t distributed error terms. As explained in Hamilton (1994), given an autoregressive process of order p (AR(p))

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + u_t, \tag{2.5}$$

where $\mathbb{E}[u_t] = 0$. If the condition

$$\mathbb{E}[(u_t u_s)] = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{otherwise,} \end{cases}$$

does not hold, then the unconditional variance changes over time and can be described as

$$u_t^2 = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + w_t, \tag{2.6}$$

where w_t is a white noise s.t. $\mathbb{E}[w_t] = 0$ and

$$\mathbb{E}[(w_t w_s)] = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{otherwise.} \end{cases}$$

Then, u_t is called an Autoregressive Conditional Heteroskedastic process of order m (*ARCH(m)).* An alternative representation supposes that

$$u_t = \sqrt{(h_t)} v_t, \tag{2.7}$$

where $\mathbb{E}[v_t] = 0$ and $\mathbb{E}[v_t^2] = 1$. If we represent h_t by the values of u_t^2, i.e.,

$$h_t = c + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2, \tag{2.8}$$

then the conditional variance follows an *ARCH(m)* process and can be represented as

$$\mathbb{E}(u_t^2|u_{t-1},u_{t-2}..)=c+\alpha_1 u_{t-1}^2+\cdots+\alpha_m u_{t-m}^2. \tag{2.9}$$

In this case, the conditional variance depends on an infinite number of lags of u_{t-j}^2 such that

$$h_t=c+\phi(L)u_t^2, \tag{2.10}$$

where (L) is the lag operator and $\phi(L)=\sum_{j=1}^{\infty}\phi_j L_j$. It follows that

$$\phi_j L_j=\frac{\alpha(L)}{(1-\delta(L))},$$

while it is assumed that $(1-\delta(L))$ has roots outside the unit circle. If Equation (2.10) is multiplied by $(1-\delta(L))$ and terms are rearranged we obtain the expression

$$h_t=k+\delta_1 h_{t-1}+\cdots+\delta_r h_{t-r}+\alpha_1 u_{t-1}^2+\cdots+\alpha_m u_{t-m}^2 \tag{2.11}$$

which yields the generalised autoregressive conditional heteroskedasticity model

$$u_t\sim\text{GARCH}(r,m).$$

2.5.3 Impulse Response Function

The Impulse Response Function (IRF) is the effect of a one unit increase in a regressor at time $t-s$ on the independent variable at time t if all other innovations are kept constant for all t. Lagging Equation (2.2) by one period returns

$$y_{t-1}=a_0+a_1 y_{t-2}+b_0 x_{t-1}+b_1 x_{t-2}+u_{t-1}, \tag{2.12}$$

By substituting Equation (2.12) in Equation (2.2) we obtain

$$y_t=a_0(1+a_1)+a_1^2 y_{t-2}+b_0 x_t+(b_1+a_1 b_0)x_{t-1}+a_1 b_1 x_{t-2}+u_t+a_1 u_{t-1}. \tag{2.13}$$

The Impulse Response Function is obtained by repeating this process recursively for any given number of lags. For each time period, the IRF formulae are reported in Table 2.4.

t	IRF
0	b_0
1	$b_1+a_1b_0$
2	$a_1(b_1+a_1b_0)$
3	$a_1^2(b_1+a_1b_0)$
...	...

Table 2.4: Impulse Response Function formulas

The reasoning behind the application of IRF is twofold: *i*) to quantify how much a change in one of the covariates affects the CDS of a firm, and *ii*) to analyse how this effect propagates through time.

2.5.4 Redundant Variables test

The Redundant Variables test allows to check whether any number of covariates in a regression model have jointly coefficient equal to zero. The test assumes the formula

$$F = \frac{(R^2_{UR} - R^2_R)/m}{(1-R^2_{UR})/(df)}, \tag{2.14}$$

where R^2_{UR} and R^2_R correspond to the coefficient of determination of the unrestricted and restricted models, respectively, m is the number of redundant variables tested and df are the degrees of freedom. Under $H_0 : \beta_0 = \beta_1 = 0$ hence x_0,x_1 are irrelevant. If evidence against the Null hypothesis is found, then the coefficients are statistically different from zero.

2.6 Results

As described in Section 2.3, our expectation is to discover a positive relation between the firm's credit default swaps and the corresponding leverage ratio, share price volatility and CDXIG index. Conversely, we anticipate to uncover a negative link with crude oil returns.

In Sections 2.6.1 and 2.6.2 we present an analysis of the estimation results for the oil producers and servicing companies subsets, respectively. Section 2.6.3 reports the IRF results. To understand how the relationship between WTI returns and the firms' CDS changed during the period in analysis, in Section 2.6.4 we examine the two quantities' Pearson correlation computed in a rolling fashion over a sliding window period of 1 year. Contextually, we analyse the coefficients $\beta^{WTI_{t-1}}$ obtained by regressing the model over the same 1-year sliding window period. Lastly, we test the robustness of the model in Section 2.6.5.

Tables 2.9 and 2.10 in Section 2.8 report the coefficients, standard errors and p-values resulting from the model estimation. The estimated β_1^{LEV}, β_1^{VOL}, $\beta_1^{WTI_{t-2}}$ parameters are also reported but their values are not discussed given their function as control variables.

2.6.1 Oil Producing Firms

Firstly, as displayed in Table 2.5, we find evidence in favour of the choice to model the data via GARCH(1,1) with Student-t distribution in six out of the eleven regressions.

Oil Producing Companies	GARCH Coefficient	St. Error	p-value	Oil Servicing Companies	GARCH Coefficient	St. Error	p-value
Anadarko Petroleum	-0.007	0.203	0.972	Enbridge	0.774	0.112	0.000
Apache	-0.018	0.171	0.917	Ensco	0.787	0.060	0.000
Chesapeake Energy	0.720	0.089	0.000	Halliburton	-0.031	0.101	0.758
ConocoPhillips	0.757	0.085	0.000	Nabors Industries	0.982	0.011	0.000
Devon Energy	0.947	0.035	0.000	Transocean	0.881	0.073	0.000
Encana	0.796	0.101	0.000	Valero Energy	0.818	0.092	0.000
Marathon Oil	-0.019	0.041	0.643	Weatherford	0.858	0.108	0.000
Murphy Oil	0.566	0.120	0.000				
Noble Energy	0.896	0.059	0.000				
Pioneer Nat. Res.	0.227	0.162	0.162				
Whiting Petroleum	0.227	1.342	0.865				

Table 2.5: GARCH coefficients, standard errors and p-values resulting from the estimation of Equation (2.3)

Table 2.9 in Section 2.8 reports the coefficients, standard errors and p-values resulting from the model estimation. The change in leverage ratio, represented by the coefficient β_0^{LEV}, is found to have a positive impact on the firms' CDS and is statistically significant

in all instances. The magnitude of the coefficients range between 0.467 (Noble Energy and Pioneer Natural Resources) and 1.786 (Chesapeake Energy) with an average of 0.872. It is worth noting that, whilst there is an inverse relation between the firm's market capitalisation and the corresponding β_0^{LEV} estimate in the case of Chesapeake Energy, Pioneer Natural Resources and Whiting Petroleum, the opposite can be observed for Anadarko Petroleum, Devon Energy and Murphy Oil. We plot the estimated β_0^{LEV} coefficients against the companies' market capitalisation in Figure 2.5a. The regression line displays that a higher market capitalisation leads to a lower β_0^{LEV}, implying that larger oil producing enterprises' CDS spreads are less sensitive to changes in their Debt-Equity composition.

Moreover, changes in share price volatility, represented by the coefficient β_0^{VOL}, have a positive effect on the CDS in seven out of eleven cases. Surprisingly, only one of these coefficients is statistically significant. This implies that the market value of the shares of an oil producer might not be a good indicator of its CDS performance. Further, the CDXIG index coefficient β_0^{CDXIG} is found to be significant at the 5% confidence level for ten firms and at the 10% level for one firm. However, the magnitude of β_0^{CDXIG} is very low in all cases. This implies that, whilst the CDXIG Index is an important determinant for CDS prices, as it encompasses the general market sentiment, the broader CDS market fluctuations do not have meaningful impacts on the firm-specific credit default swaps.

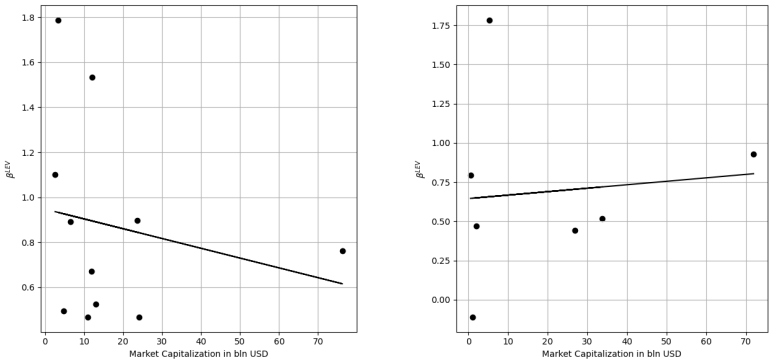
Lastly, lagged WTI returns, represented by the coefficient $\beta_0^{WTI_{t-1}}$, can be considered a significant determinant for eight oil producing firms' CDS spreads at the 5% confidence level. The coefficient magnitude varies between a minimum of -0.45 to a maximum of 0.078, with an average of -0.21. If we do not consider the non-significant estimated $\beta_0^{WTI_{t-1}}$ coefficients belonging to Chesapeake, Noble Energy and Whiting Petroleum, the average $\beta_0^{WTI_{t-1}}$ grows to -0.29. Oil producers are affected by strong entry barriers and high costs to operate new oilfields. Namely, the research and development of proprietary technologies, and land and drilling rights are capital intensive. Moreover, the extraction of oil runs at high fixed operating costs, which is one of the main reasons why producers are reluctant to halt the extraction of crude, once started. The results corroborate these features, reflecting the risk that the production costs might exceed the market price of the final product and the break-even point is not reached.

Nine out of the eleven producers' CDS display an inverse sensitivity to oil returns changes. Thus, an increase in oil returns, leads to the reduction of the following week's CDS spreads. Chesapeake and Noble Energy are the only two firms yielding opposite results but are, however, not statistically significant. This is in line with the assumptions presented in the Section 2.3, confirming how crucial crude prices are in order to evaluate an oil producer's ability to face its financial commitments, hence, its risk of default. Interestingly, if we analyse how the coefficient $\beta_0^{WTI_{t-1}}$ varies based on the size of the firms, we can observe that ConocoPhillips, Anadarko Petroleum and Marathon Oil display the highest sensitivity (in absolute value) to oil returns, despite being the first, third and fourth largest firms in our subset of oil producers. Given that these firms are large-caps, a deeper analysis of their business model, the magnitude of refining in their operations, positioning on the market and in the production stream should be performed to uncover the reason of such strong sensitivity. Additionally, the data does not reflect if profits were made by the company's production activities or made via carry trade strategies because of the contango shape of the forward curve. The same applies to understanding the low impact of crude returns to Whiting Petroleum's CDS, given that it is the smallest of the mid-caps in the subset but the least sensitive (in absolute value) to changes of crude oil returns. Figure 2.5c, however, displays that changes in oil returns are more impactful on the CDS' of firms with higher market capitalisation.

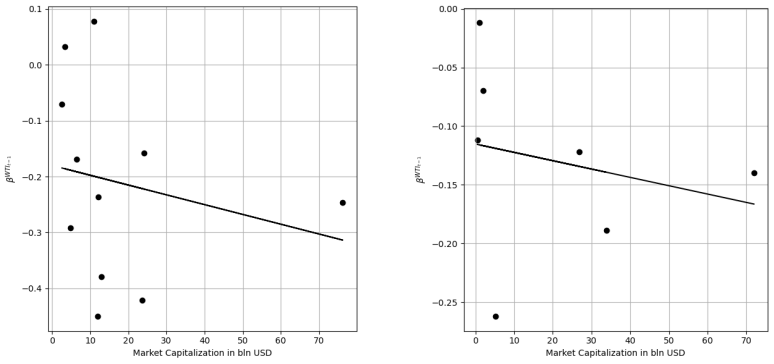
2.6.2 Oil Servicing Companies

Differently from the oil producers subset, we find strong evidence in favour of the GARCH model for oil servicing companies. As displayed in Table 2.5 in Section 2.6.1, the GARCH model is statistically significant at the 5% confidence level in six out of seven regressions.

Table 2.10 in Section 2.8 reports the coefficients, standard errors and p-values resulting from the model estimation. The coefficients significance, sign and magnitude for oil servicing firms agree with the results presented for oil producers in Section 2.6.1. Namely, the estimations suggest that changes in leverage ratios, represented by the coefficient β_0^{LEV}, are an important determinant of CDS spreads for all firms except Nabors Industries. If we exclude Nabors Industries' estimated coefficient, all estimated β_0^{LEV} are positive and range between 0.44 (Halliburton) and 1.783 (Transocean), with an average of 0.69.



(a) Producers β_0^{LEV} vs Market Cap. **(b)** Servicing firms β_0^{LEV} vs Market Cap.



(c) Producers $\beta_0^{WTI_{t-1}}$ vs Market Cap. **(d)** Servicing firms $\beta_0^{WTI_{t-1}}$ vs Market Cap.

Figure 2.5: Plot of the estimated values of β_0^{LEV} and $\beta_0^{WTI_{t-1}}$ versus Market Capitalisation

Both sign and magnitude are similar to the oil producers' regressions and in line with our expectations. On one hand, the firm with highest market cap (Enbridge) presents the second highest β_0^{LEV}. Ensco, the third smallest firm, displays one of the lowest sensitivities. On the other hand, the relation between β_0^{LEV} and market capitalisation is inverse for Valero Energy, Halliburton and Weatherford. Contrarily to oil producers, the regression line depicted in Figure 2.5b shows that β_0^{LEV} slightly increases when the market capitalisation is greater. Our analysis leads us to believe that this difference arises from the dissimilarities present in the balance sheets of the two types of companies analysed. Namely, the capital structure of oilfield servicing companies, which generally have a lower total value of assets, is significantly different from the capital structure of producing companies, leading to a higher sensitivity to an increase in leverage ratio.

Furthermore, the estimation results display that share price volatility's coefficient β_0^{VOL} is found statistically significant once at the 5% confidence level and once at the

10% level. Estimated values for β_0^{VOL} range between -0.051 and 0.146. Similarly to the oil producers dataset, share prices volatility are not a significant determinant of credit default swaps. Moreover, the CDXIG index is shown to be significant for six out of seven firms and, in parallel to the results presented in Section 2.6.1, the estimates of β_0^{CDXIG} are close to zero.

Lastly, all the estimates of $\beta_0^{WTI_{t-1}}$ for oil servicing companies are negative, suggesting that an increase in oil returns reflects in a reduction of the following week's CDS spread. The size of the coefficients ranges between -0.262 (Transocean) and -0.012 (Halliburton), with an average of -0.13. If we do not consider the non-significant estimated $\beta_0^{WTI_{t-1}}$ coefficients belonging to Ensco, Nabors and Weatherford, the average $\beta_0^{WTI_{t-1}}$ grows to -0.178. This shows that, on average, oil producers' CDS are more sensitive to crude returns changes than oil servicing companies. As previously mentioned, downstream refiners benefit from the lag between a movement in the input cost and the adjustment of the refined product price while being able to lock in their profits through the use of the so-called crackspreads. We find evidence of the significance of WTI returns as a determinant for CDS spreads in two cases at the 5% confidence level and in two cases at the 10% level. Contrarily to the relationship between β_0^{LEV} and market capitalisation, the link between $\beta_0^{WTI_{t-1}}$ and market capitalisation is more pronounced. As displayed in Figure 2.5d, changes in oil returns are more impactful on the CDS' of firms with higher market capitalisation. This is fairly surprising: as previously discussed, a larger and more integrated firm should be able to sustain a crude price drop when compared to a company of smaller size; this is especially true given their refining activities and the profits made in carry trade strategies because of the contango shape of the forward curve.

2.6.3 Long Run Relationship and Impulse Response function

As described in Section 2.5, when the coefficients of the ARDL model α_1 and β_1 in Equation (2.4) are significantly different from zero, there exist a long run relationship between the credit default swaps and their determinants allowing to estimate the Impulse Response Function.

First, as presented in Section 2.5, we applied the Redundant Variable test to check whether the estimates of α_1, β_1^{WTI} and the estimates of α_1, β_1^{LEV} are jointly different

from zero. Following the results presented in Section 2.6.1 and 2.6.2, we analysed the CDS-LEV and the CDS-WTI long run relationships. We excluded from this analysis the coefficients of the variables that were not significant as displayed in Tables 2.9 and 2.10 in Section 2.8, namely WTI in the case of Chesapeake, Noble Energy, Whiting Petroleum, Ensco, Nabors and Weatherford, and leverage ratio in the case of Nabors. Further, we did not test for the existence of a CDS-VOL and CDS-CDXIG long run relationship as β_0^{VOL} was found to be significant only for two out of eighteen firms, whilst the CDXIG Index short term effect was very close to zero for all firms. The Redundant Variables test results, reported in Table 2.6, show that there is evidence against the null hypothesis (H_0: the variables are not significant) for Apache, and Weatherford at the 10% level and at the 5% confidence level for all the other firms.

Then, we computed the Impulse Response Function of the first 5 lags. Results are reported in Table 2.11 in Section 2.8 and displayed in Figures 2.11, 2.12, 2.13 in Section 2.8. It can be noted that lagged changes in WTI returns have a meaningful impact on the firms' CDS spreads only at time t, dissipating in the following time periods. This means that, provided that all other innovations are kept constant, an increase or decrease of WTI returns will only affect the firms' following week's credit default swap value, since we adopted as explanatory variable the first lag of WTI returns, as explained in Section 2.3.1. This is valid for both producers and servicing companies.

The same applies to changes in Leverage Ratios, which have a significant impact on the firm's CDS exclusively at time t for eight producers and three servicing companies. The IRFs for Encana, Noble Energy, Whiting Petroleum, amongst the producers, and for Enbridge, Ensco, Valero Energy and Weatherford, amongst the servicing companies, show that an adjustment of the firms' Leverage Ratio impacts the firms' CDS at time t and propagates up to the fifth lag, albeit displaying a very low magnitude from the first lag onwards.

2.6.4 1-Year Rolling Correlations and Rolling Regressions

Next, we studied how the relation between WTI returns and the firm-specific CDS varies over time. We constructed a time series of 1-year rolling Pearson correlation between ΔWTI_{t-1} and each firm's ΔCDS_t. Each data point reflects the correlation between the two variables observed over the previous 52 weekly observations. In a similar manner,

	CDS−WTI		*CDS−LEVERAGE*	
Oil Producing Companies	F-Statistic	P-value	F-Statistic	P-value
Anadarko Petroleum	19.196 ***	0.000	19.049***	0.000
Apache	4.821*	0.090	5.368*	0.068
Chesapeake Energy	-	-	30.321***	0.000
ConocoPhillips	14.155***	0.001	19.57***	0.000
Devon Energy	11.028***	0.004	10.077***	0.007
Encana	25.886***	0.000	34.483***	0.000
Marathon Oil	301.089***	0.000	20.756***	0.000
Murphy Oil	31.888***	0.000	42.143***	0.000
Noble Energy	-	-	30.761***	0.000
Pioneer Nat. Res.	22.844***	0.000	28.42***	0.000
Whiting Petroleum	-	-	7.316**	0.026
	CDS−WTI		*CDS−LEVERAGE*	
Oil Servicing Companies	F-Statistic	P-value	F-Statistic	P-value
Enbridge	13.688***	0.001	18.087***	0.000
Ensco	-	-	32.004***	0.000
Halliburton	8.929**	0.012	10.003***	0.007
Nabors Industries	-	-	-	-
Transocean	7.920**	0.019	9.333***	0.009
Valero Energy	8.447**	0.015	17.115***	0.000
Weatherford	-	-	6.643**	0.036

Table 2.6: Redundant Variables F-test results. *** significance at the 1% confidence level, ** significance at the 5% confidence level, * significance at the 10% confidence level.

we constructed a time series of estimates of the regression coefficient $\beta_0^{WTI_{t-1}}$ obtained from fitting the model shown in Equation (2.4) in a rolling fashion with a sliding window size of 1-year for each firm. We chose a window of 1 year to allow for enough data to be included in each regression in order to achieve convergence.

From Figures 2.8, 2.9 and 2.10 in Section 2.8 we can observe a number of common patterns: for the producers Anadarko, Apache, ConocoPhillips, Encana, Marathon, Murphy and Pioneer Natural Resources, the 1-year rolling correlation is generally negative and weak with values reverting around -0.25 until mid-2014. It then doubles reaching levels close to -0.5 in most cases, maintaining this magnitude through 2015 and 2016. Interestingly, the correlation magnitude reduces in the period following

August 2017, reaching positive levels for ConocoPhillips, Encana and Pioneer Natural Resources. Servicing companies like Enbridge, Ensco, Halliburton and Valero are found to perform similarly. Moreover, the CDS's sensitivity to oil returns (measured via the rolling coefficients $\beta_0^{WTI_{t-1}}$) increased in magnitude throughout the period 2014-2015 and reverted to lower levels from mid- to late 2016, when the cost of a barrel of oil was stable around $50, confirming the correlation results. This pattern shows that the 1-year correlation and $\beta_0^{WTI_{t-1}}$ were higher in periods of high volatility between July 2014 and January 2016, whilst CDS spreads were spiking in 2016. However, when crude prices displayed a steady growth from mid-2017 onwards and, in parallel, CDS spreads were slowly reverting to their 2013 levels, the correlation magnitude reduced. From mid-2017, in the case of Chesapeake, Devon Energy, Encana, Noble Energy, Pioneer Natural Res., Enbridge, Ensco, Halliburton, Transocean and Valero the maximum estimated value of the correlation coefficient is 0.25, indicating a weak positive correlation between oil returns and the firms' CDS. Surprisingly, ConocoPhillips, Pioneer, Ensco, Halliburton, Transocean, Valero 's rolling $\beta_0^{WTI_{t-1}}$ estimates becomes positive in the same period.

There are, however, some exceptions. Marathon Oil's estimated rolling beta coefficient $\beta_0^{WTI_{t-1}}$ increased (in absolute value) from late 2016 until the end of 2017, almost a year and half after the other firms in the dataset. Moreover, as shown in Figures 2.8 and 2.9 in Section 2.8, Murphy Oil displayed a second sensitivity increase during late 2017 and in 2018, whilst its peers' were steadily reducing. Surprisingly, Pioneer Natural Resources, Noble Energy, Enbridge and Transocean's estimates of the rolling $\beta_0^{WTI_{t-1}}$ coefficient was positive from mid-2017.

It is slightly harder to find a specific pattern for the remaining firms in the dataset. Chesapeake displays a somewhat similar behaviour displaying, however, a positive correlation between August 2013 and November 2014. Interestingly, its 1-year rolling $\beta_0^{WTI_{t-1}}$ are mostly positive with dips during the years 2014, 2015, 2017 and 2018. It is worthwhile remembering that, as shown in Table 2.9, WTI was not a statistically significant determinant to Chesapeake's CDS. In the case of Devon Energy, the correlation increased in absolute value from close to zero to around -0.25 during the first quarter of 2013 and, contrarily to the other companies, it halved during the period between

November 2014 and October 2015. Devon Energy estimated rolling sensitivity to oil returns follows a similar pattern reaching its peak (in absolute value, i.e. lowest value) after the oil price drop started in July 2014. Transocean shows no clear pattern as the correlation oscillates around -0.3 from 2013 until 2017 before increasing to positive levels. Transocean's estimated rolling $\beta_0^{WTI_{t-1}}$ is also mostly negative with lowest points in the end of 2013 and August 2014. Its sensitivity to oil returns sporadically becomes positive in 2015 and 2016 before settling above zero in mid-2017.

In summary, the CDS spreads of 60% of the firms analysed show a higher sensitivity and stronger correlation to oil returns after the steep price drop observed as of July 2014. This relation lasted throughout 2015 and 2016, before the cost of a barrel of oil started to steadily increase. During 2017 and the first half of 2018, the increasing costs of crude oil translated to higher revenues for oil companies and was a signal for market participants, including investors and creditors, that the sector was once again recovering its strength. This translated to reduced magnitude of the CDSs sensitivity to crude returns across our dataset and, in some cases, it translated to the positive $\beta_0^{WTI_{t-1}}$ in the last period in analysis.

2.6.5 Robustness Analysis

As presented in Section 2.6, the share price volatility coefficient β_0^{VOL} resulted non-significant ($p-value > 0.05$) for all but one producing firm and one servicing company. As a robustness check, we tested our approach with a different explanatory variables specification, namely we removed the share price volatility. The new model assumes the form

$$\begin{aligned}\Delta CDS_t = &\alpha_0 + \alpha_1 CDS_{t-1} + \beta_0^{LEV}\Delta LEV_t + \beta_1^{LEV} LEV_{t-1} + \\ &\beta_0^{WTI_{t-1}}\Delta WTI_{t-1} + \beta_1^{WTI_{t-2}} WTI_{t-2} + \beta_0^{CDXIG} CDXIG_t + u_t. \end{aligned} \tag{2.15}$$

To facilitate commenting on the results, we named Equation (2.4) Model 1 and Equation (2.15) Model 2.

To understand which model is preferred for each firm, we compared the two model's Akaike Information Criterion (AIC) and Schwartz Information Criterion (BIC). Let n denote the number of parameters in the model, $L(\Theta)$ the value of the likelihood function evaluated at the estimated parameters and T the number of observations, the AIC is

classically defined as

$$AIC = -2ln(L(\Theta)) + 2n, \tag{2.16}$$

whilst the BIC is computed as

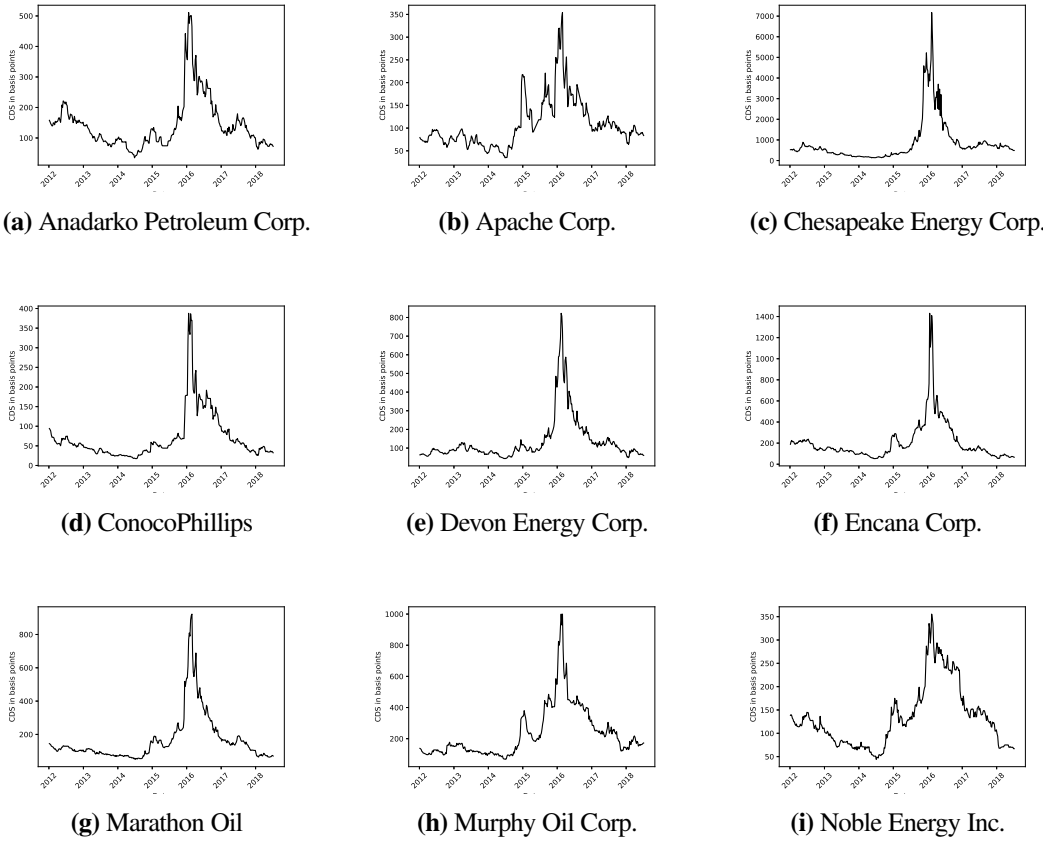
$$BIC = -2ln(L(\Theta)) + n\ln(T). \tag{2.17}$$

The lower model with lower AIC (or BIC) is preferred. As showed in Table 2.12 in Section 2.8, the BIC results suggest that Model 2 is preferred for 17 firms. Similarly, Model 2 achieves the lowest AIC for six producers and five servicing companies. We report the estimation results in Tables 2.13 and 2.14 in Section 2.8. However, it can be noted that the two models estimated β coefficients are equal up to three decimal places. This is due to the fact that the share price volatility was one of the framework's control variables and the size of the estimated β_0^{VOL} and β_1^{VOL} coefficients was very close to zero. As such, we can conclude that there is no statistical difference between Model 1 and Model 2.

2.7 Conclusions

In this chapter we analysed the relationship between lagged WTI returns and the credit default swaps of 11 oil producers and 7 oil servicing companies. We presented a structural model where CDXIG Index, lagged WTI returns, firm-specific leverage ratio and share price volatility are used to explain CDS spreads between 2012 and 2018. We find that there exist a positive relationship between leverage ratio and CDS spreads for both producers and oil servicing companies; however, whilst oil producers with larger market capitalisation are less sensitive to changes in leverage ratios compared to smaller firms, the opposite is found for oil servicing firms. Further, whilst significant in all cases, an increase (or decrease) of the CDXIG Index has a very low effect on the firms CDS spreads. On the other hand, share price volatility is not significant for both subsets. Lagged WTI returns are found significant in over 60% of the firms in the dataset. Moreover, oil producers display, on average, higher sensitivity to crude returns than servicing companies. Surprisingly, we find that the CDS' of producers and servicing companies with larger market capitalization are more sensitive to changes in oil returns. We then showed the existence of a CDS-WTI and CDS-leverage ratio long run relationship; however, the Impulse Response Function reveals that the effect of a change in lagged WTI returns on the credit default swaps dissipates after the first lag. The same applies for leverage ratio with the exception of three producers and four servicing companies where the impact of a change in leverage ratio protracts up to the 5th lag, despite having very low intensity past the first lag. Lastly, we analysed how the relationship between credit default swaps and lagged WTI returns changes over the period in analysis via rolling regressions and rolling Pearson correlation. We found that, for over 60% of the dataset, there was a weak inverse correlation and negative $\beta^{WTI_{t-1}}$ coefficients in the period leading up to 2014. Both quantities grew in magnitude during periods of high volatility, namely from mid-2014 until 2016. On the other hand, when oil prices were steadily increasing in 2017 and 2018, the strength of this relationship started to reduce and, in some cases, both correlation and $\beta^{WTI_{t-1}}$ became positive.

2.8 Figures and Tables

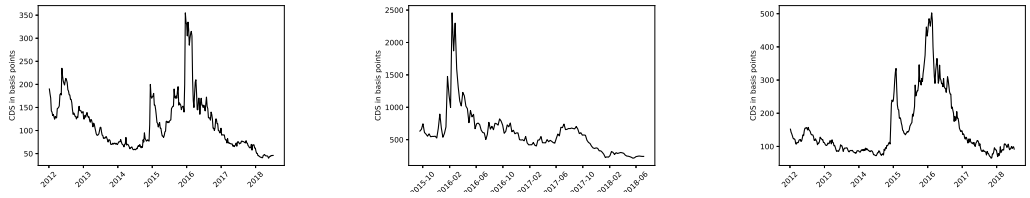


(a) Anadarko Petroleum Corp. **(b)** Apache Corp. **(c)** Chesapeake Energy Corp.

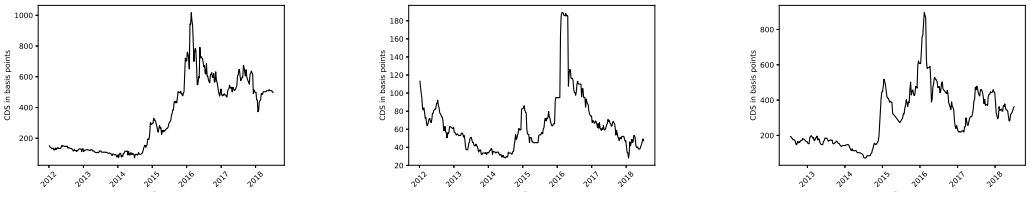
(d) ConocoPhillips **(e)** Devon Energy Corp. **(f)** Encana Corp.

(g) Marathon Oil **(h)** Murphy Oil Corp. **(i)** Noble Energy Inc.

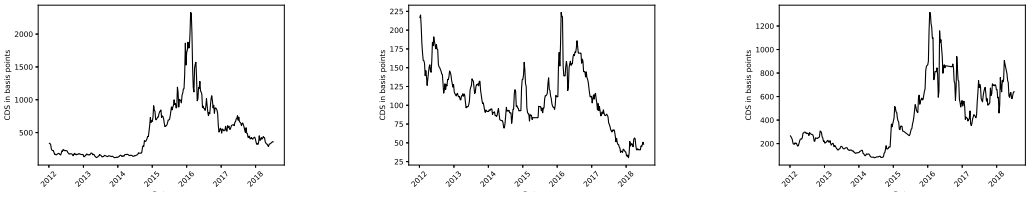
Figure 2.6: Evolution of oil producers and servicing firms' CDS over the period January 2012 to June 2018.



(a) Pioneer Natural Resources **(b)** Whiting Petroleum **(c)** Enbridge Inc.

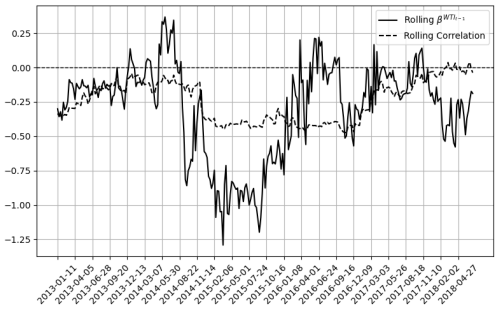


(d) Ensco Plc. **(e)** Halliburton Company **(f)** Nabors Inc.

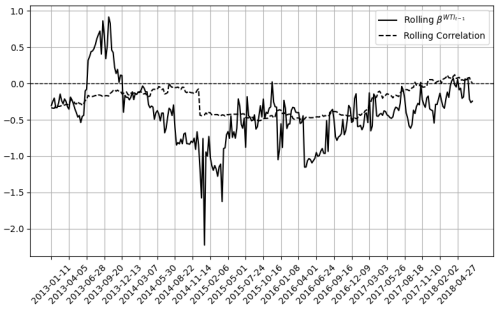


(g) Transocean Inc. **(h)** Valero Energy Corp. **(i)** Weatherford Int.

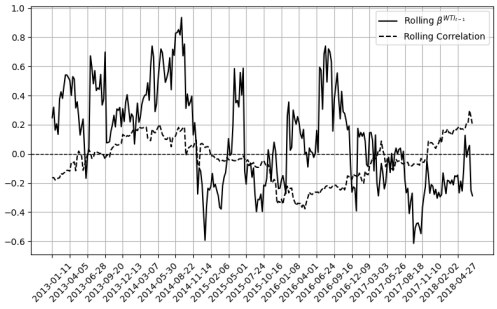
Figure 2.7: Evolution of oil producers and servicing firms' CDS over the period January 2012 to June 2018.



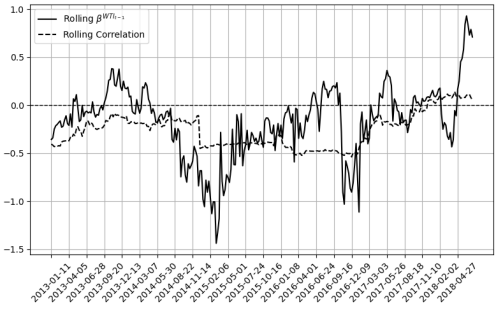
(a) Anadarko Petroleum Corp.



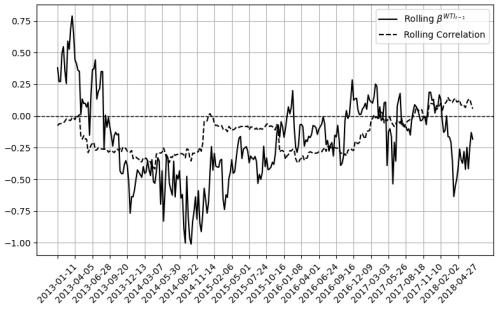
(b) Apache Corp.



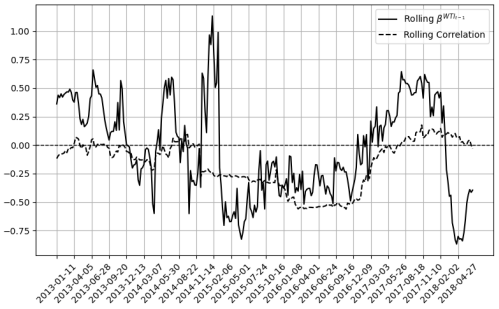
(c) Chesapeake Energy Corp.



(d) ConocoPhillips

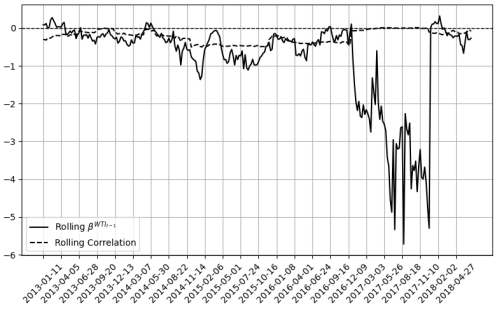


(e) Devon Energy Corp.

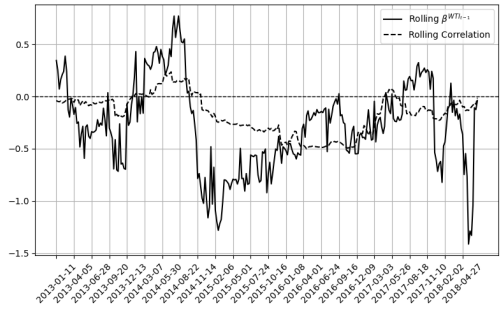


(f) Encana Corp.

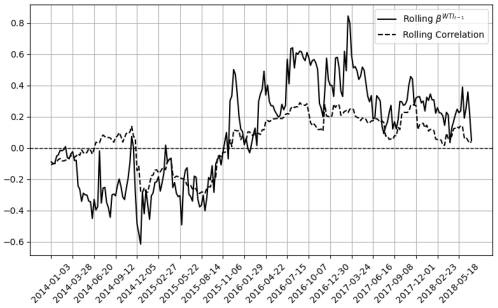
Figure 2.8: Rolling correlation (dashed line) vs rolling $\beta^{WTI_{t-1}}$ (solid line).



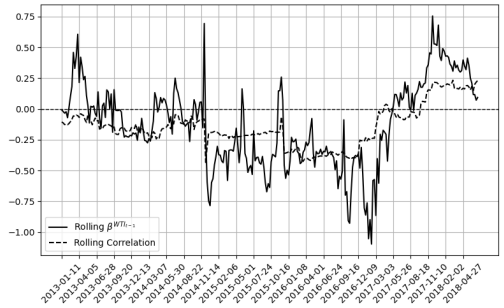
(a) Marathon Oil



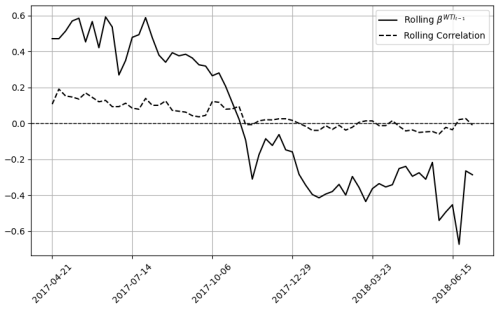
(b) Murphy Oil Corp.



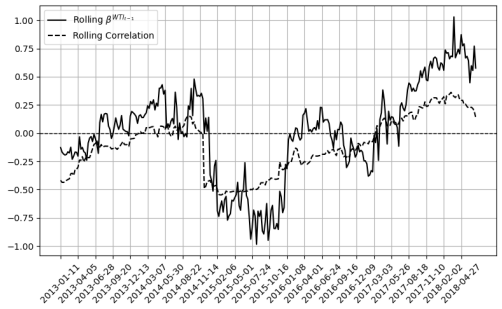
(c) Noble Energy Inc.



(d) Pioneer Natural Resources

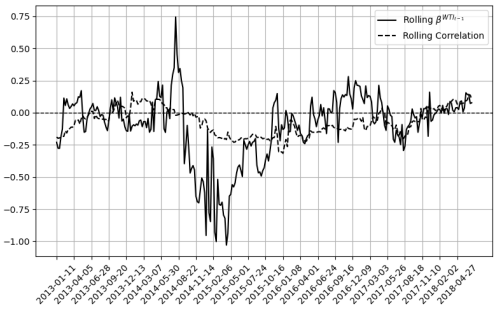


(e) Whiting Petroleum

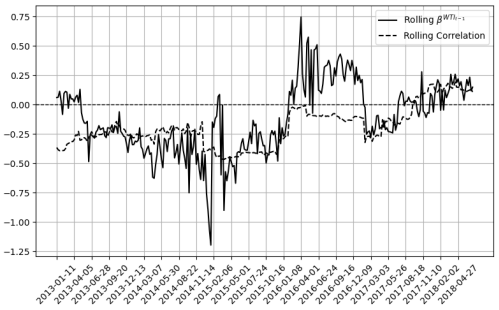


(f) Enbridge Inc.

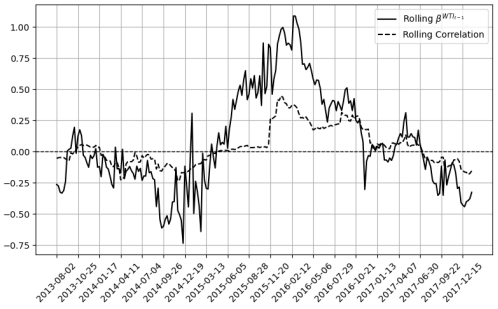
Figure 2.9: Rolling correlation (dashed line) vs rolling $\beta^{WTI_{t-1}}$ (solid line).



(a) Ensco Plc.

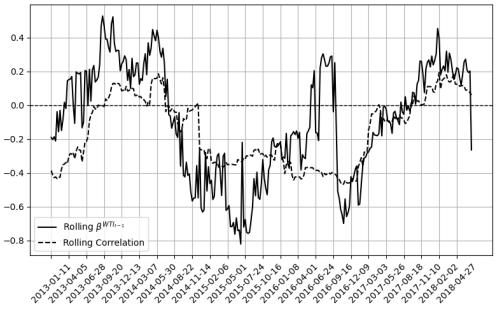


(b) Halliburton Company

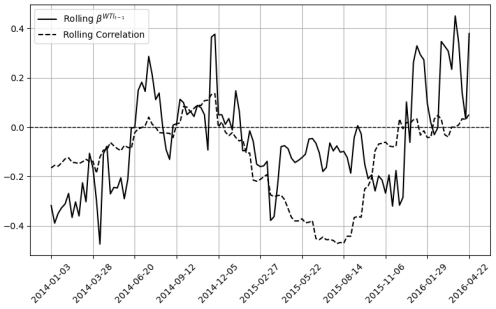


(c) Nabors Inc.

(d) Transocean Inc.

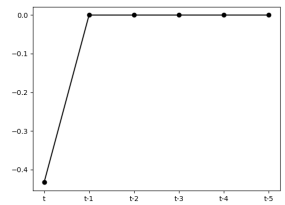


(e) Valero Energy Corp.

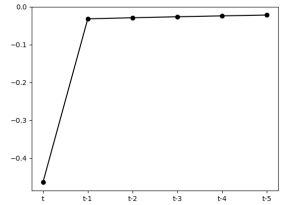


(f) Weatherford Int.

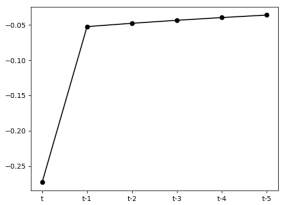
Figure 2.10: Rolling correlation (dashed line) vs rolling $\beta^{WTI_{t-1}}$ (solid line).



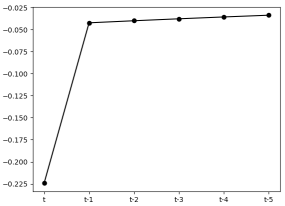
(a) Anadarko Petroleum Corp.



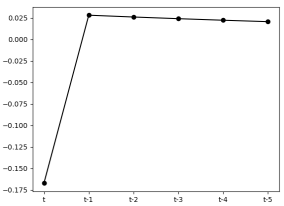
(b) Apache Corp.



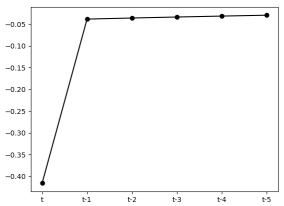
(c) ConocoPhillips



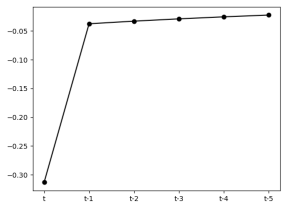
(d) Devon Energy Corp.



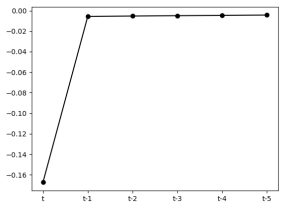
(e) Encana Corp.



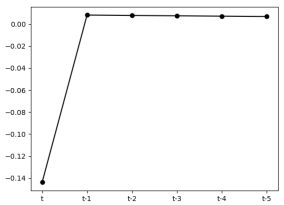
(f) Marathon Oil



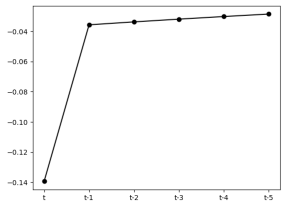
(g) Murphy Oil Corp.



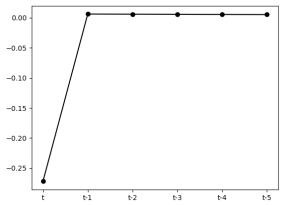
(h) Pioneer Natural Resources



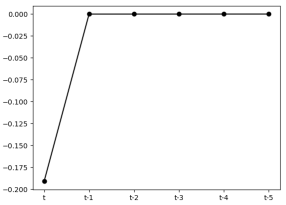
(i) Enbridge Inc.



(j) Halliburton Company

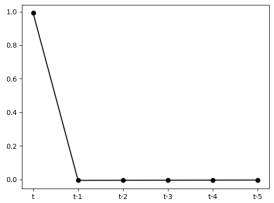


(k) Transocean Inc.

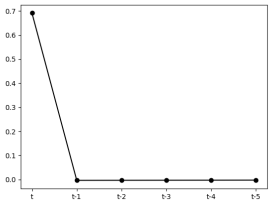


(l) Valero Energy Corp.

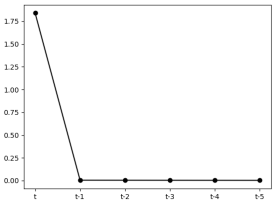
Figure 2.11: ΔWTI_{t-1} Impulse Response Function on ΔCDS_t.



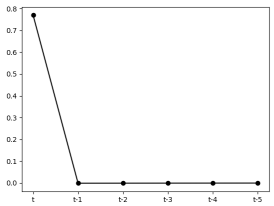
(a) Anadarko Petroleum Corp.



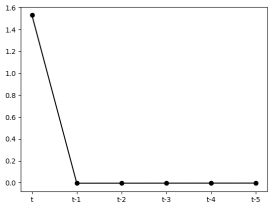
(b) Apache Corp.



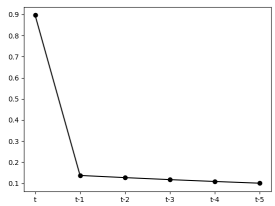
(c) Chesapeake Energy Corp.



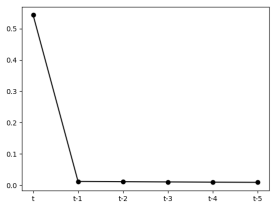
(d) ConocoPhillips



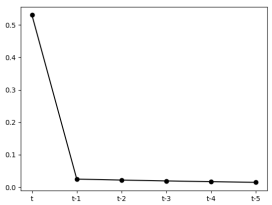
(e) Devon Energy Corp.



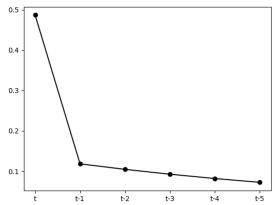
(f) Encana Corp.



(g) Marathon Oil

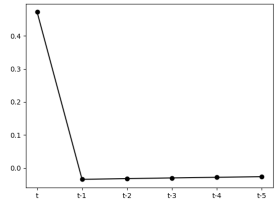


(h) Murphy Oil Corp.

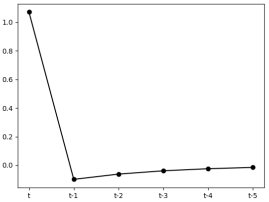


(i) Noble Energy Inc.

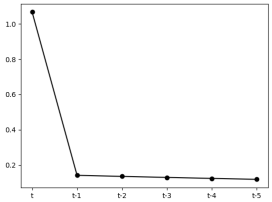
Figure 2.12: ΔLEV_t Impulse Response Function on ΔCDS_t.



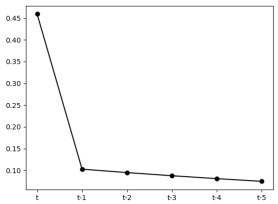
(a) Pioneer Natural Resources



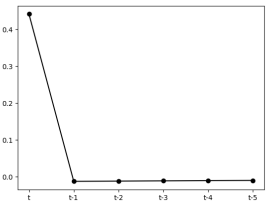
(b) Whiting Petroleum



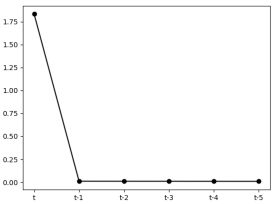
(c) Enbridge Inc.



(d) Ensco Plc.

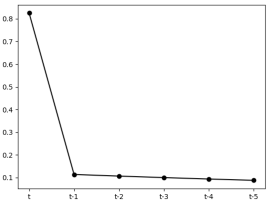


(e) Halliburton Company



(f) Transocean Inc.

(g) Valero Energy Corp.



(h) Weatherford Int.

Figure 2.13: ΔLEV_t Impulse Response Function on ΔCDS_t.

	Anadarko Petroleum		Apache		Chesapeake Energy		ConocoPhillips		Devon Energy	
Variable	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value
CDS	-2.277	0.180	-2.641*	0.085	-2.643*	0.084	-1.914	0.326	-2.072	0.256
Δ*CDS*	-15.363***	0.000	-15.489***	0.000	-2.826**	0.055	-6.463***	0.000	-16.629***	0.000
LEV	-1.848	0.357	-2.510	0.113	-1.065	0.729	-1.232	0.660	-2.382	0.147
Δ*LEV*	-7.023***	0.000	-17.092***	0.000	-11.560***	0.000	-5.115***	0.000	-6.410***	0.000
VOL	-3.453***	0.009	-2.819**	0.056	-2.224	0.198	-2.268	0.182	-3.893***	0.002
Δ*VOL*	-16.728***	0.000	-3.784***	0.003	-7.600***	0.000	-5.431***	0.000	-3.068***	0.029

	Enbridge		Encana		Ensco		Halliburton		Marathon Oil	
Variable	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value
CDS	-2.225	0.197	-2.788**	0.060	-1.242	0.655	-2.000	0.287	-2.095	0.247
Δ*CDS*	-5.786***	0.000	-15.476***	0.000	-5.097***	0.000	-17.232***	0.000	-21.191***	0.000
LEV	-2.542	0.106	-1.799	0.381	-0.983	0.760	-2.022	0.277	-1.853	0.355
Δ*LEV*	-18.697***	0.000	-16.539***	0.000	-9.236***	0.000	-10.837***	0.000	-16.801***	0.000
VOL	-3.006**	0.034	-2.362	0.153	-1.451	0.558	-4.302***	0.000	-2.135	0.231
Δ*VOL*	-3.258**	0.017	-4.842***	0.000	-17.696***	0.000	-4.515***	0.000	-4.523***	0.000

Table 2.7: ADF test results. Δ*CDS* and Δ*WTI* are percentage changes whilst Δ*LEV* and Δ*VOL* are first differences.
*** significance at the 1% confidence level, ** significance at the 5% confidence level, * significance at the 10% confidence level.

	Murphy Oil		Nabors		Noble Energy		Pioneer Nat. Res.		Transocean	
Variable	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value
CDS	-2.533	0.108	-2.093	0.247	-1.223	0.664	-2.357	0.154	-1.483	0.542
Δ*CDS*	-8.414***	0.000	-13.743***	0.000	-5.790***	0.000	-16.309***	0.000	-4.811***	0.000
LEV	-2.101	0.244	-2.046	0.267	-0.610	0.869	-1.230	0.661	-0.770	0.828
Δ*LEV*	-15.506***	0.000	-4.301***	0.000	-10.783***	0.000	-19.935***	0.000	-16.818	0.000
VOL	-2.235	0.194	-2.230	0.196	-1.647	0.459	-3.135**	0.024	-1.912	0.326
Δ*VOL*	-4.850***	0.000	-8.452***	0.000	-5.347***	0.000	-8.957***	0.000	-8.634***	0.000

	Valero Energy		Weatherford Int.		Whiting Petroleum	
Variable	ADF Statistic	p-value	ADF Statistic	p-value	ADF Statistic	p-value
CDS	-2.836*	0.053	-2.176	0.215	-1.346	0.608
Δ*CDS*	-16.002***	0.000	-5.883***	0.000	-12.945***	0.000
LEV	-2.410	0.139	-1.788	0.387	-0.450	0.901
Δ*LEV*	-11.724***	0.000	-14.026***	0.000	-10.774***	0.000
VOL	-3.838***	0.003	-1.572	0.498	-2.253	0.188
Δ*VOL*	-17.261***	0.000	-11.188***	0.000	-9.554***	0.000

Non Firm-Specific Data		
Variable	ADF Statistic	p-value
WTI	-1.342	0.610
Δ*WTI*	-16.417***	0.000
CDXIG	-3.094	0.027

Table 2.8: ADF test results. Δ*CDS* and Δ*WTI* are percentage changes whilst Δ*LEV* and Δ*VOL* are first differences.
*** significance at the 1% confidence level, ** significance at the 5% confidence level, * significance at the 10% confidence level.

	CDS_{t-1}			ΔLEV			LEV_{t-1}			ΔVOL		
Company name	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value
Anadarko Petroleum	-0.093***	0.016	0.000	0.896***	0.096	0.000	0.097***	0.033	0.003	-0.063	0.044	0.154
Apache	-0.088***	0.020	0.000	0.671***	0.099	0.000	0.057**	0.026	0.028	0.023	0.069	0.741
Chesapeake Energy	-0.104***	0.018	0.000	1.787***	0.117	0.000	0.195***	0.036	0.000	0.054	0.063	0.397
ConocoPhillips	-0.077***	0.018	0.000	0.762***	0.117	0.000	0.058	0.032	0.064	0.022	0.066	0.734
Devon Energy	-0.053***	0.018	0.003	1.533***	0.118	0.000	0.079	0.053	0.137	0.208**	0.088	0.018
Encana	-0.072***	0.015	0.000	0.892***	0.087	0.000	0.202***	0.035	0.000	-0.069	0.070	0.330
Marathon Oil	-0.068***	0.013	0.000	0.525***	0.071	0.000	0.050*	0.028	0.072	0.062	0.055	0.264
Murphy Oil	-0.118***	0.022	0.000	0.496***	0.096	0.000	0.087***	0.016	0.000	0.031	0.062	0.618
Noble Energy	-0.113***	0.019	0.000	0.468***	0.079	0.000	0.171***	0.032	0.000	-0.049	0.057	0.388
Pioneer Nat. Res.	-0.062***	0.013	0.000	0.468***	0.067	0.000	-0.005	0.017	0.774	-0.034	0.048	0.486
Whiting Petroleum	-0.363***	0.062	0.000	1.101***	0.180	0.000	0.306***	0.102	0.003	0.027	0.198	0.893

	VOL_{t-1}			$CDXIG$			ΔWTI_{t-1}			WTI_{t-2}		
Company name	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value
Anadarko Petroleum	-0.009	0.012	0.434	0.002***	0.000	0.000	-0.421***	0.081	0.000	-0.040**	0.019	0.032
Apache	0.015	0.014	0.285	0.001***	0.000	0.000	-0.450***	0.086	0.000	-0.072***	0.023	0.002
Chesapeake Energy	-0.008	0.015	0.620	0.003***	0.000	0.000	0.033	0.091	0.721	-0.041**	0.023	0.068
ConocoPhillips	0.012	0.017	0.467	0.001***	0.000	0.000	-0.247***	0.085	0.004	-0.068***	0.018	0.000
Devon Energy	-0.010	0.017	0.556	0.001***	0.000	0.005	-0.237**	0.095	0.013	-0.054**	0.023	0.020
Encana	-0.007	0.013	0.602	0.002***	0.000	0.000	-0.169**	0.084	0.043	0.016	0.021	0.449
Marathon Oil	-0.007	0.012	0.546	0.001***	0.000	0.000	-0.379***	0.074	0.000	-0.069**	0.028	0.014
Murphy Oil	0.039*	0.016	0.018	0.002***	0.000	0.000	-0.292***	0.084	0.001	-0.072***	0.028	0.009
Noble Energy	-0.075***	0.020	0.000	0.000*	0.000	0.085	0.078	0.073	0.282	-0.028*	0.016	0.085
Pioneer Nat. Res.	-0.024*	0.014	0.081	0.002***	0.000	0.000	-0.158**	0.074	0.032	-0.016	0.011	0.128
Whiting Petroleum	-0.087**	0.034	0.012	0.009***	0.002	0.000	-0.071	0.218	0.746	-0.574***	0.181	0.002

Table 2.9: Coefficients, standard errors and p-values resulting from the estimation of Equation (2.4) for oil producing firms. Intercept included but not reported. *** significance at the 1% confidence level, ** significance at the 5% confidence level, * significance at the 10% confidence level.

	CDS_{t-1}			ΔLEV			LEV_{t-1}			ΔVOL		
Company name	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value
Enbridge	-0.050***	0.015	0.001	0.929***	0.120	0.000	0.196***	0.052	0.000	0.056	0.058	0.329
Ensco	-0.077***	0.016	0.000	0.471***	0.068	0.000	0.137***	0.025	0.000	-0.025	0.049	0.620
Halliburton	-0.056***	0.013	0.000	0.440***	0.071	0.000	0.017	0.023	0.468	0.002	0.057	0.966
Nabors	-0.004	0.014	0.775	-0.111	0.095	0.246	-0.015	0.027	0.579	0.068	0.054	0.205
Transocean	-0.039**	0.019	0.043	1.783***	0.140	0.000	0.082***	0.029	0.005	-0.013	0.058	0.822
Valero Energy	-0.043***	0.014	0.001	0.519***	0.093	0.000	-0.032	0.021	0.126	0.146**	0.057	0.010
Weatherford	-0.061***	0.023	0.007	0.794***	0.127	0.000	0.158**	0.070	0.023	-0.051	0.077	0.510

	VOL_{t-1}			$CDXIG$			ΔWTI_{t-1}			WTI_{t-2}		
Company name	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value
Enbridge	0.016	0.014	0.250	0.001***	0.000	0.000	-0.140*	0.079	0.074	-0.009	0.016	0.563
Ensco	0.003	0.012	0.775	0.001***	0.000	0.003	-0.070	0.057	0.221	-0.023	0.019	0.219
Halliburton	-0.017	0.013	0.171	0.001***	0.000	0.000	-0.122*	0.063	0.052	-0.045**	0.020	0.022
Nabors	-0.016	0.021	0.437	0.000***	0.000	0.639	-0.012	0.091	0.891	-0.022	0.023	0.329
Transocean	-0.001	0.015	0.949	0.001***	0.000	0.005	-0.262***	0.094	0.005	-0.004	0.042	0.928
Valero Energy	0.028	0.018	0.130	0.001***	0.000	0.000	-0.189***	0.073	0.009	-0.009	0.010	0.375
Weatherford	0.007	0.020	0.727	0.002***	0.001	0.017	-0.112	0.114	0.326	0.006	0.019	0.767

Table 2.10: Coefficients, standard errors and p-values resulting from the estimation of Equation (2.4) for oil servicing firms. Intercept included in the estimations but not reported. *** significance at the 1% confidence level, ** significance at the 5% confidence level, * significance at the 10% confidence level.

	CDS-WTI						CDS-LEVERAGE					
Oil Producing Companies	t	t-1	t-2	t-3	t-4	t-5	t	t-1	t-2	t-3	t-4	t-5
Anadarko Petroleum	-0.433	0.000	0.000	0.000	0.000	0.000	0.992	-0.005	-0.005	-0.004	-0.004	-0.003
Apache	-0.464	-0.032	-0.029	-0.026	-0.024	-0.022	0.691	-0.003	-0.003	-0.003	-0.002	-0.002
Chesapeake Energy	-	-	-	-	-	-	1.840	0.005	0.005	0.004	0.004	0.003
ConocoPhillips	-0.273	-0.052	-0.048	-0.043	-0.040	-0.036	0.770	-0.002	-0.002	-0.002	-0.001	-0.001
Devon Energy	-0.224	-0.042	-0.040	-0.038	-0.036	-0.034	1.531	-0.005	-0.004	-0.004	-0.004	-0.004
Encana	-0.167	0.028	0.026	0.024	0.022	0.021	0.896	0.137	0.127	0.118	0.109	0.101
Marathon Oil	-0.416	-0.038	-0.036	-0.033	-0.031	-0.029	0.543	0.012	0.011	0.011	0.010	0.009
Murphy Oil	-0.313	-0.038	-0.033	-0.029	-0.026	-0.022	0.531	0.025	0.022	0.019	0.017	0.015
Noble Energy	-	-	-	-	-	-	0.487	0.118	0.105	0.093	0.082	0.073
Pioneer Nat. Res.	-0.167	-0.006	-0.005	-0.005	-0.005	-0.004	0.472	-0.034	-0.032	-0.030	-0.028	-0.026
Whiting Petroleum	-	-	-	-	-	-	1.070	-0.099	-0.062	-0.039	-0.025	-0.015
	CDS-WTI						CDS-LEVERAGE					
Oil Servicing Companies	t	t-1	t-2	t-3	t-4	t-5	t	t-1	t-2	t-3	t-4	t-5
Enbridge	-0.144	0.008	0.008	0.008	0.007	0.007	1.067	0.142	0.136	0.130	0.125	0.119
Ensco	-	-	-	-	-	-	0.459	0.103	0.095	0.088	0.081	0.075
Halliburton	-0.139	-0.036	-0.034	-0.032	-0.030	-0.029	0.442	-0.013	-0.012	-0.011	-0.011	-0.010
Nabors Industries	-	-	-	-	-	-	-	-	-	-	-	-
Transocean	-0.272	0.006	0.006	0.006	0.005	0.005	1.832	0.011	0.010	0.010	0.010	0.009
Valero Energy	-0.191	0.000	0.000	0.000	0.000	0.000	0.512	-0.056	-0.053	-0.051	-0.049	-0.047
Weatherford	-	-	-	-	-	-	0.825	0.113	0.106	0.100	0.093	0.088

Table 2.11: First five lags of Impulse Response Function of lagged WTI returns and Leverage Ratios.

Oil Producers					Oil Servicing Companies				
	Model 1		Model 2			Model 1		Model 2	
Company name	AIC	BIC	AIC	BIC	Company name	AIC	BIC	AIC	BIC
Anadarko Petroleum	-2.497	-2.350	**-2.544**	**-2.420**	Enbridge	-2.669	-2.522	**-2.675**	**-2.550**
Apache	-2.380	-2.233	**-2.389**	**-2.264**	Ensco	-2.781	-2.634	**-2.792**	**-2.667**
Chesapeake Energy	-2.308	-2.161	**-2.318**	**-2.193**	Halliburton	-2.830	-2.683	**-2.833**	**-2.709**
Conocophillips	-2.455	-2.308	**-2.464**	**-2.340**	Nabors	**-2.470**	-2.303	-2.455	**-2.314**
Devon Energy	**-2.299**	-2.152	-2.291	**-2.167**	Transocean	-2.488	-2.341	**-2.499**	**-2.375**
Encana	-2.522	-2.375	**-2.531**	**-2.406**	Valero Energy	**-2.781**	-2.634	-2.776	**-2.652**
Marathon Oil	**-2.586**	-2.439	-2.583	**-2.459**	Weatherford	-2.687	-2.449	**-2.705**	**-2.503**
Murphy Oil	**-2.456**	-2.309	-2.447	**-2.323**					
Noble Energy	**-2.909**	**-2.743**	-2.873	-2.733					
Pioneer Natural Resources	-2.574	-2.427	**-2.578**	**-2.453**					
Whiting Petroleum	**-2.047**	-1.735	-2.022	**-1.758**					

Table 2.12: Akaike Information Criterion (AIC) and Schwartz Information Criterion (BIC) resulting by the estimation of Model 1 and 2.
The AIC and BIC of the preferred model are reported in bold.

	CDS_{t-1}			ΔLEV			LEV_{t-1}		
Company name	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value
Anadarko Petroleum	-0.100***	0.017	0.000	0.986***	0.097	0.000	0.091***	0.034	0.008
Apache	-0.080***	0.019	0.000	0.674***	0.097	0.000	0.053**	0.026	0.037
Chesapeake Energy	-0.105***	0.016	0.000	1.767***	0.118	0.000	0.193***	0.034	0.000
Conocophillips	-0.069***	0.014	0.000	0.765***	0.117	0.000	0.050*	0.030	0.096
Devon Energy	-0.053***	0.017	0.002	1.548***	0.119	0.000	0.096**	0.045	0.034
Encana	-0.072***	0.015	0.000	0.895***	0.086	0.000	0.198***	0.035	0.000
Marathon Oil	-0.092***	0.012	0.000	0.537***	0.072	0.000	0.066**	0.028	0.018
Murphy Oil	-0.093***	0.020	0.000	0.512***	0.102	0.000	0.068***	0.013	0.000
Noble Energy	-0.071***	0.016	0.000	0.451***	0.079	0.000	0.126***	0.029	0.000
Pioneer Natural Res.	-0.060***	0.013	0.000	0.471***	0.067	0.000	-0.007	0.017	0.699
Whiting Petroleum	-0.334***	0.064	0.000	1.092***	0.186	0.000	0.160**	0.077	0.039
	$CDXIG$			ΔWTI_{t-1}			WTI_{t-2}		
Company name	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value
Anadarko Petroleum	0.002***	0.000	0.000	-0.418***	0.086	0.000	-0.045*	0.020	0.023
Apache	0.001***	0.000	0.000	-0.444***	0.085	0.000	-0.073***	0.023	0.001
Chesapeake Energy	0.003***	0.000	0.000	0.041	0.090	0.648	-0.037*	0.021	0.075
Conocophillips	0.001***	0.000	0.000	-0.250***	0.084	0.003	-0.073***	0.017	0.000
Devon Energy	0.001***	0.000	0.007	-0.213**	0.095	0.025	-0.044**	0.021	0.034
Encana	0.002***	0.000	0.000	-0.172**	0.084	0.040	0.021	0.019	0.255
Marathon Oil	0.002***	0.000	0.000	-0.382***	0.073	0.000	-0.089***	0.027	0.001
Murphy Oil	0.002***	0.000	0.000	-0.298***	0.084	0.000	-0.089***	0.029	0.002
Noble Energy	0.000*	0.000	0.072	0.056	0.074	0.451	0.017	0.012	0.176
Pioneer Natural Res.	0.002***	0.000	0.000	-0.186***	0.072	0.010	-0.022**	0.010	0.027
Whiting Petroleum	0.005***	0.001	0.000	-0.174	0.226	0.441	-0.633***	0.183	0.001

Table 2.13: Coefficients, standard errors and p-values resulting from the estimation of Equation (2.15) for oil producing firms. Intercept included but not reported. *** significance at the 1% confidence level, ** significance at the 5% confidence level, * significance at the 10% confidence level.

	CDS_{t-1}			ΔLEV			LEV_{t-1}		
Company name	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value
Enbridge	-0.041***	0.013	0.001	0.980***	0.122	0.000	0.210***	0.050	0.000
Ensco	-0.076***	0.014	0.000	0.479***	0.068	0.000	0.139***	0.025	0.000
Halliburton	-0.067***	0.012	0.000	0.445***	0.069	0.000	0.032	0.022	0.160
Nabors	-0.014	0.013	0.252	-0.090	0.102	0.380	-0.023	0.028	0.414
Transocean	-0.040**	0.019	0.031	1.789***	0.136	0.000	0.082***	0.028	0.004
Valero Energy	-0.031***	0.011	0.006	0.536***	0.089	0.000	-0.048**	0.020	0.017
Weatherford	-0.056**	0.022	0.011	0.784***	0.127	0.000	0.136**	0.064	0.033

	$CDXIG$			ΔWTI_{t-1}			WTI_{t-2}		
Company name	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value	Coeff	Std. Error	P-Value
Enbridge	0.001***	0.000	0.000	-0.137*	0.077	0.076	-0.008	0.016	0.635
Ensco	0.001***	0.000	0.001	-0.071	0.057	0.217	-0.021	0.017	0.224
Halliburton	0.001***	0.000	0.000	-0.125*	0.063	0.045	-0.041***	0.019	0.035
Nabors	0.000	0.000	0.855	-0.045	0.079	0.569	-0.032	0.020	0.110
Transocean	0.001***	0.000	0.003	-0.264***	0.093	0.004	-0.004	0.042	0.919
Valero Energy	0.001***	0.000	0.000	-0.174**	0.072	0.016	-0.001	0.008	0.935
Weatherford	0.001**	0.001	0.023	-0.113	0.115	0.325	0.001	0.015	0.920

Table 2.14: Coefficients, standard errors and p-values resulting from the estimation of Equation (2.15) for oil servicing firms. Intercept included but not reported. *** significance at the 1% confidence level, ** significance at the 5% confidence level, * significance at the 10% confidence level.

Chapter 3

Machine Learning in presence of mixed signs: the case of crude oil prices in 2020

Abstract

Oil markets experienced a truly unprecedented situation in 2020. Crude consumption declined at the start of the year, following a stagnant worldwide economic situation in the second half of 2019. At the same time, Saudi Arabia and Russia, the top producing countries after the United States, did not commit to production cuts. Furthermore, the lockdown introduced worldwide as a measure to stop the COVID-19 pandemic sent gasoline and jet fuel consumption crashing. As a consequence, storage facilities in the US soon reached their full capacity, causing the WTI first nearby Future to plunge to negative levels on April 20th 2020 for the first and only time in history. In this chapter we first test the viability of two price prediction frameworks over three different 3-months periods enclosed in the first half of 2020. Employing Variational Mode Decomposition, 5-minute WTI and Brent prices are decomposed into modes, which are then used as inputs to forecast the one-period-ahead price via Generalised Additive Model (GAM) and Feed-Forward Neural Network (FFNN). We propose a Recursive Forecasting Methodology (RFM) to compute the aforementioned forecasts by using modes generated exclusively from past data. The results show that forecasting via FFNN is more accurate in five out of six cases. With respect to WTI, highest accuracy is obtained when the frameworks were trained using both positive and negative prices, while the test data was strictly positive. Contrarily, the methodologies forecasting ability is highly affected by the presence of negative prices in the test dataset of WTI in the period February-April 2020. The lower accuracy resulting from predicting Brent prices during the same period suggests the existence of a spillover effect (between the two benchmark indices) of the structural instability caused by the unprecedented WTI price drop. The application of the two frameworks to 5-minute WTI prices observed between May 2023 and April 2024, a

longer time period where no structural instabilities were detected, leads to lower RMSE and MAPE. The results obtained tend to approach the performance of the baseline Exponential Smoothing for all datasets. However, the presence of negative prices in the WTI February - April 2020 testing dataset deteriorates tenfold the accuracy of all methodologies, including ES, displaying the limitations of the frameworks to capture the structural instability generated by the unforeseen drop of WTI prices to negative levels.

3.1 Introduction and literature review

During the first half of 2020, amid the Coronavirus pandemic, the global economy entered a stage of deep distress with local authorities enforcing lockdown measures and banning non-essential travel to avoid the spread of the virus. The disappearance of daily commutes, leisure travel and the closure of factories around the world reduced the world consumption of crude oil and its refined products by a third almost overnight. At the same time, crude future prices began to drop as market participants expected storage facilities to rapidly fill up. Alarms were further raised when the main delivery point for oil in the United States - Cushing, Oklahoma - reached 80% of its capacity. In an attempt to create a floor for oil prices, the US government authorised the acquisition of some 75 million barrels to fill up the American national petroleum reserves and backed a landmark deal by OPEC+ to curb production by as much as 10%. Despite these efforts, the West Texas Intermediate contracts for delivery in May 2020 traded for as low as -$38 on April 20th, i.e., as soon as the maximum storage capacity in Cushing was reached. For the first time in history, traders were paying money to sell their barrels of oil. Benchmarks around the world also came under pressure with Brent plummeting to $19, the lowest since the late 90s. However, contracts for later delivery were trading at higher prices as they included a premium for the increasing storage costs, displaying a market feature which was labelled as "super-contango."

Despite first-nearby contracts returning to positive levels within the following trading day and lockdown measures slowly being lifted in many countries, the future of the oil sector was very uncertain: the market was still oversupplied with a lot of oil that had no demand to meet; producing countries like UAE, Saudi Arabia, Iraq, Qatar and Algeria, whose budgets rely on petrodollars were expected to cut their government spending being unable to meet their production break-even point. Furthermore, independent crude producers around the globe were faced with an unprecedented crisis. In the US alone 32 oil patches and 25 oilfield services companies filed for Chapter 11 during the first three quarters of 2020.[1]

The focus of this chapter is the prediction of WTI and Brent in those crucial days.

[1]Source: the Haynes Boone Energy Bankruptcy Reports and Surveys. `https://www.haynesboone.com/publications/energy-bankruptcy-monitors-and-surveys`

To wit, we examine oil five-minute prices over 3 different (but overlapping) time periods: *i*) January 1st 2020 to March 31st 2020; *ii*) February 1st 2020 to April 30th 2020; *iii*) April 1st 2020 to June 30th 2020. The time periods were carefully chosen in order to test the methodologies in a period of "standard" positive prices (*i*), a period where negative WTI prices appeared in the test dataset (*ii*) and a period where negative WTI prices are part of the training dataset (*iii*). Further, we extended the analysis to WTI prices ranging between May 1st 2023 and April 30th 2024. The introduction of a larger dataset allows us to have a further comparison with the forecasts of the 2020 datasets. Moreover, whilst being more computationally expensive, training the models on a greater number of observations should lead to a lower sensitivity to noise in the data, thus better capturing the underlying distribution of the data (including outliers) and improve generalisation by reducing overfitting. We applied Variational Mode Decomposition (VMD) to map the non-stationary WTI and Brent prices time series to functions, called modes, which are then used as covariates to train a Generalised Additive Model (GAM) and a Feed-Forward Neural Network (FFNN) for forecasting purposes. We proposed a Recursive Forecasting Methodology which allows to predict oil prices using modes generated exclusively from past data. The accuracy of the forecasts is measured via RMSE and MAPE. Additionally, we implemented Exponential Smoothing (ES) as a baseline to compare the performance of the frameworks. The results obtained tend to approach the performance of the baseline Exponential Smoothing for all datasets. However, the presence of negative prices in the WTI February - April 2020 testing dataset deteriorates tenfold the accuracy of all methodologies, including ES, displaying the limitations of the frameworks to capture the structural instability generated by the unforeseen drop of WTI prices to negative levels.

Empirical Mode Decomposition (EMD; Huang et al., 1998), is an empirical multi-resolution technique used to perform a joint space-spatial frequency decomposition of a signal by successive removal of elemental intrinsic mode functions (IMF), which represents the oscillatory modes of the original signal going from high to low frequency ranges. EMD has been applied in many fields, including the prediction of financial time series: for instance Premanode & Toumazou (2013) used it to forecast FX, whilst Zhang et al. (2008), Zhang et al. (2009), An et al. (2013) and Lisi & Nan (2014) applied EMD

to predict crude oil and electricity prices in various settings.

Conversely, Variational Mode Decomposition (VMD) is a more recent technique introduced by Dragomiretskiy & Zosso (2014). It is used to decompose a signal into a predetermined amount of modes, which oscillate around their respective central frequency. Compared to EMD, VMD is preferable for its capability to separate tones of similar frequencies achieving a better signal characterisation and effectiveness in de-noising the underlying time series, (Dragomiretskiy & Zosso, 2014).

On one hand, Generalised Additive Models, (GAM; Hastie & Tibshirani, 1986), are a popular regression approach that model the dependence of a response variable on a set of covariates using a flexible specification of the additive predictor via smooth functions. In our context, the independent variables correspond to the modes extracted via the VMD framework. GAMs relax the parametric assumptions of linear and generalised linear models by modelling the effects of continuous covariates using smooth functions. This is appealing for non-stationary and non-linear time series forecasting. On the other hand, Feed-Forward Neural Network (FFNN) learns from examples: sets of inputs are fed into the network, the corresponding outputs are then calculated and compared to the actual/real data. In a recurrent manner, the gradient of the loss function is computed and the coefficients used to calculate the outputs are adjusted until a chosen stopping criterion is met. Similarly to GAMs, once the structure of the FFNN is determined, there is no need to pre-specify a functional form that describes the underlying relationship between the dependent and independent variables.

In recent years, EMD and VMD hybrid models have been adopted in the prediction of financial time series and compared extensively. Nava et al. (2018) used Support Vector Regression to forecast financial time series decomposed via EMD. Hong (2011) used EMD to forecast five-minute crude oil prices. Wang et al. (2018) analyse the relationship between Internet Concern (a measure of investor attention based on internet data) and oil price volatility. The Bivariate Empirical Mode Decomposition (BEMD) method is used the time series, whilst Extreme Learning Machine (ELM) is applied to forecast price volatility in various settings, thus showing how incorporating Internet Concern in the model improves accuracy. Lahmiri (2015) compares the prediction of

California electricity prices and Brent crude oil using Generalised Regression Neural Networks (GRNN) fed with modes obtained via EMD and via VMD. Also, Lahmiri (2016) shows the superior forecast accuracy of a VMD ensemble model for daily WTI prices, Canadian/US exchange rate and Chicago Board Options Exchange NASDAQ 100 Volatility Index (VIX). In their work, the VMD-GRNN hybrid model is compared with an EMD-GRNN ensemble, with an Auto-Regressing Moving Average (ARMA) model, and with a Feed-Forward Neural Network (FFNN) trained with the past five observations. Similarly, Gyamerah (2020) shows that training a GRNN model using modes obtained via VMD produces more accurate forecasts compared to using IMFs extracted via EMD in the case of one-minute interval Bitcoin prices. Lastly, Zhu et al. (2019) forecast daily carbon prices from the Shenzhen and Hubei Province in China by applying VMD to the original time series. An evolutionary clustering algorithm is adopted to create virtual modes which are individually used to predict the carbon price and then combined via the Induced Ordered Weighted Averaging operator to produce the final forecast.

The remainder of the chapter is organised as follows. In Section 3.2 we introduce the adopted methodologies, the Recursive Forecasting Methodology and the related performance measures. The data analysed and the respective results are presented in Sections 3.3, 3.3.2 and 3.3.3. In Section 3.4 we summarise and conclude the chapter.

3.2 Methodology

In this section we describe the methodologies applied in this paper, namely Empirical Mode Decomposition, Variational Mode Decomposition, Generalised Additive Models, Feed-Forward Neural Networks and the Recursive Forecasting Methodology proposed.

3.2.1 Empirical Mode Decomposition

Empirical Mode Decomposition (EMD) allows to decompose a complex signal (such as non-linear non-stationary time series) into a finite and often small number of functions called Intrinsic Mode Functions (IMF). Following Huang et al. (1998), each IMF has the same number of zero crossing and extrema and is symmetric with respect to its local mean. The recursive procedure, called sifting algorithm (Huang et al., 1998), performs the following steps:

- Find all the local maxima M_i $(i=1,2,...)$ and minima m_k $(k=1,2,...)$ in the signal $S(t)$;
- Interpolate the lower and upper envelopes via cubic spline: $M(t)=f_M(M_i,t)$ and $m(t)=f_m(m_i,t)$;
- Compute the envelope mean $\mu_t=M(t)-m(t)$;
- Compute the series $Z(t)=S(t)-\mu_t$;
- Repeat the previous steps until: i) μ_t approaches zero, ii) the numbers of local extrema and zero-crossings differ by at most 1, or iii) the pre-defined maximum number of iterations is reached. Then, if $Z(t)$ meets the IMF conditions of same number of zero crossings and extrema and is symmetric around its local mean, $IMF_i(t)=Z(t)$ with residual $r(t)=S(t)-IMF_i(t)$;
- If $Z(t)$ does not meet the required criteria then $S(t)$ is replaced by $Z(t)$;
- Iterate the previous steps computing $IMF_1,IMF_2,...$ and $r_1,r_2,...$ until $IMF_i(t)_1^{n-1}$ such that $r_i(t)_1^{n-1}$ does not have more than two local extrema.

Lastly, the signal $S(t)$ can be expressed as

$$S(t) = \sum_{j=1}^{N} IMF_j(t) + r_N(t). \tag{3.1}$$

Lahmiri (2016) and Gyamerah (2020) train their prediction models using IMFs extracted via EMD and modes extracted via VMD. They show that the VMD hybrid model leads to more accurate forecasts. As a result, in this research we applied EMD exclusively in the process of choosing the optimal number of modes k to be extracted via VMD. This procedure is described in Section 3.2.2.

All the Empirical Mode Decomposition calculations are run via the `EMD` R package (Kim & Oh, 2021).

3.2.2 Variational Mode Decomposition

VMD is a technique that aims to decompose a time series into a discrete number k of modes where each has limited bandwidth in spectral domain. Each mode k is required to be mostly located around a center pulsation ω_k obtained during the decomposition process (Dragomiretskiy & Zosso, 2014). Constraints like the recursive sifting procedure, hard-band limits, and lack of mathematical theory in other decomposition methods can be solved using VMD (Isham et al., 2018). Allowing to decompose a time series f in a pre-determined number of modes k, VMD has been adopted to deal with noise in the signal of time series data (Dragomiretskiy & Zosso, 2014). A signal is decomposed via the following recursive algorithm (Dragomiretskiy & Zosso, 2014):

1. a Hilbert transform is applied on each mode u_k to retrieve a unilateral frequency spectrum;

2. each mode's frequency spectrum is shifted to baseband by mixing it with an exponential tuned to the respective estimated center frequency;

3. the bandwidth is estimated via Gaussian smoothness of the demodulated signal.

As a result, the constrained problem that follows can be described as

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \tag{3.2}$$

subject to

$$\sum_k u_k = f, \tag{3.3}$$

where f is the signal, t is the time script, k is the number of modes, u_k is k-th the mode, $j^2 = -1$, δ is the Dirac distribution, ω is the frequency, $*$ represents the convolution and $\|\cdot\|_2^2$ is the squared L2 norm of the gradient. Higher order k denote lower frequency modes (Dragomiretskiy & Zosso, 2014).

Further, the number of modes k extracted via VMD has to be specified by the user. To determine the optimal k value, we followed the procedure described in Lahmiri (2016) and Gyamerah (2020). Namely, we first decompose the data via EMD and obtain N IMFs. Then, we run the VMD-FFNN and VMD-GAM frameworks for values of k in the range $k = [N-4, N+4]$ for each dataset. The optimal value of k is then chosen via cross-validation of the performance metrics described in Section 3.2.6. Variational Mode Decomposition is run via the `VMDecomp` R package (Mouselimis, 2022).

3.2.3 Generalised Additive Models

Generalised Additive Models (GAM; Hastie & Tibshirani, 1986) model the dependence of a response variable on a set of covariates using a flexible specification of an additive predictor as follows (Wood, 2017)

$$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) + ... + f_k(x_{ki}), \tag{3.4}$$

where $g(\cdot)$ is a link function that connects the expected value of the dependent variable to the predictors, $\mu_i = \mathbb{E}(Y_i)$ and the response variable Y_i belongs to some exponential family. X_i^* defines a vector of parametric predictors, whilst each f_i represents a smooth function of the continuous predictor x_j, for $j = 1,...,k$.

In our case the model assumes the form

$$g(\mu_i) = \beta_0 + f_1(M_{1,i}) + f_2(M_{2,i}) + ... + f_j(M_{k,i}) \tag{3.5}$$

where $M_{j,i}$ represents the i-th observation of the j-th mode extracted via VMD, for $j = 1,...,k$.

GAMs provide flexibility in the model specification, avoiding assumptions typical of parametric linear models and allowing the relaxation of linearity constraints of generalised linear models. Each smooth function can be written as a linear combination of regression parameters and known basis function as

$$f_j(x_{j,i}) = \sum_{q_j=1}^{Q_j} b_{j,q_j}(x_{j,i})\beta_{q_j}, \quad j = 1,...,k, \tag{3.6}$$

where $b_{j,q_j}(\cdot)$ denotes the q_j-th basis function evaluated at the i-th observations of the j-th covariate in the model and β_{q_j} represents the corresponding regression parameter (Wood, 2017). The number of basis function is chosen to be large enough to capture the relationship between the dependent variable and the covariates (Wood, 2017). Note that the smooth function representation in Equation (3.6) allows to write the model in (3.4) as a linear model. The degree of smoothness of the function is determined by adding a "wiggliness" penalty to the least-squares estimation, which corresponds to the integrated squared second order derivative of the smooth function. Thus, the model parameters are estimated by minimising the following penalised least-squares criterion:

$$\|y - X\beta\|^2 + \sum_j \lambda_j \int [f_j''(x_j)]^2 dx. \tag{3.7}$$

where X contains the parametric predictors and the basis evaluated at the continuous variables, β contains the intercept and all the regression coefficients associated to the smooth functions. In Equation (3.7), $\lambda_j \geq 0$ determines the smoothness of the estimated function. When $\lambda_j \to 0$, f is close to a spline without penalty, resulting in higher wiggliness. On the other hand, the estimated function approaches a straight line whenever

$\lambda_j \to \infty$. The penalty for each smooth function can be written as a quadratic form

$$\lambda_j \int [f_j''(x_j)]^2 dx_j = \beta_j^T S_j \beta_j, \tag{3.8}$$

where

$$S_j = \lambda_j \int b_j b_j^T dx_j. \tag{3.9}$$

and b_j is a vector of the basis functions evaluated at the covariates. The minimisation problem becomes

$$\|y - X\beta\|^2 + \beta^T S \beta \tag{3.10}$$

where $S = diag(0, \lambda_1 S_1, ..., \lambda_k S_k)$ is the overall penalty matrix. Parameter estimation proceeds in a two-stages iterative approach. In the first step, given a fixed value of λ_k, the regression coefficients are estimated using a (weighted) least-squares approach. In the second step, given the updated values of β, the smoothing parameters are estimated using generalised cross-validation (Craven & Wahba, 1979; Golub et al., 1979). For details on the estimation approach, we refer the reader to Wood (2017).

Thin plate regression splines are the basis used to approximate each f_i. They are considered an ideal smoother as they are constructed by defining how much weight to give to the conflicting goals of matching the data and making f_i smooth, and finding the function that best satisfies the resulting smoothing objective (Wood, 2017). The standard GAMs implementation is the one from the R `mgcv` package by Simon Wood (Wood, 2020).

3.2.4 Feed-Forward Neural Networks

A Feed-Forward Neural Network (FFNN) can be thought of a parallel distributed processor composed of elementary units called neurons. They allow the storage of prior experiential knowledge in the form of synaptic weights acquired through the intermediate steps of the learning process. The adaptability of the system enables the adjustment of the synaptic weights and the system structure. The network architecture consists of one input layer, one or more hidden layer(s), containing the neurons, and one output layer.

The outputs are calculated as:

$$\hat{Y} = \phi(WX + b) \tag{3.11}$$

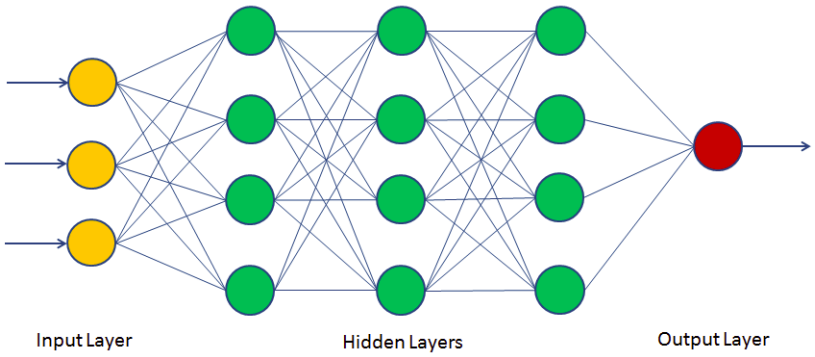


Figure 3.1: Example of the structure of a neural network

where X represent the input layer, W is a vector of weights, b is the bias and ϕ is a chosen activation function.

The network learns by adjusting the values of the weights that connect each layer to the following one. Its objective is the minimisation of the chosen loss function, in our case given by the sum of squared differences between the output $\hat{y}_i$ and the target value y_i:

$$L(\hat{y}_i, y_i) = \frac{1}{2}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2, i = 1,...,N. \tag{3.12}$$

The adjustment starts from the output layer and propagates backwards towards the input layer, in a gradient descent fashion:

$$w_i^{new} = w_i^{old} + \alpha \Delta L(\hat{y}_i, y_i) \tag{3.13}$$

where w_i are the weights of the corresponding i-th input, α represents the learning rate, which determines the speed of change, and $\Delta L(\hat{y}, y)$ is the partial derivative of the error function with respect to the i-th input weight w_i for each layer.

Normalising the input data facilitates the convergence of the algorithm and overcomes the possible weakness of the chosen loss function - namely being influenced by the magnitude of the inputs.

In our context, the structure of our the neural network consists of the input layer, one hidden layer and the output layer. The inputs layer size corresponds to k, the number modes extracted via VMD. The number of nodes in the hidden layer is arbitrarily set

to $k-1$, leaving the optimisation of this parameter via cross-validation for future works on the subject. The output consists of one node, corresponding to the predicted price. Results are compared by choosing the optimal value k via cross-validation based on the performance criteria described in Section 3.2.6. All computations were done via the `neuralnet` R package (Fritsch et al., 2019).

3.2.5 Recursive Forecast Methodology (RFM) for testing

As presented in Section 3.1, VMD-hybrid frameworks are used in the literature to forecast a number of financial time series (Lahmiri, 2015, 2016; Gyamerah, 2020; Zhu et al., 2019). The authors divide their data into training and testing sets, following the classical 80%-20% split. The models are trained by decomposing the training data via VMD into a certain number k of modes, determined following the procedure described in Section 3.2.2. The test data is then decomposed via VMD in the same number of k modes. Finally, for each time t, their value is fed into the trained model to obtain the one-period-ahead price forecast and the performance metrics are computed. For instance, let

$$\begin{aligned}
&Y_t \ (t=0,...,T)\text{: Price time series,}\\
&Y_t^{TRAIN} \ (t=0,...,j)\text{: Training data,}\\
&Y_t^{TEST} \ (t=j+1,...,T)\text{: Testing data,}
\end{aligned}$$

where the data is classically split 80% for training and 20% for testing. The model is trained on the k modes extracted via VMD of Y_t^{TRAIN}

$$M_{t,i}^{TRAIN} \ (t=0,...,j; \ \ i=1,...,k).$$

Then, Y_t^{TEST} is also decomposed in k modes

$$M_{t,i}^{TEST}, \ (t=j+1,...,T; \ \ i=1,...,k).$$

Feeding to the trained model

$$\mathscr{M}_{j+1}^{TEST} = [M_{j+1,1}^{TEST},...,M_{j+1,k}^{TEST}],$$

the values of the k modes at time $j+1$, the forecast for $\hat{Y}_{j+2}$ is obtained. Doing so for $t=j+1,...,T-1$ returns the forecast time series $\hat{Y}_t$ $(t=j+2,...,T)$. By decomposing the entire testing dataset, the resulting modes will include, in some sense, future information encompassed in the testing data trajectory. Thus, the forecast is computed using modes generated from future data given that the values of $M_{t,i}^{TEST}$ are conditional to Y_t^{TEST} $(t=j+1,...,T)$. Thus,

$$M_{\tau,i}^{TEST} \ (j+1<\tau<T),$$

depends on both

$$Y_t^{TEST} \ (t=j+1,...,\tau),$$
$$Y_s^{TEST} \ (s=\tau+1,...,T).$$

Hence, the predicted price are not reliable as they include future information on the data trajectory.

In the real world, only past data is available at time t when predicting the price at $t+1$. Once trained, the framework requires the value of k modes at time t as inputs to return the forecast $\hat{Y}_{t+1}$. Thus, we split Y_t as

$$Y_t^{TRAIN} \ (t=0,...,j)\text{: Training data,}$$
$$Y_t^{GAP} \ (t=j+1,...,J)\text{: Gap data,}$$
$$Y_t^{TEST} \ (t=J+1,...,T)\text{: Testing data.}$$

Then, we extract k modes via VMD of Y_t^{TRAIN} and Y_t^{GAP}, namely

$$M_{t,i}^{TRAIN} \ (t=0,...,j; \ \ i=1,...,k)\text{: Training data modes,}$$
$$M_{t,i}^{GAP} \ (t=j+1,...,J; \ \ i=1,...,k)\text{: Gap data modes.}$$

The framework is trained on $M_{t,i}^{TRAIN}$; in this phase, the Gap data remains unseen by the model. By feeding $\mathscr{M}_J^{GAP}$, the vector containing the value of the modes at time J, to the

trained model the first forecast $\hat{Y}_{J+1}$ is computed. Then, in a recursive manner, thus the name Recursive Forecasting Methodology, we append one observation of Y_t^{TEST} to Y_t^{GAP}

$$Y_t^{GAP} \ (t = j+1,...,J+1).$$

and decompose it into k modes

$$M_{t,i}^{GAP} \ (t = j+1,...,J+1; \ \ i = 1,...,k).$$

The forecast for $\hat{Y}_{J+2}$ is obtained by feeding $\mathscr{M}_{J+1}^{GAP}$ to the trained model. The recursive loop ends when

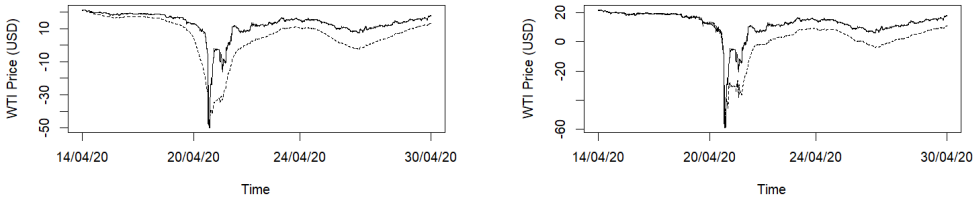
$$\hat{Y}_t \ (t = J+1,...,T)$$

is obtained. Lastly, we computed the evaluation metrics presented in Section 3.2.6. A synthesis of the RFM is given in Algorithm 1. Following this procedure, the forecast $\hat{Y}_t$ is computed using modes generated exclusively from past data.

Figure 3.2 shows a plot of the WTI February-April realised prices against the predicted values obtained via the VMD-FFNN hybrid model. On the left, the forecast was generated using the testing methodology from the literature, and, on the right, it was generated using the RFM. Note that, for this dataset, $Y_t^{TRAIN} \in \mathbb{R}^+$ and $Y_t^{TEST} \in \mathbb{R}$. Using the standard methodology, the value of $M_{t,i}^{TEST}$ between 14/04/2020 and 19/04/2020 are lower as they reflect the price drop observed on 20/04/2020. The resulting forecast suffer from a clear downward bias as shown in Figure 3.2a. Furthermore, between 14/04/2020 and 19/04/2020 prices were always positive. The Root Mean Squared Error (see Equation (3.14)) computed on this subset should be in line with the results obtained for a dataset with strictly positive values in the testing data. Namely, as presented in Table 3.6 in Section 3.5, the forecast of the WTI January-March dataset returns $RMSE = 0.816$. However, using the standard methodology, the RMSE between 14/04/2020 and 19/04/2020 is 3.35.

As shown in Figure 3.2b, if prices are predicted recursively as proposed in this section, the forecasts is far more accurate between 14/04/2020 and 19/04/2020 producing $RMSE = 0.567$. However, as it will be discussed later in Section 3.3, for both methodologies the accuracy is highly reduced when negative prices are predicted. In this context,

we allocated 80% of the data to the training dataset, 1% as Gap data, and the remaining 19% for testing. The size of the Gap dataset was chosen arbitrarily. We reserved the estimation of this parameter via cross-validation for further works on the subject.



(a) Classical Methodology **(b)** Recursive Forecast Methodology

Figure 3.2: Example plot of the WTI February-April true data against the forecast obtained using the VMD-FFNN hybrid model generated using the testing methodology from the literature, on the left, and using the Recursive Forecasting Methodology, on the right. The solid line represents the true value whilst the dashed line represents the predicted value. Values are in USD.

3.2.6 Performance evaluation

To evaluate the accuracy of the predictions, we compute the standard Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE). They are calculated as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}, \tag{3.14}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|. \tag{3.15}$$

Smaller values of MAPE and RMSE indicate a more accurate prediction model.

3.2.7 The Diebold-Mariano test

The Diebold-Mariano (DM; Diebold & Mariano, 1995) test assesses the existence of a statistical difference between two forecasts. We define e_t and r_t as the residuals of the forecasts f and g respectively, $\overline{d}$ as the average difference of the squared residuals, and γ_k as the autocovariance at lag k, that is,

$$e_t = y_t - f_t, \qquad r_t = y_t - g_t, \tag{3.16}$$

Algorithm : Recursive Forecast Methodology (RFM) for testing

Input: a time series of prices

Output: a time series of predicted prices and its RMSE and MAPE

1: Import the price trajectory Y_t
2: Split Y_t into Y_t^{TRAIN} $(t=0,...,j)$, Y_t^{GAP} $(t=j+1,...,J)$, and Y_t^{TEST} $(t=J+1,...,T)$
3: $N \leftarrow$ Number of modes obtained from applying EMD to Y^{TRAIN}
4: Initialise $\hat{Y}^{nn}$,$\hat{Y}^{gam}$
5: **for** k in the range $[N-4,\ N+4]$ **do**
6: $M_{t,i}^{TRAIN} \leftarrow k$ modes resulting from applying VMD to Y_t^{TRAIN}
7: $nn \leftarrow$ FFNN training, Target Y_{t+1}^{TRAIN}, Covariates $M_{t,i}^{TRAIN}$
8: $gam \leftarrow$ GAM training, Target Y_{t+1}^{TRAIN}, Covariates $M_{t,i}^{TRAIN}$
9: **for** τ in the range $[J+1:T-1]$ **do**
10: $M_{t,i}^{GAP} \leftarrow k$ modes resulting from applying VMD to Y_t^{GAP}
11: $\hat{Y}_\tau^{nn} \leftarrow$ Forecast via nn, inputs $M_{\tau-1,i}^{GAP}$, the value at time $\tau-1$ of the k modes (see line 10)
12: $\hat{Y}_\tau^{gam} \leftarrow$ Forecast via gam, inputs $M_{\tau-1,i}^{GAP}$, the value at time $\tau-1$ of the k modes (see line 10)
13: $\hat{Y}^{nn} \leftarrow \hat{Y}_\tau^{nn}$
14: $\hat{Y}^{gam} \leftarrow \hat{Y}_\tau^{gam}$
15: Append Y_τ^{TEST} to Y_t^{GAP}.
16: **end for**
17: Compute $RMSE^{nn}$ $(Y_t^{TEST},\hat{Y}^{nn})$ and $MAPE^{nn}$ $(Y_t^{TEST},\hat{Y}^{nn})$
18: Compute $RMSE^{gam}$ $(Y_t^{TEST},\hat{Y}^{gam})$ and $MAPE^{gam}$ $(Y_t^{TEST},\hat{Y}^{gam})$
19: Compute the Diebold Mariano test between $\hat{Y}^{nn}$ and $\hat{Y}^{gam}$
20: **end for**
21: **return** performance metrics and optimal value of k

$$d_t = (e_t)^2 - (r_t)^2, \qquad \overline{d} = \frac{1}{T}\sum_{t=1}^{T} d_t, \tag{3.17}$$

$$\gamma_k = \frac{1}{T}\sum_{t=k+1}^{T} (d_t - \overline{d})(d_{t-k} - \overline{d}). \tag{3.18}$$

The Diebold-Mariano test can be defined as

$$DM = \frac{\overline{d}}{\sqrt{[\gamma_0 + 2\sum_{t=1}^{T}\gamma_k]/T}}. \tag{3.19}$$

Under the Null hypothesis H_0

$$H_0 : E(\overline{d}) = 0, \tag{3.20}$$

that is, the expected accuracy of forecasts f and g are equal, the DM test statistics follows a standard normal distribution $DM \sim \mathcal{N}(0,1)$.

3.2.8 Exponential Smoothing

Employing Exponential Smoothing (Holt, 2004), a forecast is computed as the weighted average of past observations, where the weights decrease exponentially over time. Thus, the older the observations, the lower the associated weight. It is defined as

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t. \tag{3.21}$$

where $0 < \alpha < 1$ is the smoothing parameter. If we expand Equation (3.21) backwards in time we obtain

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + ... + (1-\alpha)^n \hat{y}_{t-n-1}. \tag{3.22}$$

Hence, the influence of past observations on the one-period-ahead forecast decreases exponentially with time. Given the length of our datasets and the exponential decaying of the value attached to older observations, we set $\hat{y}_{t-n-1} = y_{t-n}$. Experimentally, we set $\alpha = 0.7$. Given its structure, Exponential Smoothing allows to forecast non-stationary time series without differencing the data, thus, making the resulting performance metrics comparable with those produced by the proposed frameworks.

3.3 Data and results

3.3.1 Forecasting the 2020 datasets

The data are composed of five-minute WTI and Brent future oil prices in 3 different time periods: *i*) January 1st 2020 to March 31st 2020; *ii*) February 1st 2020 to April 30th 2020; *iii*) April 1st 2020 to June 30th 2020. All data was obtained from Bloomberg. Table 3.1 summarises the descriptive statistics and number of observations of each dataset.

Dataset	Number of Observations	Min	25%	Median	Mean	75%	Max
WTI January-March	17100	19.45	33.58	51.06	46.46	55.48	65.48
WTI February-April	16943	-38.96	21.34	28.86	33.15	50.01	54.50
WTI April-June	17441	-38.96	21.88	29.57	28.54	37.27	41.57
Brent January-March	16658	21.75	36.77	55.67	51.22	62.18	70.89
Brent February-April	15802	16.01	27.33	34.01	39.15	54.43	59.85
Brent April-June	16536	16.04	29.05	33.83	33.31	39.70	43.90

Table 3.1: Data summary statistics.

Figure 3.4 in Section 3.5 displays a plot of the six datasets. The data included between the two dotted vertical lines is reserved to the Gap dataset, which separates the training dataset on its left from the testing dataset on its right. As discussed in Section 3.2.5, 80% of the data was included in the training dataset, the Gap dataset comprised 1% of the observations and the testing dataset included the remaining 19% of the data. The number of modes in which the datasets are decomposed was chosen following the procedure described in Section 3.2.2, namely, we applied EMD to the training data. For each dataset the we obtained $N = 10$ IMFs via EMD. We then varied k, the number of modes to be extracted via VMD, in the range $[N-4,N+4]$. For each value of k, the modes obtained were then used as inputs to train the FFNN and the GAM models, as explained in Section 3.2. Figures 3.5, 3.6, 3.7, 3.8, 3.9, 3.10 in Section 3.5 display the training data decomposition performed by VMD for the best performing value of k. Higher order IMFs represent the low frequency components of the decomposed signals, whilst the k-th order IMF depicts the residual trend component. This displays the efficacy of the decomposition, as described in Dragomiretskiy & Zosso (2014). We then followed the recursive methodology presented in Section 3.2.5 to compute the forecasts via GAM and FFNN for all values

Dataset	Statistic	p-value	Dataset	Statistic	p-value
WTI January-March	-8.466	0.000***	Brent January-March	8.650	0.000***
WTI February-April	-3.922	0.000***	Brent February-April	-5.387	0.000***
WTI April-June	-19.589	0.000***	Brent April-June	-16.936	0.000***

Table 3.2: Diebold-Mariano test results. The forecast residuals of the two methodologies are compared for each dataset. *** significance at the 1% confidence level, ** significance at the 5% confidence level, * significance at the 10% confidence level.

of k for all datasets. The Diebold-Mariano test confirms the statistical difference of the two forecasts for all datasets. A summary of the DM test results is presented in Table 3.2.

A collection of the performance metrics for each dataset and framework is presented in Tables 3.6 and 3.7 in Section 3.5. Although the difference is very small (up to the second decimal place), the VMD-FFNN approach outperforms the VMD-GAM hybrid in all cases with the sole exception of the January-March Brent dataset. The results show that it is possible to forecast Brent prices with an average error margin as little as \$0.586 (VMD-GAM: $MAPE = 0.012,\ RMSE = 0.586$) and \$0.560 (VMD-FFNN: $MAPE = 0.012, RMSE = 0.560$) in the period April-June 2020. The forecast accuracy is slightly lower in the periods January-March 2020 (VMD-GAM: $MAPE = 0.025, RMSE = 0.866$; VMD-FFNN: $MAPE = 0.027, RMSE = 0.923$) and February-April 2020 (VMD-GAM: $MAPE = 0.032, RMSE = 1.025$; VMD-FFNN: $MAPE = 0.031, RMSE = 0.987$), where the average error margin is close to \$1 . It must be noted that the Brent January-March and the Brent February-April datasets include the steep price drop observed between March 6th and March 9th in the training data, which might have led to a lower forecast accuracy.

With regards to WTI, the average forecast error for the April-June period is \$0.823 (VMD-GAM: $MAPE = 0.02,\ RMSE = 0.823$) and \$0.738 (VMD-FFNN: $MAPE = 0.017,\ RMSE = 0.738$), whilst for the January-March dataset is \$0.836 (VMD-GAM: $MAPE = 0.03,\ RMSE = 0.836$) and \$0.816 (VMD-FFNN: $MAPE = 0.03, RMSE = 0.816$). These results are in line with the Brent datasets forecasts. Furthermore, the information carried by the wider range of data values included in the training set in the period April-June led to a slightly higher accuracy of the predictions. However, the MAPE and RMSE increases substantially when the negative prices appear in the test dataset (VMD-GAM: $MAPE = 2.29,\ RMSE = 11.526$; VMD-FFNN:

$MAPE = 2.279$, $RMSE = 11.482$). Both frameworks display an inability to capture the structural instability given by prices dropping to negative levels. Interestingly, if we compute the performance metrics over the period April 13th to April 19th, the average prediction error is in line with the results produced for the Brent datasets (VMD-GAM: $MAPE = 0.027$, $RMSE = 0.585$; VMD-FFNN: $MAPE = 0.026$, $RMSE = 0.568$). However, the inaccuracy significantly increase over the period April 20th and 21st 2020 (VMD-GAM: $MAPE = 13.029$, $RMSE = 20.602$; VMD-FFNN: $MAPE = 12.957$, $RMSE = 20.502$). Furthermore, given that the methodologies continue to predict negative values throughout April 22nd, when WTI was already trading above \$0, the forecast accuracy in the subset April 22nd to April 30th 2020 is still low (VMD-GAM: $MAPE = 0.784$, $RMSE = 11.341$; VMD-FFNN: $MAPE = 0.782$, $RMSE = 11.308$).

Thus, we can recognise three separate situations in the case of WTI: *i*) positive-positive, when training and testing is done exclusively on positive prices resulting in a similar accuracy to the forecast of Brent during the same period; *ii*) negative-positive, when the methodologies are trained with prices of both sign and testing is done using positive values and the accuracy is slightly higher compared to *i*; *iii*) positive-negative, when the testing set includes values of both signs but training set comprises strictly positive prices and the approaches loose their viability. Conversely, Brent displays the highest accuracy in the prediction of the April-June. As such, the results show that the choice of dataset adopted in the training of the two methodologies affects the corresponding forecast performance. As shown, the prediction of WTI is greatly affected by the presence of negative prices observed in April 2020. However, the structural dissimilarities between the American and the global crude oil markets also influence the results. Namely, whilst WTI is mined and delivered on American soil, Brent is extracted in the North sea and transported via underwater pipes and cargo ships leading to great storage flexibility. As discussed, the OPEC+ production cut delay in late 2019 impacted both benchmarks; however, the price drop was more marked for WTI due to the large quantities of shale output by American miners without demand to match it (due to COVID-19 restrictions).

Our results outperform similar works present in the literature: Lahmiri (2015) and Lahmiri (2016) achieve $RMSE = 3.44$ and $RMSE = 2.10$ when forecasting Brent

and WTI daily prices, respectively, via a VMD-GRNN hybrid model, suggesting the advantage of employing VMD on higher frequency data.

For the practitioner interested in applying the two approaches in the same time periods, Table 3.3 reports the maximum differences between the forecast and the actual observed price.

Dataset	VMD-FFNN	VMD-GAM	Dataset	VMD-FFNN	VMD-GAM
WTI January-March	1.776	2.006	Brent January-March	1.673	3.191
WTI February-April	37.910	41.738	Brent February-April	7.217	7.047
WTI April-June	0.790	2.060	Brent April-June	1.589	1.615

Table 3.3: Highest possible prediction error for the two methodologies applied to each dataset. Values are in USD.

3.3.2 Forecasting 2023-2024 data

We extended the analysis to WTI prices ranging between May 1st 2023 and April 30th 2024, a time period where no structural instabilities were detected. The introduction of a larger dataset allows to have a further comparison with the findings presented in Section 3.3. Moreover, whilst being more computational expensive, training the models using an greater number of observations should lead to a lower sensitivity to noise in the data, better capture the underlying its distribution (including outliers) and improving generalisation by reducing overfitting. The data was obtained via Thomson Reuters' Eikon Python API and is displayed in Figure 3.3. The data included between the two dotted vertical lines is reserved to the Gap dataset, which separates the training dataset on its left from the testing dataset on its right. After pre-processing, the dataset consists of 71726 observations. The summary statistics are reported in Table 3.4.

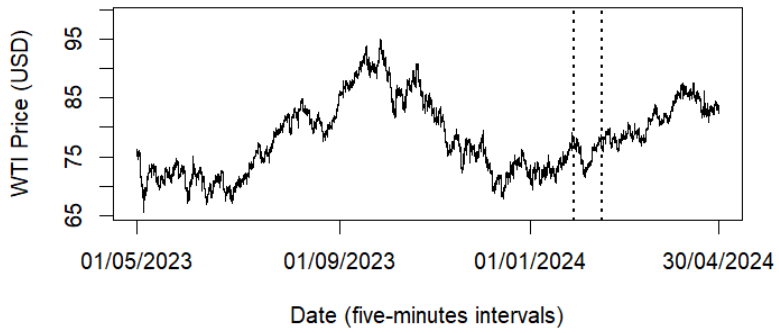


Figure 3.3: WTI May 2023 - April 2024 five-minute prices (in USD). The data included between the two dotted vertical lines is reserved to the Gap dataset, which separates the training dataset on its left from the testing dataset on its right.

Dataset	Number of Observations	Min	25%	Median	Mean	75%	Max
WTI May 2023 - April 2024	71726	65.53	72.88	77.72	78.24	82.77	94.90

Table 3.4: WTI May 2023 - April2024 Data summary statistics.

We follow the same methodology presented in Section 3.2. Employing EMD to the dataset leads to the extraction of $N = 10$ IMFs. Thus, VMD is applied to the data to extract $k = [6,14]$ modes. The VMD-GAM and VMD-FFNN frameworks are trained for

all values of k and the optimal model is chosen via cross-validation of the performance metrics, which are reported in Table 3.8 in Section 3.5. Figure 3.11 in Section 3.5 display the training data decomposition performed by VMD for the best performing value of k. For all values of k, the lowest MAPE and RMSE are reported in bold. The DM test, reported in Table 3.5, confirms the statistical difference of the two forecasts for all datasets.

The VMD-GAM framework performs slightly better than the VMD-FFNN model, returning $MAPE = 0.007$ and $RMSE = 0.659$ for $k = 9$, and $MAPE = 0.007$ and $RMSE = 0.683$ for $k = 8$, respectively.

In comparison with the results presented in Section 3.3.1, the May 2023 - April 2024 datasets forecasts are, on average, \$0.07 more accurate than the best performing WTI dataset (April - June 2020). The absence of structural instabilities observed in 2020, the longer period analysed and the 'completeness' of the dataset make it preferable to apply the frameworks to this dataset.

	Diebold-Mariano test		Highest prediction error	
Dataset	Statistic	P-value	VMD-FFNN	VMD-GAM
WTI January-March 2024	17.475	0.000***	2.579	2.653

Table 3.5: Diebold-Mariano test results and highest possible prediction error of the WTI May 2023 - April 2024 dataset. The forecast residuals of the two methodologies are compared for each dataset. *** significance at the 1% confidence level, ** significance at the 5% confidence level, * significance at the 10% confidence level. Highest possible prediction error values are in USD.

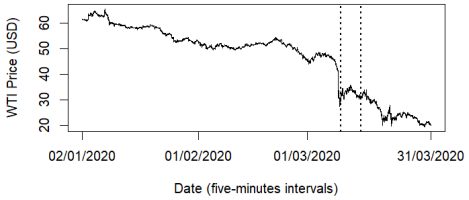
3.3.3 Comparison with Exponential Smoothing

In Table 3.9 in Section 3.5 we present the RMSE and MAPE generated by the two hybrid models, for the best performing value of k, and by ES. The results show that the proposed methodologies tend to approach the performance of ES for all datasets. However, the presence of negative prices in the WTI February - April 2020 dataset deteriorates tenfold the accuracy of all methodologies, displaying the limitations of the frameworks to capture the structural instability generated by the unforeseen drop of WTI prices to negative levels.

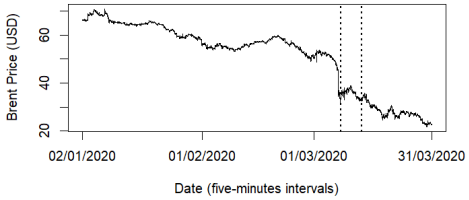
3.4 Conclusions

In this paper we applied Variational Mode Decomposition coupled with Generalised Additive Models and Feed-Forward Neural Network to predict WTI and Brent 5-minute crude oil prices in different time periods. The choices of datasets revolve around the unprecedented event of WTI prices reaching negative levels in April 2020 during the COVID-19 pandemic. We proposed a Recursive Forecasting Methodology to ensure that the forecasts are computed using modes generated exclusively from past data. The results showed that the two techniques are able to forecast WTI oil prices in the positive-positive and negative-positive scenario. However, both approaches fail in the positive-negative case, namely if the frameworks are trained on positive values and tested on prices of mixed sign. Further, the prediction accuracy of Brent oil prices is the highest in the period April - June 2020 and the lowest in the period February-April, displaying a spillover effect caused by the structural instability observed for WTI prices during April 2020. Further, training the models on a larger amount of data which does not comprise of the structural instabilities observed in 2020 produces more accurate results in the period May 2023 - April 2024. The two proposed hybrid frameworks tend to approach the performance of ES for all datasets. However, the presence of negative prices in the WTI February - April 2020 testing dataset deteriorates tenfold the accuracy of all methodologies, including Exponential Smoothing, displaying the limitations of the frameworks to capture the structural instability generated by the unforeseen drop of WTI prices to negative levels.

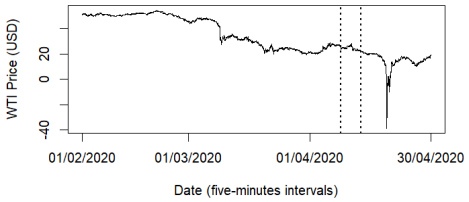
3.5 Figures and Tables



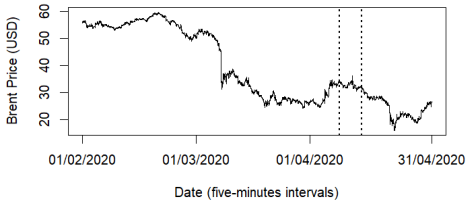
(a) WTI prices observed between 01/01/2020 and 31/03/2020



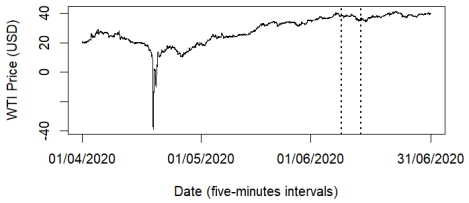
(b) Brent prices observed between 01/01/2020 and 31/03/2020



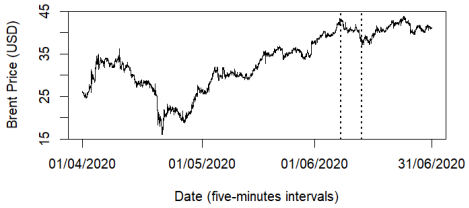
(c) WTI prices observed between 01/02/2020 and 30/04/2020



(d) Brent prices observed between 01/02/2020 and 30/04/2020



(e) WTI prices observed between 01/04/2020 and 30/06/2020



(f) Brent prices observed between 01/04/2020 and 30/06/2020

Figure 3.4: WTI and Brent five-minute prices in USD. The data included between the two dotted vertical lines is reserved to the Gap dataset, which separates the training dataset on its left from the testing dataset on its right.

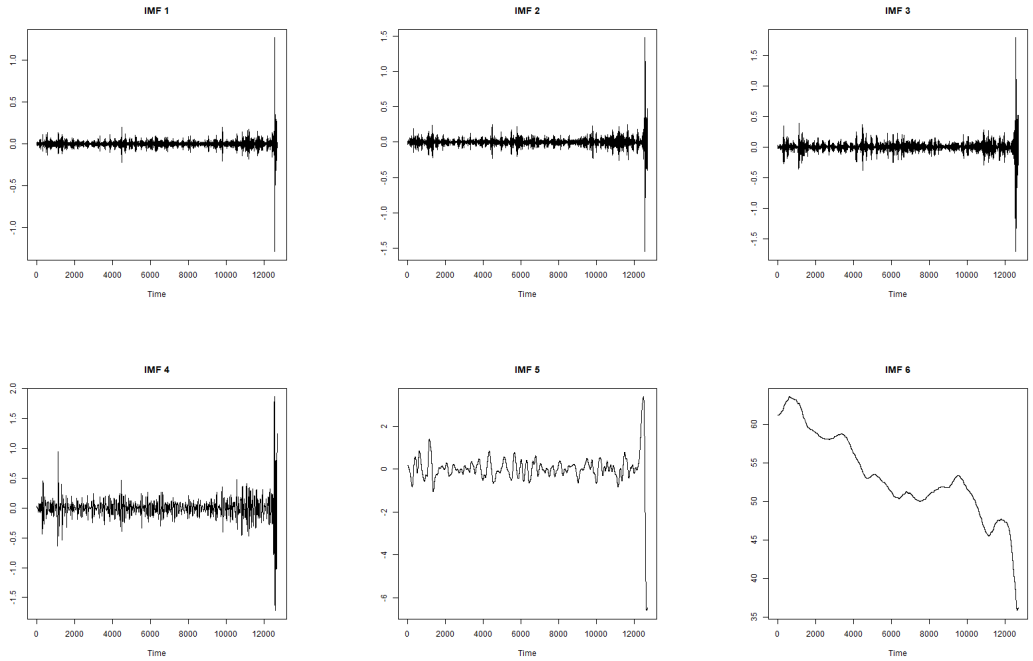


Figure 3.5: VMD decomposition of the WTI January-March 2020 Dataset

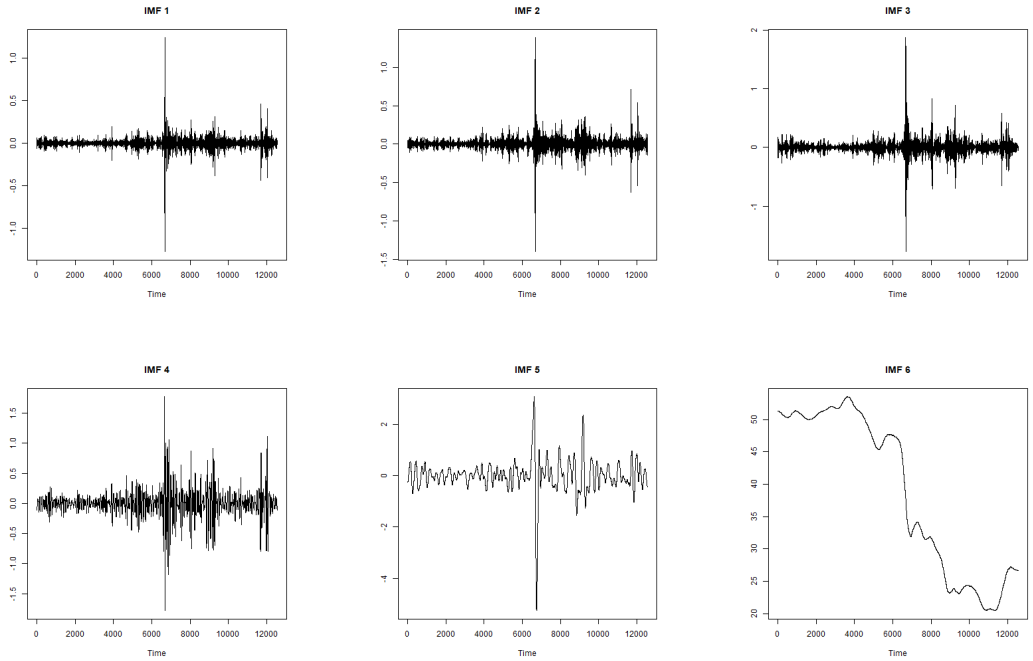


Figure 3.6: VMD decomposition of the WTI February-April 2020 Dataset

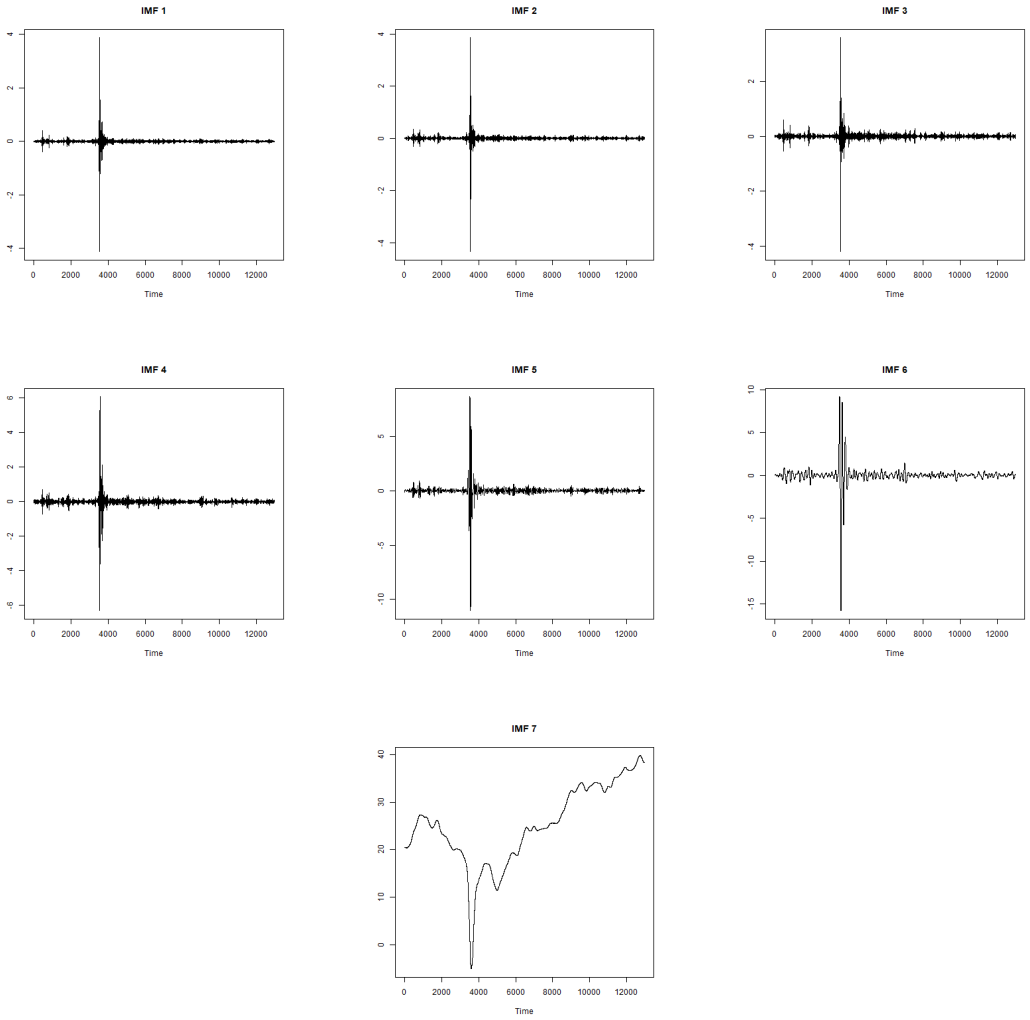


Figure 3.7: VMD decomposition of the WTI April-June 2020 Dataset

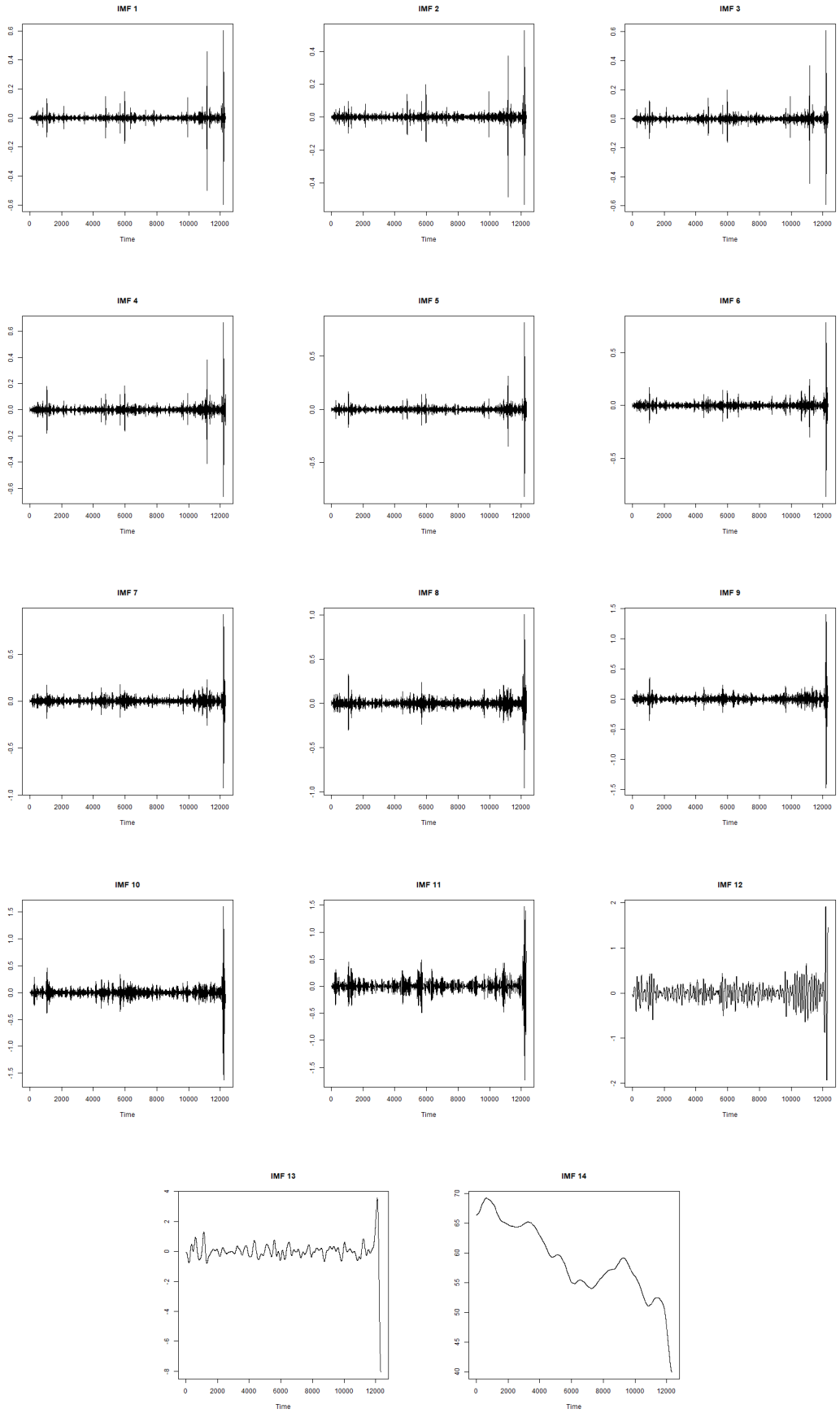


Figure 3.8: VMD decomposition of the Brent January-March 2020 Dataset

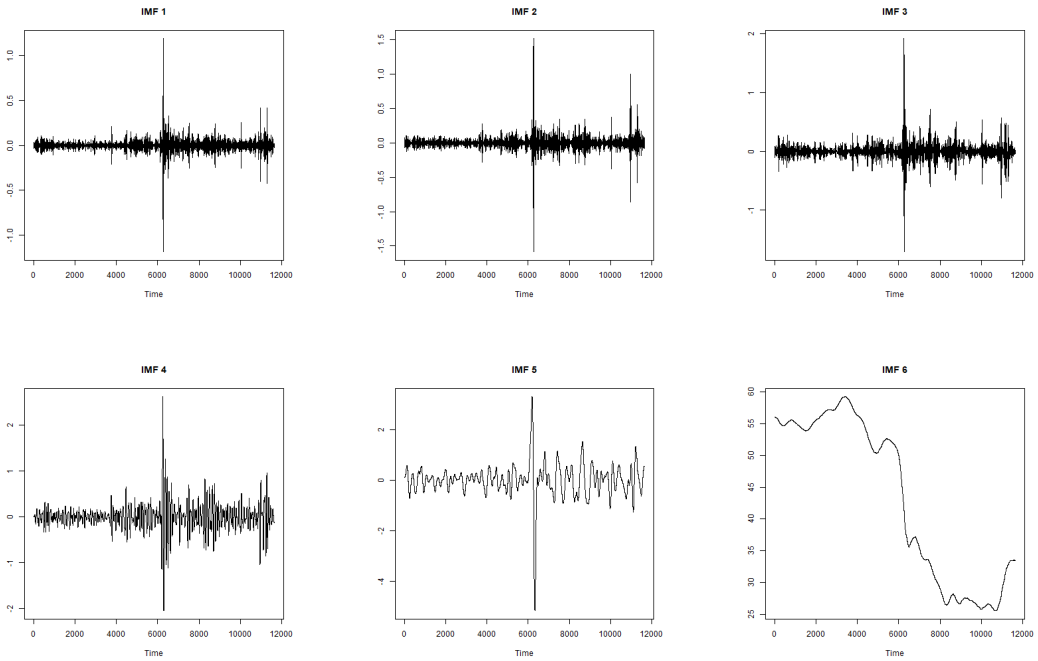


Figure 3.9: VMD decomposition of the Brent February-April 2020 Dataset

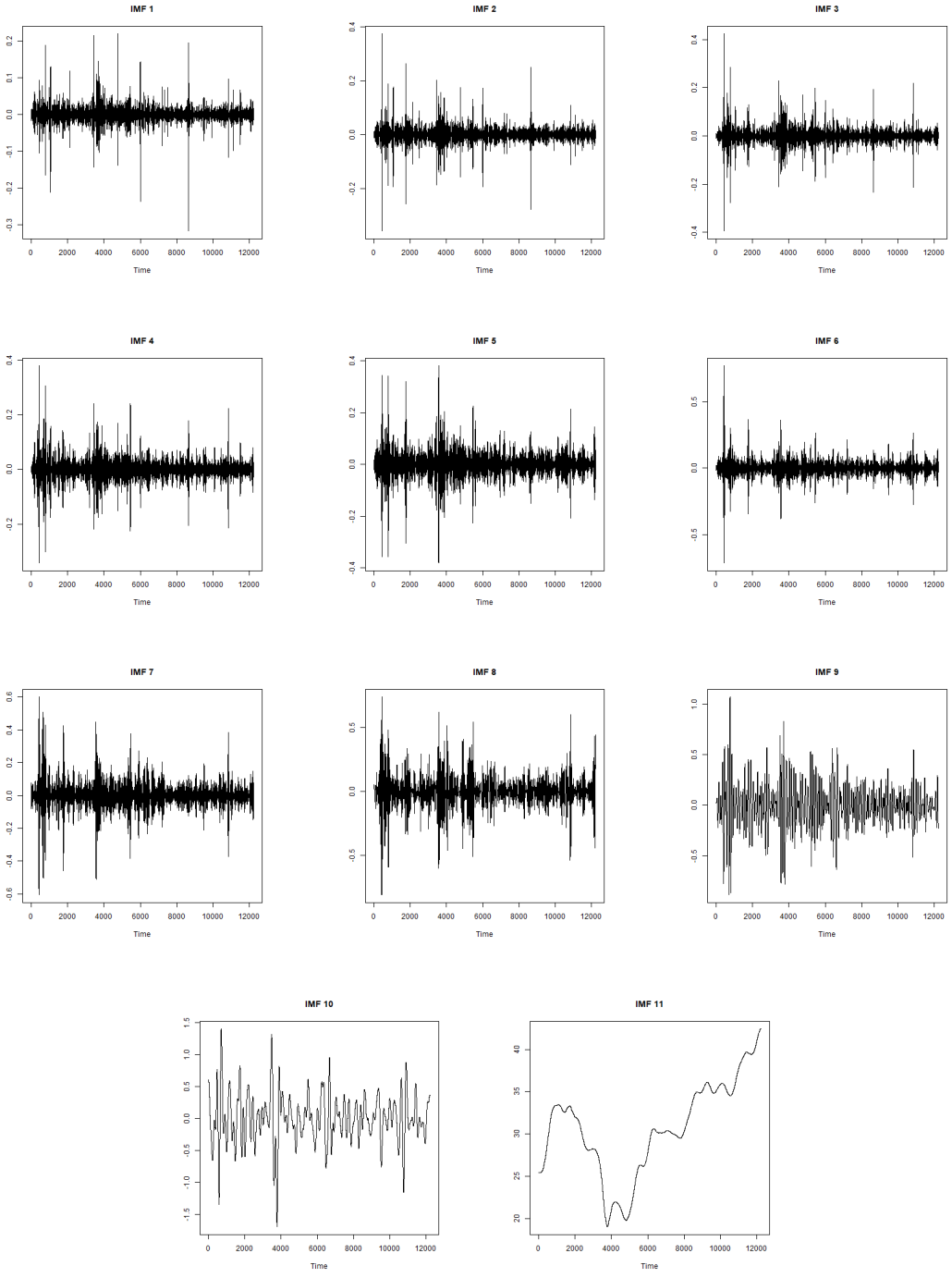


Figure 3.10: VMD decomposition of the Brent April-June 2020 Dataset

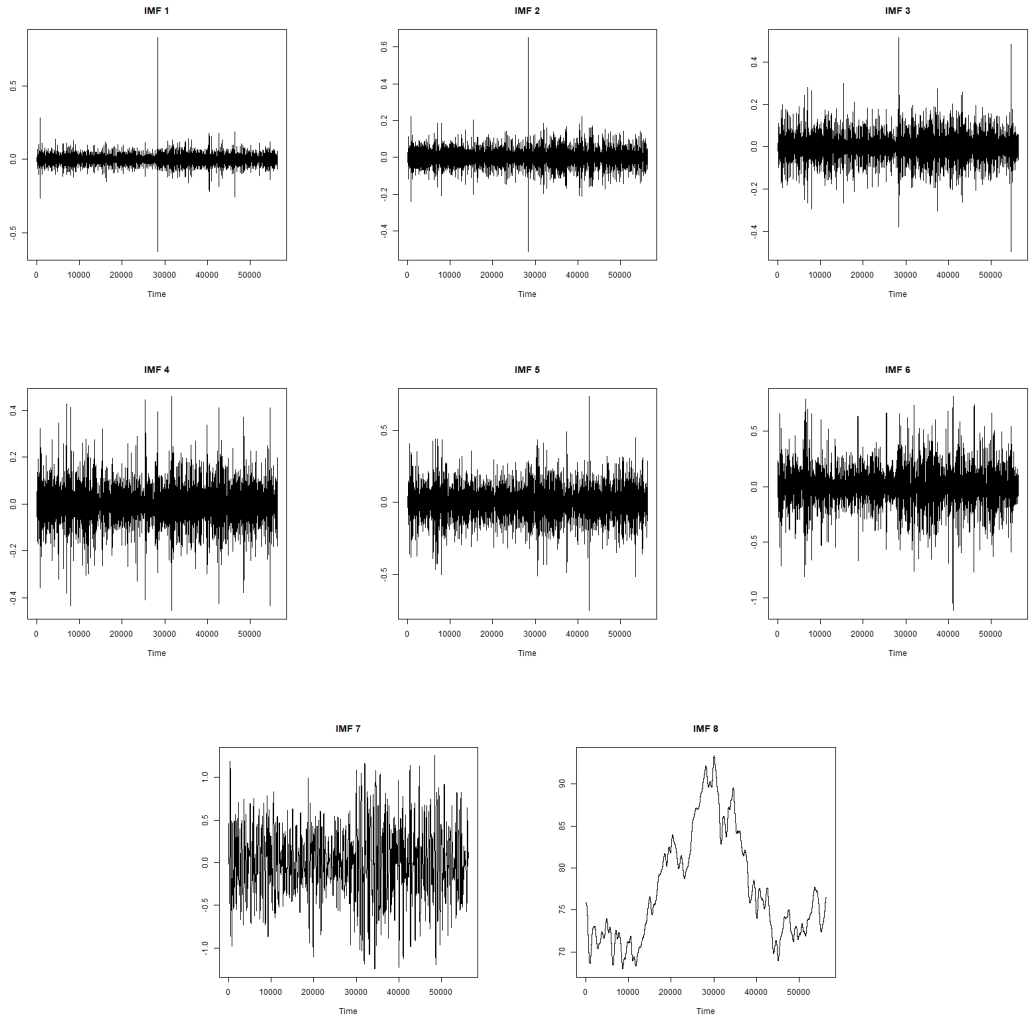


Figure 3.11: VMD decomposition of the WTI May 2023 - April 2024 Dataset

Dataset	WTI January-March 2020				WTI February-April 2020				WTI April-June 2020			
Model	VMD-GAM		VMD-FFNN		VMD-GAM		VMD-FFNN		VMD-GAM		VMD-FFNN	
k	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
6	**0.030**	**0.836**	**0.030**	**0.816**	**2.290**	**11.526**	**2.279**	**11.482**	0.020	0.836	0.020	0.835
7	0.033	0.890	0.033	0.893	2.398	11.640	2.380	11.548	**0.020**	**0.823**	**0.017**	**0.738**
8	0.033	0.891	0.035	0.952	2.399	11.745	2.431	12.044	0.022	0.921	0.019	0.820
9	0.033	0.895	0.035	0.970	2.431	12.206	2.380	12.013	0.022	0.898	0.020	0.830
10	0.032	0.884	0.034	0.970	2.441	12.276	2.430	12.253	0.022	0.909	0.022	0.983
11	0.032	0.892	0.034	0.962	2.455	12.409	2.459	12.448	0.021	0.883	0.020	0.847
12	0.033	0.892	0.038	1.128	2.460	12.388	2.431	12.156	0.021	0.882	0.024	1.045
13	0.033	0.891	0.034	0.963	2.434	12.373	2.370	12.176	0.021	0.886	0.019	0.817
14	0.034	0.925	0.038	1.064	2.440	12.478	2.365	12.022	0.022	0.893	0.019	0.812

Table 3.6: Performance metrics for the three WTI datasets. The lowest values are reported in bold.

Dataset	BRENT January-March 2020				BRENT February-April 2020				BRENT April-June 2020			
Model	VMD-GAM		VMD-FFNN		VMD-GAM		VMD-FFNN		VMD-GAM		VMD-FFNN	
k	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
6	0.051	1.451	0.053	1.517	**0.032**	**1.025**	**0.031**	**0.987**	0.012	0.601	0.012	0.590
7	0.055	1.584	0.058	1.697	0.043	1.299	0.039	1.162	0.013	0.617	0.012	0.602
8	0.045	1.384	0.043	1.346	0.045	1.342	0.043	1.261	0.013	0.608	0.012	0.595
9	0.036	1.147	0.036	1.174	0.053	1.498	0.050	1.445	0.012	0.600	0.012	0.563
10	0.033	1.088	0.036	1.149	0.058	1.590	0.054	1.447	0.012	0.590	0.012	0.567
11	0.031	1.028	0.031	1.053	0.059	1.591	0.057	1.620	0.012	0.590	**0.012**	**0.560**
12	0.029	0.995	0.034	1.165	0.059	1.600	0.054	1.543	0.012	0.587	0.012	0.581
13	0.026	0.875	0.032	1.117	0.061	1.635	0.061	1.743	0.012	0.587	0.012	0.604
14	**0.025**	**0.866**	**0.027**	**0.923**	0.062	1.652	0.055	1.464	**0.012**	**0.586**	0.012	0.585

Table 3.7: Performance metrics for the three Brent datasets. The lowest values are reported in bold.

Dataset	WTI May 2023 - April 2024			
Model	VMD-GAM		VMD-FFNN	
k	MAPE	RMSE	MAPE	RMSE
6	0.007	0.694	0.007	0.696
7	0.007	0.669	0.007	0.693
8	0.007	0.673	**0.007**	**0.683**
9	**0.007**	**0.659**	0.007	0.729
10	0.007	0.662	0.007	0.730
11	0.008	0.683	0.007	0.721
12	0.008	0.712	0.008	0.812
13	0.008	0.750	0.008	0.770
14	0.008	0.739	0.008	0.756

Table 3.8: Performance metrics of the WTI May 2023 - April 2024 dataset. The lowest values are reported in bold.

Dataset	VMD-GAM			VMD-FFNN			ES	
	k	MAPE	RMSE	k	MAPE	RMSE	MAPE	RMSE
WTI January - March 2020	6	0.030	0.836	6	0.030	0.816	**0.004**	**0.152**
WTI February - April 2020	6	2.290	11.526	6	2.279	11.482	**0.027**	**0.649**
WTI April - June 2020	7	0.020	0.823	7	0.017	0.738	**0.002**	**0.087**
Brent January - March 2020	14	0.025	0.866	14	0.027	0.923	**0.003**	**0.149**
Brent February - April 2020	6	0.032	1.025	6	0.031	0.987	**0.003**	**0.122**
Brent April - June 2020	14	0.012	0.586	11	0.012	0.56	**0.002**	**0.097**
WTI May 2023 - April 2024	9	0.007	0.659	8	0.007	0.683	**0.001**	**0.066**

Table 3.9: MAPE and RMSE produced via VMD-GAM, VMD-FFNN and ES for each dataset. For the VMD hybrid models, we report the number of modes k which generated the best performing model. In bold are the lowest RMSE and MAPE.

Chapter 4

A Sentiment Analysis Approach to Oil Prices and Oil Price Volatility

Abstract

We designed and implemented a pipeline architecture that connects world news, their sentiment and the fluctuations of oil markets; allowing us to quantify the specific connection between world news and those fluctuations. In this order, we studied how re-weighting the sentiment scores of oil-related words co-occurring with war-related nouns can lead to improved predictability of next-day returns, historical and conditional volatility of WTI prices. Experiments are run over a two-years period that covers the Russian invasion of Ukraine. An extensive experimental validation (via Mutual Information, Pearson Correlation and Granger causality) exhibited a connection between sentiment from world news and lagged WTI historical and conditional volatility. Further, we showed how the implemented sentiment measures can outperform standard sentiment analysers when forecasting next-day WTI volatility. Lastly, we analysed which stand-alone noun included in our ad-hoc sentiment measure affects the forecast accuracy the most.

4.1 Introduction

A growing body of literature has established a link between sentiment *signals* extracted from automated Natural Language Processing (NLP) tools and asset returns (and their volatility) in financial markets. Similarly, oil markets are just as frequently discussed in online news articles, newspapers, media and social media. A recent research by Loughran & McDonald (2020), among others, confirms the intuition that oil markets are reactive to oil related news. On the other hand, specific aspects of oil supply and demand and how they affect returns and volatility make it, perhaps, the most challenging market to be analysed with NLP tools. This aspect is also brought out by the recent survey by Bastos-Santos et al. (2023).

In this chapter we wish to verify whether calculating an ad-hoc sentiment scores of crude oil-related news articles can effectively capture market sentiment, generating a signal which can be used to predict WTI returns and volatility in a more accurate way than an off-the-shelf NLP tool. Furthermore, we analysed which stand-alone noun, included in our ad-hoc sentiment measure, affects the forecast accuracy the most. The results show that the terms “peace”, “ceasefire” and “negotiation/s/negotiated” have a higher impact on the forecast accuracy. Contextually, omitting the nouns “war/s” leads to a stronger Granger causality of next-day historical and conditional volatility. Thus, we identify that war is now intrinsic in the current state of affairs and news regarding the ongoing conflicts do not bring novelty anymore. Contrarily, news article focused on “peace“, “negotiations” and “ceasefire” are unanticipated, leading to higher predictability of oil volatility.

The remainder of this chapter is organised as follows: in Section 4.2 we discussed the existing literature on the topic, in Section 4.3 we introduced the main features of the crude oil markets as of 2024, in Section 4.4 we described the data analysed and we presented the framework architecture in Section 4.5. Results are reported in Section 4.6, in Section 4.6.3 we present the findings of the “leave-one-out” estimations , whilst Section 4.7 concludes this chapter.

The present study is currently under consideration by Information Sciences.

4.2 Literature Review

The connection between news and financial markets is intuitive: prices should reflect expectations, and news should be a main driver in defining expectations. Several studies have sought to explore, quantify and to some extent operationalise this intuitive connection. Recent articles found that current Machine Learning techniques are indeed able to mine sentiment from news, headlines and media displaying a robust connection with prices. For example, Wan et al. (2021) look at news related to individual large companies and find a spillover effect between a firm specific sentiment and entities belonging to the same financial network industry sector instead of a financial network constructed from news co-occurrence. Loughran & McDonald (2020) provide an overview of the existing approaches from a financial perspective, focusing on the construction and use of lexicons in finance. They discuss readability as an attribute of a corpus, highlighting the importance of defining what is really being measured through statistics.

The two main Natural Language Processing techniques deployed this type of analysis are Sentiment Analysis and Topic Modelling. In a nutshell, the former aims at finding a signal in texts that summarises expectations, while the latter is about finding sets of words that describe the actual topic of a text. Both rely on several enabling NLP technologies, namely tokenisation and word embedding, which parse a text and convert it into an abstract representation via real-valued vectors which encode the meaning of words ensuring that terms with similar meaning are close to each other in the vector space.

As reported by Bastos-Santos et al. (2023), there are three different approaches to computing sentiment scores: lexicon-driven methodologies, Machine Learning algorithms and hybrid techniques. Lexicon-based methodologies gather keywords and pre-assign sentiment values in a dictionary, which is applied to obtain a sentiment score for the text in analysis. The two most used finance-specific dictionaries are Henry (2008) and the Loughran & McDonald (2011). Conversely, if a subset of the corpus is already labelled, Machine Learning algorithms (such as CNN, LSTM and RNN) can be trained to predict the sentiment contained in articles. Lastly, hybrid approaches use a lexicon-based score as input of a Machine Learning approach and have been shown to improve the prediction Accuracy. Sentiment analysis is now robust and available in off-the-shelf

software modules, e.g., the SpaCy Python module[1], for inclusion in pipelines for the online analysis of news and markets.

A recent work by Deveikyte et al. (2022) applied Sentiment Analysis to streams of news and investigates the connection between the obtained signal and the next-day stock market volatility. To model the reality of human traders being influenced by a multiplicity of news outlets, the authors study articles written by experts and published on Bloomberg as well as posts and headlines from social media (Twitter). They analyse the sentiment of their corpus in terms of properly lagged stock returns and volatility and find evidence of a correlation between news headlines sentiment and stock market returns, displaying the forecasting superiority of the former over the latter. However, this feature does not apply to stock volatility. Further, they show a strong negative correlation between negative sentiment on Twitter and next-day volatility. Deveikyte et al. (2022) account for a more straightforward lexicon adopted by Twitter members, allowing for a more detectable optimisms/pessimism about markets compared to professional articles published on Reuters or similar platforms. This implies a greater difficulty in extracting a signal from a text written on purpose to avoid sentiment language.[2]

In their research, Atkins et al. (2018) extract topics via Latent Dirichlet Allocation and adopt them as inputs for a simple Naïve Bayes Classifier to forecast volatility directional changes of US stocks and stock indices. The authors show how signals obtained by news articles have higher predictability power when the target variable is the asset's volatility rather than the closing price of the asset itself, achieving an average Accuracy of 56% for the former and a poorer 49% for the latter.

4.3 The Case of Oil Markets

It is important to highlight the special characteristics of oil markets, which today exhibit three important features:

1. The impact of geopolitical turmoil at the time of writing with two major armed conflicts shaking the world, one involving Russia, the second world oil producer

[1] Please see `https://spacy.io/`.

[2] See for instance this recent op-ed on the relation between impartiality in conflict reporting and the usage of 'loaded' terms such as *invasion, unprovoked aggression* and *resistance*. `https://www.aljazeera.com/opinions/2022/3/9/is-absolute-impartiality-always-a-necessity-in-journalism`

and the subject of sanctions decided against it by a number of OECD countries; and the war in the Middle East, home to many OPEC (Oil Producing Exporting Countries) countries;

2. The increasing drive towards emissions reduction by oil companies, under the pressure of shareholders in particular;

3. The large output of shale oil by the US which has become the first oil producer and has started being a major exporter, making irrelevant the discussion of 'peak oil' and other features of oil markets discussed in important papers of the years 2000s and 2010s.

We identified the conflicts in Ukraine and in the Middle East as the pivotal geopolitical events that occurred during the period in analysis given the involvement of oil producing countries and the disruption caused to the global demand-supply chain. Remarkably, prices have not risen since the beginning of the Middle East war in October 2023 – showing the limits of the economic analysis and the additional information provided by approaches such as Machine Learning techniques to forecast oil prices.[3]

In a recent piece of work Loughran et al. (2019) investigate the specific topic of how crude oil-specific news influence its prices and trading strategies profitability. Their findings shed an interesting light on how the oil markets, with their reliance on geographic, geopolitical and logistic aspects, could be analysed with specific NLP channels. They effectively illustrate the trade-off between focusing textual analysis on oil-related news articles and the subsequent need for such specific methods. Their evaluation of the sentiment embedded in crude oil-related news is based on the construction of a list of specific keyword/modifier patterns and assigning each of them a score which would be counter-intuitive to the general public but is, in reality, an accurate reflection of how such news would impact oil returns and volatility. Each co-occurrence was searched within a range of [-4,4] from the keyword position. They create a total of four dictionaries of oil related keywords and sentiment modifiers, two for co-occurrences with negative tones and two with positive tones. For instance, sentiment modifiers like *"weak"* related to

[3]At the time of this writing, WTI index (as well as Brent) values have been mean-reverting around the $80 level, with essentially no change compared to the previous period.

"demand" and, on the opposite end of the spectrum, *"strong"* associated to *"supply"* were deemed as *"negative tone."* As a matter of fact, if the market is producing excess capacity of oil, a rise in supply unmatched by a growth in demand leads to a buildup of inventories which must be safely stored in safe ground storage facilities or in cargoes for a fee paid by the producers. In this scenario prices are expected to decrease in order to re-balance supply and demand. If these properties are not taken into account, a standard sentiment analyser would deem, for instance, an *"increase in supply"* as positive. An extreme case was experienced in April 2020 when storage facilities were operating at full capacity and the demand was at historical lows due to COVID-19 restrictions, triggering a collapse of WTI front month Future prices to -$37 for the very first time in history. Conversely, when *"weak"* relates to *"supply"* or *"strong"* relates to *"demand,"* a price increase is expected. Another interesting example is *"weather/cold,"* which could sound like a negative piece of news, but is in fact a forecast of a likely increase in consumption and, hence demand. Further, Loughran et al. (2019) introduced a dictionary of stand-alone nouns containing words with geopolitical meaning like *"attack"* and *"bomb,"* and terms that refer to weather phenomena such as *"storm"* and *"hurricane."*

Thanks to this pipeline, they find evidence of two types of connection between news and oil prices, which they term *overreactions.* The first is short-term overreaction they describe as "Dow Jones, oil- related" news; the second is a kind of general industry's reaction to news about world events.

4.4 Data

4.4.1 Market Data

WTI prices, the main crude oil benchmark for North America, were obtained via Thomson-Reuters' Eikon portal for the period ranging from February 2nd 2022 to December 31st 2023. WTI returns are classically computed as

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}} \tag{4.1}$$

where S_t represents the spot price a given date *t*. We employed the Augmented Dickey-Fuller (ADF) test to check the stationarity of the prices and returns time series. According to the ADF test, if the null hypothesis (H0) fails to be rejected, the time series is said to have a unit root, highlighting its non-stationarity. On the other hand, if the null hypothesis is rejected, the data is deemed as stationary. Results of the ADF test are displayed in Table 4.1. In the case of WTI prices, since the statistic is greater than the critical value, we fail to reject the null hypothesis at the 5% confidence level. Conversely, we find evidence against the null hypothesis in the case of log returns (p-value = 0.000), which suggest the series is stationary.

Target Variable	ADF Statistics	P-value
S_t	-1.404	0.579
R_t	-10.828	**0.000**

Table 4.1: Results for the Augmented Dickey-Fuller Test for presence of a unit root.

Further, we computed two daily volatility measures, the historical volatility and conditional volatility of oil returns. Using a rolling window of 25 data points, the annualised historical volatility was computed for each date t as

$$\sigma = \sqrt{252} \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R_i - \overline{R})^2}. \tag{4.2}$$

To compute the conditional volatility we employed a *GARCH(1,1)* model. GARCH, introduced by Bollerslev (1986), is an extension of the Autoregressive Conditional Het-

eroskedasticity (ARCH) model. GARCH is often used when dealing with financial data like prices returns and interest rates since their variance changes through time and depends on its own past values. The WTI daily returns, annualised historical and conditional volatility are plotted in Figure 4.1 in Section 4.8. Table 4.2 reports the summary statistics of the WTI Returns, annualised WTI historical and conditional volatility time series.

Variable	No. of observations	mean	std	min	25%	50%	75%	max
WTI Returns	480	0.000	0.027	-0.121	-0.017	0.002	0.018	0.084
WTI Historical Volatility	480	0.403	0.128	0.169	0.330	0.385	0.446	0.877
WTI Conditional Volatility	480	0.411	0.095	0.253	0.354	0.393	0.440	0.783

Table 4.2: Summary statistics of the three target variables WTI daily returns, WTI historical and conditional volatility.

4.4.2 Crude-oil-related news articles

Our corpus consisted of a total of 42172 news articles that included the *"crude oil"* keyword in the period in analysis. Similarly to WTI prices, the articles were extracted from Thomson-Reuters' Eikon portal. The dataset appeared to be rather evenly distributed over time with about 75 articles per day, with a lower amount released during weekends and holidays. Raw news articles were pre-processed by removing unnecessary html tags and stop words. We omitted the terms "up" and "down" from the NLTK[4] Python package's stop words list as these are often used in crude oil related news articles to express an increase or decrease of specific quantities or asset values. Each article was divided in sentences which were then tokenised.

[4]Please see `https://www.nltk.org/`.

4.5 Methodology

In this section we discussed the methodologies applied, namely Sentiment Analysis, Correlation, Mutual Information, Granger Causality, Latent Diritchlet allocation, Logistic Regression, Feed-Forward Neural Networks and Scoring Measures.

4.5.1 The Sentiment Analyser

The VADER (Valence-Aware Dictionary for sEntiment Reasoning) sentiment analyser by Hutto & Gilbert (2014), which is now part of the general-purpose NLTK package[5] was the software of choice to obtain a baseline sentiment score. The VADER project starts with a focus on capturing text sentiment, specifically *intensity,* in text from microblogs available on the Web. Sentiment seeding is initiated by manual annotation and consensus scoring of words/lexical features. Then, VADER deploys a combination of qualitative and quantitative methods, with rules that capture grammatical- and syntax-conventions to score input texts. While words are assigned a value between -3 (extremely negative) and +3 (extremely positive), texts are assigned an overall score normalised between -1 and +1. In our work we have adopted VADER's [-1,+1] interval–and the meaning of the values–so as to keep the results comparable to those of the 'plain' VADER analyser, which provides a baseline sentiment scoring for our analysis.

4.5.2 Sentiment Scoring

We implemented Loughran et al. (2019)'s four co-occurrences dictionary methodology and adopted their idea of a stand-alone nouns dictionary. However, for the latter we created three different venues:

1. the original Loughran et al. (2019) version, called NOUNS1. It comprises of 75 stand-alone words;

2. an extension of the above, called NOUNS2, which includes the terms *"war/s," "peace," "ceasefire"* and *"negotiation/s/negotiated,"* which we considered pertinent for the geopolitical events that took place in 2022-2023;

[5]Please see `https://www.nltk.org/`.

3. a smaller dictionary, called NOUNS3, consisting only of the the seven war-related keywords presented above.

Example of co-occurrences and stand-alone nouns with expected negative effect on crude oil prices (weight = -1)				
Keyword	Sentiment Modifier	Keyword	Sentiment Modifier	Stand-alone nouns
Buying	Constraint	Drilling	Exceed	Discoveries
Consumption	Crash	Inventory	Grow	Overproduction
Demand	Poor	Output	High	Oversupply
Economy	Reduce	Production	Soar	Surplus
Import	Restraint	Supply	Strong	Glut
Example of co-occurrences and stand-alone nouns with expected positive effect on crude oil prices (weight = 1)				
Keyword	Sentiment Modifier	Keyword	Sentiment Modifier	Stand-alone nouns
Drilling	Constraint	Buying	Exceed	Attack
Inventory	Crash	Consumption	Grow	Dispute
Output	Poor	Demand	High	Outage
Production	Reduce	Economy	Soar	Storm
Supply	Restraint	Import	Strong	Tension

Table 4.3: Example of keyword/sentiment modifier co-occurrences and nouns with positive and negative tones

An example of keywords, sentiment modifiers and stand alone nouns included in the aforementioned dictionaries are reported in Table 4.3. The co-occurrences included in the negative- and positive-tones dictionaries were assigned a weight of either -1 or +1, respectively. These values were chosen to reflect market sentiment specific to the oil markets, which we expected to be normally misinterpreted by a generic finance sentiment analyser. In a similar manner, we allocated a -1/+1 weight to all dictionary terms based on the anticipated effect of news article containing such words in relation to crude prices and volatility. For instance, term “peace” was assigned a -1 while, symmetrically, expressions like “war” were assigned to +1. All the possible co-occurrences of keywords and sentiment modifiers were searched within each sentence of all articles. Upon inspection, we found only one instance of a modifier at distance four from the keyword, hence reduced the search range to the interval [-3,3]. Similarly, we extracted all the stand-alone nouns for the three dictionary versions introduced above. Lastly, the compound VADER scores were obtained using the NLTK Python module. This procedure led to the creation of five different sentiment measures for each news article:

1. the sum of all keyword/sentiment modifier co-occurrences weights, denoted as MV_t;

2. the sum of all keyword/sentiment modifier co-occurrences and nouns weights based on NOUNS1, denoted as $TV1_t$;

3. the sum of all keyword/sentiment modifier co-occurrences and nouns weights extracted from NOUNS2, denoted as $TV2_t$;

4. the sum of all keyword/sentiment modifier co-occurrences and nouns weights extracted from NOUNS3, denoted as $TV3_t$;

5. the compound sentiment score obtained using VADER, labelled $VADER_t$.

Next, we follow Gabrovsek et al. (2016) and aggregate sentiment scores for every measure at each date t as follows

$$\text{Sent}_t = \frac{\text{Npos}_t - \text{Nneg}_t}{\text{Npos}_t + \text{Nneg}_t + \text{Nneu}_t + 3},$$

where Npos_t, Nneg_t, and Nneu_t correspond to the number of positive, negative and neutral sentiment news articles at date t, respectively. The denominator is adjusted with the Laplace correction for a three-way classifier. As classical in the literature, the sentiment scores obtained for articles released during weekends and holidays were incorporated in the first following trading day. Finally, the four daily sentiment signals were plotted against the delayed (next-day) WTI returns, historical volatility and GARCH conditional volatility in Figures 4.2, 4.3 and 4.4 in Section 4.8.

4.5.3 Evaluation metrics

We assessed our results with three different methods, namely Pearson's correlation, Mutual Information and Granger Causality.

4.5.3.1 Correlation

Pearson's correlation measures the strength of the linear correlation between two datasets. It takes values between -1 and +1, where a value of +1 indicates perfect positive correlation (when a linear equation with a positive coefficient describes the relationship between

the two sets), whilst a value of -1 implies a perfect negative correlation. Lastly, values close to zero indicate a lack of correlation between the two variables.

4.5.3.2 Mutual Information

The Mutual Information measure (MI, or simply I) captures the amount of information about a variable gained by knowing the other variable's behaviour. To compute it we need to determine the distribution of each variable, let's say $Pr(A)$ and $Pr(B)$ and their joint distribution $Pr(A,B)$. Then the MI between A and B is defined as

$$I(A;B) = \sum_{a \in A} \sum_{b \in B} Pr(a,b) \log \frac{Pr(a,b)}{Pr(a) \cdot Pr(b)}. \tag{4.3}$$

The normalised MI denoted NMI[6], lies in [0,1] and is often used to assess the connection between variables.

4.5.3.3 Granger Causality

The Granger's causality test, introduced by Granger (1969), is used to determine the existence of "casual links" between variables, i.e., whether a time series X is useful to forecast another time series Y. Given a generic model

$$Y_t = \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t \tag{4.4}$$

where X_{t-j} and Y_{t-j} represent the two state variables and ε_t is a white noise series. Causality between X_t and Y_t exists if $a_j \neq 0$. Under the null hypothesis (H0), a variable X does not Granger-cause the target variable Y. If evidence against H0 is found (p-value ≤ 0.05) then X is said to 'Granger cause' Y, meaning that knowing the past values of X improve the ability to forecast Y. The p-value measures the degree of significance of this link. In this research, we analyse whether the sentiment extracted from crude oil-related news articles can be used to forecast next WTI returns, historical and conditional volatility.

[6]Normalisation was performed by the respective Scikit-Learn function. `https://scikit-learn.org/stable/modules/generated/sklearn.metrics.normalized_mutual_info_score.html`

4.5.4 Latent Dirichlet allocation

Latent Dirichlet allocation (LDA) is a generative process which, given a corpus, extracts a set of latent topics by assuming how the documents were generated. This is done by defining a joint probability distribution over the existing corpus and the unknown topics. A posterior (conditional) distribution of the topics given the corpus is then computed using this joint probability distribution. Topics are assumed to be established before the data was created and are defined as a distribution across a pre-existing set of words. LDA operates as follows: firstly, a distribution over topics is generated for each document in the corpus; secondly, terms from the vocabulary are selected based on the chosen topic distribution. As a result, a document is depicted as a distribution over topics and each topic is described as a distribution over words. The topics are allocated to all documents, however, their distribution differs, e.g., topic 1 may be found in document 1, not found in document 2 and so on.

In the literature, LDA has been applied to inference topics of financial news which are then used as inputs for statistical and machine learning approaches to forecast the change between days of a specific financial instrument. For instance, Atkins et al. (2018) achieved a 56% Accuracy using topic distributions to forecast U.S stock volatility directional changes via a Naive Bayes classifier. On the other hand, they showed that their model performed worse than 'a random prediction' when the target variable was the stock closing price. Similarly, Deveikyte et al. (2022) applied a comparable pipeline extracting a sentiment signal financial news and tweet coupled with topics modelled via LDA. Their machine learning architecture was used to forecast directional changes in FTSE100 volatility with very good prediction Accuracy. We describe our prediction architecture in the following sections.

4.5.5 Logistic Regression

Logistic regression is a model that allows the posterior probabilities to be modelled via a linear combination of the covariates. Differently from a standard linear regression, it maps the inputs to probabilities via a cost function that allows the outputs to be in the

range [0,1]. This cost function is called Sigmoid (or Logistic) function and is defined as

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{4.5}$$

By providing an easy to assess decision boundary, logistic regression is a methodology often used for binary classification purposes. Classically, the results are matched to either class if the posterior probability is below or above the threshold value 0.5.

4.5.6 Feed-Forward Neural Networks

On the other hand, in the Feed-Forward Neural Networks (FFNN) the inputs flow in a unilateral fashion from the input layer to the output layer via one or more hidden layers. The learning process comes from updating the weights at each node by a fraction (determined by the learning rate) of the gradient of the loss function in order to minimise the difference between the outputs and the real values; this procedure is called gradient descent. Similarly to the logistic regression, when the output node activation function is the Sigmoid, Equation (4.5), the results are in the range [0,1], allowing for binary classification. The structure of the network consisted of the input layer of size equal to the number of input features, 2 hidden layers and 1 output layer of size 1. We set a 0.3 dropout rate to avoid overfitting and achieve robust results. The hyperparameters for the neural network were obtained by cross-validation through the *GridSearchCV()* function included in Scikit-Learn package. Finally, the weights and biases were optimised via the *Adam* optimiser from PyTorch.

4.5.7 Scoring Measures

To assess the forecasting ability of our prediction models, we computed Accuracy, Precision, Recall, $F1$ Score and $F2$ Score. Table 4.4 defines the variable names used for the computation of such measures.

Variable	Meaning	Variable	Meaning
TP	True Positives	FP	False Positives
TN	True Negatives	FN	False Negatives

Table 4.4: Definition of the Variables used in Equations (4.6), (4.7), (4.8) and (4.9)

Accuracy represents the number of correctly predicted labels over the total number of

predictions. It is computed as

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4.6}$$

Precision describes the ratio of correctly predicted positive labels over the total number of predicted positives

$$\text{Precision} = \frac{TP}{TP+FP} \tag{4.7}$$

Recall is the ratio between the correctly predicted positives over the total number of real positives

$$\text{Recall} = \frac{TP}{(TP+FN)} \tag{4.8}$$

Lastly, the $F1$ and $F2$ measures are the harmonic means of Precision and Recall. The $F1$ score provides an equal balance between the two and is a more appropriate measure than Accuracy if there is an unevenness between the amount of data points attributed to each class. On the other hand, the $F2$ measure gives more importance to Recall over Precision. They are derived from the generic F_β formula as

$$F_\beta = \frac{1+\beta^2}{\frac{\beta^2}{Recall} + \frac{1}{Precision}} \tag{4.9}$$

by setting $\beta = 1$ or $\beta = 2$ accordingly. When comparing the forecast results, we focused on a mix of Accuracy and $F2$ measure to highlight the amount of correctly predicted labels and at the same time account for Precision and Recall.

4.6 Results

In Section 4.6.1 we analysed the evaluation metrics applied to each of the computed sentiment measures wrt. next-day WTI returns, historical and conditional volatility. In Section 4.6.2 we discussed the application of LDA and presented the FFNN and Logistic Regression forecast performance metrics results. Lastly, in Section 4.6.3 we studied which noun included in the NOUNS3 dictionary affects the performance metrics the most.

4.6.1 Evaluation metrics analysis

Variable	Meaning
S_{t+1}	WTI price at day *t+1*
R_{t+1}	WTI returns at day *t+1*
MV_t	Total Modifiers Value at day *t*
$TV1_t$	Total Value at day *t* of the sum of the keyword/sentiment modifiers co-occurrences and nouns weights following the NOUNS1 dictionary
$TV2_t$	Total Value at day *t* of the sum of the keyword/sentiment modifiers co-occurrences and nouns weights following the NOUNS2 dictionary
$TV3_t$	Total Value at day *t* of the sum of the keyword/sentiment modifiers co-occurrences and nouns weights following the NOUNS3 dictionary
$VADER_t$	VADER score aggregated at day *t*

Table 4.5: Definitions of the Variable Names used in Table 4.6.

We assessed the importance of adjusting the weights to specific terms and word patterns when computing sentiment scores of crude oil-related news articles via the Normalised Mutual Information (NMI) test, the Pearson correlation and the Granger causality of the five sentiment scoring systems against lagged WTI oil returns R_{t+1}, historical and conditional volatility of returns, $HistVol_{t+1}$ and $CondVol_{t+1}$ respectively. Results are summarised in Table 4.6 where, as detailed in Table 4.5 in Section 4.5.2, MV_t is the Modifiers Value aggregated at day *t* calculated by searching for co-occurrences based on four tone modifier dictionaries, $TV1_t$, $TV2_t$ and $TV3_t$ represent the scores obtained by adding to MV_t the weights of the stand-alone nouns found in the news articles based on the three variants of the Nouns dictionaries, namely NOUNS1, NOUNS2 and

Sentiment	Variable	Corr Score	p-value	NMI Score	G.C. F-test	p-value
MV_t	R_{t+1}	0.050	0.272	0.954	1.022	0.313
MV_t	$HistVol_{t+1}$	-0.273	**0.000**	0.954	4.684	**0.031**
MV_t	$CondVol_{t+1}$	-0.286	**0.000**	0.954	7.120	**0.008**
$TV1_t$	R_{t+1}	0.012	0.800	0.967	0.020	0.887
$TV1_t$	$HistVol_{t+1}$	-0.235	**0.000**	0.968	0.456	0.500
$TV1_t$	$CondVol_{t+1}$	-0.206	**0.000**	0.968	1.751	0.186
$TV2_t$	R_{t+1}	0.020	0.670	0.968	0.097	0.756
$TV2_t$	$HistVol_{t+1}$	-0.214	**0.000**	0.968	0.342	0.559
$TV2_t$	$CondVol_{t+1}$	-0.185	**0.000**	0.968	0.999	0.318
$TV3_t$	R_{t+1}	0.051	0.263	0.962	1.039	0.309
$TV3_t$	$HistVol_{t+1}$	-0.236	**0.000**	0.962	3.672	**0.056**
$TV3_t$	$CondVol_{t+1}$	-0.240	**0.000**	0.962	4.651	**0.032**
$VADER_t$	R_{t+1}	0.023	0.614	0.977	0.160	0.689
$VADER_t$	$HistVol_{t+1}$	-0.023	0.621	0.977	0.447	0.504
$VADER_t$	$CondVol_{t+1}$	-0.065	0.155	0.977	7.042	**0.008**

Table 4.6: Evaluation metrics of the relationship between the five daily sentiment scores against next-day WTI returns, WTI historical volatility and WTI conditional volatility for the period 02-Feb-2022 until 31-Dec-2023. In bold are the p-values that show significance at the 10% confidence level.

NOUNS3. Lastly, $VADER_t$ is the VADER score calculated on the whole corpus and aggregated at each date *t*.

The NMI scores are reasonably high for all methods. On one hand, the correlation between the WTI returns and all five sentiment measures is not significant. On the other hand, the results display evidence of negative correlation with historical and conditional volatility, with values ranging between -0.286 and -0.185, in the case of the MV_t, $TV1_t$, $TV2_t$ and $TV3_t$ sentiment measures. This implies that an increase in daily aggregate sentiment score translates into a mild reduction in the next-day oil volatility. Based on the way we computed our four sentiment measures, an increase in either of them will be due to an increase (or decrease) of co-occurrences or nous with positive (negative) weight. For instance, a higher frequency of the co-occurrence "supply/reduction," which was assigned a weight of 1, would lead to the increase of sentiment measures and is expected to reflect

in higher oil prices. Consequently, based on the correlation results, a greater sentiment score would translate into a mild decrease in the historical and conditional volatility.

Note that before the early 2000s, higher commodity prices were most often occurring with high price volatility. This feature then disappeared, with large inventories (possibly built by precautionary demand) reducing the relationship between high prices and high volatility. Furthermore, in the case of crude oil, there is no real concern on supply as the United States has consistently increased his production of shale oil since 2015 (and has become the first world producer as of 2018). It can be argued that oil markets have shown different features since the early 2000s with no more "normal behaviour" being exhibited, and the massive arrival of shale oil produced in the US changing the usual geopolitical picture. Similarly, "peak oil", defined as the moment in time when 50% of world-wide oil reserves would have been depleted, was discussed extensively in the literature in the early 2000s since practitioners and academics have tried to predict the reaching period of this threshold. In the recent years, and especially after the 2015 Paris Agreement, the energy markets changed their centre of attention to the reduction of emissions by scaling down their consumption of fossil fuels, aiming to limit the temperature growth below 1.5°C above pre-industrial levels. Thus, the focus shifted from "peak oil" to "peak demand", or the point in time when the crude oil demand will start reducing. Historically, crude oil displayed high volatility levels for high spot price. This feature is sometimes called "Inverse-Leverage Effect" as the term "leverage effect" describes the property of high volatility when the stock market collapses. Interestingly, the VADER_t measure at all significance levels, suggests a lack of connection between the sentiment it provides and WTI returns and volatility.

The Granger causality results display some interesting features. With p-values below 0.05, there is evidence of causality between MV_t and both next-day historical and conditional volatility; and between TV3_t and VADER_t and conditional volatility. Lastly, TV3_t is shown to Granger cause historical volatility at the 10% confidence level (p-value = 0.056). No sentiment measure displays any predictive power towards next-day WTI returns, coupled with the fact that TV1_t and TV2_t do not show causality with respect to either next-day historical or conditional volatility. As a reminder, TV1_t is computed incorporating the weights of keyword/sentiment modifier co-occurrences (aggregated

into MV_t) and the weights of stand-alone nouns introduced by Loughran et al. (2019). We then extended the stand-alone nouns dictionary with the addition of war-related terms and computed $TV2_t$. Whilst the original stand-alone nouns dictionary is comprised of words related to "attack," "outage" and "discovery," it can be argued that this outcome originates from noise created by a subset of these terms. However, we found that MV_t and $TV3_t$ improve the forecasting power of the models.

4.6.2 Topic Modelling and Forecasting

We then focused on the discovery of hidden topics from our oil-related news articles by implementing the Latent Dirichlet allocation (LDA). We evaluated whether topics augmented with our sentiment measures extracted from crude oil-related news articles can improve the ability to forecast WTI returns, WTI historical and conditional volatility directional changes between date t and $t+1$. We defined a directional change as

$$y_{t+1} = \begin{cases} 0 & \text{if } X_t - X_{t+1} \leq 0 \\ \\ 1 & \text{if } X_t - X_{t+1} > 0 \end{cases} \tag{4.10}$$

where X_t is the value of the target variable at date t. We applied LDA to our corpus comprising 42172 news articles. Since the number of topics K is a hyperparameter specified by the user, we opted to tune it by cross-validation comparing the forecasting Accuracy results. Thus, we set $K = [4,5,6,8]$ and constructed four respective LDA models. Each model provided us with a feature vector of length K_i of topic distributions for all news articles. For all unique dates, we aggregated the topic distributions as follows

$$W_{i,t} = \frac{1}{n} \sum_{i=1}^{n} w_{i,j,t} \tag{4.11}$$

where $W_{i,t}$ represents the aggregated distribution of the i-th topic at date t, $w_{i,j,t}$ is the distribution of the i-th topic on the j-th article at date t; lastly, n corresponds to the number of news articles released on date t. Following the results presented in Section 4.5.3.3, we concatenated every aggregated feature vector $W_{i,t}$ with one of the sentiment measures that displayed Granger causality towards the target variables, namely MV_t, $TV3_t$ and

VADER_t. Thus, for each of the four LDA models, we constructed three vectors to be used as inputs for the predictive models, namely Logistic Regression (LR) and Feed-Forward Neural Network.

For each model we outline the number of topics, the forecasted target variable, Accuracy, Precision, Recall, $F1$ and $F2$ measures. The highest Accuracy scores are outlined in bold. Tables 4.10, 4.11 and 4.12 in Section 4.8 report the forecast metrics of the neural networks models. When predicting next-day WTI conditional volatility directional changes, the MV_t sentiment measure coupled with the 4 topics LDA model is able to achieve an Accuracy of 65% with an $F2$ measure equal to 0, which displays limitations of this specific model prediction. On the other hand the 5 topics LDA model perform similarly with an Accuracy of 64% and an $F2$ measure of 0.365. Comparably, for the TV3_t measure, despite the 4 topics model producing a higher Accuracy, it is preferable to pick the 5 topics LDA model which returns an Accuracy of 63% with an $F2$ measure equal to 0.37. Lastly, the sentiment signal extracted via VADER has higher predictability power when augmented with a 4 topics LDA model to forecast conditional volatility, returning an Accuracy of 64% and an $F2$ measure of 0.179.

Results for the logistic regressions are reported in Tables 4.13, 4.14 and 4.15 in Section 4.8. The MV_t measure is again able to forecast next-day WTI conditional volatility directional changes with highest Accuracy (65%) if the 6 topics model is used. However, the 4 topics model outputs an Accuracy of 59% with an $F2$ measure of 0.479. Similarly, the TV3_t measure prediction Accuracy is the highest when forecasting conditional volatility with the 6 topics LDA model. However, given the very low $F2$ measure, it is preferable to pick the 4 topics model which returns a 59% Accuracy and an $F2$ score of 0.455. Lastly, if VADER is deployed with the 8 topics model it is possible to predict conditional volatility changes with an Accuracy of 57% and $F2 = 0.13$ whilst historical volatility can be forecast with the same Accuracy and an $F2 = 0.337$. These outcomes are in line with the Granger Causality results presented in 4.6, namely that the three measures have higher ability to forecast the next-day directional changes of WTI conditional volatility over the WTI returns' and WTI historical volatility's. The only contradicting result is showed by logistic regression using VADER's sentiment measure

which displays a similar Accuracy when forecasting the two volatilities exhibiting, however, a higher $F2$ measure in the case of historical volatility.

In summary, despite achieving a similar $F2$ score, the neural network and the logistic regression perform best if the MV_t and our TV3_t measures are used over the off-the-shelf VADER sentiment signal. This feature is more visible in the logistic regression where VADER's Accuracy is similar to the other sentiment signals but the $F2$ is over 10 basis point lower. The FFNN returns a higher Accuracy compared to the logistic regression models but lower $F2$. This suggests there is a need of adjusting for crude oil-related co-occurrences and, given the recent adverse geopolitical events, for war-related nouns when computing sentiment analysis of news articles focused on crude oil. When compared to the literature, our results are in line with both Deveikyte et al. (2022) and Atkins et al. (2018).

4.6.3 Leave-One-Out Estimation

As explained in Section 4.5.2, the TV3_t measure was obtained as the sum of the weights of all keywords/sentiment modifier co-occurrences and stand-alone nouns found in each news articles, aggregated per day. As a reminder, the NOUNS3 dictionary included the words "war/s," "ceasefire," "negotiation/s/negotiated," and "peace." In this section we analyse which stand-alone nouns, included in NOUNS3 and used in the computations of the TV3_t measure, affect the forecast Accuracy and F2-measure the most. Namely, we aim to understand whether news regarding war, ceasefire, negotiation or peace would lead to lower-accuracy forecast.

To do so, we computed the sentiment measures TV3_t in a "leave-one-out" experiment (see Bishop (2006) for its applications in Machine Learning) excluding one stand-alone noun (and its plural/past tense forms, if present) at a time. We named the new measures $\text{TV3}^{\text{WAR}}{}_t$, $\text{TV3}^{\text{CEASEFIRE}}{}_t$, $\text{TV3}^{\text{NEGOTIATION}}{}_t$ and $\text{TV3}^{\text{PEACE}}{}_t$, where the superscript indicates which noun was removed. The excluded noun that generates the highest difference in Accuracy and F2-measure is considered to have the greatest importance.

Following the methodology presented in Section 4.5, we computed the Pearson Correlation, NMI and Granger Causality between the four measures and next-day WTI returns, historical and conditional volatility, respectively. Results, displayed in Table 4.7, are broadly in line with the findings presented in Table 4.6 in Section 4.6. The correlation

between next-day returns and the four "leave-one-out" measures is always close to zero, thus, not significant. However, the correlation between the four sentiment measures and the two volatility quantities is always negative with values ranging between -0.21 and -0.23, and growing to -0.31 and -0.33 for $TV3^{WAR}{}_t$. In other words, in a surprising finding, excluding the noun "war" increases correlation with the two volatility measures. As discussed in Section 4.6.1, an increase in daily aggregate sentiment score would translate into a mild reduction in next-day oil volatility. Further, the NMI is high for all measures, with differences only at the third decimal place. Lastly, none of the "leave-one-out" sentiment measures Granger-cause next-day returns, in line with previous findings. We find evidence against the null hypothesis (i.e. no Granger causality) for conditional volatility at the 5% confidence level in all cases. Furthermore, there is evidence of Granger causality between next-day historical volatility and $TV3^{WAR}{}_t$, $TV3^{CEASEFIRE}{}_t$ and $TV3^{PEACE}{}_t$,.

Interestingly, the $TV3^{WAR}{}_t$ measure, which was computed by excluding the nouns "war/s", returns the lowest p-values for both next-day historical and conditional volatility. This finding suggests a stronger predictability power. We will discuss this interesting result in detail later in this section.

Sentiment	Variable	Corr. Score	p-value	NMI Score	G. C. F-test	p-value
$TV3^{WAR}_t$	R_{t+1}	0.059	0.197	0.962	1.416	0.235
$TV3^{WAR}_t$	$HistVol_{t+1}$	-0.314	**0.000**	0.962	6.847	**0.009**
$TV3^{WAR}_t$	$CondVol_{t+1}$	-0.329	**0.000**	0.962	6.930	**0.009**
$TV3^{CEASEFIRE}_t$	R_{t+1}	0.050	0.277	0.963	0.964	0.327
$TV3^{CEASEFIRE}_t$	$HistVol_{t+1}$	-0.228	**0.000**	0.963	2.913	*0.088*
$TV3^{CEASEFIRE}_t$	$CondVol_{t+1}$	-0.231	**0.000**	0.963	4.188	**0.041**
$TV3^{NEGOTIATION}_t$	R_{t+1}	0.050	0.271	0.961	0.978	0.323
$TV3^{NEGOTIATION}_t$	$HistVol_{t+1}$	-0.231	**0.000**	0.961	2.557	0.110
$TV3^{NEGOTIATION}_t$	$CondVol_{t+1}$	-0.232	**0.000**	0.961	4.162	**0.042**
$TV3^{PEACE}_t$	R_{t+1}	0.052	0.253	0.959	1.077	0.300
$TV3^{PEACE}_t$	$HistVol_{t+1}$	-0.210	**0.000**	0.959	2.764	*0.097*
$TV3^{PEACE}_t$	$CondVol_{t+1}$	-0.212	**0.000**	0.959	3.870	**0.050**

Table 4.7: Evaluation metrics of the relationship between the four daily "leave-one-out" sentiment scores against next-day WTI returns, historical volatility, and conditional volatility for the period 02-Feb-2022 until 31-Dec-2023. In bold are the p-values that show significance at the 5% confidence level. Results in italics are significant within the 10% confidence level.

Then, we trained the Neural Network four times, maintaining the structure presented in Section 4.5.6. In each training phase, the topic models extracted via LDA, as presented in Sections 4.5.4 and 4.6.2, were augmented with one of the four "leave-one-out" sentiment measures and fed to the Neural Network to predict next-day WTI returns, historical and conditional volatility, respectively. The Neural Network performance metrics results for each of the four"leave-one-out" measure are displayed in Tables 4.16, 4.17, 4.18 and 4.19.

The importance of the excluded noun is revealed by a decrease in Accuracy and F2-metrics; the higher the reduction, the greater the impact on the forecast accuracy. Hence, we computed the difference between the performance metrics of the Neural Network predictions (see Section 4.6.2) and the performance metrics of each of the four newly-trained FFNN frameworks. For each topic model and for each target variable, we report in Table 4.8 the sentiment measures that produced the greatest Accuracy and F2-metric differences. The overall highest differences per target variable among the four different topic models are in boldface.

Highest Accuracy Difference								
Variable	4 Topics Model		5 Topics Model		6 Topics Model		8 Topics Model	
Return	$TV3_t^{PEACE}$	**13.68%**	$TV3_t^{WAR}$	8.43%	$TV3_t^{CEASEFIRE}$	6.32%	$TV3_t^{NEGOTIATION}$	8.42%
Historical Volatility	$TV3_t^{NEGOTIATION}$	6.32%	$TV3_t^{CEASEFIRE}$	8.41%	$TV3_t^{CEASEFIRE}$	1.05%	$TV3_t^{NEGOTIATION}$	**10.53%**
Conditional Volatility	$TV3_t^{WAR}$	-1.05%	$TV3_t^{WAR}$	4.21%	$TV3_t^{PEACE}$	**8.42%**	$TV3_t^{WAR}$	-4.20%
Highest F2 Difference								
Variable	4 Topics Model		5 Topics Model		6 Topics Model		8 Topics Model	
Return	$TV3_t^{PEACE}$	0.1	$TV3_t^{WAR}$	0.1	$TV3_t^{CEASEFIRE}$	**0.29**	$TV3_t^{WAR}$	0.05
Historical Volatility	$TV3_t^{WAR}$	-0.05	$TV3_t^{CEASEFIRE}$	**0.33**	$TV3_t^{PEACE}$	0.22	$TV3_t^{NEGOTIATION}$	0.15
Conditional Volatility	$TV3_t^{PEACE}$	0.26	$TV3_t^{NEGOTIATION}$	**0.28**	$TV3_t^{WAR}$	-0.15	$TV3_t^{PEACE}$	0.23

Table 4.8: Highest Accuracy and F2-measure differences between the FFNN model trained on $TV3_t$ and the same trained with a "leave-one-out" sentiment measures. For each target variable, we report the sentiment measure name and the difference value. In bold are the overall highest differences per target variable among the four different topic models.

The exclusion of the noun "peace" reflects in the greatest reduction of forecast accuracy of both next-day returns and conditional volatility by 13.7% and 8.42%, respectively. Similarly, withholding the terms "negotiation/s/negotiated" leads to a 10.53% lower Accuracy of historical volatility forecast. The omission of "ceasefire" translates in a reduction of the F2-measure for the next-day returns and historical volatility forecasts by 0.33 and 0.29, respectively. Lastly, "negotiation/s/negotiated" appears to affect the F2-measures of the conditional volatility prediction by 0.28.

In a perhaps surprising finding, omitting the terms "war/s" does not lead to the greatest reduction of any target variable's forecast performance metrics. To understand this behaviour, we displayed in Table 4.9 the number of occurrences of each noun in the corpus; "war/s" has, by far, the highest frequency, appearing in 3893 occasions, i.e., 26 times more than "ceasefire", 4 time more often than "negotiation/s/negotiated" and almost 9 times more than "peace."

"War/s"	"Ceasefire"	"negotiation/s/negotiated"	"Peace"	Total
3893	147	905	445	5390

Table 4.9: Number of occurrences of selected stand-alone nouns in the news articles corpus.

These frequencies are understandable given that the period we examined starts on February 2nd 2022, a mere 22 days before the Russian army crossed Ukrainian borders. Moreover, our news articles corpus, which ends on December 31st 2023, includes the first three months of the Gaza-Israel conflict, which began on October 7th 2023. Thus, war has been raging during 97% of the time interval considered, making its presence a new 'standard' geopolitical state of affairs. As a consequence, after an initial period, news regarding the ongoing conflicts are not considered a novelty anymore. This should explain why the omission of the terms "war/s" from the sentiment measure computation does not affect the forecast accuracy as much as the omission of the other nouns present in the dictionary. As shown earlier, we find evidence of a stronger Granger causality between the TV3_t^{WAR} measure and both next-day historical and conditional volatility, which corroborates our interpretation. Moreover, omitting the terms "war/s" leads to a higher correlation between the sentiment measure and the two volatility time series.

Another observation is that the presence of "peace", "negotiation/s/negotiated" and "ceasefire" in the news has a higher effect on the precision of the forecasts. Contrarily to "war", news about peace treaties, negotiations and ceasefire are unanticipated, leading to a stronger variation of oil prices and impacting the forecast accuracy of our framework.

4.7 Conclusions

The technical core of our analysis has been the experimental exploration of four different sentiment measures to forecast next-day directional changes for WTI returns, historical volatility and conditional volatility, against a baseline provided by the off-the-shelf sentiment analyser. Each measure was given as a dictionary of crude-oil related keywords/sentiment modifier co-occurrences and stand-alone nouns; to each pair and to each noun we assigned a +1/-1 weight to model the expected effect (on next-day WTI prices) of news containing such words. This simple, human-centred methodology allowed us to account for properties of the crude oil market that, counter-intuitively for a general-purpose sentiment analyser, reflect on prices changes. We found that our sentiment measures MV_t, which is computed considering only the co-occurrences weights, and $\mathrm{TV3}_t$, which is computed based on the co-occurrences and the war-related nouns dictionary (NOUNS3), reliably 'Granger-cause' historical and conditional volatility. In the same period, the baseline VADER sentiment analyser only found a Granger cause for conditional volatility. Furthermore, the best-performing MV_t model achieved an Accuracy $= 64\%$ and $F2 = 0.36$; the highest test results for the $\mathrm{TV3}_t$ model are Accuracy $= 59\%$ and $F2 = 0.479$, whilst the VADER model produces Accuracy $= 57\%$ and $F2 = 0.33$ when forecasting the target variables using either an ad-hoc neural network or logistic regression. Further, we show that the forecast Accuracy and F2 measure are greatly affected by the omission of the terms "peace", "negotiation/s/negotiated" and "ceasefire", since their presence in news articles brings novel information that reflect in oil prices changes. Conversely, we find that the omission of the nouns "war/s" increases correlation and produces stronger Granger causality between the sentiment measure and next-day historical and conditional volatility. This feature is a result of what can be considered the new "standard" geopolitical state of affairs. The present results are currently under consideration by Information Sciences.

4.8 Figures and Tables

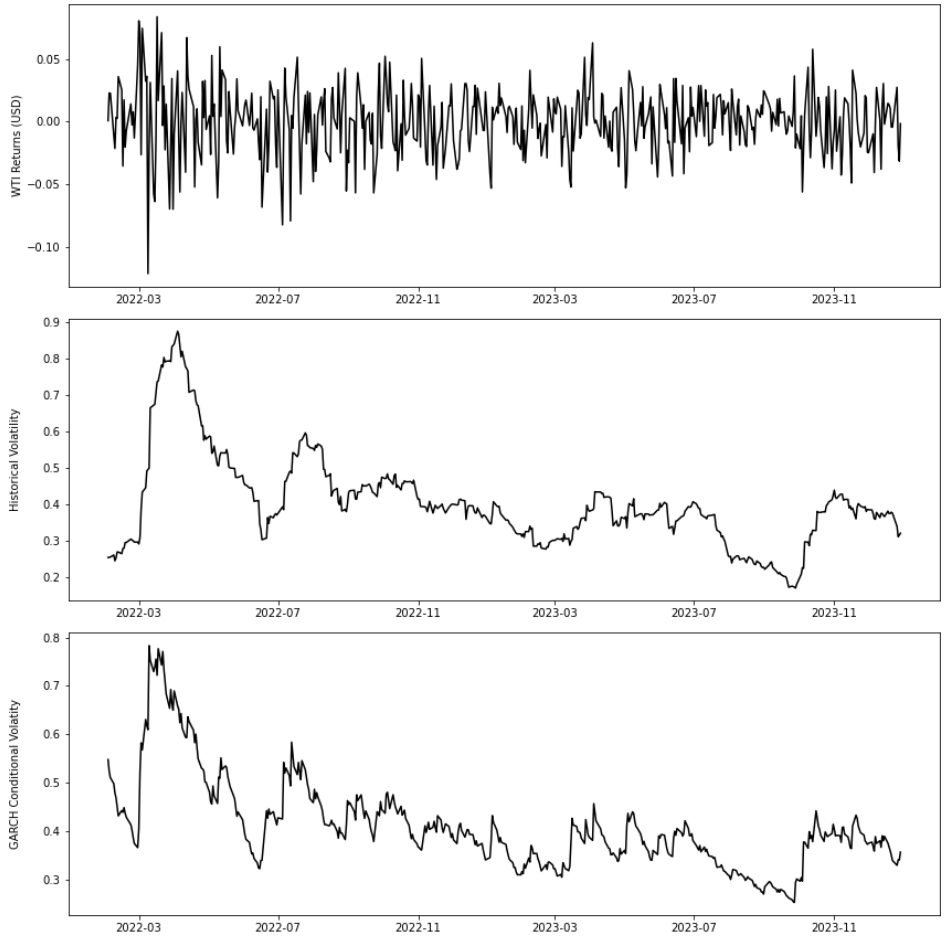


Figure 4.1: WTI daily returns in USD, annualised daily historical and conditional volatility of WTI returns for the period from 02-Feb-2022 to 31-Dec-2023.

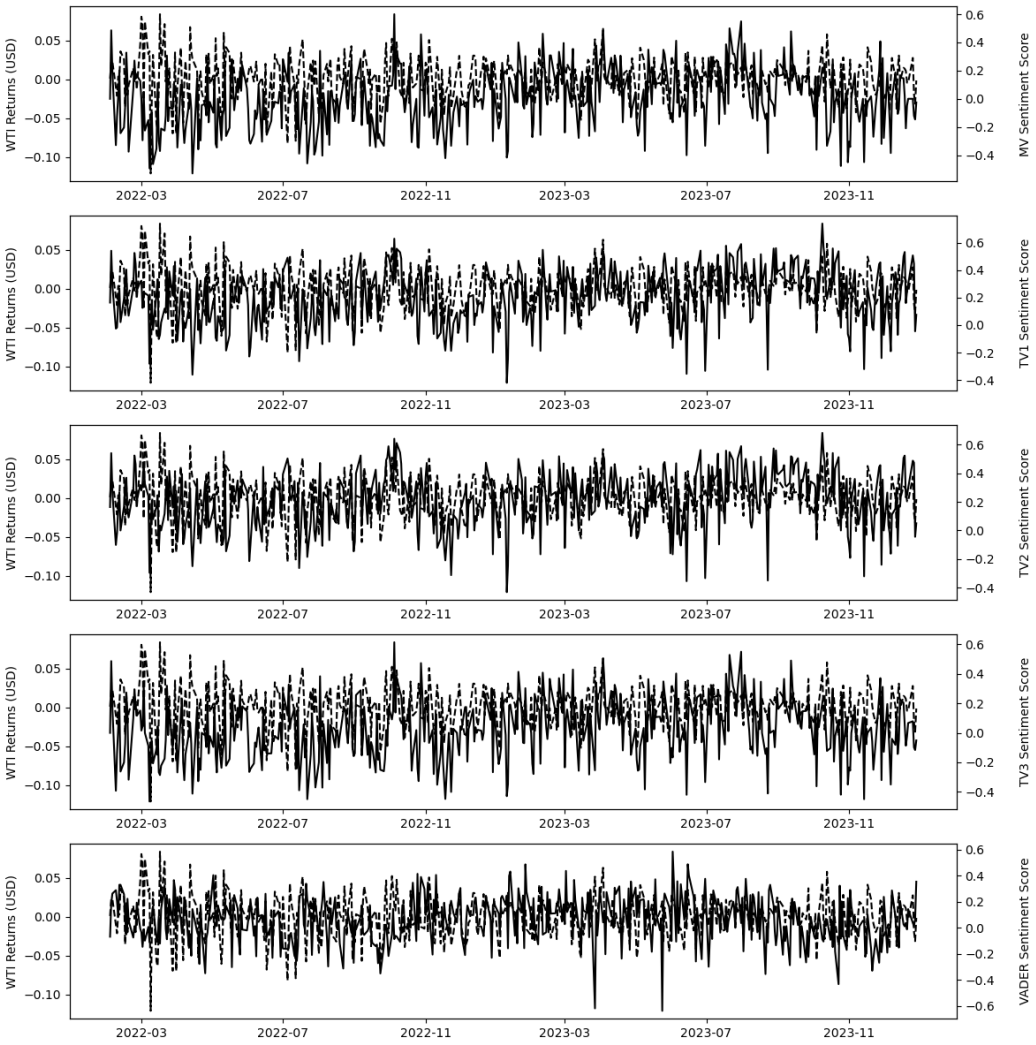


Figure 4.2: WTI daily returns in USD (dashed) plotted against the five sentiment measures (solid) for the period from 02-Feb-2022 to 31-Dec-2023. The sentiment measures are: (from top to bottom) MV,TV1,TV2,TV3 and VADER.

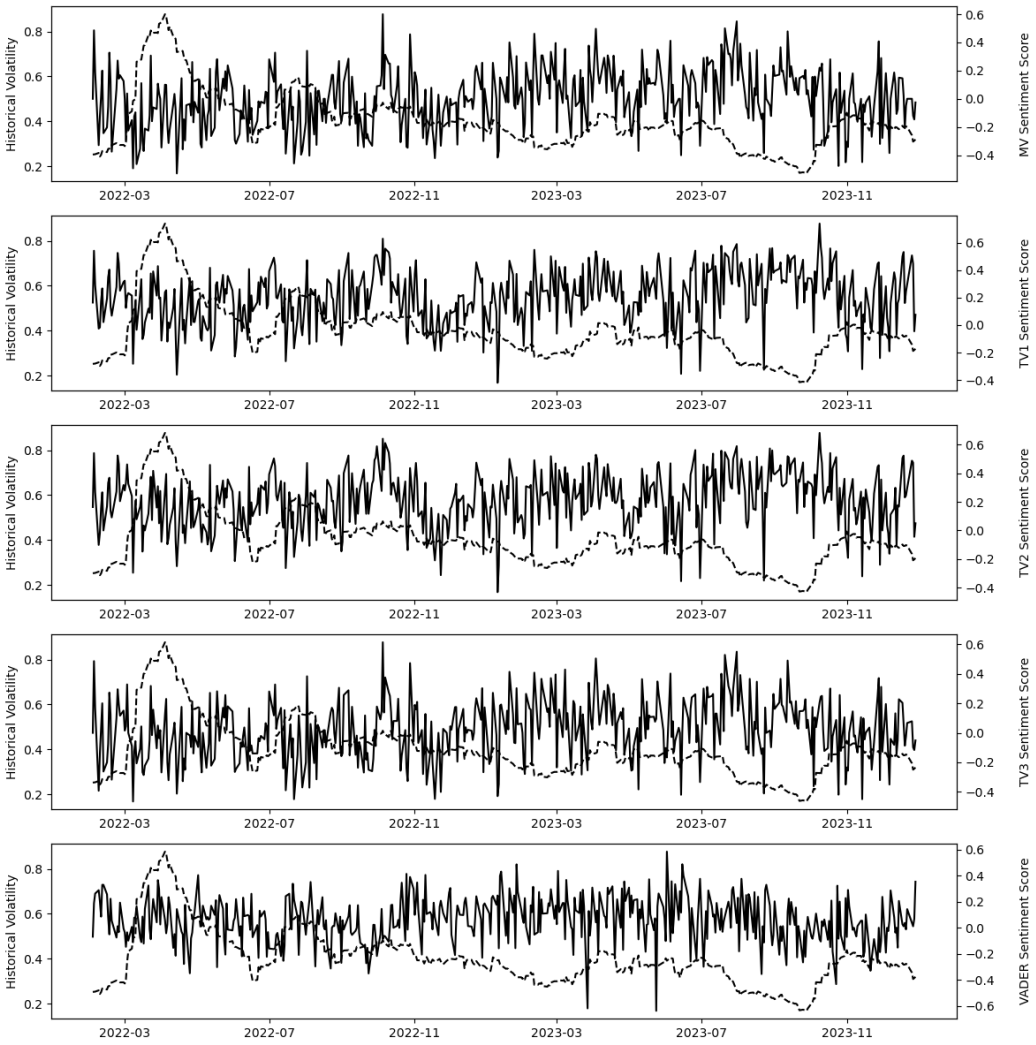


Figure 4.3: Annualised daily historical volatility of WTI returns (dashed) plotted against the five sentiment measures (solid) for the period from 02-Feb-2022 to 31-Dec-2023. The sentiment measures are: (from top to bottom) MV,TV1,TV2,TV3 and VADER.

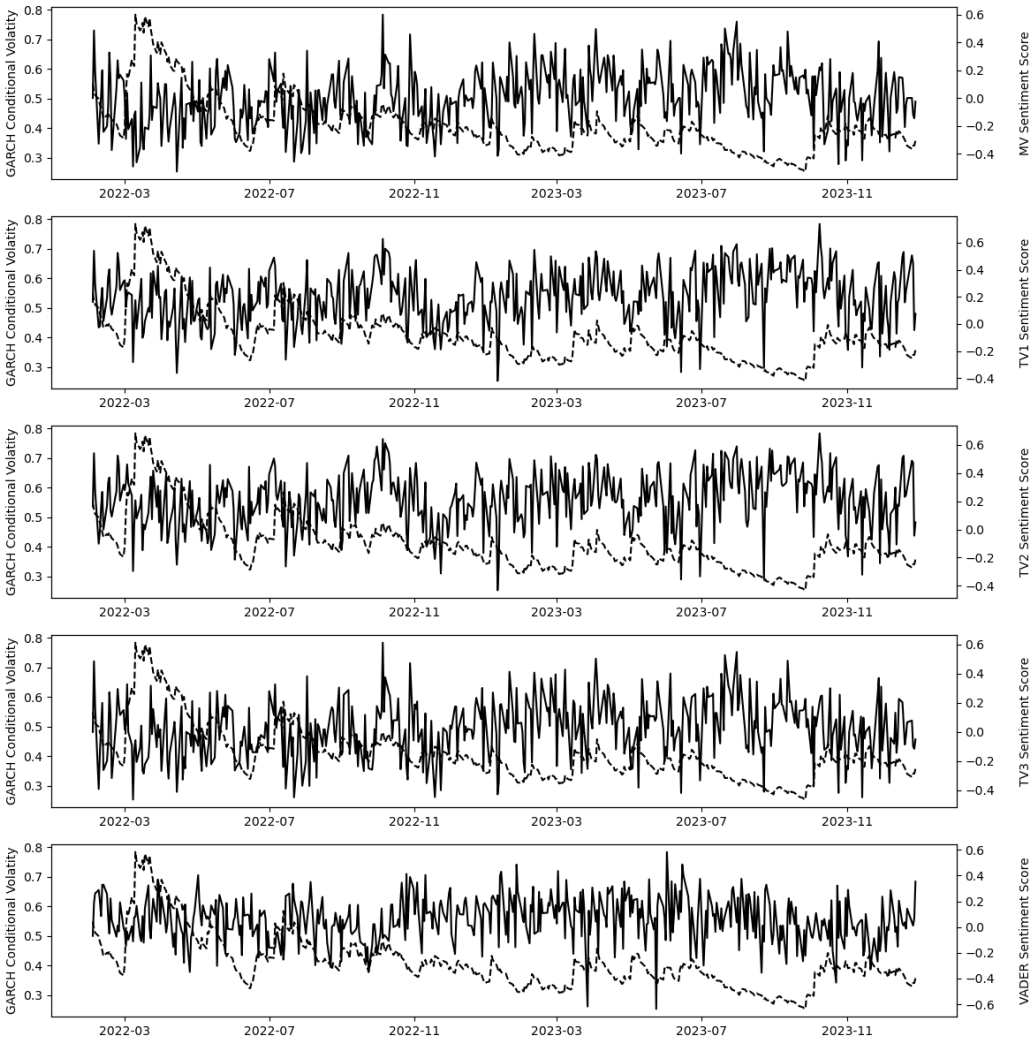


Figure 4.4: Annualised daily conditional volatility of WTI returns (dashed) plotted against the five sentiment measures (solid) for the period from 02-Feb-2022 to 31-Dec-2023. The sentiment measures are: (from top to bottom) MV,TV1,TV2,TV3 and VADER

FFNN - MV_t						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	43.15%	0.378	0.395	0.432	0.392
	Historical Volatility	53.68%	0.381	0.471	0.537	0.449
	Conditional Volatility	**65.26%**	0.000	0.000	0.653	0.000
5	Returns	46.31%	0.467	0.438	0.463	0.443
	Historical Volatility	48.42%	0.452	0.422	0.484	0.428
	Conditional Volatility	**64.21%**	0.219	0.438	0.642	0.365
6	Returns	49.47%	0.400	0.462	0.495	0.448
	Historical Volatility	48.42%	0.476	0.426	0.484	0.435
	Conditional Volatility	**61.05%**	0.156	0.333	0.611	0.272
8	Returns	47.36%	0.511	0.451	0.474	0.462
	Historical Volatility	43.15%	0.500	0.389	0.432	0.407
	Conditional Volatility	**58.94%**	0.125	0.267	0.589	0.217

Table 4.10: Performance metrics of the Feed Forward Neural Network forecast. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA and augmented with the MV_t sentiment measure.

FFNN - $TV3_t$						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	46.31%	0.289	0.406	0.463	0.376
	Historical Volatility	51.57%	0.571	0.462	0.516	0.480
	Conditional Volatility	**66.31%**	0.031	0.500	0.663	0.125
5	Returns	50.52%	0.511	0.479	0.505	0.485
	Historical Volatility	54.73%	0.476	0.488	0.547	0.485
	Conditional Volatility	**63.15%**	0.250	0.421	0.632	0.370
6	Returns	44.21%	0.444	0.417	0.442	0.422
	Historical Volatility	49.47%	0.452	0.432	0.495	0.436
	Conditional Volatility	**57.89%**	0.031	0.100	0.579	0.069
8	Returns	54.73%	0.556	0.521	0.547	0.527
	Historical Volatility	43.15%	0.381	0.364	0.432	0.367
	Conditional Volatility	**57.89%**	0.156	0.278	0.579	0.240

Table 4.11: Performance metrics of the Feed Forward Neural Network forecast. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA and augmented with the $TV3_t$ sentiment measure.

FFNN - $VADER_t$						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	47.36%	0.556	0.455	0.474	0.472
	Historical Volatility	49.47%	0.476	0.435	0.495	0.442
	Conditional Volatility	**64.21%**	0.063	0.333	0.642	0.179
4	Returns	45.26%	0.489	0.431	0.453	0.442
	Historical Volatility	52.63%	0.524	0.468	0.526	0.478
	Conditional Volatility	**53.68%**	0.063	0.125	0.537	0.104
6	Returns	38.94%	0.422	0.373	0.389	0.382
	Historical Volatility	51.57%	0.476	0.455	0.516	0.459
	Conditional Volatility	**52.63%**	0.219	0.259	0.526	0.250
8	Returns	50.52%	0.356	0.471	0.505	0.442
	Historical Volatility	43.15%	0.333	0.350	0.432	0.347
	Conditional Volatility	**58.94%**	0.250	0.348	0.589	0.323

Table 4.12: Performance metrics of the Feed Forward Neural Network forecast. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA and augmented with the VADER sentiment measure.

LOGISTIC REGRESSION - MV_t						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	50.53%	0.474	0.400	0.434	0.413
	Historical Volatility	56.84%	0.533	0.190	0.281	0.219
	Conditional Volatility	**58.95%**	0.410	0.500	0.451	0.479
5	Returns	**53.68%**	0.510	0.556	0.532	0.546
	Historical Volatility	50.53%	0.463	0.738	0.569	0.660
	Conditional Volatility	47.37%	0.327	0.531	0.405	0.472
6	Returns	53.68%	0.507	0.822	0.627	0.731
	Historical Volatility	50.53%	0.460	0.690	0.552	0.628
	Conditional Volatility	**65.26%**	0.429	0.094	0.154	0.111
8	Returns	**53.68%**	0.515	0.378	0.436	0.399
	Historical Volatility	**53.68%**	0.478	0.524	0.500	0.514
	Conditional Volatility	48.42%	0.356	0.656	0.462	0.561

Table 4.13: Performance metrics of the Logistic regression forecast. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA and augmented with the MV_t sentiment measure.

LOGISTIC REGRESSION - $TV3_t$						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	51.58%	0.486	0.400	0.439	0.415
	Historical Volatility	56.84%	0.545	0.143	0.226	0.168
	Conditional Volatility	**58.95%**	0.405	0.469	0.435	0.455
5	Returns	52.63%	0.500	0.511	0.505	0.509
	Historical Volatility	**53.68%**	0.485	0.762	0.593	0.684
	Conditional Volatility	49.47%	0.346	0.563	0.429	0.500
6	Returns	51.58%	0.493	0.800	0.610	0.711
	Historical Volatility	51.58%	0.469	0.714	0.566	0.647
	Conditional Volatility	**64.21%**	0.250	0.031	0.056	0.038
8	Returns	49.47%	0.457	0.356	0.400	0.372
	Historical Volatility	**55.79%**	0.500	0.548	0.523	0.537
	Conditional Volatility	49.47%	0.367	0.688	0.478	0.585

Table 4.14: Performance metrics of the Logistic regression forecast. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA and augmented with the $TV3_t$ sentiment measure.

LOGISTIC REGRESSION - $VADER_t$						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	49.47%	0.478	0.711	0.571	0.648
	Historical Volatility	**54.74%**	0.429	0.071	0.122	0.086
	Conditional Volatility	42.11%	0.338	0.750	0.466	0.603
5	Returns	47.37%	0.468	0.822	0.597	0.714
	Historical Volatility	**52.63%**	0.467	0.500	0.483	0.493
	Conditional Volatility	40.00%	0.338	0.813	0.477	0.634
6	Returns	47.37%	0.474	1.000	0.643	0.818
	Historical Volatility	51.58%	0.450	0.429	0.439	0.433
	Conditional Volatility	**56.84%**	0.235	0.125	0.163	0.138
8	Returns	47.37%	0.462	0.667	0.545	0.612
	Historical Volatility	**56.84%**	0.520	0.310	0.388	0.337
	Conditional Volatility	40.00%	0.338	0.813	0.477	0.634

Table 4.15: Performance metrics of the Logistic regression forecast. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA and augmented with the VADER sentiment measure.

FFNN - $TV3_t^{WAR}$						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	37.89%	0.400	0.360	0.379	0.367
	Historical Volatility	60.00%	0.452	0.559	0.600	**0.534**
	Conditional Volatility	**67.37%**	0.031	1.000	0.674	0.139
5	Returns	42.11%	0.378	0.386	0.421	0.385
	Historical Volatility	51.58%	0.524	0.458	0.516	**0.470**
	Conditional Volatility	**58.95%**	0.156	0.294	0.589	0.250
6	Returns	42.11%	0.400	0.391	0.421	0.393
	Historical Volatility	54.74%	0.524	0.489	0.547	**0.495**
	Conditional Volatility	**55.79%**	0.156	0.250	0.558	0.223
8	Returns	51.58%	0.444	0.488	0.516	**0.478**
	Historical Volatility	43.16%	0.476	0.385	0.432	0.400
	Conditional Volatility	**62.11%**	0.250	0.400	0.621	0.357

Table 4.16: Performance metrics of the "leave one out" Neural Network forecast obtained using topics extracted via LDA augmented with the $TV3_t^{WAR}$ sentiment measure. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA

FFNN - $TV3_t^{CEASEFIRE}$						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	46.32%	0.511	0.442	0.463	0.455
	Historical Volatility	57.89%	0.357	0.536	0.579	**0.487**
	Conditional Volatility	**68.42%**	0.094	0.750	0.684	0.313
5	Returns	49.47%	0.467	0.467	0.495	**0.467**
	Historical Volatility	45.26%	0.310	0.361	0.453	0.349
	Conditional Volatility	**63.16%**	0.250	0.421	0.632	0.370
6	Returns	45.26%	0.422	0.422	0.453	0.422
	Historical Volatility	50.53%	0.429	0.439	0.505	**0.437**
	Conditional Volatility	**61.05%**	0.188	0.353	0.611	0.300
8	Returns	52.63%	0.422	0.500	0.526	**0.482**
	Historical Volatility	45.26%	0.452	0.396	0.453	0.406
	Conditional Volatility	**65.26%**	0.344	0.478	0.653	0.444

Table 4.17: Performance metrics of the "leave one out" Neural Network forecast obtained using topics extracted via LDA augmented with the $TV3_t^{CEASEFIRE}$ sentiment measure. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA

FFNN - $TV3_t^{NEGOTIATION}$						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	43.16%	0.489	0.415	0.432	0.428
	Historical Volatility	50.53%	0.476	0.444	0.505	**0.450**
	Conditional Volatility	**68.42%**	0.156	0.625	0.684	0.391
5	Returns	45.26%	0.444	0.426	0.453	0.429
	Historical Volatility	49.47%	0.500	0.438	0.495	**0.449**
	Conditional Volatility	**58.95%**	0.125	0.267	0.589	0.217
6	Returns	51.58%	0.578	0.491	0.516	0.506
	Historical Volatility	54.74%	0.595	0.490	0.547	**0.508**
	Conditional Volatility	**61.05%**	0.219	0.368	0.611	0.324
8	Returns	41.05%	0.378	0.378	0.411	0.378
	Historical Volatility	45.26%	0.405	0.386	0.453	0.390
	Conditional Volatility	**65.26%**	0.250	0.471	0.653	**0.400**

Table 4.18: Performance metrics of the "leave one out" Neural Network forecast obtained using topics extracted via LDA augmented with the $TV3_t^{NEGOTIATION}$ sentiment measure. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA

FFNN - $TV3_t^{PEACE}$						
No. of topics	Variable	Accuracy	Precision	Recall	F1	F2
4	Returns	37.89%	0.289	0.325	0.379	0.317
	Historical Volatility	52.63%	0.548	0.469	0.526	**0.483**
	Conditional Volatility	**65.26%**	0.063	0.400	0.653	0.192
5	Returns	49.47%	0.511	0.469	0.495	**0.477**
	Historical Volatility	47.37%	0.571	0.429	0.474	0.451
	Conditional Volatility	**61.05%**	0.125	0.308	0.611	0.238
6	Returns	49.47%	0.533	0.471	0.495	**0.482**
	Historical Volatility	50.53%	0.405	0.436	0.505	0.429
	Conditional Volatility	**55.79%**	0.156	0.250	0.558	0.223
8	Returns	48.42%	0.467	0.457	0.484	**0.459**
	Historical Volatility	47.37%	0.524	0.423	0.474	0.440
	Conditional Volatility	**58.95%**	0.313	0.370	0.589	0.357

Table 4.19: Performance metrics of the "leave one out" Neural Network forecast obtained using topics extracted via LDA augmented with the $TV3_t^{PAECE}$ sentiment measure. In bold are the highest Accuracy scores for each model based on the number of topics extracted via LDA

Bibliography

An, N., Zhao, W., Wang, J., Shang, D., & Zhao, E. (2013). Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. *Energy*, 49, 279–288.

Atkins, A., Niranjan, M., & Gerding, E. (2018). Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science*, 4.

Bastos-Santos, M. V., Morgado-Dias, F., & Silva, T. (2023). Oil sector and sentiment analysis—a review. *Energies*, 16, 4824.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag.

Black, F. & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3), 637–654. Number: 3.

Blanco, R., Brennan, S., & Marsh, I. W. (2005). An Empirical Analysis of the Dynamic Relation between Investment-Grade Bonds and Credit Default Swaps. *The Journal of Finance*, 60(5), 2255–2281. Number: 5.

Blose, L. & C.P. Shieh, J. (1995). The impact of gold price on the value of gold mining stock. *Review of Financial Economics*, 4(2): 125-139.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.

Boone, H. (2020). Energy bankruptcy reports and surveys.

Boone, H. (2021). Energy bankruptcy reports and surveys.

Campbell, J. Y. & Taksler, G. B. (2002). Equity Volatility and Corporate Bond Yields. *National Bureau of Economic Research Working Paper Series*, No. 8961.

Collin-Dufresne, P., S. Goldstein, R., & Martin, J. (2001). The Determinants of Credit Spread Changes. *The Journal of Finance*, 56, No. 6.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31.

Deveikyte, J., Geman, H., Piccari, C., & Provetti, A. (2022). A sentiment analysis approach to the prediction of market volatility. *Frontiers in Artificial Intelligence*, 5.

Diebold, F. X. & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.

Domanski, D., Kearns, J., Lombardi, M., & Shin, H. (2015). Oil and debt. *BIS quarterly review*.

Dragomiretskiy, K. & Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3), 531–544.

Ericsson, J., Jacobs, K., & Oviedo, R. (2009). The Determinants of Credit Default Swap Premia. *Journal of Financial and Quantitative Analysis*, 44(01), 109–132. Number: 01.

Figuerola Ferretti, I. & Cervera, I. (2018). Recent credit risk and bubble behavior in the corporate energy sector.

Fritsch, S., Guenther, F., Wright, M., Suling, M., & Mueller, S. (2019). *neuralnet: Training of Neural Networks*. R package version 1.44.2.

Gabrovsek, P., Aleksovski, D., Mozetic, I., & Grcar, M. (2016). Twitter sentiment around the earnings announcement events. *PLOS ONE*, 12.

Geman, H. & Ohana, S. (2009). Forward Curves, Scarcity, and Price Volatility in Oil and Natural Gas Markets. *Energy Economics*, 31, 576–585.

Geman, H. & Vergel Eleuterio, P. (2013). Investing in fertilizer mining companies in times of food scarcity. *Resources Policy*, 38(4), 470–480. Number: 4.

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2), 215–223. Publisher: [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality].

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438.

Gyamerah, S. A. (2020). On forecasting the intraday bitcoin price using ensemble of variational mode decomposition and generalized additive model. *Journal of King Saud University - Computer and Information Sciences*.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hastie, T. & Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3), 297–310.

Hendry, D. F., Pagan, A. R., & Sargan, J. D. (1984). Dynamic specification. *Handbook of econometrics*, 2, 1023–1100.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication (1973)*, 45(4), 363–407.

Holt, C. (2004). Forecasting seasonals and trends by exponential weighted moving averages. *International Journal of Forecasting*, 20, 5–10.

Hong, L. (2011). Decomposition and forecast for financial time series with high-frequency based on empirical mode decomposition. *2010 International Conference on Energy, Environment and Development - ICEED2010*, 5, 1333–1340.

Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C.-C., & Liu, H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454, 903–995.

Hutto, C. J. & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. In E. Adar, P. Resnick, M. D. Choudhury, B. Hogan, & A. Oh (Eds.), *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM 2014, Ann Arbor, Michigan, USA, June 1-4, 2014* (pp. 216–225). New York: The AAAI Press.

Isham, M. F., Leong, M., Lim, M., & Ahmad, Z. A. (2018). Variational mode decomposition: Mode determination method for rotating machinery diagnosis. *Journal of Vibroengineering*, 20.

Kaldor, N. (1939). Speculation and Economic Stability. *The Review of Economic Studies*, 7(1), 1–27. Number: 1.

Kim, D. & Oh, H.-S. (2021). *EMD: Empirical Mode Decomposition and Hilbert Spectral Analysis*. R package version 1.5.9.

Lahmiri (2015). Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices. *IEEE Systems Journal*, 11(3), 1907–1910.

Lahmiri, S. (2016). A variational mode decompoisition approach for analysis and forecasting of economic and financial time series. *Expert Systems with Applications*, 55, 268–273.

Lisi, F. & Nan, F. (2014). Component estimation for electricity prices: Procedures and comparisons. *Energy Economics*, 44, 143–159.

Longstaff, F. & Schwartz, E. S. (1995). A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. *The Journal of Finance*, 50(3), 789–819. Number: 3.

Loughran, T. & McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1), 35–65.

Loughran, T. & McDonald, B. (2020). Textual analysis in finance. *Annual Review of Financial Economics*, 12(1), 357–375.

Loughran, T., McDonald, B., & Pragidis, I. (2019). Assimilation of oil news into prices. *International Review of Financial Analysis*, 63, 105–118.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, 29(2), 449–470. Number: 2.

Mouselimis, L. (2022). *VMDecomp: Variational Mode Decomposition using R*. R package version 1.0.1.

Nava, N., Di Matteo, T., & Aste, T. (2018). Financial time series forecasting using empirical mode decomposition and support vector regression. *Risks*, 6, 7.

Pesaran, H. & Shin, Y. (1995). An autoregressive distributed lag modeling approach to co-integration analysis. *Econometncs and Economic Theory in the 20st Century: The Ragnar Frisch Centennial Symposium*, 31.

Premanode, B. & Toumazou, C. (2013). Improving prediction of exchange rates using differential emd. *Expert Systems with Applications*, 40(1), 377–384.

Strong, J. S. (1991). Using oil share portfolios to hedge oil price risk. *Quarterly Review of Economics and Business*, 31(1), 48+. Number: 148.

Tufano, P. (1998). The Determinants of Stock Price Exposure: Financial Engineering and the Gold Mining Industry. *The Journal of Finance*, 53(3), 1015–1052. Number: 3.

Wan, X., Yang, J., Marinov, S., Calliess, J.-P., Zohren, S., & Dong, X. (2021). Sentiment correlation in financial news networks and associated market movements. *Scientific Reports*, 11, 3062.

Wang, J., Athanasopoulos, G., Hyndman, R. J., & Wang, S. (2018). Crude oil price forecasting based on internet concern using an extreme learning machine. *International Journal of Forecasting*, 34(4), 665–677.

Wood, S. (2017). *Generalized Additive Models: An Introduction with R, Second Edition*. Chapman & Hall. CRC Texts in Statistical Science. CRC Press.

Wood, S. N. (2020). *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.* R package version 1.8-33. `https://CRAN.R-project.org/package=mgcv`.

Zhang, X., Lai, K., & Wang, S.-Y. (2008). A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Economics*, 30(3), 905–918.

Zhang, X., Yu, L., Wang, S., & Lai, K. K. (2009). Estimating the impact of extreme events on crude oil price: An emd-based event analysis method. *Energy Economics*, 31, 768–778.

Zhu, J., Wu, P., Chen, H., Liu, J., & Zhou, L. (2019). Carbon price forecasting with variational mode decomposition and optimal combined model. *Physica A: Statistical Mechanics and its Applications*, 519, 140–158.