UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI "M.FANNO"

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

OIL PRICE AND SHALE OIL RIG NEXUS: AN EVALUATION OF OIL PRICE RESILIENCE

RELATORE:

CH.MO PROF. MASSIMILIANO CAPORIN

CO-RELATORE:

CH.MO PROF. FULVIO FONTINI

LAUREANDO: ROCCO ROMANIELLO

MATRICOLA N. 1189774

ANNO ACCADEMICO 2019 – 2020

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

The candidate declares that the present work is original and has not already been submitted, totally or in part, for the purposes of attaining an academic degree in other Italian or foreign universities. The candidate also declares that all the materials used during the preparation of the thesis have been explicitly indicated in the text and in the section "Bibliographical references" and that any textual citations can be identified through an explicit reference to the original publication.

Firma dello studente

Acknowledgements

I would first like to thank my supervisor, Professor Massimiliano Caporin, whose expertise was invaluable, for his continuous support, patience, and immense knowledge. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

Then, I would like to thank my parents, Rosa e Leonardo, for their continuous encouragement and support throughout all these years. To them goes my sincere gratitude, I could not have gotten this far without all the sacrifices you made, and I will always be grateful for this.

To all my family, Thank you. No matter the distance, you are always with me.

Index

List of Figures

Table

List of Abbreviations

WTI: West Texas Intermediate

COP: Change in the WTI oil price

CSO: Change in Shale oil rig count

CNSO: Change in Non-shale oil rig count

COP_MINUS: Change in the WTI oil price, only negative shocks

COP_PLUS: Change in the WTI oil price, only positive shocks

CSO_4WSUM: Sum of the previous 4 weeks for the Change in shale oil rig count

CSO_13WSUM: Sum of the previous 13 weeks for the Change in shale oil rig count

CNSO_4WSUM: Sum of the previous 4 weeks for the Change in non-shale oil rig count

CNSO_13WSUM: Sum of the previous 13 weeks for the Change in non-shale oil rig count

COP_4WSUM: Sum of the previous 4 weeks for the Change in the WTI oil price

COP_13WSUM: Sum of the previous 13 weeks for the Change in the WTI oil price

COP_MINUS4WSUM: Sum of the previous 4 weeks for the Change in the WTI oil price, negative shocks only

COP_MINUS13WSUM: Sum of the previous 13 weeks for the Change in the WTI oil price, negative shocks only

COP_PLUS4WSUM: Sum of the previous 4 weeks for the Change in the WTI oil price, positive shocks only

COP_PLUSS13WSUM: Sum of the previous 13 weeks for the Change in the WTI oil price, positive shocks only

EFFR: Effective Federal Funds Rate

TEDRATE: TED Spread

DGS10: 10-Year Treasury Constant Maturity Rate

DTWEXBGS: Trade weighted U.S. Dollar Index

GSPC: S&P 500

VIX: Volatility index

STLFSI2: ST. Louis Fed Financial Stress Index

NFCI: National Financial Conditions Index

Introduction

Oil can be considered as one of the most important commodities in the commodity markets, as its movements affect not only the global oil industry but the whole world economy. Moreover, it has recently attracted the attention of the financial markets by reaching for the first time a negative territory. Historically, oil is characterized by having high volatility, experiencing drops and spikes in price throughout the decades. The oil price has plummeted from a maximum of \$108/bbl in June 2014 to a low of \$28/bbl in February 2016, reaching in April 2020 a negative price of \$-36.9/bbl. The changes in oil price have reflected in changes in drilling activities, visible through the number of active rigs. A rig can be defined as a machine that trough the rotation of its drill pipe drills a new well to explore for, develop and produce oil. Together with the conventional drilling technique, the combined usage of horizontal drilling and hydraulic fracturing has greatly expanded the ability of producers to profitable recover oil from low-permeability geologic plays, mostly shale, expanding the US supply of petroleum. This study aims to verify if , and how, the oil prices' fluctuations have an impact on rig counts and drilling activity, in order to detect what is the relationship between them. Starting from 2011, Baken Hughes has started publishing weekly reports on U.S drilling activities, providing with accurate data on the rig counts for each basin in U.S, classified into Shale oil rig count and Non-shale oil rig count. These rigs are expected to be related to the oil price, yet this nexus has hardly been quantitatively explored. Only recently Khalifa et al. (2017) have empirically showed the presence of a delayed and dynamic relationship between changes in oil price and in rig counts, however without the specification between rig type. Given the recent event in the oil market, a better understanding of the behaviour of the rig count as response of the oil price fluctuations could be highly relevant to analysts, investors, oil companies, commercial and investment banks and policy makers. In details, this work aims to empirically test if the oil rig counts – oil nexus is affected by the nature of the oil extraction, i.e. classifying the rigs between shale and non-shale. Moreover, it is investigated the presence of asymmetry in the relationship between oil price and rig counts by considering negative and positive oil prices, defined as shocks. Finally, it is investigated whether the rig counts present a different resilience with respect to negative shocks. The dissertation is then organized as the following:

Chapter 1 provides a brief introduction on the oil industry and its importance for the world economy, a description of the oil formation process and reserve estimation, where the latter plays a fundamental role in defining the production rate and consequentially the U.S. Oil supply.

Chapter 2 is a review of literature, which considers the generic role of the oil price in the economy, an extensive analysis related to drilling activities and finally a focuses on the strand of literature related to the changes in oil rig activity to changes in oil prices.

Chapter 3 provides a description of the data, introducing the WTI crude oil price and the rig counts time series. After a statistical analysis, the relationship between the variables is investigated to obtain insights about the dynamics as well as the presence of asymmetry and resilience.

Chapter 4 conclude the analysis by introducing five different Vector AutoRegression (VAR) models, useful to investigate and understand the nexus between the considered variables. These models confirm the analysis performed in chapter 3, displaying the presence of a lagged and asymmetric relationship.

1. Introduction to the Oil Industry

1.1. History of Oil

Petroleum, in its different form, has played a central role since millennia. From it, it is possible to obtain several fuels which are fundamental for the functioning of nowadays society. The word *petroleum* derives from the Latin words *petra* and *oleum*, which mean *rock oil* and it is used to refer to a mixture of naturally occurring hydrocarbons deriving from geological formations (Speight, 2011). The use of petroleum and its derivatives has a long history. For instance, proofs of the use of *bitumen*[1] can be found in ancient civilizations like Sumerians and Babylonians[2]. Another example has been found in the Bible, where Noah is said to have use *asphalt* to caulk the Ark. In more recent time, petroleum was introduced into Europe after Arabian scientists developed techniques for its distillation. As documented by Marco Polo around 1270, Persian civilization had already an entrenched commercial petroleum industry, in particular for *kerosene* since it was used as an illuminate. For what concerns the Americas, the use of the *bitumen* was mainly for decoration and bonding purposes (Speight, 2011). A breakthrough moment in the petroleum industry has been the discovery of oil in Titusville, Pennsylvania by Colonel Edwin L. Drake by drilling for the first time an oil well for 21 metres in 1859. This discovery has defined the begin of the modern petroleum industry, started from the United States. Since then, the oil industry started to acquire increasing importance for the functioning of the society. This importance has been exacerbated by the industrial revolution which brought the invention of gasoline engine and diesel engine and the consequently switch from coal to oil to fuel steam engines. By the beginning of the 20th century, petroleum was used to fuel every type of aircraft and until today it is still the dominant source of energy used by mankind.

Nowadays, Oil is the most consumed primary energy source in the World, satisfying around 33% of the total consumption in 2018[3]. During the last two decades, oil importance kept growing at a steady level, both for the increasing demand from emerging countries and for new technologies adopted that allowed to *unconventionally* extract oil in a new, economical way. Historically the demand for oil is characterized by being multisectoral. For instance, oil is demanded in sectors like Road transportation, Petrochemicals, Residential/Agricultural/Commercial, Aviation, and Electricity Generation. The OPEC's "*World Oil Outlook 2040*" (WOO) report that, in the *OECD* countries during 2017, half of the

---

[1] Highly viscous liquid or semi-solid form of petroleum
[2] Speight (2011) reports that bitumen was used both for construction and decoration purposes.
[3] BP Statistical Review of World Energy 2019, p.9

demand came from the road transportation sector, followed by the petrochemicals with 14.38%. On the supply-side, the major suppliers of oil are United States (16.2%), Saudi Arabia (13%) and Russia (12.1%)[4].

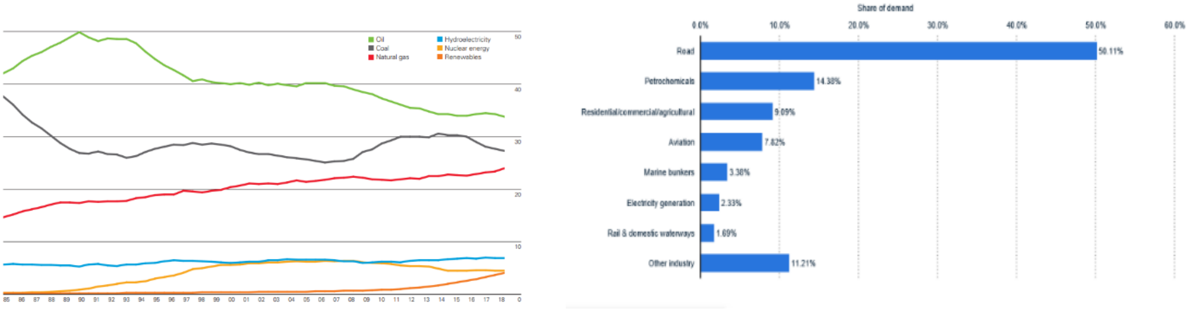

Figure 1-2: (Left) Share of global primary energy consumption by fuel (BP, 2019) and (Right) Distribution of oil demand in the OECD by sector (WOO, 2018)

1.2. Oil formation and Reserves estimation

To understand the importance of the petroleum in our society might be of interest to understand the formation process of the oil and the related production processes necessary to obtain the finite products. The widest accepted theory regarding the origin of petroleum is the biogenic theory. Under this theory, oil and natural gas are the result of compression and heating of plants and micro-organisma like plankton, over geological time, into different atomic arrangements depending on the specific original organic matter and the specific nonorganic sediments (Carmalt, 2016). A central role is played by the temperature since allows the rearrangement of the molecular chains of carbons. As the temperature increase with depth, the more *cracking* of atoms occurs obtaining different products: first heavy oils, then lighter oils and later gases. A concept widely known in the oil industry is the "*oil window*" which represent the depth at which is possible to obtain liquid hydrocarbons. This window corresponds to a depth from 2,130 m to 5,500m where the temperature is between 65 and 150 C°. However, temperature itself is not enough to create an oil formation. The other two key-components are migration and trap. Once the organic matter is deposited, the process of migration starts. Basically, it consists in the seeping of the liquid hydrocarbon through porous rocks[5] due to its larger volume and lower density than the organic matter. This process continues until the

[4] The data provided by BP refers to 2018, as a share of total crude oil production excluding biomass and coal derivatives. BP Statistical Review of World Energy 2019, p.16.
[5] Porosity and permeability are two fundamental features for the rock to become a *source rock*.

hydrocarbons find layers able to trap them in porous rock formation, commonly called *reservoirs* (Figure 3).

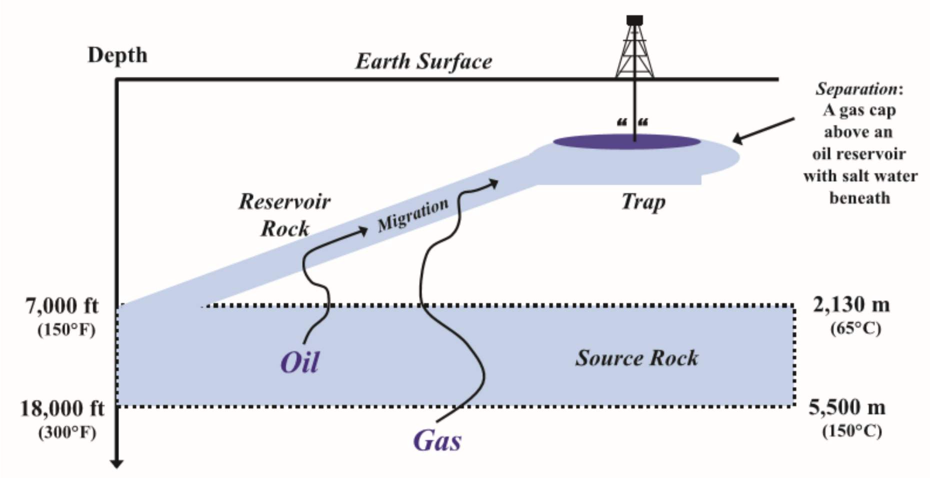

Figure 3: Formation of Oil and Gas deposits. Source: Inkpen & Moffett, (2011).

A reservoir is characterized by having the right degree of permeability (considered as capacity of transmit fluids) and porosity (capacity to store fluids). Among the different reservoir structure, the simplest is the anticline and the dome, each of them characterized by a convex upper surface (Figure 4).

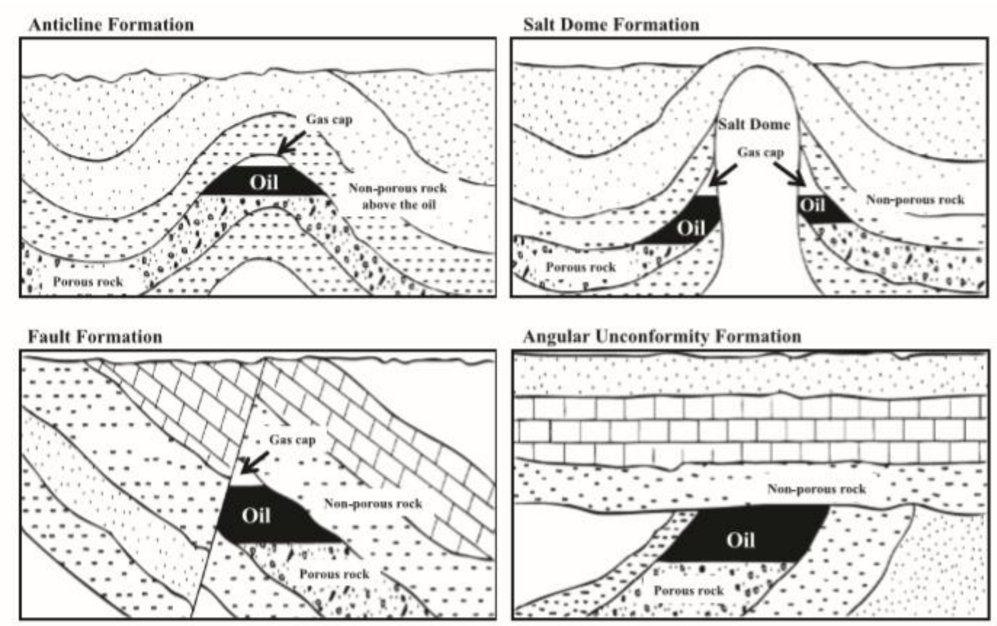

Figure 4: Oil reservoirs and cap structures. (Inkpen & Moffett, 2011)

Given the scarcity of petroleum, defined by its nature, the price of the related derivative products depends, broadly speaking, on the classic supply-demand dynamics. Due to this reason, it is fundamental to evaluate with high degree of certainty the quality and the quantity of the reserves before starting any exploration or production process. Reserves valuation represents a fundamental concept in the oil and gas industry. No other industry puts that much emphasis on an asset which has yet to be developed and cannot be touched or seen. In fact, the total quantity of oil and gas contained in any reservoir is uncertain. Furthermore, only part of this *oil-in-place* (OIP) is physically recoverable and depending on the state of technology and the costs required for the extraction, only a portion of OIP is economically available and it is generally called *reserves*. The quantification of reserves however is endogenous and depends on two elements: resource estimation and the feasibility of the extraction, that is if it is economically convenient to extract, given the actual oil prices and technologies available. At this point, to better understand how reserves are quantified and how the production rate is determined, is fundamental to distinguish between the geological and economical point of views which are driven by different beliefs. In fact, depending on the perspective adopted, the causality direction changes. Following the market prospective, the production ratio which is defined as the rate of change of the reserve over time determines the stock of reserve itself. Therefore the market, by mean of demand-supply dynamics, set the quantity of oil to extract through the price, *caveat* that reserves' estimates are changing depending on the endogenous variables' movements. Conversely, for geologist the production rate is adjusted toward the optimal amount of reserve/production rate (RPR) that allows constant drilling activities. This is because a too high production rate affects the pressure within the basin leading to a potential spike in the extraction costs, i.e. cultivation cost, making the drilling process not economically feasible. A comprehensive framework adopted for the classification of mineral reserves and resources considering the related uncertainty and feasibility is the McKelvey box (McKelvey, 1973). The model considers both the vertical and the horizontal dimension. The former is related to the price and the cultivation cost, that is the economical perspective while the latter refers to the degree of certainty of the existence of reserves.

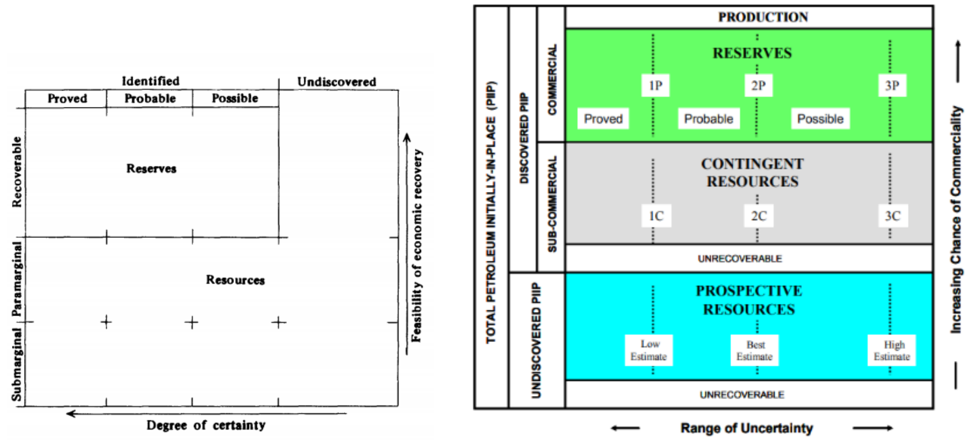

Figure 5-6: (Left) Original McKelvey box (1973) and a modified version (Right) from Society of Petroleum Engineers (2011)

From a modified, more specific box provided by the *Society of Petroleum Engineers* (2011) we can immediately notice that the total petroleum in-place can be divided in discovered and undiscovered. The latter is the oil that can be hypothetically or speculative found within the resource and it is related to exploration activities. The former instead refers to known or at least inferred quantity and includes four subcategories: production, reserves, contingent resources and unrecoverable (Inkpen & Moffett, 2011). Emphasis must be placed on contingent resources defined as quantity potentially recoverable depending on the condition of commercial markets or the technology required not yet available. Direct consequence is that the separation line between reserves and contingent resources can easily move up or down, creating uncertainty and variability in the quantification of reserves. A further specification is on the category of reserves: proved, probable and possible. The proved one are characterized by having at least a 90% certainty of being recovered. Probable reserves are based on reasonable evidence (between 50% and 90%) of recoverability of hydrocarbons in a feasible way while possible are reserves whose existence is based on evidence that must be proven (< 50%)[6]. Another useful tool widely adopted in the industry to understand the exploratory cycle and to analyse the marginal decrease in the productivity is the so-called "*creaming curve*". It considers the cumulative discoveries against the cumulative wells over time, as shown in the Figure 7**.**

[6] The classification and definition of reserves are required by law and overseen by *Securities and Exchange Commission (SEC)*, to avoid speculation in the business valuation

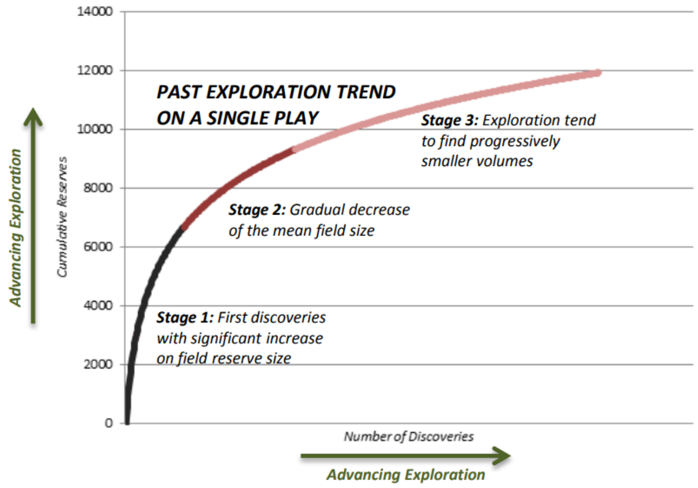

Figure 7: Creaming curve representation, (Bohorquez M.O, 2014)

It shows an initial steep as larger fields are found, and as exploration activity goes further, the average size of discoveries will fall ending up with a flat curve. Estimations of reserves is crucial not only for the evaluation of the feasibility of drilling but to guarantee a reliable forecasts of future oil supply. If we consider a geological time scale, geologists might argue that hydrocarbons could reproduce, hence can be considered as renewable source, but in term of human lifetime this source must be considered as not renewable. Therefore the crude oil, being a non-renewable fossil fuel, is characterized by having a depletion rate generally higher than the rate of reproduction. A definition of depletion rate is given by HÖÖK *et al.* (2014) as "*the rate at which is produced in a field or region expressed as a fraction of either the ultimately recoverable resources*[7] *(URRs) or the remaining reserves".* A logical consequence of the fact that oil quantity is in a finite amount is that sooner or later the world oil output will reach a maximum, declining inevitably thereafter. This concept is at the basis of the *Hubbert peak theory*, theorized by Dr. M. King Hubbert, a geophysicist that correctly forecasted the 1970 peak oil for the US production 14 years earlier. The basic concept of the Peak oil assumes that the oil production is following an exponential growth and because of this the exhaustion of a finite resource would come relatively and surprisingly soon. The production curve of a finite resource assumed by Hubbert resemble a bell-shaped curve and its cumulative production a logistic function. However, this model neglects several key variables which could have a huge impact on the world's recoverable reserves. Among them, as in the McKelvey box, oil prices play a central role for the determination of resources that are economically recoverable. Moreover, technology progresses could increase the reserves both because of technological

[7] URR are an estimate of the total amount of oil that could ever be recovered and produced

upgrades and because the increase in price may make a technology economically convenient to use.

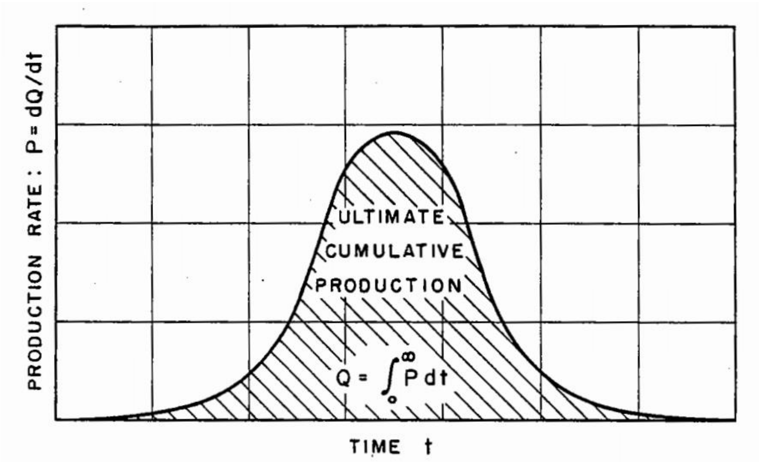

Figure 8: The Hubbert curve. (Hubbert M. K ,1956)

1.3. Conventional and Unconventional drilling techniques

The estimation process of the reserves analysed so far represent just the first step of the life cycle of an oil field that is generally composed by several stage: Exploration, Development and Production (Inkpen & Moffett, 2011). The feasibility evaluation lead to a definition of the expectations about the future value of the resource, setting a time paths for the resource price and the quantity that will be extracted. The last step of the exploration phase is the Drilling activity which consists in an initial extraction stage performed by drilling a wildcat well into the underground to confirm the presence of a petroleum reservoir. Once the rig has been set up, the drilling process can start. Now it is of great importance to highlight that, besides the classic drilling techniques, the introduction of new drilling methods as the *directional* and *horizontal fracking* has defined a milestone in the oil industry.

The "*conventional*" drilling technique widely adopted in the industry consists in the use of vertical well that aims directly at a target beneath it. Once the final depth has been reached, through the casing pipe (the part of the well which goes underground) a small pipe is run into allowing the oil and gas to flow up the well (Speight, 2011).

Conversely to the vertical drilling which is applied for high permeability rocks, in the last decades "*unconventional"* techniques have allowed to extract oil from layers of sedimental rock characterized by low permeability, yielding a product commonly called *tight or shale oil*[8]. One of the possible applicable technique which allow to enhance the permeability of layers is the *hydraulic fracturing* or *fracking*. This method has been used starting from 1950 in combination with vertical wells but has gained importance once it has been used in combination with another unconventional method, *directional drilling*, allowing to drill along the geological strata producing sufficient oil and gas to become an attractive approach. Once the well has been drilled, the permeability of the strata is increased by fracking through a high-pressure water injection[9] allowing oil and gas to flow to the well. This method presents several drawbacks, or at least issues to be addressed. The first one is the higher costs in monetary and energy terms. The amount of energy required increase as the depth increase, requiring higher cost. Moreover, this technology requires the use of several wells to extract, increasing the need of water to be injected. As a direct consequence of the use of water is that once it comes back to the surface, it must be treated as chemical waste, increasing the total operational cost[10].

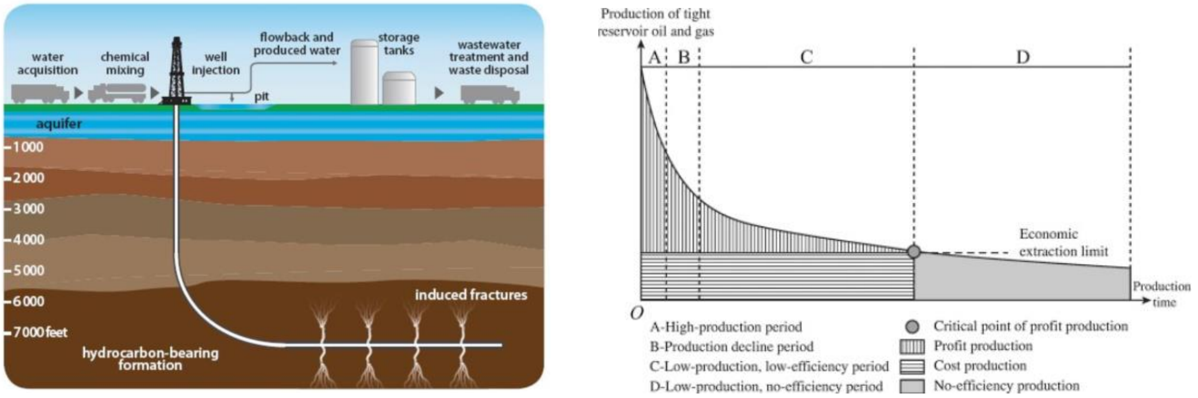

Figure 9-10: (Left) Horizontal drilling and fracking (Carmalt, 2016) and (Right) Production rate behavior of tigh reservoir oil (Zou, 2017)

Furthermore, compared to the conventional oil, tight oil deposits have a lower production capability and is characterized by having a high initial production rate which declines faster. Consequently, tight oil extraction is generally faster, more expensive, with a higher sensitivity of the project profitability to the oil prices. Depending on the oscillation of the oil price, such projects might be not profitable. Another consideration that must be done is that the risks faced by the conventional and unconventional extractions are different. While the former faces the

[8] The word *tight* is used to identify the low permeability of the layer. In the oil industry is generally used tight oil rather than shale when refers to production and resources. However, tight differs from shale since it can be extracted from not only shale formation but also from carbonates and sandstone.

[9] In more specific terms, the water is mixed by sand or ceramic grains with additional chemicals (also called "fracking fluid") to increase the water effectiveness

[10] The analysis does not consider the related higher environmental costs connected by the injection and the chemical waste.

risk of drilling "*dry holes*" that is, drilling a well without hydrocarbon fuels or without enough commercial quantities, the unconventional bear the risk of extraction rates. As stated above the production rate fall fast, therefore the risk is to not drill the well in the best horizontal layer or that the flow coming from the well is not as expected. So far, we have seen that from the petroleum is possible to obtain different products depending on several variables (type of layers, temperature, pressure and extraction techniques). However, a primary classification can be performed considering the crude oil *American Petroleum Institute* (API) *gravity* which defines its quality by measuring the density. Therefore, depending on the type of crude oil different mix of products are obtained. Crude oil with low API gravity (API < 22.3°) can be considered as *heavy crude oil*, yielding a lower-valued product due also to its higher sulphur content. Oppositely, a lighter (and sweeter[11]) crude oil is characterized by having higher API gravity (> 31.1°), yielding products of higher value[12]. The necessity of a classification arose because of the need to provide to refineries a guide to processing conditions (Speight, 2011). In fact, the differences in quality impact the processes required by refineries to obtain the final products. The lower the API gravity, the more processes needed, the higher the cost and the lower the final value of the products.

1.4. Transportation and refining processes

Before the refining process however, crude oil must be first moved from oil fields and transported to refineries. Most crude oil require treatments before the transportation since it often contains quantities of gas, saltwater or even sand (Speight, 2011). Moreover, crude oil contains excessive quantities of water that must be removed since the maximum tolerable amount of water is between 0.5 and 2% to be moved by pipeline. Transportation represents one of the most important constituents of the oil industry value chain and because of its importance, it requires a high level of coordination among the agents that operate throughout all the process. The distance represents a key variable for the determination of the form of transport since for a shorter distance it might be more efficient the use of truck while for medium could be preferred rail and for the longest distances might be dominant the use of tank or pipeline. Depending on the distance and consequently on the mean adopted, transportation costs have a huge impact on

[11] It is defined as *sweet* because of the low level of sulphur (do not exceed 0.5%). Wlazlowski, S., Hagströmer, B., & Giulietti, M. (2011). Causality in crude oil prices. *Applied Economics*, *43*(24), 3337-3347.

[12]http://www.petroleum.co.uk/api

the price of the products. Inkpen & Moffett (2011) identified two transportation segments: upstream and downstream. Firstly, oil is moved to refineries by tankers or pipeline (upstream segment) while after the refining process the derivative products are moved via more traditional transports (downstream segment: truck, railroad, tankers, pipeline).

Oil pipelines are made of steel or plastic, are buried underground, and depending on the type of oil crude the speed at which the oil is moved changes between 1 and 6 meters per second. Pipelines can be used to collect oil from wells on land or offshore and for the transmission of crude oil over long distances to refineries or refined products to markets. Crude oil tankers instead are ships constructed with the only purpose of moving crude oil for long distances. The most common crude oil tankers are the "*Very Large Crude Carriers*" (VLCCs) and the "*Ultra Large Crude Carriers*" (ULCCs) which are able to move huge quantity of oil over the globe.

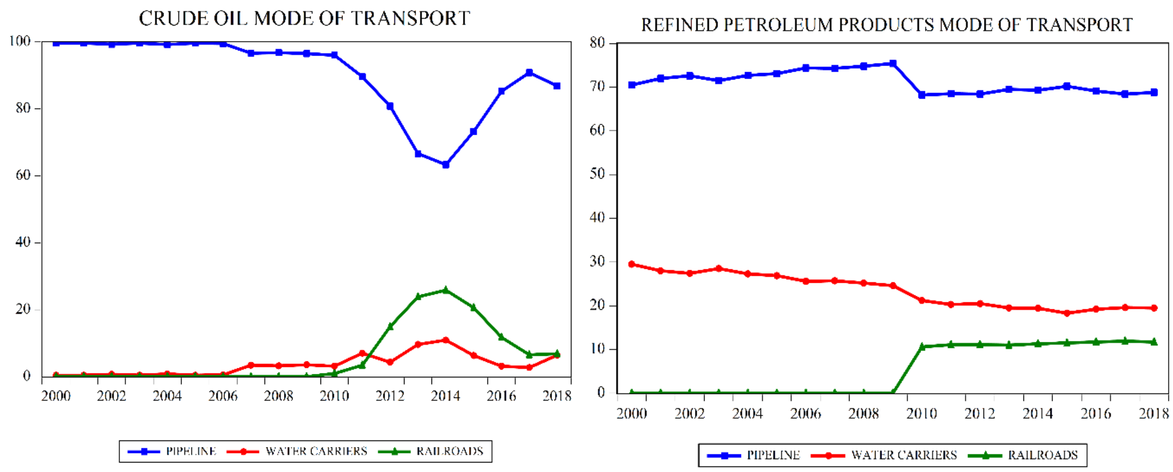

Figure 11-12: U.S. Department of Energy, Energy Information Administration (EIA) updated at Nov. 14, 2019

In the United States, during 2018 around 86% of the crude oil has been moved through pipelines (Figure 11), while the downstream transportation made extensive use of each mode of transport available. Even if the process is still dominated by pipelines (68%, Figure 12), railroad and smaller tanker truck are often used to deliver refined products from refineries to markets. Interesting to notice that pipelines are used to move several products (*batch transportation*), both for the up and downstream, and due to the physical differences, each product is shipped sequentially following the quality grade of the products.

As already stated above, different types of crude oils yield different products. However, crude oil itself has no value on the market without being refined and therefore transformed into products such as diesel, propane, gasoline etc. A refinery is an installation that process crude oil to obtain finished petroleum products by using heat, pressure, catalyst and chemicals. The output obtained by refining processes will depend on the unique molecular composition of crude oil. Despite the physical differences of crude oils, the core refining process is *distillation* or *separation* which aims to separate the oil into its fraction or boiling ranges of the various component hydrocarbons (Speight, 2011). Crude oil is heated and different products boil off at different temperatures with lighter products recovered at lower temperatures until heavy gas oil is obtained at 1000° as last product. A widely benchmark used in the industry to measure the daily processing capacity of crude oil distillation units is the crude oil refining capacity which is the amount oil produced in a refinery each day. The U.S. crude oil refinery capacity shows a steady increase over time with approximately 19 million barrels per day in 2019. The increase in the production has put pressure on refineries to keep raising capacity in order to process all the oil extracted, but in the recent years the growth has been limited.

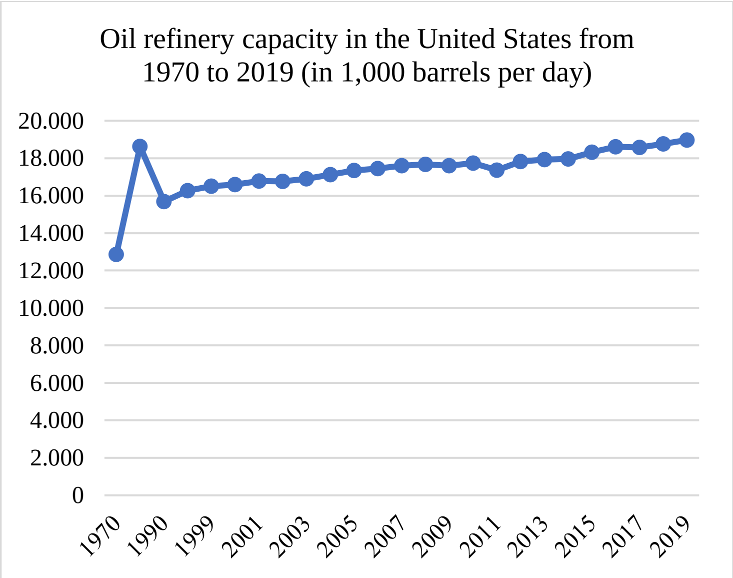

Figure 13: BP Statistical Review of World Energy (2020)

Once the derivative products have been refined, they are moved following the downstream segment through different mode of transports toward the related markets. A comprehensive view of the transportation and refining process is given by the map below (Figure 14) which shows the complexity of the transport system with higher concentration of infrastructures and refineries nearby sedimentary basins.

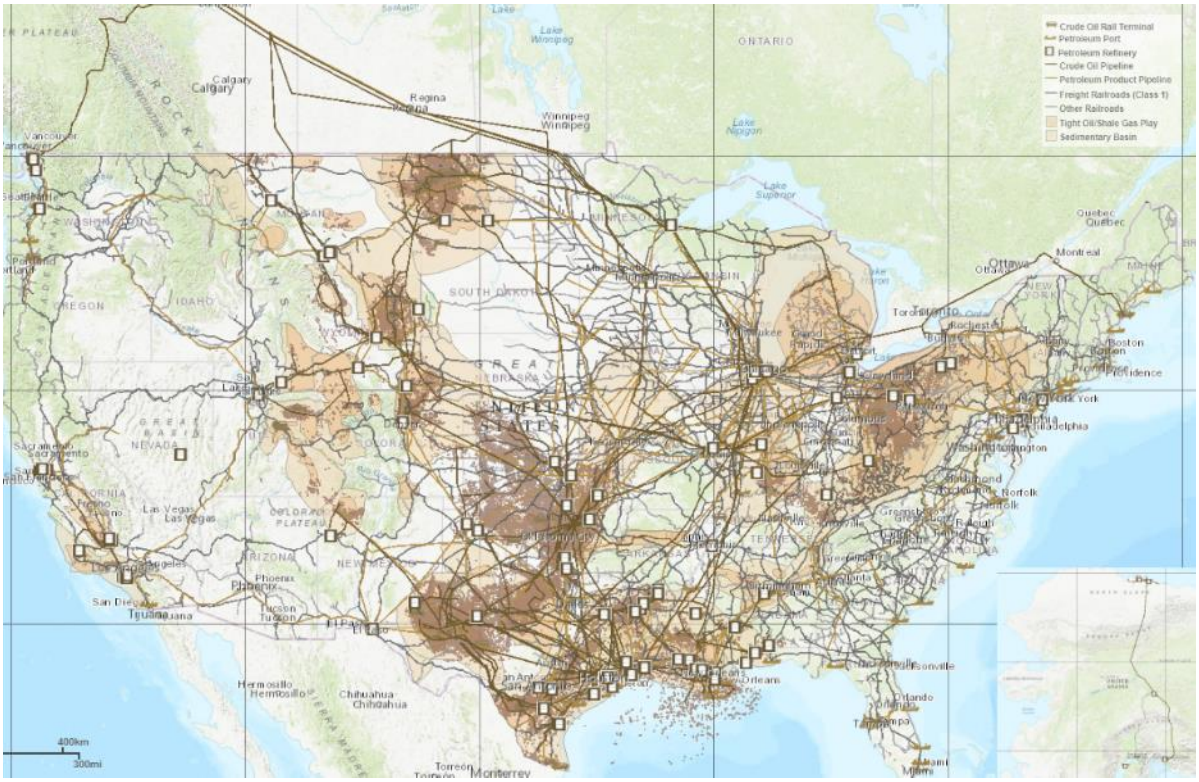

Figure 14: U.S Oil industry infrastructure, EIA (2020)

This brief overview on the oil industry has been necessary to understand why the oil is still a fundamental resource for the nowadays society. Despite its use since millennia, the oil production has kept growing to match the increasing demand boosted by the adoption of techniques which improved the recovery rate of crude oil and by the oil price volatility that has contributed to increase the exploration and production (E&P) activities as well as the improvement of available extraction methods. The extensive use of hydraulic fracturing techniques for the shale oil and the weekly publications of the U.S. rig count provided by Baker Hughes starting from 2011 open up to interesting studies to verify what is the relationship between oil prices and the change in the rig count for both shale and non-shale rig. The following chapters will try to understand whether there is and how strong is the relationship between those variables.

2. Literature review

During the last decade, the importance of the shale oil production, over the total US oil supply, has been significantly increasing. Therefore, it could be extremely interesting to study the relationship between the crude oil prices and the oil rig count. Since 2011, the number of US rig count has been weekly published by Baker Hughes, as a service to the petroleum industry, becoming a fundamental business barometer for the drilling industry and its suppliers. Therefore, a review of literature is important in order to understand what has been already explored and where this work can be located.

Due to the importance of the oil price in the economy, there exist an extensive literature about the effects of the oil price on the economic performance. Empirical studies have shown how the increasing oil prices can negatively affect the economic performance of a countries that are oil-importer. Hamilton (1983) has firstly shown how the oil shocks have significantly contributed to the 1972 U.S recession while Mork (1989) and Mork *et al.* (1994) have extended the study to analyse the asymmetric response of the GNP to oil price increases and decreases. Moreover, Huntington (1998) have proven that a significant part of the observed asymmetric responses are due to adjustments within the energy sector and not within the rest of the economy.

Other empirical studies have focused on costs and revenues to evaluate the economic profitability of oil and gas firms or specific projects and additional studies have been concerned with the evaluation of well field using the so called "*economic limit analysis*"[13] (Kaiser (2010,2012); Gülen *et al.* (2013); Howard and Harp Jr (2009); Henriques and Sadorsky (2011)).

Another strand of the literature is related to drilling activities. More in detail, Osmundsen *et al.* (2010) applied econometric analyses to determine fundamental factors for variation in drilling productivity over time and among different wells. Osmundsen *et al.* (2012) focused on the effects of various types of drilling experience (or learning) on exploration drilling productivity by means of econometrics models finding out that congestion externalities and depletion effects dominate learning effects while the experience of the drilling facility is found to have no significant effect on productivity for the sample average well. By considering the background uncertainty of the investment decision in drilling activities, due to the fluctuations of the oil price, Fattouh *et al.* (2016), Dahl and Duggan (1998) and Baffes *et al.* (2015) have studied the response of oil-producer countries to oil shocks, demonstrating that under uncertainty it is

[13] Typically used to evaluate when the operation of a well is no longer profitable, defining the point in time at which the use of the well is uneconomic.

optimal to not change the production level of oil supply, i.e. not to cut output, because of the possibility of an increase in the oil price soon. Kellogg (2009) instead examines the relationship-specific learning by-doing finding that the productivity of an oil production company and its drilling contractor is increasing remarkably as they accumulate experience working together. Moreover, he has proven that production companies tend to work with experienced drilling rigs rather than those with which they have worked relatively little. Smith and Lee (2017) instead formulated a model that assess how the economic viability of US shale oil reserves varies with fluctuations of oil price (*"price elasticity of reserves")* finding that the volume of all major US shale oil is highly inelastic with respect to price. Therefore, a drop in the oil price decrease the reserves' volume only moderately. To this extend Mohn (2008) has instead observed that positive oil price shocks are reserve enhancing, due to response both in effort and efficiency of exploration, by means of a Vector error-correction model (VECM) specification. Furthermore, he has been observed that oil companies are willing to accept higher exploration risk as oil price increase, leading to a lower success rates and higher expected discovery size. Belonging to the same strand of literature, Mohn and Osmundsen (2011) have found that the exploration activities are negatively correlated to oil price price volatility as well as background risk. Moreover, they observed that exploration responses to a fall in oil price is significant while the symmetric response (rise of oil price) is negligible, showing an asymmetric effect. Indeed, the response is characterized by having an instantaneous adjustment to price drops while having a slower adjustment for price increases.

However, the most relevant strand of literature for this work is the one related to the changes in oil rig activity to changes in oil prices, underscoring the importance of lags. Toews and Naumov (2015) propose a three-dimensional VAR model for the upstream sector of oil and gas industry in order to identify the dynamic effects of oil price on drilling activity and the costs related to it. The main findings consist in a positive relationship that runs from the change in oil price to the global drilling activity[14] (lagged by 3-4 quarters with a permanent effect), and costs of drilling[15] (lagged by 6 quarters, with a 2-years transitory effect) while the shocks from the opposite side do not affect the price of oil permanently. Ringlund *et al.* (2008) instead analyse, considering different non-OPEC regions, how oil rig activity is affected by the crude oil price by means of dynamic regression models augmented with latent components to capture trend and seasonality. Despite the observed general positive relationship between the considered variables, the authors have observed that the strength of this relationship varies

[14] An increase of 1% in the oil price leads to an increase of the global drilling activity by 1%.
[15] An increase of 1% in the oil price leads to an increase of the costs of drilling by 0.5%.

across regions. Anderson *et al.* (2014) show in their study based on existing wells in Texas that oil production does not respond to price incentives while do respond strongly to prices, by reformulating the classical Hotelling's classic model. Kellogg (2014) instead, combining information on well-level drilling with expected oil price volatility, tests the sensitivity of firms' investment decisions to changes in the uncertainty of their economic environment. A key finding is that firms reduce their drilling activity whenever the volatility rise. Moreover, estimates the effect of lagged price changes on drilling activity with the highest impact on rig counts after 3 months. Apergis *et al.* (2016) focus on the dynamic relationship between the development of oil rigs and oil production by considering data on the six major oil production regions in the U.S. (Bakken, Eagle Ford, Marcellus, Niobrara, Haynesville and Permian regions). They apply econometric techniques, including the error correction modelling, to examine the link among oil production, rig count and crude oil prices. The results show a long-run equilibrium relationship in each of the six regions, with a major impact for the Permian region. Moreover, both rig count and crude oil prices yield a positive and statistically significant coefficients with respect to the total oil production. Khalifa *et al.* (2017) study the relationship between changes in oil price and changes in rig counts, while accounting for other major economic and financial variables. The model adopted includes the Vector Auto Regressive model with "eXogenous" variables (VAR-X) as well as quantile-on-quantile approach and the quantile regression, subsetting the estimated periods. The authors show that there are evidence suggesting the presence of a non-linear link between the considered variables, with the changes of rig counts being affected by the change of oil price up to 3 months. However, it is important to notice that in the latter work, the rigs have not been split between shale oil rig and non-shale oil rig.

This work aims to examine the time series empirical relationship between the change in the rig count, split by type, i.e. considering both the shale and non-shale rig count in the U.S, and the change in oil price accounting also for other several exogenous variables which could be helpful for the estimation. Moreover, it is of significant interest to understand whether there is the presence of an asymmetric response in the rig count driven by the change in oil price. Given the recent event in the oil market, a better understanding of the behaviour of the rig count as response of the oil price fluctuations could be highly relevant to analysts, investors, oil companies, commercial and investment banks and policy makers.

3. Data description

Before introducing the methodology applied for the analysis and the empirical results, a description of the data adopted is introduced in order to obtain useful insights on the economic *rationale* of the relationship. In particular, focus will be on the relevant variables, both the Shale and Non-shale oil rig count, the WTI oil crude spot price and finally a set of potentially relevant economic and financial covariates that could affect either the evolution of the rig counts or the relationship between them and the oil price.

3.1 West Texas Intermediate (WTI) oil price

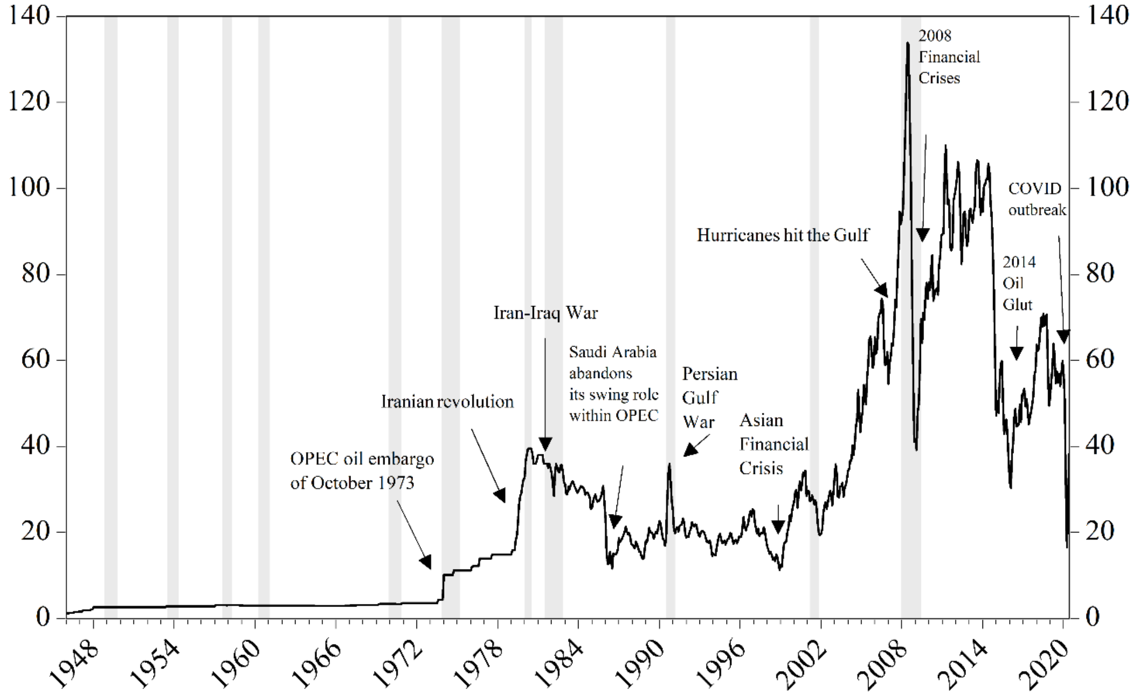

Figure 15: Monthly WTI oil spot price, source: Federal Reserve Bank of St. Louis

As shown in the first chapter, the demand for oil is the demand for its derivative products. In fact, once the oil has been produced, it must be transported to a refinery in order to gain a real market value. However, the observed market prices are the result of many factors, with the most important being the demand and supply fundamentals, studied by Hamilton (2009) in a relevant paper.

Following the author, the dynamics of the oil demand are governed by two key variables, the price and the GDP per capita. As observed by Hamilton (2009), the changes in oil price tend to be permanent and are governed by different regimes at different times, therefore are difficult to predict. Defining the price elasticity of demand as the reaction of quantity demanded to change in prices, the author has proven that the price elasticity of the demand for oil, despite the difficulties in computation, appears to be very low[16]. For what concern the supply side, it reflects the capability of oil exploration and development not only to find the oil but also to recover it and to move it. The supply of oil reflects also the several agreements reached by the largest producers of oil[17], cutting or increasing the production, that lead to a change in oil prices. Historically the supply shocks have had a huge impact on the oil prices, leading to a higher price volatility. For instance, the restrain in quantities imposed by OPEC in October of 1973, during times of growing market demands and political instability, drove the price upward from 4$ per barrel to nearly 11$ by the end of April 1974. Other supply shocks identified by (Inkpen & Moffett, 2011) are the one driven by the disruptive man-made forces such as the Persian Gulf War of 1991 or by the sudden forces of nature as the hurricanes in 2006 that damaged the production facilities and platforms of the Mexican Gulf, leading to a higher upward pressure on price.

Looking at the Figure 15 it is possible to identify two crude price eras: pre-1970 and post-1970. Prior to the 70's the crude oil markets were regional, with a stable price which fluctuated between 2.5$ /bbl and 4.3$/bbl. However, in the fall of 1973, the Oil embargo enacted by OPEC exercised enormous pressure on the pricing power, leading to a spike in the oil prices. After an initial period of high price levels around 30$, oil price began to fall in 1980, reaching a monthly level of 12$ after the oil had begun to be traded on futures markets. The figure provides a chronological order of event until the last drop in oil prices due to the global pandemic that affected the demand in April 2020.

[16] The paper reports two estimates for both the short- run and the long-run price elasticity. Respectively, Dahl (1993) through a survey obtain -0.07 and -0.30 while Cooper (2003) through an annual time-series regression reports -0.05 and -0.21.
[17] United States, Russia and Saudi Arabia account for more than 40% of the total production in 2018

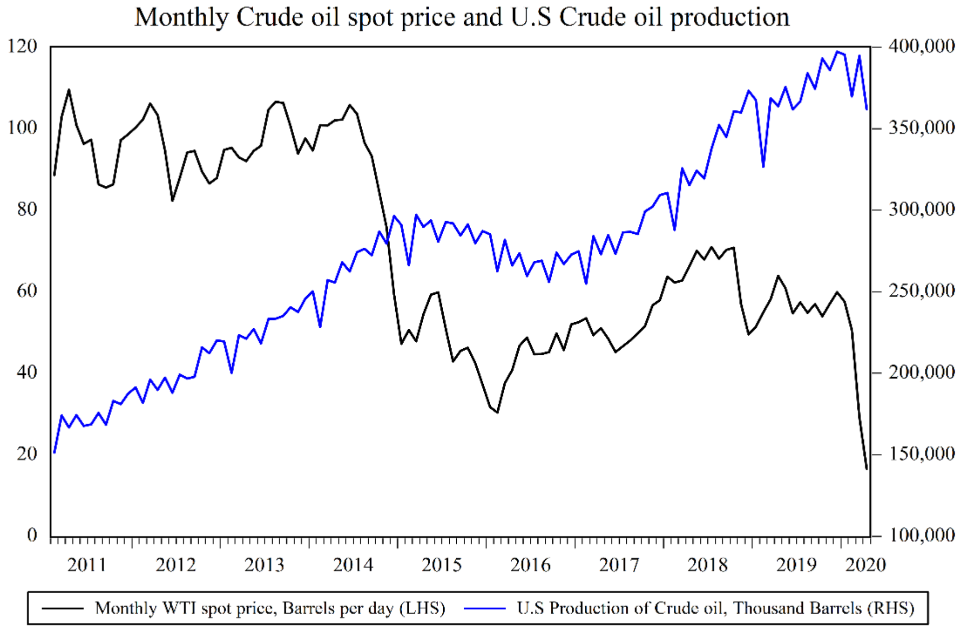

Figure 16: Monthly crude oil spot price and U.S crude oil production, Source: EIA, (2020)

Since this dissertation focuses only on data starting from 04 February 2011, considered as a reference date in which shale oil production started to expand at a faster pace, the only two relevant events are the 2014 oil glut and the oil price drop due to the Sars-Cov 2, visible on the Figure 16. By looking at the time series, it is possible to observe how the 2014 oil price drop has affected the long run value of the oil. Even though the explosion of the shale oil production is considered as the main determinant for the 2014 oil drop, others might be considered as relevant for the drop. The OPEC decision to maintain output levels instead of lowering them has played a key role in the decline of oil price and so has the slowdown of global real economic activity since 2011, as observed by Prest (2018) and Ellwanger *et al.* (2017). Another relevant event for the oil industry has been the pandemic that has globally affected the economy. In this case however, the drop in the oil price has been a combination of both demand and supply shocks. In fact, with the spread of Coronavirus, the demand for oil has fallen globally, forcing the OPEC members to reunite in Vienna on 5 March 2020 in order to review the oil production rate. At the summit, OPEC agreed to cut oil production by an additional 1.5 million barrels per day (bpd) through the second quarter of the year and called on Russia and other non-OPEC members of OPEC+ to abide by the OPEC decision. But the following day Russia rejected the demand, marking the end of the unofficial partnership, with oil prices falling 10% after the announcement. The immediate reaction of Saudi Arabia, the *de facto* leader of OPEC, has been the threat to open the oil spigots in the world and pushing the production of bpd to maximum levels. On the demand side instead, the worsening of the lock-down and the absence of space to store barrels of oil caused a further and dramatic drop in the oil price, reaching for the first time a negative price on 20 April 2020 of -36.9$/bbl. However, it is important to specify that the negative price has concerned the futures markets. In fact, the sale of crude oil can occur in three different type of transactions: spot transactions, futures markets and contract arrangements. A spot transaction takes place whenever the good is sold for cash and it is immediately available. A futures contract instead is a promise to deliver a quantity of an underlying commodity, at a predetermined place, price and time in the future. The West Texas Intermediate futures contracts are exchange on the New York Mercantile Exchange (NYMEX), and serve for several purposes. They can be used as reference to obtain information about expectations of oil supply and demand conditions as well as a mean for price discovery. Moreover, they can be used to hedge risks or speculate on the financial markets. Generally, contracts adopted for the latter purpose are liquidate before the expiration by opening an opposite trading position. This dynamic has exacerbated the oil market, already pressured by an excessive supply and a low level of demand, leading to a drop in the one-month futures contracts in a negative territory[18] as it is possible to observe in Figure 16. At the moment of writing, the WTI spot crude oil price has closed at 41.46$/bbl.

Finally, the analysis has been performed on weekly observation of the West Texas Intermediate (WTI) crude spot oil price, retrieved from the Federal Reserve Bank of St. Louis database, starting from 04 February 2011 to 03 April 2020, for a total of 479 observations.

3.2. Baken Hughes rig count

The combined usage of horizontal drilling and hydraulic fracturing has greatly expanded the ability of producers to profitable recover oil from low-permeability geologic plays, mostly shale. Starting from 2011, Baken Hughes has started publishing weekly reports on U.S drilling activities. The report contains time series of rig counts for each basin in U.S, providing detailed data on the shale oil rig, classifying them into Oil, Gas and Miscellaneous. The basins can be classified with the respect to the region to which they belong:

[18] Despite the historicity of the negative oil price, it is not irrational. Agents on the financial markets were willing to pay a price in order to avoid the higher storage costs that they would have paid if they would not be able to close their long positions. This open up to a "Contango" situation in which the prices of futures are higher than the spot prices, leading to potential profitable opportunities whenever there is the possibility to store barrels (generally to carry on these trading are used Very Large Crude Carrier to store the barrels).

- Northeast Region: *Marcellus, Utica*
- Gulf Coast Region: *Haynesville, Eagle Ford*
- Mid-Continent Region: *Fayetteville, Ardmore Woodford, Arkoma Woodford, Cana Woodford, Granite Wash, Mississippian*
- Southwest Region: *Barnett, Permian*
- Rocky Mountain Region: *DJ-Niobrara, Williston*

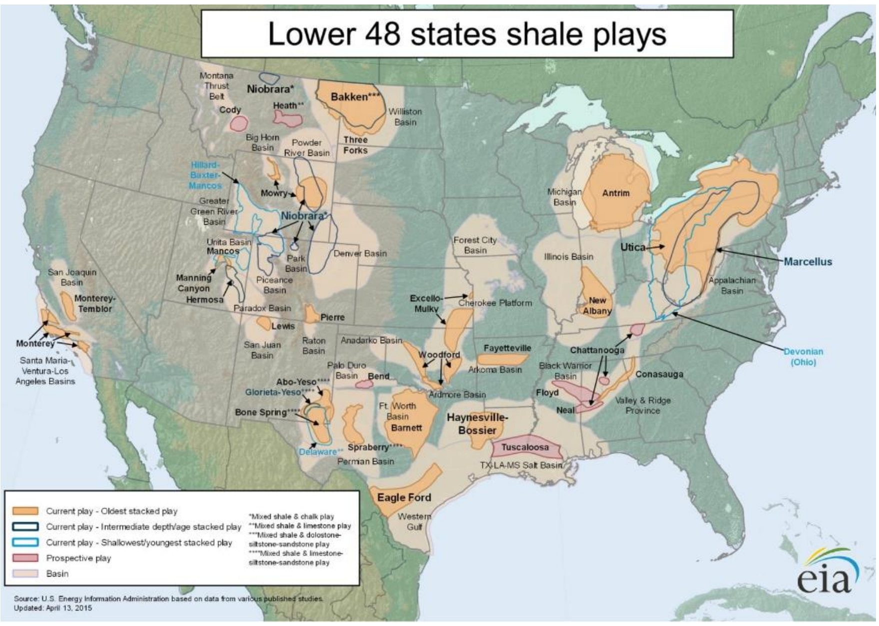

Figure 17: U.S Shale plays, source: EIA (2015)

The northeast region includes shale plays located in the Illinois, Michigan and Appalachian Basins. The latter one includes the Marcellus shale plays with a total area of 95,000 square miles following the estimates of the U.S Energy Information Administration. The Gulf Coast region instead includes the Haynesville shale play that is located in Texas and Louisiana and the Eagle Ford shale play located in the Texas Maverick Basin. Combined these two plays have an area of 14,752 square mile. The Mid-Continent region includes the Fayetteville, Woodford and Cana Woodford shale plays with an estimated area of 14,388 square miles. The Southwest instead is located in the Permian basin and contains the Barnett shale play with a total area of 10,462 square miles. Finally, the Rocky Mountain region includes the Williston and DJ-Niobrara shale plays with a combined area of 37,033 square mile.

The innovations brought by the combination of horizontal drilling and hydraulic fracturing is also visible in proved reserves. In Figure 18 is possible to observe the impact of the shale oil revolution which led to a large increase of the proved reserves deriving from tight-oil bearing formations such as the Williston Basin and the Bakken shale.

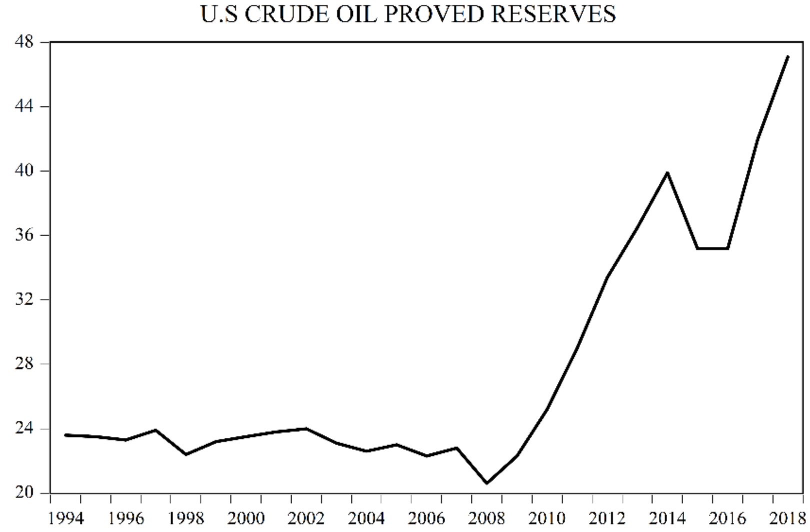

Figure 18: U.S. Energy Information Administration, Annual report of Domestic oil and gas reserves,1994-2018

After the initial increase started from 2008, reserves have fallen due to the 2014 oil price drop to later recover starting from 2016, showing an upward trend. At the end of 2018, *EIA* has estimated that the United States had 47,053 million barrels of crude oil proved reserves, where 41% of them in Texas and 12% in North Dakota (Figure 19).

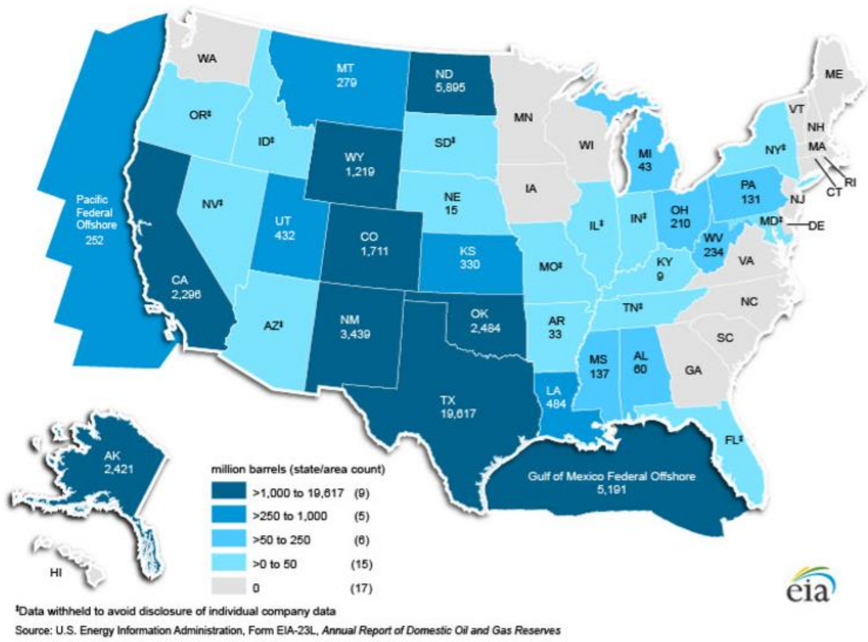

Figure 19: Crude oil and lease condensate proved reserves by state/area, 2018

Even though the production of each basin is not object of the study, it might be useful to understand how each basin is contributing to the whole U.S oil supply and how the rig have been moved during the observed time. Baken Hughes provides monthly reports on the production for the major basins, together with a report on the number of wells drilled, completed and drilled but uncompleted (DUC). This allows to analyse dynamics related to the U.S oil supply that are fundamental despite not being the focus of this study. By looking at the Figure 20, it is possible to observe a common behaviour, that is after an initial phase in which the production is at a general low level, despite the huge number of rig count, the production per rig increases exponentially[19] while the rig count gradually decrease. The reason behind this behaviour is that typically the oil does not begin to flow right after a rig drills a well in a shale formation. Moreover, even though the hydraulic fracking is adopted to complete wells in order to begin the oil flowing, the fracking step does not have to happen immediately after the well is drilled. Therefore, the production can be defined as a lagged outcome of the rig count which effect are generally visible after several months. It is also important to notice that the increasing in the production is also a result of the incremental experience as observed in Osmundsen *et al.* (2012). As it is possible to observed, the main contributors for the U.S oil production are the Permian Basin, the Eagle Ford basin and the Niobrara (despite its low level of rig count). Moreover, the absence of a direct and contemporaneous connection between the rig count and oil production is due to the fact that new drilling will not increase the production if the wells are not being completed. Hence, only the completion of a well could drive up the production even though the drilling activity falls.

[19] This phenomenon has been amplified by the implementation of the latest technology for the horizontal drilling, incorporating new materials and smaller drill that have speeded up drilling allowing for deeper and greater horizontal penetration of shale rock.

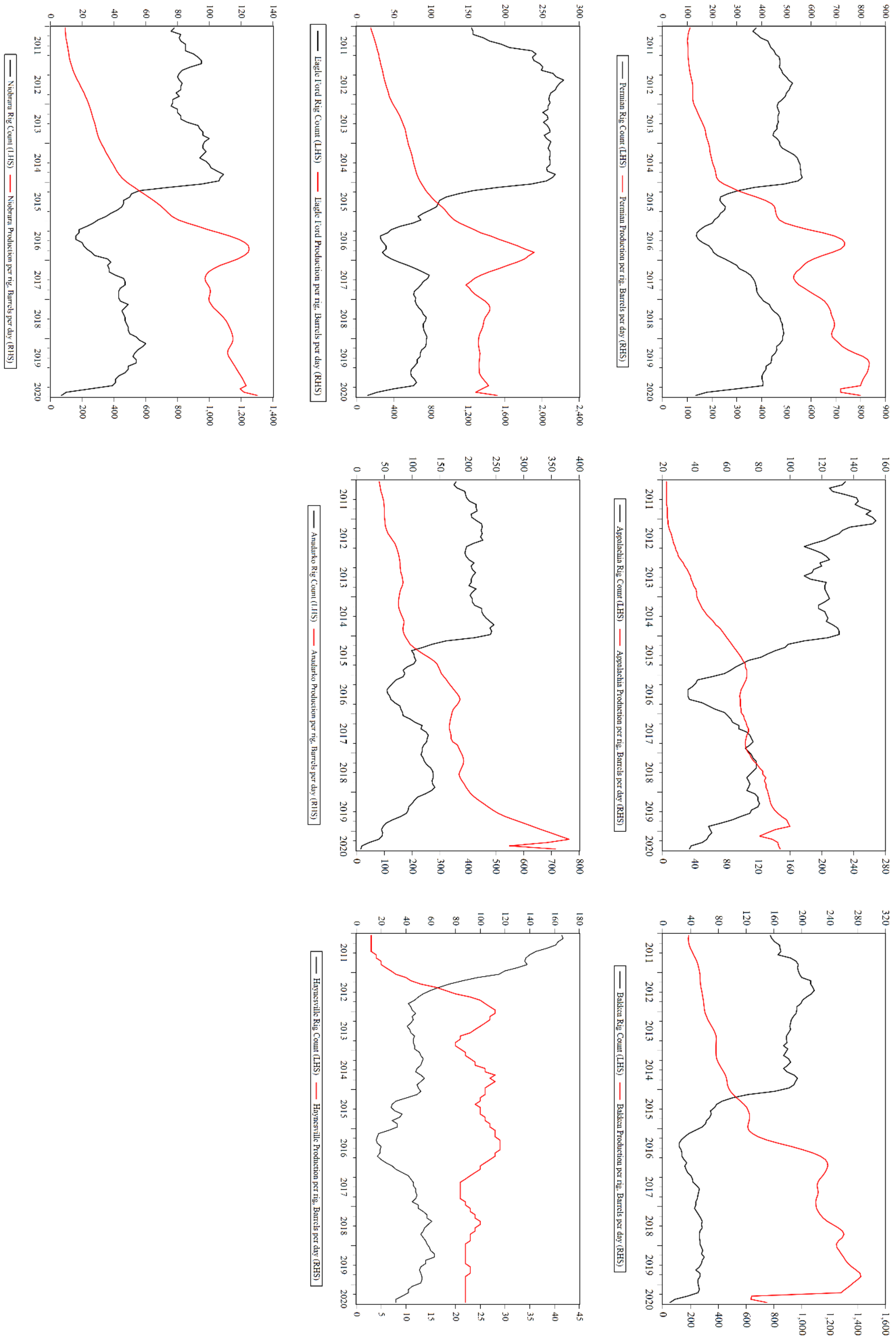

Figure 20: Rig count and Production by basin, source: EIA (2020)

Therefore, the behaviour of the oil production is strictly related to the DUC count. In particular, an useful tool is the ratio of DUCs to completions (*D/C ratio*) given by the ratio of the total number of drilled but uncompleted wells over the monthly number of completed wells.

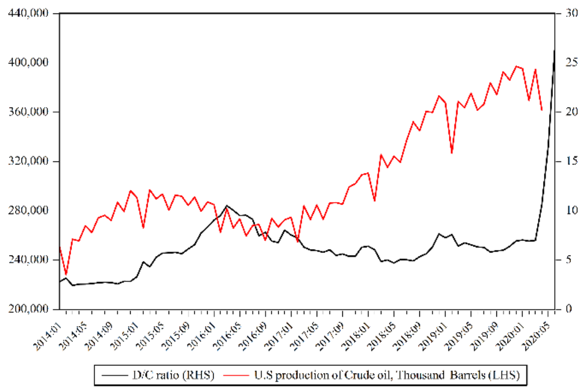

Figure 21: D/C ratio and U.S. Production of oil, source: EIA (2020)

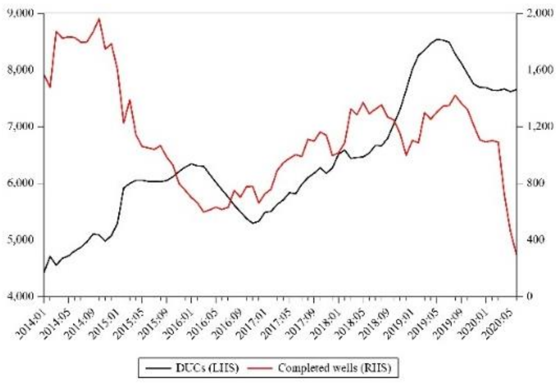

Figure 22: DUCs vs completed wells, source: EIA (2020)

By looking at the Figure 21 is possible to observe how the production and the D/C ratio are related. An increase in the D/C ratio can be due to an increase in the DUC or to a decrease in the completed wells. During periods in which the D/C ratio increase, the production tends to decrease and vice versa. By decomposing the D/C ratio, plotting the two component is possible to observe what has driven the fluctuation of the ratio (Figure 22). The initial spike of the ratio is due to a contemporaneous decrease in completed wells and an increase in the DUCs, to later decrease due to an increase in completed wells, while the DUCs have kept growing. The recent drop in oil price has affected the number of completed wells, leading to a decrease of two third in six months. A rolling correlation of the crude oil production and the D/C ratio, with a 12-month window, confirms the relationship highlighted before.

3.3. Descriptive statistics of the variables

Finally, after a brief introduction of what might drive the change in the rig count, it is possible to present the data set used. The number of rig count has been classified in shale and non-shale oil rig, summing up the number of rigs for each basin introduced before, starting from 04 February 2011 until 03 April 2020, with 497 observations.

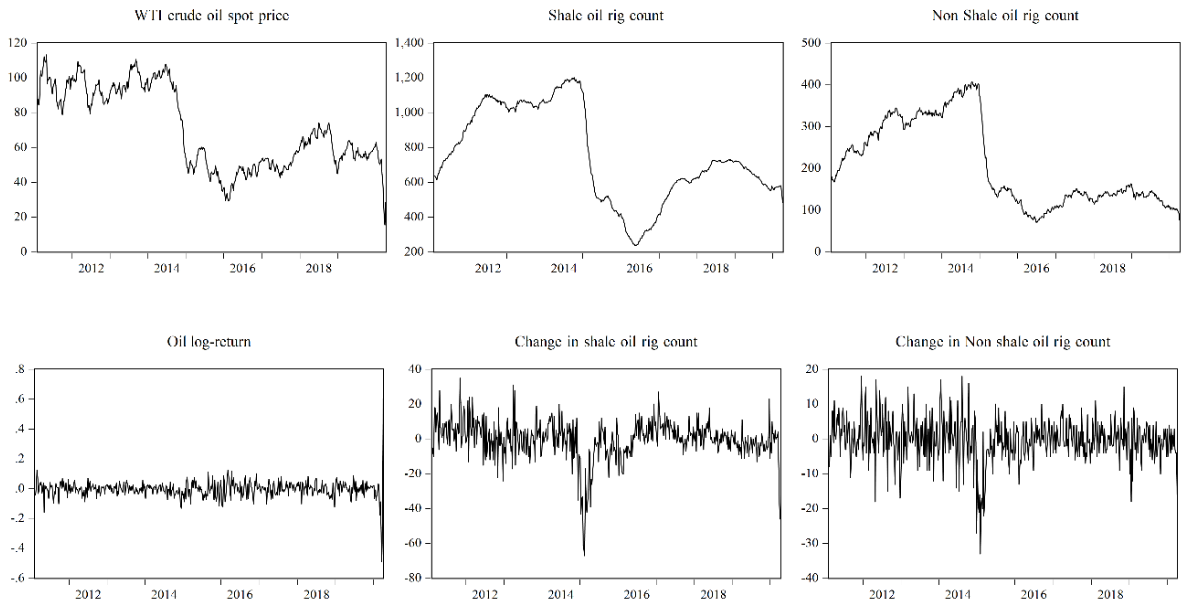

Figure 23: Plots of the variables, both in levels (top) and first difference (bottom)

The figure 23 shows the evolution of the above considered variables. The upper part displays the variables in levels while the bottom represents the variables in first difference. As already discussed above, the WTI shows a drop in price both in 2014 and 2020. Consequentially, both the shale and the non-shale rigs are characterized by a drop in the same period, lagged by weeks. Considering their first difference is possible to observe that, despite the 2014 huge drop in oil price, the related change has not been as severe as the 2020 drop. This is probably due to the fact the latter one has been more rapid and sudden with respect to the former. Considering instead the rig count, is possible to observe a huge drop following the 2014 oil glut while the effect of the 2020 drop is not yet observable since the last observation considered for the study is the beginning of April, missing the lagged response of the rig counts.

	WTI price	Shale oil rig	Non Shale oil rig
Mean	70.71000	752.5511	208.2965
Median	63.29000	714.0000	151.0000
Maximum	113.3900	1202.000	408.0000
Minimum	15.48000	237.0000	71.00000
Std. Dev.	23.21268	269.2769	100.9565
Skewness	0.152374	0.030662	0.509889
Kurtosis	1.619396	1.872899	1.730733
Jarque-Bera	39.89547	25.42925	52.90928
Probability	0.000000	0.000003	0.000000
Sum	33870.09	360472.0	99774.00
Sum Sq. Dev.	257559.9	34659796	4871876.
Num. of Observations 479			

	Δlog WTI	Δ Shale rig	Δ Non shale rig
Mean	−0.002366	−0.320084	−0.215481
Median	0.001048	1.000000	0.000000
Maximum	0.605431	35.00000	18.00000
Minimum	−0.487559	−67.00000	−33.00000
Std. Dev.	0.057656	11.68059	6.495734
Skewness	0.763826	−1.503714	−0.676710
Kurtosis	39.03739	9.526430	5.538121
Jarque-Bera	25912.13	1028.475	164.7866
Probability	0.000000	0.000000	0.000000
Sum	−1.130879	−153.0000	−103.0000
Sum Sq. Dev.	1.585673	65080.03	20126.81
Num. of Observations: 478			

Figure 24: Descriptive statistics of the variables

Figure 24 provides the descriptive statistics of the variables, both in level and in first difference. By looking at the table on the left hand side, it is possible to observe that the minimum value for both the shale and non-shale are located in the 2016, where the shale reaches first the minimum, followed by the non-shale after almost one month[20]. Similar situation is for the maximum level reached by the rig count, both on October with a week of delay for the non-shale. Moreover, from the table it is possible to observe that the shale rig count is highly negatively skewed while the non-shale rig is only moderately skewed. Furthermore, the shale rigs are more volatile, that is, are more subject to changes rather than the non-shale oil rigs. A reason for this higher volatily can be found in the nature of the shale rig, characterized by being more cost-sensitive and faster in the realization. In order to evaluate the presence of links across the three most relevant variables and to verify whether the initial insights obtained by the descriptive statistics is meaningful, could be useful to look at the contemporanous correlations (Figure 25) and the rolling correlations among their first difference.

Correlations	Δlog WTI	Δ Shale rig	Δ Non shale rig
Δ log WTI	1.0000		
Δ Shale rig	−0.0515	1.0000	
Δ Non shale rig	−0.0828	0.3417	1.0000
Num. of Observations: 478			

Figure 25: Contemporaneous correlations coefficients

The rig counts seem to be positive correlated among each other while are negatively correlated with the oil log-return, that is, as the wti increases both the shale and non-shale decrease. We can reformulate the relationship in a differet way. Ideally, assuming the rig counts are a proxy

[20] The shale oil rig count reaches its minimum value on May 13, 2016 while the non-shale rig on June 24, 2016.

of the productivity, as the rigs increase, the production increases leading to a reduction of the WTI oil price. However, these relationships are weak and are varying over time, challening the stability of the correlations across the sample in the future.

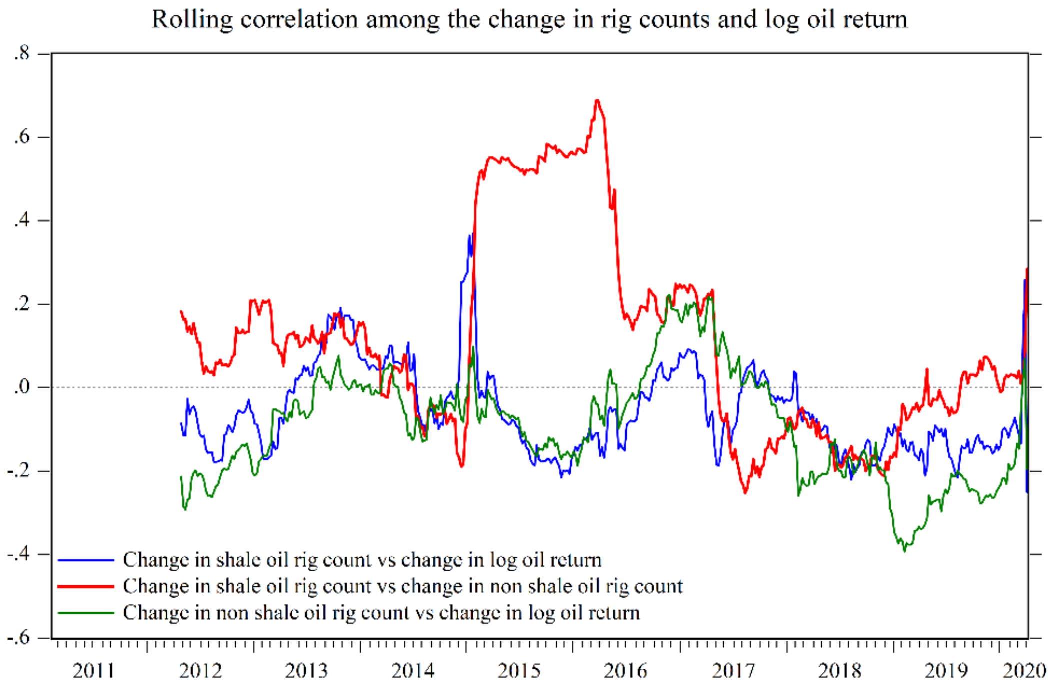

Figure 26: Rolling correlation among the change in rig counts and oil log-returns

To capture the relationship of the variables over time it is possible to compute the rolling correlation over a 64-week rolling window displayed in Figure 26. Focusing on the relationship among the rig count, the correlation has been positive during the first period in which the WTI price was on high level, to later experience a brief negative phase followed by a jump to a higher, stronger level during the 2014 price drop. In the recent years the correlations has been weakening, reaching a level close to zero prior the 2020 oil drop. Considering instead the relationship between the rig counts and the oil log-return, for both type has been varying and instable over time, with higher negative correlation of the non-shale oil with respect to the shale for all the considered period apart from the period which goes from 2015 to 2017. This result confirm what has been stated above, the shale oil is generally responding faster to the changes in oil price rather than the non-shale, even though the significance of the results is arguable. In fact, due to the nature of the business and the time required for the realization of the wells and, consequentally, the realization of the profit, it is rational to think that the relationship between the variables might require some time, i.e weeks. Thus, it is useful to analyse how these variables are related, considering the presence of lags. To shed some light on the lead-lag relationship of the considered variables, it is possible to use the cross-correlation. This tool allows to measure the strenght of linear dynamic dependence among stationary[21] time series, tracking their movements. In particular, those time series may be related to past lags, therefore the sample cross correlation function (CCF) is helpful for identifying lags of one variable that could be useful predictors of the other one. Assuming two time series (y_t and x_t), the sample *CCF* is defined as the set of sample correlations between x_{t+h} and y_t for $h = 0, \pm1, \pm2, \pm3$ and so on. Whenever h takes on negative value, it represents the correlation between the lagged value of x with respect to the y at time t. In this case, the value of x_{t+h} with negative h are said to be predictors of y_t, that is x *lead* y, provided that the value are significant. Conversely, whenever h takes on positive value, x *lags* y. The dotted line represents the 95% confidence interval for evaluation of white noise series. If the considered series are white noise, then the correlation coefficient are within the bounds.

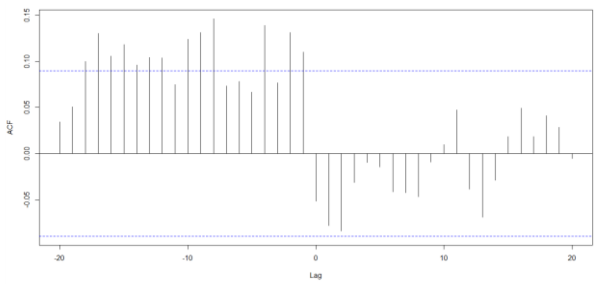

Figure 27: Cross-correlation among oil log-returns vs change in shale oil rig counts

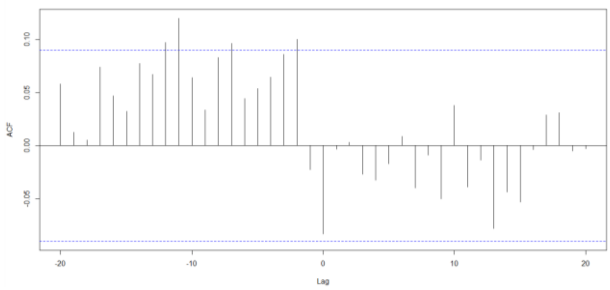

Figure 28: Cross-correlation among oil log-returns vs change in non-shale oil rig counts

By considering as x the oil log-return, it is possible to visualize the cross-correlation with the shale and non-shale oil rigs. On the left hand side of Figure 27 is shown the relationship between the oil log return and the change in shale oil rig count while on the right hand side is considered the change in non-shale oil rig count (Figure 28). The cross-correlation has been computed considering 20 lags. For the shale rig mostly of them are significant at 5% level, that is, the lagged value of oil returns are affecting the change in shale oil rig count, with a lagged effect that can last until 18 weeks. The non-shale rigs instead are characterized by having a weaker relationship, with significant value only at lag 1, 7, 11 and 12 (LHS of Figure 28). The coefficient are reported in the table 1, where the lags represents the weeks and the asterisks

[21] The concept of stationarity will be explained in the next chapter together with an extensive description of econometrics methodologies.

identify the 5% statistically significant coefficeints. Moreover it is interesting to notice that the contemporaneous cross-correlation (Lag at t_0) it is not significant for both series.

Lag	ΔShale rig vs lags of Oil returns	Oil returns vs lags of ΔShale rig	ΔNon-shale rig vs lags of Oil returns	Oil returns vs lags of ΔNon-shale rig
0	−0.051466	−0.051466	−0.082792	−0.082792
1	0.109915*	−0.078129	−0.022433	−0.003057
2	0.131113*	−0.084080	0.099924*	0.003355
3	0.076569	−0.031368	0.085681	−0.026550
4	0.138655*	−0.009409	0.064184	−0.032488
5	0.066768	−0.014404	0.053825	−0.017109
6	0.077920	−0.041586	0.044457	0.008955
7	0.073392	−0.042156	0.096127*	−0.039434
8	0.146216*	−0.046842	0.082564	−0.008933
9	0.131082*	−0.008891	0.033680	−0.049950
10	0.124117*	0.009631	0.063768	0.037861
11	0.074539	0.047104	0.119641*	−0.038887
12	0.103466*	−0.038720	0.096717*	−0.013606
13	0.104223*	−0.069075	0.066804	−0.077472
14	0.096105*	−0.029092	0.077228	−0.043563
15	0.118309*	0.018489	0.032437	−0.053057
16	0.105692*	0.049244	0.046944	−0.003746
17	0.130287*	0.018195	0.073958	0.029030
18	0.099597*	0.040944	0.005442	0.030980
19	0.050626	0.028299	0.012664	−0.004871
20	0.034466	−0.005417	0.057919	−0.002635

Table 1: Cross-correlations coefficients

This analysis confirms what the economic sense would lead to think.In fact, are the oil log-returns that affect the change in rig count, where the effect on the shale last longer than the non-shale[22].

In order to verify the existence of linear dynamic dependence in the data it is possible to use the *multivariate Portmanteau test* test statistic, a generalized model for multivariate series proposed by Hosking (1980), deriving from the *Portmanteau test* for univariate time series that allows to test the null hypothesis $H_0 : \boldsymbol{\rho_1} = \cdots = \boldsymbol{\rho_m} = 0$ (that is, testing for zero cross-correlations) against the alternative hypothesis $H_a : \boldsymbol{\rho_i} \neq 0$ for some *i* satisfying $1 \leq i \leq m$, with m being a positive integer. In particular, the multivariate Ljung-Box test statistic is defined as:

$$Q_k(m) = T^2 \sum_{\ell=1}^{m} \frac{1}{T-\ell} tr\left(\hat{\boldsymbol{\Gamma}}_\ell' \, \hat{\boldsymbol{\Gamma}}_0^{-1} \hat{\boldsymbol{\Gamma}}_\ell \hat{\boldsymbol{\Gamma}}_0^{-1}\right)$$

Where $tr(\mathbf{A})$ is the trace of the matrix $\mathbf{A}$ and T is the sample size. Under the null hypothesis that $\boldsymbol{\Gamma_\ell} = \mathbf{0}$ for $\ell > 0$ and the condition that the multivariate time series are normally distributed,

[22] The shale is characterized by having 13 coefficients which are statistically significant while the non-shale only 4.

$Q_k(m)$ is asymptotically distributed as $\chi^2_{mk^2}$, that is, as a chi square distribution with mk^2 degrees of freedom. If there exists a certain linear dynamic depedence among the variables, the test statistic should reject the null hypothesis of no cross-correlations. Running the test on the considered variables, the null hypothesis is rejected since all *p-values* are smaller than 0.05, confirming the presence of serial correlations.

Since all these preliminary tests have shown the presence of a linear dynamic relationship in the data, it is worth to further investigate this relationship and eventually evaluate whether this relationship is symmetric with respect to the oil shocks and how the rig counts respond to those shocks.

3.4. Further investigation on the linear relationship among the variables

As already stated several time, one of the objectives of this work is to identify the potential presence of asymmetric responses of the change in rig counts to oil shocks. A measure that could help to detect this presence is the exceedance correlation. This method, initially proposed by Longin and Solnik (2001), focus on the correlations across variables, conditioning on a specific section of their joint density. However, to evaluate the presence of asymmetry due to the oil shocks, the model can be adapted as following:

$$\rho^{+}_{\Delta WTI} = Corr(\Delta Y_t, \Delta WTI_t |\, \Delta WTI > 0)$$

$$\rho^{-}_{\Delta WTI} = Corr(\Delta Y_t, \Delta WTI_t | \Delta WTI < 0)$$

Where $Y_t = \begin{bmatrix} \Delta CSO_t \\ \Delta CNSO_t \end{bmatrix}$ represents the change in shale oil rig count and non-shale oil rig count at time t. Thus, instead of conditioning on a specific section of their joint density as proposed in the original paper, this measure is computed for a given threshold, that is 0.

The exceedance correlation $\rho^{-}_{\Delta WTI}$ measures the association between Y_t and ΔWTI_t whenever the oil shocks are negative, i.e. $\Delta WTI_t < 0$. On the other hand, $\rho^{+}_{\Delta WTI}$ monitors the correlation between Y_t and ΔWTI_t whenever the latter takes positive value, i.e the case of positive shocks. Even though it is considered still a contemporaneous relationship, the exceedance correlations enables to verify whether the impact of oil prices on changes in rig counts, depending on the sign of the oil shocks, is the same, i.e symmetric, or not. In the latter case, this might result in an asymmetric response.

Starting from the case in which the oil shocks are positive, it is already possible to notice the different coefficient for both shale and non-shale rig count. The negative contemporaneous correlation for positive shocks seems to be stronger, reinforcing the assumption of the negative relationship between the oil shocks and the rig counts. As the WTI increase, the number of rig count decreases, and vice versa. The correlation coefficient between rig counts instead is very close to the one computed for the usual correlation.

Exceedance correlations | Δ logWTI > 0	Δ log WTI	Δ Shale rig	Δ Non shale rig
Δ log WTI	1.00000		
Δ Shale rig	−0.25691	1.00000	
Δ Non shale rig	−0.19231	0.34352	1.00000

Figure 29: Exceedance correlations for positive oil shocks

By considering the exceedance correlations for those observations which are linked to a negative oil shock, it is possible to observe that the coefficients are different from the case of positive oil shocks.

Exceedance correlations | Δ logWTI < 0	Δ log WTI	Δ Shale rig	Δ Non shale rig
Δ log WTI	1.00000		
Δ Shale rig	0.16805	1.00000	
Δ Non shale rig	0.05628	0.35744	1.00000

Figure 30: Exceedance correlations for negative oil shocks

In particular, the negative oil shocks are correlated with a lower strength to the shale and non-shale rig count. The biggest difference is for the non-shale oil, which seems to be more "resilient" to negative shocks. Therefore, the results obtained by the exceedance correlations suggest that the relationship between the variables might not be simply linear. The presence of asymmetry questions the fixed relationship between changes in oil prices and rig counts, which could be due to binding contracts, rigor of responses as well as delays.

An interesting modification of the exceedance correlation is the one proposed by Caporin et al. (2018) and Khalifa et al. (2017). The latter considers the exceedance correlation, conditioning only on the change in rig count quantiles. Similar to this, the exceedance correlations conditioning the change in oil-log return quantiles is considered as the following:

$$\rho_{\tau}^{+} = Corr(\mathrm{Y_t}, \Delta WTI_t |\, F(\Delta WTI_t) > \tau), \tau \geq 0.5$$

$$\rho_{\tau}^{-} = Corr(\mathrm{Y_t}, \Delta WTI_t | F(\Delta WTI_t) < \tau), \tau \leq 0.5$$

Where $F(\Delta WTI_t)$ is the cumulative density function (CDF) of the oil-log returns and τ is the pre-specified quantile. The exceedance correlation ρ_{τ}^{-} measures the association between ΔY_t and ΔWTI_t when the oil-log return is located in its lower τ empirical quantile, with τ taking values below 0.5, that is below the median. On the other hand, ρ_τ^+ evaluates the correlation between the two variables when the oil-log return takes values above the median, for values of τ above 0.5. As in the paper proposed by Khalifa (2017), the exceedance correlations are computed for the contemporaneous case as well as for the cases where lags are considered.

Thus, it is evaluated the exceedance correlations:

$$\rho_\tau^+ = Corr(\Delta Y_t, \Delta WTI |\, F(\Delta WTI_t) > \tau\,)$$

$$\rho_\tau^- = Corr(\Delta Y_t, \Delta WTI | F(\Delta WTI_t) < \tau$$

for ΔWTI= $[\Delta WTI_t, \Delta WTI_{t-1}, \Delta WTI_{t-4}, \sum_{j=1}^{4} \Delta WTI_{t-j}]$ corresponding to the contemporaneous, lag 1, lag 4, and the sum of the 4 previous oil returns, i.e. monthly returns. The exceedance correlation analysis allows to verify whether the impact of oil prices on changes in rig counts, both shale and non-shale, depending on the changes in oil prices, being associated with a decreases in the rig counts (upper quantiles) or increases in rig counts (lower quantiles). The figure 31 shows that the exceedance correlations change when moving from the contemporaneous relationship to the case where the oil returns lags are considered. By considering the contemporaneous ρ_τ^-, the change in shale oil shows a homogenous behavior, while the non-shale seems to react with a weaker response to the change in oil returns. Considering instead the value above the median, both series show the same pattern, with a stronger response for the last quantiles. Evaluating instead the series considering the first lag, the results obtained differ. For the lower quantiles the shale oil shows a stronger response with respect to the lag 0 while the non-shale rigs show a weaker correlation considering its contemporaneous coefficient. The upper quantiles present an interesting result. Considering the last quantiles, both the shale and non-shale present a lower positive correlation with respect to their lag 0. Similar results are obtained when considering the 4^{th} lag. For the values below the median, the shale oil rigs are characterized by having a stronger relationships with respect to the previous lags analysed, while the non-shale show a weaker link over quantiles, decreasing faster than the shale oil. For the upper quantiles instead, both series present a correlation coefficient which is around the zero. To further investigate whether the use of lagged returns leads to an increase in the exceedance correlations, the sum of the previous 4 lags is considered, $\sum_{j=1}^{4} \Delta WTI_{t-j}$. The results confirm what it was expected, that is, in the lower quantiles the values for both series are bigger, i.e there is a stronger correlations (negative), with respect to the contemporaneous case. Different is the case for the upper quantiles, where the series present a weaker exceedance correlations than lag 0, apart from the 50^{th} and the 60^{th} percentile of the non-shale rigs[23].

In general, the use of lagged oil returns leads to an increase in the exceedance correlations. Moreover, it is possible to observer how the behaviour of the series differs considering the lower and upper quantiles. For the former, the relationships are positive and quite stable, increasing as the lags increase, apart from the 4^{th} lag of the non-shale rigs. In the latter case, where the quantiles are above the median, it is visible a more heterogenous behaviour across all the cases considered, with the exceedance correlations being closer to zero for the lagged returns, taking positive or negative value moving toward the upper tail. The marked difference between the exceedance correlations for the upper and lower quantiles, together with the previous analysis which considered positive and negative contemporaneous shocks, suggests that the relationship between the rig counts and the oil price returns is not simply linear, presenting evidences of asymmetric responses.

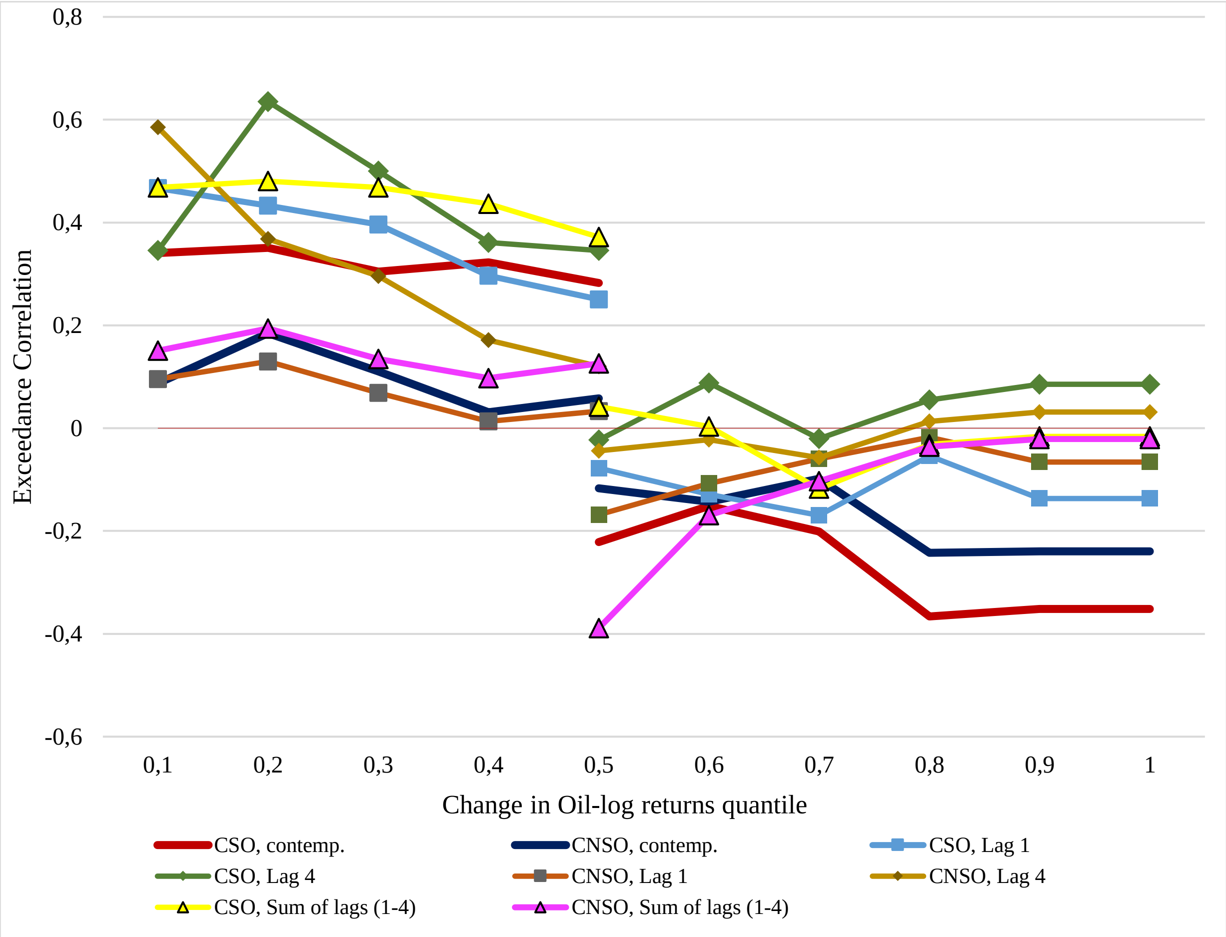

Figure 31: Exceedance Correlation, conditioning the change in oil-log return quantiles

[23] Moving toward the upper tail, after the 6^{th} decile, shale and non-shale rigs show a quite similar response to the change in oil returns.

The figure presented above has shown that, even considering the contemporaneous relationship of the variables, the correlation coefficients vary depeding on the sing of the oil shocks. Therefore, a complementary analysis of the rolling correlation of the full sample is introduced, in order to evaluate how the variables behave if we split the sample conditioning on the nature of the oil returns, positive or negative.

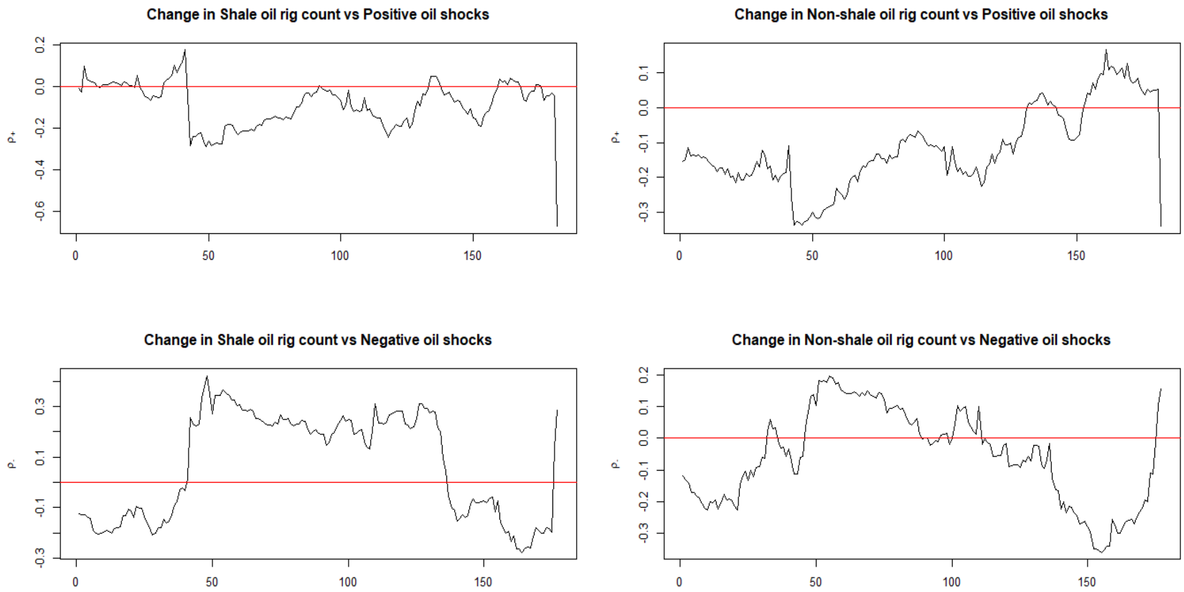

Figure 32: Rolling correlation, split by sign of oil shocks

The results, visible in the figure above are in line with what found so far. In case of positive shocks , the shale oil rig count is characterized by a negative stronger relationship with respect to the non-shale rig count, considering that almost all the coefficients are below zero. Considering instead the negative shocks, the positive relationship is more visible in the shale oil rigs rather than in the non-shale. However, the relationship is instable over time, considering also the low coefficient computed for the exceedance correlation in case of negative shocks, mostly for the non-shale case.

To shed some further light on both the asymmetric response and to the lead-lag relationship of the variables, a further analysis considering the cross-correlation is introduced. The first analysis is conducted considering the sign of the oil shocks, for both rig count. By considering as x the negative oil-log returns, both the shale e non-shale rigs show a positive and significant coefficients, for lagged value of the oil returns. Moreover, it is interesting to notice that the contemporaneous cross-correlation (Lag at t_0) it is not significant for both series. If instead the positive oil returns are considered, the cross-correlations of the lagged oil returns are not significant, that is x *does not lead* y. In this case, the contemporaneous correlation t_0 is significant for both shale and non-shale. In particular, for the former seems that *x lags y*, for two lags.

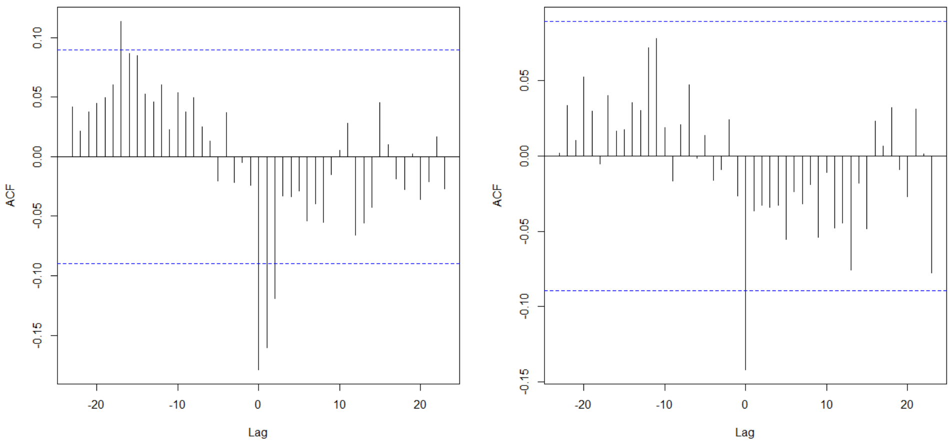

Figure 33: Cross-correlation for positive oil shocks, Shale oil (left) and Non-shale oil (right)

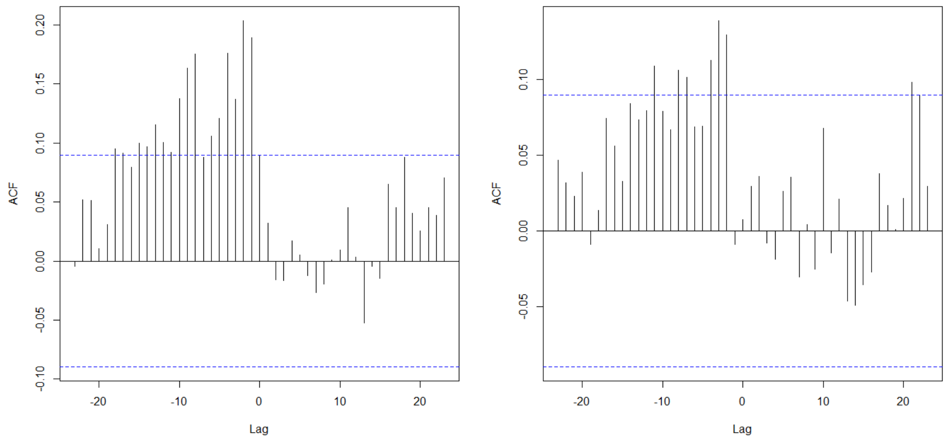

Figure 34: Cross-correlation for negative oil shocks, Shale oil (left) and Non-shale oil (right)

Since the interest is to investigate the impact of lags on the rig counts, it might be useful to introduce a cross-correlation analysis which considers the sum of the 4 previous oil returns ($\sum_{j=1}^{4} \Delta WTI_{t-j}$) and the sum of the 13 previous oil returns, i.e the quarter returns ($\sum_{j=1}^{13} \Delta WTI_{t-j}$).

In both cases, it is possible to observe significant values. Taking into consideration the sum of the 4 lags, it is shown a positive correlation, with a significant value also at lag 0, for the shale. The absence of a contemporaneous correlation for the non-shale might signal a delay in the response of this type of rig, which is in line with what observed so far. Considering the quarterly returns, both series show a stronger correlation, with higher value for the shale oil rigs.

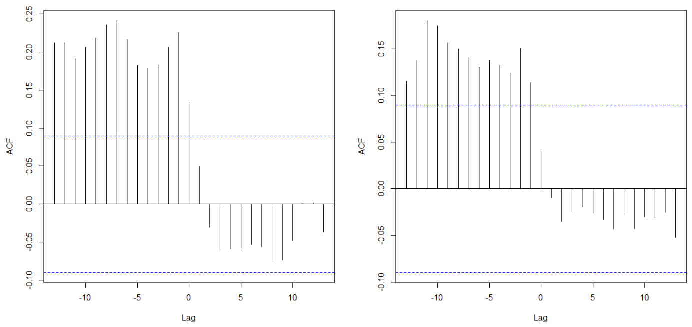

Figure 35: Cross-correlation for the sum of 4 previous oil returns, Shale rigs (left), Non-shale rigs (right)

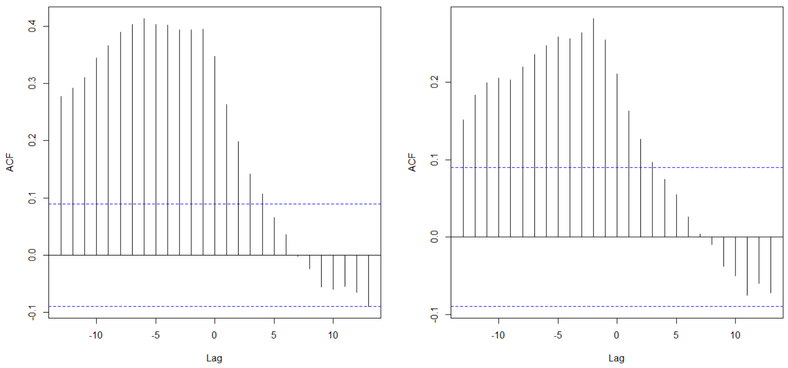

Figure 36: Cross-correlation for previous quarter of oil returns, Shale rigs (left), Non-shale rigs (right)

As pointed out by Pierce, variables may be functionally related and yet uncorrelated or, more often, they could be correlated yet not causally related. In fact, the former effect arises because correlation is a measure only of linear association while the latter arises because of common association of both with a third factor (Pierce 1977). Moreover, as shown in his paper, both the sample cross-correlations and the regression of one variable (let us say y), on the past, present and future x, or viceversa, might be misleading if the autocorrelation in the series is not properly taken into account.

Therefore, to detect the presence of causality in the variables, the approach proposed by Pierce (1977) will be applied in this work. As suggested in his paper, to detect and assess the relationships between variables, two steps are fundamental:

- Fitting univariate time-series models to each series of interest
- Assessing the causality patterns among variables

The causal relationships among variables, should be verified only after having explained y_t on the basis of its own past history $\overline{y_t}$, that is fitting univariate time-series. Then a variable will cause another, let's say x causes y, only if, after the fitting process, some more history remains to be explained by x_s, $s < t$, i.e. by $\overline{x_t}$. This means to relate x to the part of y_t which cannot be explained by its own past which correspond to the innovation in the univariate time series[24]. Therefore, it is required to *whiten* y_t (Pierce, 1977).

Once the series have been estimated, the residuals obtained are free from autocorrelation. This procedure removes the bias implicit in the sample cross-correlation, allowing to obtain a more reliable cross-correlation computed on those residuals, called *residual cross correlations* with the following formula:

$$\widehat{r_k} = r_{\hat{u}\hat{v}}\,(k) = \sum \hat{u}_{t-k}\,\hat{v}_t / \left[\sum \hat{u}_t^2 \sum \hat{v}_t^2\right]^{\frac{1}{2}} .$$

3.4.1. Estimation of ARMA models and residual cross-correlations

Series	***Estimated ARMA models***
$\boldsymbol{CSO\ (4,4)}$	$(1 - 0.79L + 0.4L^2 - 0.63L^3 + 0.15L^4)y_t = (1 - 0.58L + 0.51L^2 - 0.53L^3 + 0.24L^4)\varepsilon_t$
$\boldsymbol{CNSO\ (4,5)}$	$(1 - 1.32L + 0.83L^2 - 1.24L^3 + 0.79L^4)y_t = (1 - 1.24L + 0.84L^2 - 1.3L^3 + 0.84L^4)\varepsilon_t$
$\boldsymbol{COP\ (4,5)}$	$(1 - 1.06L + 0.87L^2 - 0.21L^3 + 0.42L^4)y_t = (1 - 0.84L + 0.47L^2 + 0.05L^3 + 0.35L^4 + 0.02L^5)\varepsilon_t$
$\boldsymbol{\sum_{i=1}^{4} COP_{t-j}\ (5,6)}$	$(1 - 1.1L - 0.58L^2 + 1.02L^3 - 014L^4 - 0.09L^5)y_t = (1 + 0.21L^1 - 061L^2 + 0.27L^3 - 0.71L^4 + 0.04L^5 + 0.87L^6)\varepsilon_t$
$\boldsymbol{\sum_{i=1}^{13} COP_{t-j}\ (2,13)}$	$(1 - 1.31L + 0.31L^2)y_t = (1 - 0.926L^{13})\varepsilon_t$

Table 2: Estimated ARMA models, minimizing the AIC criteria

All the series considered for the cross-correlation analysis have been modelled and estimated. The model for each univariate time series has been chosen by minimizing the AIC criteria[25]. Once the model is estimated, the residuals of each time series is taken into consideration for the cross-correlation analysis proposed by Pierce(1977). All the univariate time-series are characterized by having residuals which do not present autocorrelation.

---

[24] Assuming that a time series is invertible and stationary, an AR(p) is defined as: $y_t = \sum_{i=1}^{p} \Phi_i y_{t-1} + \nu_t$ which is equivalent to: $\phi(L)y_t = \nu_t$ where ν_t represent a white noise residual which can be called innovation.

[25] For further information about the criteria, see next chapter.

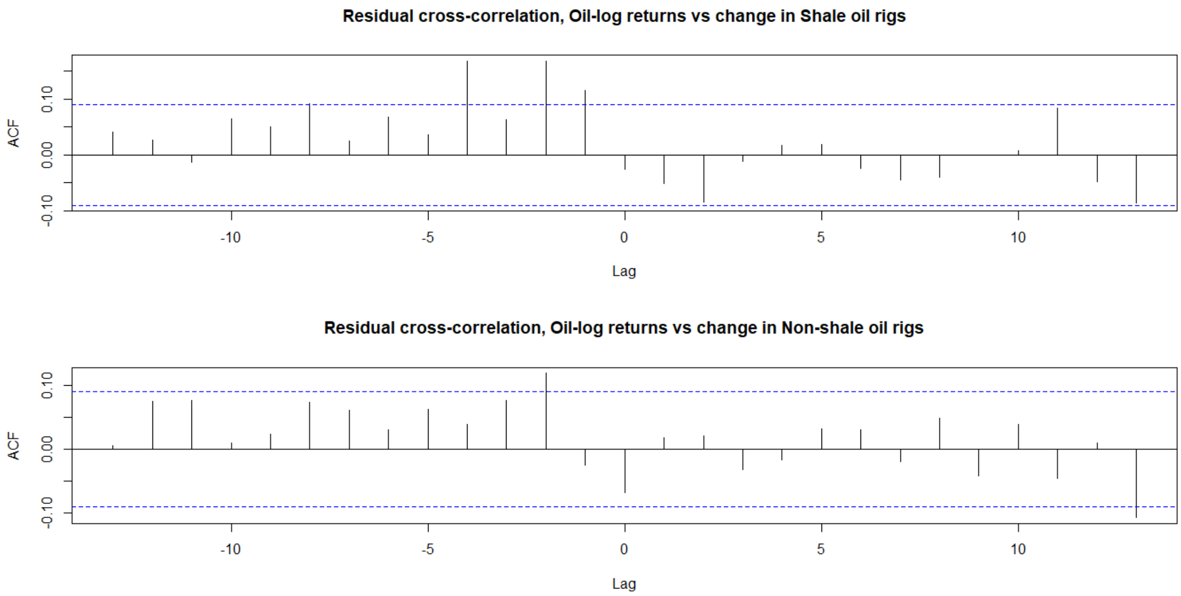

Figure 37: Residual cross-correlation, Oil returns against Shale rig count (top) and Non-shale rig count (bottom)

By looking at the figure 37, is immediate to notice the differences with respect to the figures 27-28 . The residual cross-correlation for the shale oil rig count highlight as significant only the $-2^{nd}, -3^{rd}, -4^{th}$ and -8^{th} lag while the cross correlation for non-shale rigs is found to be significant only for the second lag. Therefore, it is already possible to state the absence of instantaneous causality. Moreover, it is the oil price which causes changes in the rig counts, with a lagged effect, excluding the possibility of the other way around, confirming once again the results obtained so far.

This analysis can be extended, including the sum of the previous 4 weeks oil returns and the previous quarter (13 weeks).

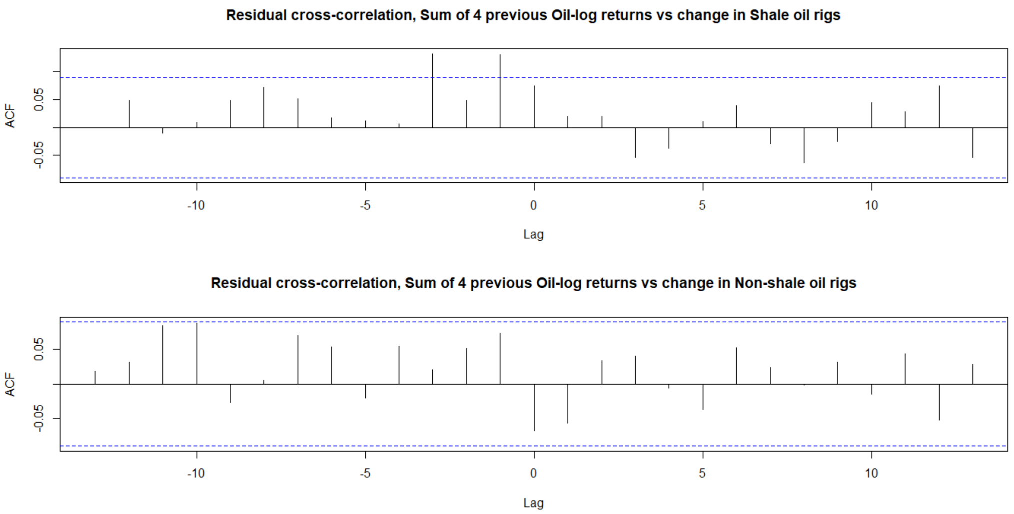

Figure 38: Residual cross-correlation considering the 4 previous oil returns

For the former case (figure 38), it is found that only the shale oil rig count shows a relationship with the oil returns (lag -1 and -3) while the shale oil rig count does not show significant coefficients.

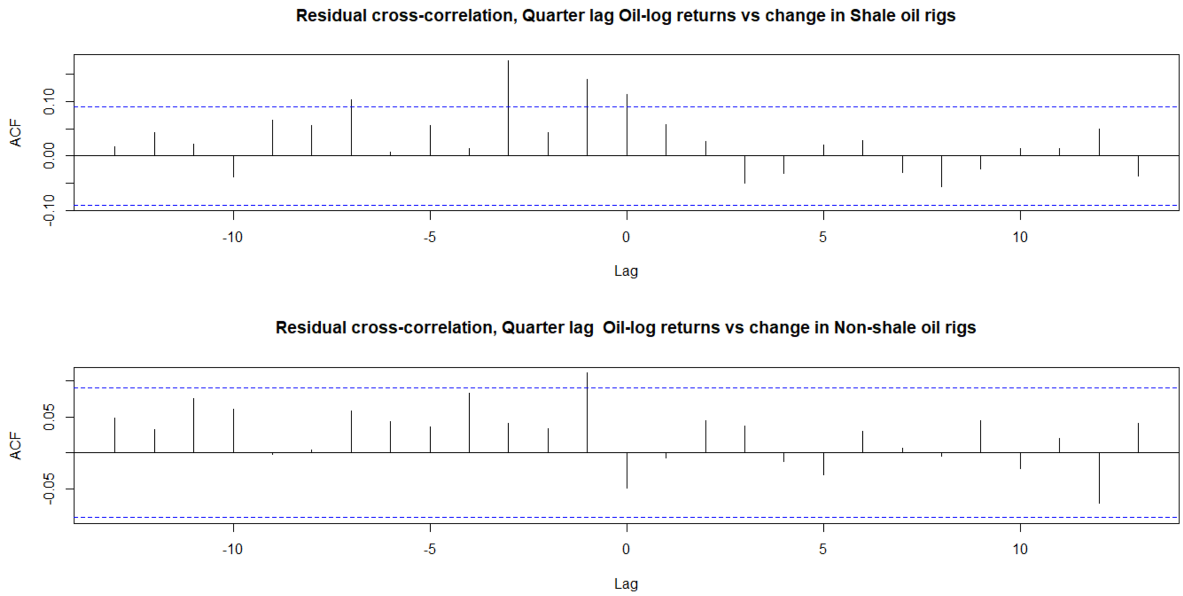

Figure 39: Residual cross-correlation considering the 13 previous oil returns

For the latter case, as showed in the figure 39, both the shale and non-shale rigs are characterized by being caused by the quarter oil returns. More in detail, the shale rigs show a stronger relationship for lag 0,-1 and -3 while the Non-shale only at lag -1 despite its very weak significance.

The overall analysis carried so far to further investigate the presence of a lagged relationship and the presence of asymmetry in the response has confirmed what the economic theory would suggest. So far, it has been observed that both the Shale rig count and the Non-shale rig count are caused by the oil-log returns and not the contrary. Moreover, this relationship is lagged, both the quarterly and the monthly oil returns could help understanding the change in the rig counts.

Another interesting result is the asymmetric response of the rig counts to change in oil returns. Whenever the oil returns are positive, i.e. positive shocks, both rig counts tend to decrease, with a higher impact on the shale oil rig count. Moreover, the asymmetry seems to be stronger for the negative shocks, and even more when considering lagged values. The results obtained so far suggest that the relationship between rig counts and oil returns is neither fixed nor simply linear. The reasons for these results might be few. Among them, the nature of the industry whose results are not instantaneously visible, as well as the presence of binding contracts and delay due to natural and economic reasons.

3.5. Potential economic and financial covariates variables

The last variables considered for this work are a set of economic and financial variables that might help explaining the relationship between the rig counts and the relationship which links them with the oil price. There are three type of variables which have been considered: Economic variables, indicators of financial stress and those related to the bond/credit market.

For the first category it has been identified the Trade-Weighted U.S. Dollar index (DTWEXBGS) and the S&P 500 index (GSPC). The former is known as the Broad index, created by the FED, measuring the value of the United States dollar, based on its competitiveness versus trading partners. This index is adopted to determine the purchasing value of the dollar as well as to summarize the effect of dollar appreciation and depreciation against foreign currencies, giving importance (weight) to currencies most widely adopted for international trades. It is useful for this analysis since it allows to monitor the effect of the reference currency (Dollar) used in pricing the oil relative to other currencies. The S&P 500 Index contains 500 of the largest stocks traded on the New York Stock Exchange and Nasdaq, and it is generally used to monitor the overall health of large American companies.

Moving to the indicators of financial stress, it is possible to identify the St. Louis Fed Financial Stress Index (STLFSI2), the CBOE Volatility Index (VIX) and the National Financial Conditions index (NFCI). The first one is an index that measure the degree of financial stress in markets, provided by the Federal Reserve Bank of St. Louis. The average value of the index is designed to be zero whenever the financial market conditions are normal while a value below zero represents a financial market stress below the average, and viceversa. The VIX instead, represents a market index representing the market's expectations for volatility over the next coming 30 days, providing a measure of market risk and investors' sentiments. The Chicago Fed's National Financial Conditions Index instead provides a comprehensive weekly update on U.S. financial conditions in money markets, debt and equity markets and the traditional and "shadow" banking systems. It is a weighted average of 105 indicators of risk, credit and leverage in the financial system. Whenever it takes on zero value, the U.S financial system operates at its historical average levels of risk, credit and leverage. The considered variables concerning the bond/credit market are the Effective Federal Funds rate (EFFR), the 10-year

Treasury Constant maturity rate (DGS10) and the Ted spread (TEDRATE). The EFFR consists of domestic unsecured borrowings in U.S dollars by depository institutions while the DGS10 represent the annual average for the interest rate on 10-year Treasury bonds. Finally, the TED spread is defined as the difference between the interest rate on interbank loans and on short-term U.S. government debt, the so-called T-bills[26]. The term "TED" is an acronym formed from *T-Bill* and *ED*, the ticker symbol for the Eurodollar futures contract (**T**reasury-**E**uro**D**ollar rate). The TED spread is a fundamental indicator for the perceived credit risk in the general economy since the T-bills are considered risk-free while the LIBOR reflects the credit risk of lending to commercial banks. An increase in the spread reflects higher perceived risk of default on interbank loans from lenders. All the data have been collected from the Federal Reserve Bank of St. Louis database, covering the period which goes from 4 February 2011 until 03 March 2020, with 497 observations.

Starting from a descriptive analysis of the variables, it is already possible to observe the presence of a linear dynamic relationship between the considered variables and the oil price. This relationship is lagged, that is, it is required some time to observe the impact of the oil log-returns on the rig counts, both shale and non-shale. Moreover, the shale rig counts are more affected than the non-shale, due to their business and natural cycle. The rig counts show higher positive correlation during oil drop, this is potentially due to the fact that they might react in the same way, or at least in a similar way, whenever the oil shocks are negative. This opens to interesting considerations that must be verified by mean of specific tools, eventually considering both positive and negative shocks to evaluate potential asymmetry in the reactions. Therefore, the next chapter will be dedicated to the exploration of this linear dynamic relationship, to assess what might drive the rig counts and at which extent.

[26] $TED\ spread = 3\ Month\ LIBOR\ rate - 3\ Month\ Treasury\ bill\ Interest\ rate$

4. Econometrics framework and Methodology

The analysis carried so far has showed the presence of a dynamic linear relationship between the considered variables requiring, therefore, a multivariate time series analysis. This branch of time-series considers not only the dependency coming from the past value of each time-series but also the inter-dependency among them to model and explain the interactions and co-movements among a group of time-series variables. One of the most used tools for this analysis is the Vector AutoRegression (VAR), a multivariate extension of the univariate autoregression.

This simple tool made popular by Sims (1980) provides a systematic way to capture rich dynamics in multiple time series and represent an alternative to the use of multivariate simultaneous equation model (Lütkepohl, 2011). In a VAR model, each variable is a linear function of the past values of itself and the past values of all the other variables. A simple scheme proposed by (Lütkepohl, 2005) illustrates the steps required in order to perform a structural analysis.

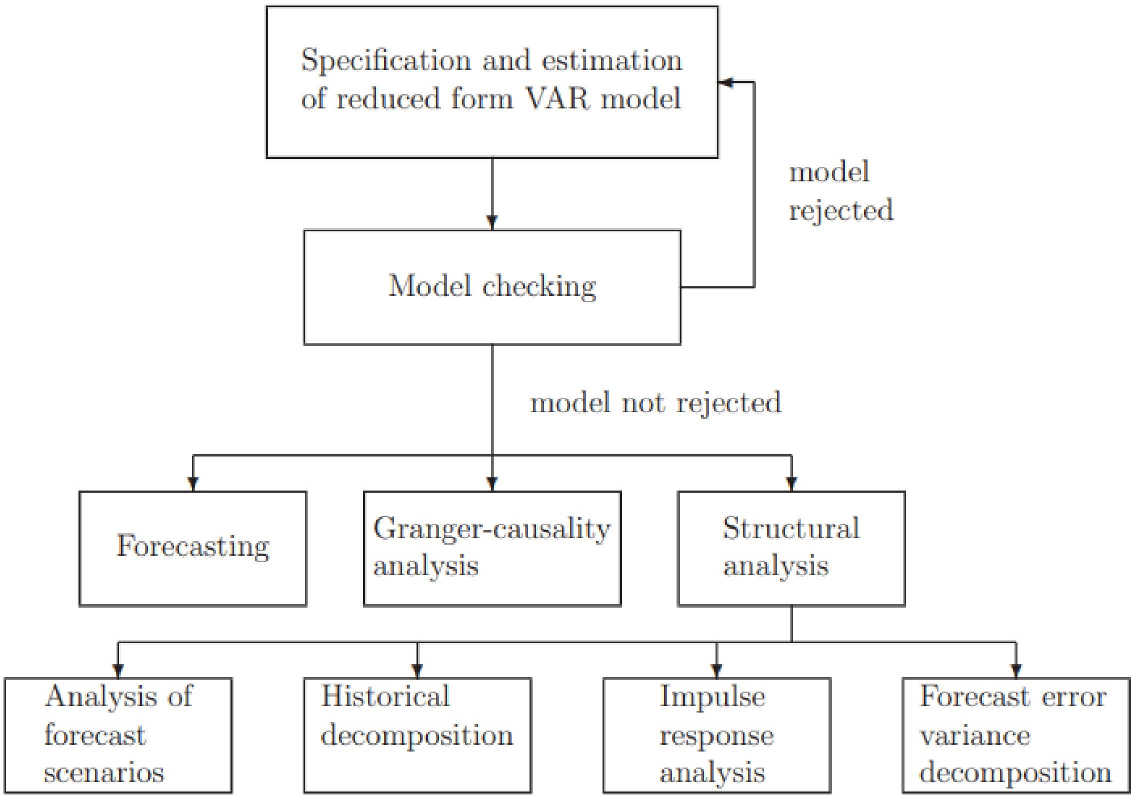

Figure 40: VAR analysis scheme, (Lütkepohl, 2011)

4.1. VAR(p) processes

Considering a generic VAR(p) model where p signals the order of the model[27], the equation of interest is:

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \qquad t = 0, \pm 1, \pm 2, \pm 3, \ldots, \qquad (4.1)$$

Where $y_t = (y_{1t}, \ldots, y_{Kt})'$ is a $(K \times 1)$ random vector, the A_i are fixed $(K \times K)$ coefficient matrices, $\nu = (\nu_1, \ldots, \nu_K)'$ is a fixed $(K \times 1)$ vector of intercept terms allowing for the possibility of a nonzero mean, $E(y_t) \neq 0$ while $u_t = (u_{1t}, \ldots, u_{Kt})'$ is a K-dimensional *white noise* or *innovation process*. The latter are characterized by a zero mean and a non-singular covariance matrix, that is, $E(u_t) = 0, E(u_t u_t') = \Sigma_u$ and $E(u_t u_s') = 0$ for $s \neq t$.

Using the lag operator notation, the equation (4.1) can be written as:

$$y_t = \nu + \boldsymbol{A} L y_t + u_t$$

$$(\boldsymbol{I} - \boldsymbol{A}L) y_t = \nu + u_t$$

$$\boldsymbol{A}(L) y_t = \nu + u_t \qquad \text{where } \boldsymbol{A}(L) = \boldsymbol{I} - \boldsymbol{A}L \rightarrow I_k - A_1 z - \cdots - A_p z^p \qquad (4.2)$$

An interesting feature of the Vector AutoRegression is that any VAR(p) process can be written in VAR(1) form. Assuming that y_t is a VAR(p) as in the (4.1), a corresponding Kp-dimensional VAR(1) is given by:

$$Y_t = \boldsymbol{\nu} + \boldsymbol{A} Y_{t-1} + U_t \qquad (4.3)$$

Where:

$$Y_t := \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad \boldsymbol{\nu} := \begin{bmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{A} := \begin{bmatrix} A_1 & A_2 & \ldots & A_{p-1} & A_p \\ I_K & \cdots & \cdots & 0 & 0 \\ 0 & I_K & \vdots & 0 & 0 \\ \vdots & \cdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{bmatrix}, \quad U_t := \begin{bmatrix} u_t \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The process Y_t is stable if:

$\det(I_{Kp} - \boldsymbol{A}z) \neq 0$ for $|z| \leq 1$, where $\det(I_{Kp} - \boldsymbol{A}z) = \det(I_K - A_1 z - \cdots - A_p z^p)$[28] is the *stability condition*. Therefore, the process Y_t is stable if its reverse characteristic polynomial has no roots in and on the complex unit circle. On the assumption that the process has been

[27] The VAR order is typically chosen by considering several model selection criteria.
[28] Generally, this polynomial is referred as the *reverse characteristic polynomial* of the VAR(p) process.

initiated in the infinite past ($t = 0, \pm1, \pm2, \ldots$), it generates stationary time series. Its mean vector is then given by:

$$\boldsymbol{\mu} := E(Y_t) = \left(I_{Kp} - \boldsymbol{A}\right)^{-1}\boldsymbol{\nu}$$

and the autocovariances are:

$$\Gamma_Y(h) = \sum_{i=0}^{\infty} \boldsymbol{A}^{h+i}\,\Sigma_{\mathrm{U}}\left(\boldsymbol{A}^{i}\right)', \text{ where } \Sigma_U := E(U_t U_t').$$

A fundamental proposition related to the stability condition is the *Stationarity Condition*. A stable VAR(p) process $y_t, t = 0, \pm1, \pm2, \ldots$, is *stationary*. Since stability implies stationarity, the stability condition is referred to as stationarity condition.

The stationarity represents a fundamental characteristic of stochastic processes[29]. The literature provides two definitions of stationarity, *weak stationarity* and *strong stationarity*. For the definition of the former, the first two marginal moments are important. In fact, a stochastic process is weakly stationary (or covariance stationary) if:

- $E(y_t) = \mu, \ \forall t$
- $E[(y_t - \mu)(y_{t-h} - \mu)'] = \Gamma_y(h) = \Gamma_y(-h)', \forall t$ and $\forall h = 0,1,2, \ldots$

The first condition requires that the mean vector is time invariant while the latter requires that the autocovariances of the process do not depend on *t* but just on the time period *h* the two vector y_t and y_{t-h} are apart[30]. Therefore, the weak stationarity only requires the shift-invariance (in time) of the first moment and the cross moment (the auto-covariance).

4.2. Stationarity

The concept of stationarity is fundamental in time-series analysis since this feature allows for a more meaningful interpretation of the results yielded by the model adopted (in this case, the Vector AutoRegression requires that all time series must be stationary). Causes of violation of stationarity is the presence of a trend in the mean, which can be due either to the presence of a unit root or of a deterministic trend. For the latter case, the process is called a *trend-stationary process*, if an underlying trend (only function of time) can be removed, leading to a stationary

[29] A common approach in the analysis of time series data is to consider the observed time series as part of a realization of a stochastic process (see Appendix A).
[30] The $\Gamma_y(h)$ is a matrix of finite covariances.

process[31]. Different is the violation of the stationarity due to the presence of a unit root. Recalling the characteristic equation of the (4.2), if its roots do not satisfy the stability condition, then the stochastic process is said to be a *difference stationary* process, or *Integrated.* This means that the process can be transformed into a weakly-stationary process by applying the so-called differencing transformation. These processes are characterized by having an order of integration. It represents the number of times the differencing operator must be applied to obtain a weak stationary process. A process that has to be differenced *r* times is said to be integrated of order *r*, with the notation I(*r*). Whenever the process is integrated of order 1, it is called Unit Root process. This kind of non-stationary processes can be easily transformed into weakly stationary process by taking the first difference. The first difference of a time series (generally denoted as Δ), is the series of changes from one period to the next. If y_t denoted the values of the time series y at time t, then the first difference of y at time t is equal to $y_t - y_{t-1} = \Delta y_t$[32].

The most basic methods for detecting the presence of stationarity is the data visualization, that is, determining visually whether they present some known property of stationary or non-stationary data. Several characteristics can be observed by plotting the data as seasonality, presence of a trend or level change or increase variance. A more sophisticated tool available to detect the presence of non-stationarity is the autocorrelation function (ACF) plot that allows to visualize the value of the ACF for increasing lags. Generally, time series which are non-stationary presents a slowly degradation toward zero of the ACF values over time. Conversely, a stationary time series is characterized by having values that tend to degrade to zero very quickly. However, a more rigorous approach for the detection of stationarity in time series is the use of parametric test developed to detect specific types of stationarity.

The first useful test is the augmented Dickey-Fuller test (ADF), that tests the null hypothesis that a unit root is present in the time series. The alternative hypothesis is generally either the stationarity or trend stationarity. The ADF represents an augmented version of the Dickey-Fuller test and consider a larger and more complicated set of time series models, adding lagged differences. The testing procedure of the ADF test is the same as the original Dickey-Fuller test, but applied to the model:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

[31] In the presence of a shock, trend-stationary processes are mean-reverting, that is, the series will converge again towards its mean.

[32] For further explanation, see appendix B.

Where α represent the constant, β the time trend coefficient and p are the lags of the autoregressive process. There are three main version of the test coming from the imposition of constraint on α and β:

- No constant, No trend: $\Delta y_t = \gamma y_{t-1} + \sum_{p=1}^{m} \delta_p \Delta y_{t-p} + \varepsilon_t;$
- Constant, No trend: $\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{p=1}^{m} \delta_p \Delta y_{t-p} + \varepsilon_t;$
- Constant and Trend: $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{p=1}^{m} \delta_p \Delta y_{t-p} + \varepsilon_t;$

The unit root test is then carried out under the null hypothesis that $\gamma = 0$ against the alternative of $\gamma < 0$, that is:

$$H_0: \gamma = 0 \qquad \text{vs} \qquad H_1: \gamma < 0$$

The value for the test statistic $\boldsymbol{DF}_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$ is computed and compared with the relevant critical value for the Dickey-Fuller test. If the calculated test statistic is more negative than the critical value, then the null hypothesis is rejected.

The *rationale* behind the test is that if the series is a unit process, then the lagged level of the series y_{t-1} will not provide any relevant information in predicting the change in y_t apart from the one obtained from the lagged changes and, therefore, the null hypothesis is not rejected. If instead, the process is stationary, it shows reversion to the mean thus the lagged level provides relevant information in predicting the change of the series leading to the rejection of the null hypothesis. However, as pointed out by Perman and Byrne (2006), in the practice the test tends to have low power and the rejection of the null does not always mean that there are evidences of non-stationarity.

The econometrics literature provides a “complementary” test for the Dickey-Fuller test, that is the Phillips-Perron test (1987) that tests for unit root. It addresses the problem that the DGP for y_t might have a higher order of autocorrelation than is admitted in the test equation, making y_{t-1} endogenous and therefore invalidating the Dickey-Fuller t-test. To overcome this issue, the Phillips-Perron test makes a non-parametric correction to conduct for autocorrelations and heteroscedasticity, i.e. it does not require to select the level of serial correlation as in the ADF test.

All tests use the model:

$$y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t$$

The null hypothesis restricts $\gamma = 1$, applying the same restrictions as the ADF test.

Another test that must be mentioned that allows to verify the presence of a unit root is the KPSS (Kwiatkowski–Phillips–Schmidt–Shin) test. However, conversely to the ADF test, the null hypothesis assumes stationarity around a mean or a linear trend, while the alternative hypothesis assumes a unit root process. The test is based on linear regression, divided in three parts: a random walk (r_t), a deterministic trend (βt) and a stationary error (ε_t), with the regression equation:

$$x_t = r_t + \beta t + \varepsilon_t \qquad \text{where } r_t = r_{t-1} + u_t \text{ with iid } \boldsymbol{u} \sim (0, \sigma^2)$$

The null hypothesis is:

$$H_0: \sigma^2 = 0 \text{ vs } H_a: \sigma^2 > 0$$

The restriction on β allow to test for the null hypothesis to be around a mean or a trend. The analysis will focus on the first two test since they are the most common tools adopted in the econometrics literature.

Following the methodology seen so far, a first insight on the stationarity of the considered time series can be obtained by considering Figure 23. In the upper part where are displayed the variables in levels, there are noticeable trends and changing in level. All of them show a drop around 2014, during the oil glut as already analysed in the previous chapter. Therefore, according also with the standard economic theory, the financial data seem to be integrated of order one. Moreover, by looking at the bottom part of the figure is it possible to observe that, by taking the first difference of the time series, they look stationary around their mean.

By considering instead the autocorrelation function, the variables expressed in level present a visible feature of non-stationary time series since they decay toward zero very slowly. By taking the first difference of the time series, they should present a stationary behaviour. However, as possible to observe in figure 41, the oil-log return is the only once which is characterized by a not significant autocorrelation value. Worth to notice the different behaviour of the rig counts. The change in shale oil rig count shows autocorrelation values which are significant over several lags while the non-shale oil rig count shows a weaker significance over time.

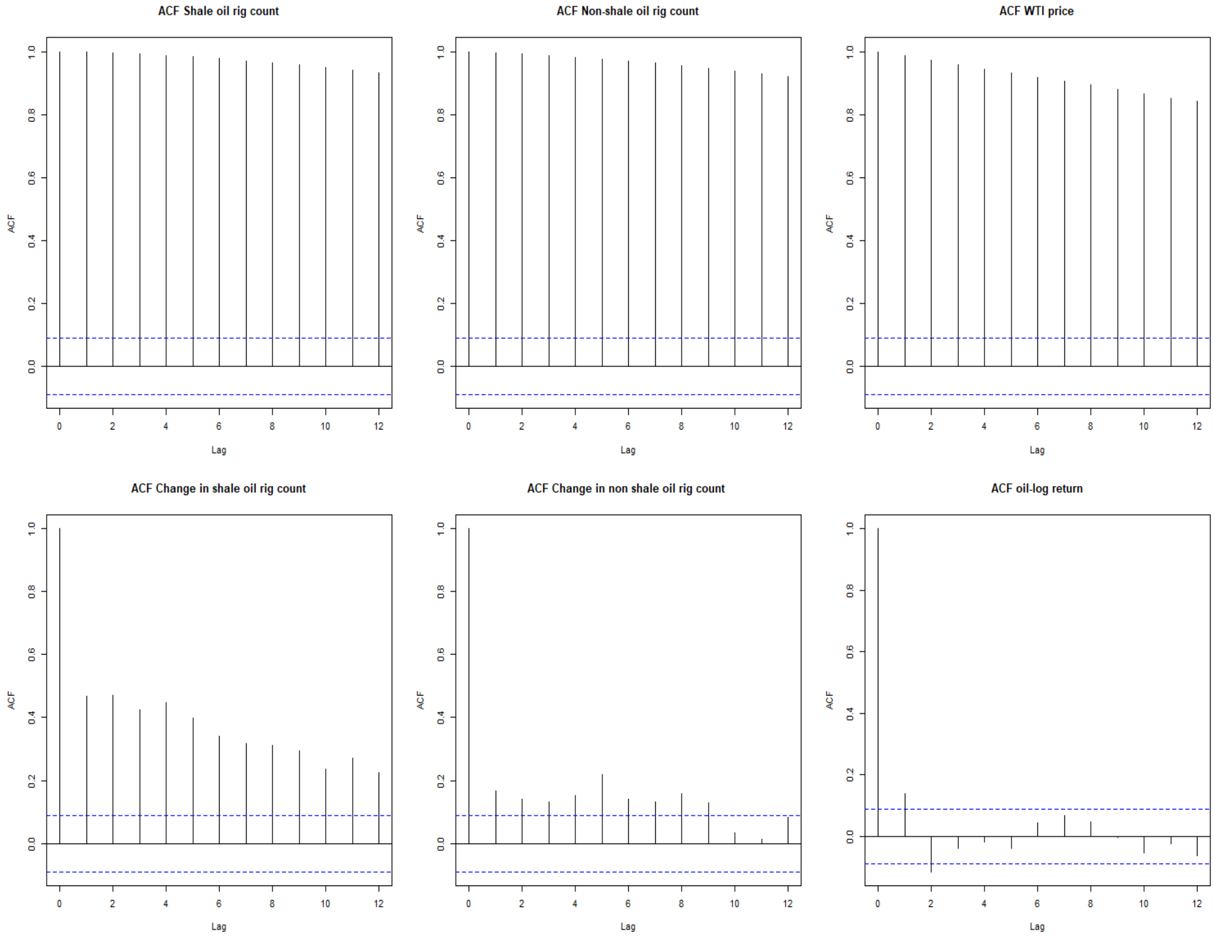

Figure 41: Autocorrelation function for the considered variables, in levels (top) and first difference (bottom)

However, a more rigorous approach is required to establish whether the series are characterized by a unit root or not. Starting the analysis from the Phillips-Perron Unit root test, it is possible to observe that the results are in line with what has been assumed so far. The tests performed on the levels, considering all three methods proposed, show that the series are characterized by a unit root, not rejecting the null hypothesis. Considering instead their first difference, the series strongly reject the null hypothesis of unit root process in all considered cases.

Variables	No Trend & No Intercept		
	Adjusted Test statistics	C.V at 5%	P-value
WTI price	-1.13598	-1.94149	0.2332
Shale oil rig	-0.57307	-1.94149	0.4688
Non shale oil rig	-0.75473	-1.94149	0.3591

Variables	Trend & Intercept		
	Adjusted Test statistics	C.V at 5%	P-value
WTI price	-2.20945	-3.41925	0.4827
Shale oil rig	-1.89649	-3.41925	0.6548
Non shale oil rig	-2.20127	-3.41925	0.4873

Variables	Intercept only		
	Adjusted Test statistics	C.V at 5%	P-value
WTI price	-1.03422	-2.86737	0.7422
Shale oil rig	-1.04949	-2.86737	0.7366
Non shale oil rig	-0.84233	-2.86737	0.8056

[1] The tables reports the critical values of the Phillips-Perron Unit Root test. The tests have been performed considering three different cases: trend and intercept, intercept only and No intercept & No trend. The null hypothesis of the test is the presence of a unit root.

Variables	No Trend & No Intercept		
	Adjusted Test statistics	C.V at 5%	P-value
ΔWTI price	-14.36018*	-1.94149	0.000
ΔShale oil rig	-16.68088*	-1.94149	0.000
ΔNon shale oil rig	-20.75750*	-1.94149	0.000

Variables	Trend & Intercept		
	Adjusted Test statistics	C.V at 5%	P-value
ΔWTI price	-14.24955*	-3.41927	0.000
ΔShale oil rig	-16.82629*	-3.41927	0.000
ΔNon shale oil rig	-20.75347*	-3.41927	0.000

Variables	Intercept only		
	Adjusted Test statistics	C.V at 5%	P-value
ΔWTI price	-14.29317*	-2.86737	0.000
ΔShale oil rig	-16.67848*	-2.86738	0.000
ΔNon shale oil rig	-20.74625*	-2.86738	0.000

[1] The tables reports the critical values of the Phillips-Perron Unit Root test. The tests have been performed considering three different cases: trend and intercept, intercept only and No intercept & No trend. The asterisks (*) denotes the 1% level of significance. The null hypothesis of the test is the presence of a unit root.

Figure 42-43: Phillips-Perron Unit root test results for variables in levels (LHS) and in first difference (RHS)

These results are confirmed by the Augmented Dickey-Fuller test, showed below. The variables on the left-hand side of the figure are in levels while on the right-hand side there are their first difference. Clearly, in all the different models proposed by the test, the time series show the presence of a unit root. The tests never reject the null hypothesis, confirming what has been observed in the autocorrelation plots. By taking the first difference, the time series reject the null hypothesis in all cases, showing the absence of a unit root.

Variables	No Trend & No Intercept		
	Test statistics	C.V at 5%	P-value
WTI price	-1.05176 (1)	-1.94149	0.2624
Shale oil rig	-0.86741 (4)	-1.94149	0.3399
Non shale oil rig	-0.82326 (5)	-1.94149	0.3590

Variables	Trend & Intercept		
	Test statistics	C.V at 5%	P-value
WTI price	-2.31689 (1)	-3.41927	0.4234
Shale oil rig	-2.30097 (4)	-3.41933	0.4321
Non shale oil rig	-2.46206 (5)	-3.41935	0.3470

Variables	Intercept only		
	Test statistics	C.V at 5%	P-value
WTI price	-1.09864 (1)	-2.86738	0.7177
Shale oil rig	-1.24931 (4)	-2.86749	0.6542
Non shale oil rig	-0.97038 (5)	-2.86743	0.7649

[1] The tables reports the critical values of the Augmented Dickey-Fuller test. The tests have been performed considering three different cases: trend and intercept, intercept only and No intercept & No trend. In parentheses are reported the optimal lag length automatically used, based on SIC. The null hypothesis of the test is the presence of a unit root.

Variables	No Trend & No Intercept		
	Test statistics	C.V at 5%	P-value
Δ WTI price	-15.4020 (1)*	-1.94149	0.000
Δ Shale oil rig	-4.20094 (3)*	-1.94149	0.000
Δ Non shale oil rig	-6.14876 (4)*	-1.94149	0.000

Variables	Trend & Intercept		
	Test statistics	C.V at 5%	P-value
ΔWTI price	-15.4054 (1)*	-3.41929	0.000
ΔShale oil rig	-4.32183 (3)*	-3.41933	0.003
ΔNon shale oil rig	-6.26129 (4)*	-3.41935	0.000

Variables	Intercept only		
	Test statistics	C.V at 5%	P-value
ΔWTI price	-15.4133 (1)*	-2.86739	0.000
ΔShale oil rig	-4.20415 (3)*	-2.86742	0.001
ΔNon shale oil rig	-6.15688 (4)*	-2.86743	0.000

[1] The tables reports the critical values of the Augmented Dickey-Fuller test. The tests have been performed on the first difference, considering three different cases: trend and intercept, intercept only and No intercept & No trend. In parentheses are reported the optimal lag length automatically used, based on SIC.The asterisks (*) denotes the 1% level of significance. The null hypothesis of the test is the presence of a unit root.

Figure 44-45: ADF test results, levels on RHS and first difference on LHS

However, as showed later by Perron (1989), not considering the potential presence of existing break while performing the ADF test, the test could lead to biased results, reducing the ability to reject a false unit root null hypothesis. A structural break can be defined as an unexpected (abrupt) change at a point in time. The change could involve a change in mean or in other parameters of the process. A structural break can reflect institutional, legislative, or technical change. It can also be generated due to economic policies or large economic shocks. Therefore, several tests have been constructed in order to deal with the presence of structural break. Since the structural breaks can have a permanent impact on the time series pattern, testing the unit root hypothesis considering the presence of structural break has several advantages. First and foremost, it prevents the results from becoming biased toward unit root. Moreover, it allows to identify the break date, yielding useful insight for the interpretation of the results. Perron (1989) provides a test which allows to test for unit root while considering the presence of structural break in both the null and the alternative hypothesis. This type of test considers the break date as exogenous. Starting from the Perron's unit root test, Zivot and Andrews (1992) suggest a sequential test in which the break date is not exogenous, i.e. fixed by the authors, but estimated within the test (endogenous structural break). The test allows for a single break in the intercept and the trend of the time series. The breakpoint is estimated by considering a series of dummy variables for each possible break date and it is chosen the one which support the most the alternative hypothesis, the most significant test statistics, that is the break date is selected where there is the strongest evidence against the null hypothesis from the ADF test. Moreover, it does not allow for a break under the null hypothesis. Waheed et al. (2006) explicitly show in their paper the models proposed by Zivot and Andrews:

- A model which allows for a one-time change in the level of the series:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \gamma DU_t + \sum_{p=1}^{m} \delta_p \Delta y_{t-p} + \varepsilon_t;$$

- A model which allows for a one-time change in the slope of the trend function:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \theta DT_t + \sum_{p=1}^{m} \delta_p \Delta y_{t-p} + \varepsilon_t;$$

- A model which combines the previous model, that is, a one-time change in the level and the slope of the trend function:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \gamma DU_t + \theta DT_t + \sum_{p=1}^{m} \delta_p \Delta y_{t-p} + \varepsilon_t;$$

Considering what said so far, by looking at the time series showed in figure 23, it is clearly visible a structural break which coincide with the oil glut of 2014, which can be considered as an abrupt economic shock. Therefore, to verify the hypothesis of unit root process while accounting for potential structural break, it is computed the Zivot and Andrews test. The results are showed below, from which is possible to obtain useful information.

	Zivot-Andrews Unit Root test				
	Break date	Lags	Test statistics	C.V at 5%	P-value
WTI price	10/03/2014	1	-4.97543*	-4.93	0.000
Shale oil rig	12/12/2014	5	-5.47475*	-4.93	0.000
Non shale oil rig	12/26/2014	8	-8.51139*	-5.08	0.000

[1] The tables reports the critical values of the Zivot-Andrews unit root test (1992) in the presence of a structural break in a series with intercept and/or linear trend. The results reported considered the smallest t-statistics with the related break date. The asterisks (*) denotes the 5% level of significance.

Figure 46: Zivot-Andrews Unit Root test results

The first thing to notice is that all time series reject the null hypothesis of unit root at 5% level of confidence. The test shows the weakness of the ADF test in the presence of a structural break and its bias toward the unit root. Another interesting finding is the break date that are estimated to be during the 2014 oil glut. In particular the oil price shows a break date at the beginning of October, followed after 10 weeks by the shale oil and 12 weeks by the non-shale oil rig counts, as observable in figure 23. The break dates are in line with what predicted from the economic theory, that is, the rig counts present a lagged response with regards of the oil price.

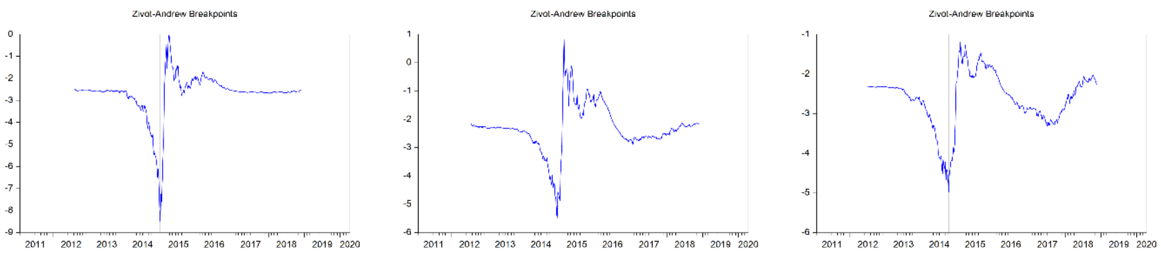

Figure 47: Zivot-Andrews test for the Non shale rig count (left), shale rig count (middle), and oil price (right)

To verify whether the results are reliable, The Perron (1989) Breakpoint Unit Root is performed and the results are shown below. For all the three variables, the test rejects the null hypothesis of unit root. The break dates are slightly different from the dates reported by Zivot-Andrews test, identifying the structural break two weeks before the WTI oil price and the shale oil rig count, and one week before for the non-shale oil rig count. However, this difference could be due to the different methodology used for the determination of the structural break, since the Perron test (1989) specified an *a priori* fixed break date while the Zivot-Andrews test (1992) endogenously determines break dates from the data.

Perron Breakpoint Unit Root test					
	Break dates	Lags	Test statistics	C.V at 5%	P-value
WTI price	9/26/2014	7	-5.2254*	-5.1757	0.04
Shale oil rig	11/28/2014	4	-5.4921*	-4.8598	0.01
Non shale oil rig	12/19/2014	8	-8.4890*	-5.1757	0.01

[1] The tables reports the critical values of the Perron (1989) unit root test in the presence of a structural break in a series with intercept and/or linear trend. The test has been performed considering three different cases: break in both trend and intercept, break in intercept only and break in trend. In the table are reported only the results which yielded the lowest t-stats, the optimal break date for each series and the optimal number of lags used for the computation of the test statistic, following the F-statistic selection. The asterisks (*) denotes the 5% level of confidence. The null hypothesis of the test is the presence of a unit root.

Figure 48: Perron Unit Root test, allowing for structural breaks

4.3. Specification and Estimation of VAR(p) model

4.3.1. Model selection criteria (Optimal Lag Order Selection)

Once the stationarity of the time series has been proved, it is necessary to determine the best p^{th} order for the VAR. The literature provides several model selection criteria which are used to choose the VAR order that minimizes them over a set of possible orders $m = 0, \dots, p_{max}$

The general form for the Optimal Lag Order Selection is:

$$\mathrm{C(m)} = \log\det(\widehat{\sum}_{\mathrm{m}}) + \mathrm{c_T}\varphi(\mathrm{m})$$

Where $\widehat{\sum}_m = T^{-1}\sum_{t=1}^{T}\hat{u}_t\,\hat{u}_t'$ is the OLS residual covariance matrix estimator for a reduced form VAR model of order m, $\varphi(m)$ is a function of the order m which penalized large VAR orders and c_T is a sequence which may depend on the sample size and identifies the specific criterion (Lütkepohl, 2011). The lag order is chosen by balancing these two terms.

Among the most popular criterion proposed by the literature are:

Akaike's information criterion (Akaike 1973,1974)

$$AIC(m) = log\,det\,(\widehat{\sum}_m) + \frac{2}{T}mK^2;$$

The Hannan-Quinn criterion (Hannan and Quinn, 1979)

$$HQ(m) = \log det(\widehat{\Sigma}_m) + \frac{2loglogT}{T} mK^2;$$

And the Schwarz criterion (Schwarz, 1978)

$$SC(m) = \log \det\left(\hat{\Sigma}_m\right) + \frac{logT}{T} mK^2;$$

In all these criteria $\varphi(m) = mK^2$ is the number of VAR parameters of order m. Moreover, considering small samples, AIC has better properties that HQ and SC. Lütkepohl (2005) shows that AIC always suggest the largest order, HQ is in the middle while SC chooses the smallest order. If the HQ and the SC criteria are both consistent, the order estimated converges in probability to the true VAR order p.

4.3.2 Estimation of VAR(p) model

Rewriting the VAR(p) model of (4.1) in a more compact form we obtain:

$$y_t = [v_0, A_1, \dots, A_p] Z_{t-1} + u_t$$

Where $Z_{t-1} = (1, y'_{t-1}, \dots, y'_{t-p})$ and given a sample of size T, the parameters can be estimated efficiently by ordinary least square (OLS) for each equation separately.

The values are then obtained by:

$$[\hat{v}_0, \hat{A}_1, \dots, \hat{A}_p] = \left(\sum_{t=1}^{T} y_t Z'_{t-1} \right) \left(\sum_{t=1}^{T} Z_t Z'_{t-1} \right)^{-1}$$

The estimator is identical to the generalized least squares (GLS) estimator if there are no restrictions imposed on the parameters. Whenever the process y_t is a normally distributed process with $u_t \sim \mathcal{N}(0, \Sigma_u)$, then the estimator is identical to the ML estimator, conditional on the initial presample values. However, whenever it is required to impose restriction on the parameters, OLS estimation may be inefficient and therefore GLS estimation are preferred.

Let $\boldsymbol{\alpha} = vec[v_1, A_1, \dots, A_p]$ and suppose that the restrictions are linear and aim to exclude some of the lagged variables from some of the equations, then those linear restrictions can be written as:

$$\boldsymbol{\alpha} = R\gamma$$

Where R is a suitable and known restriction matrix with rank M $((K^2p + 2K) \times M)$, consisting of zeros and ones and γ is the $(M \times 1)$ vector of unrestricted parameters.

Then the GLS estimator for γ is:

$$\hat{\gamma} = \left[R'\left(\sum_{t=1}^{T} Z_t Z'_{t-1} \otimes \Sigma_u^{-1}\right)R\right]^{-1} R'vec\left(\Sigma_u^{-1}\sum_{t=1}^{T} y_t Z'_{t-1}\right)$$

4.4. VAR model Diagnostic

Once the model has been estimated, it is necessary to verify whether the fitted model is appropriate. Moreover, another purpose of the model checking is to suggest for further improvements if needed. Generally, a fitted model is considered to be adequate if:

- All parameters are statistically significant at a specified level
- If the residuals have no significant serial or cross-correlations
- There is no structural changes or outlying observations
- The residuals do not violate the distributional assumption

Several are the diagnostic test considered important by the literature to assess the validity of the model.

Since a serious misspecification could arise if the residuals are correlated, The Portmanteau Test is adopted as a standard tool for checking the residual autocorrelation in VAR models. The null hypothesis of the test is that all residual auto-covariances are zero, that is, $H_0: E(u_t u'_{t-i}) = 0$ for ($i = 1,2, \dots$). The alternative hypothesis instead considers at least one autocovariance (and thus autocorrelation) is non-zero. The portmanteau statistic is given by:

$$Q_h = T\sum_{j=1}^{h} tr\,(\hat{C}_j'\hat{C}_0^{-1}\hat{C}_j\hat{C}_0^{-1}$$

Where $\hat{C}_j = T^{-1}\sum_{t=j+1}^{h} \hat{u}_t\, \hat{u}'_{t-j}$ with $\hat{u}_t$ being the mean-adjusted estimated residuals (Lütkepohl, 2005).

To verify whether the series are normally distributed, the Jarque-Bera test statistic can be used in order to measure the difference of the skewness and kurtosis of the series with those from the normal distribution. The null hypothesis is:

$$H_0: normal\ distribution, skewness\ and\ excess\ kurtosis\ are\ zero$$

$$vs$$

$$H_1: non - normal\ distribution$$

This test is defined as follow:

$$\frac{N}{6}(S^2 + \frac{(K-3)^2}{4})$$

With S,K and N denoting respectively the sample skewness, sample kurtosis and the sample size.

The reported probability is the probability that a Jarque-Bera statistic exceeds (in absolute value) the observed value under the null hypothesis. Therefore, a small probability value lead to the rejection of the null of a normal distribution.

If the model has proved to be a good model, then it can be performed the structural analysis. Generally, the interactions between economic variables are studied by considering the effects of changes in one variable on the other variables of interest. Differently, in VAR models changes in the variables are induces by nonzero residuals, that is, by shock which can have a structural interpretation, provided that structural restrictions have been placed accordingly. Therefore, the effects of nonzero residuals (so-called shocks) are traced through the system in order to study the relations among the variables, by examining the dynamic effects of shock on the other variables in the system. This analysis is known as Impulse Response Function (IRF) or Multiplier analysis. A complementary tool for investigating the impact of shocks in VAR models is the Forecast Error Variance Decompositions, which is used to analyse the proportion of the unanticipated change of a variable that is attributable to its own innovations and shocks to other variables in the system.

4.4.1. Impulse Response Function (IRF)

Recalling the VAR model equation (4.1), the impulses or shocks enter through the residual vector $u_t(u_{1t}, \dots, u_{Kt})'$. Each nonzero component of u_t corresponds to an equivalent change in the associated left-hand side variable which in turn will induce further changes in the other variables of the system in the next periods (Lütkepohl, 2011). The marginal effect of a single nonzero element in u_t can be studied and interpreted in a more convenient manner by making use of the corresponding moving average (MA) representation[33]: $y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j}$.

The marginal response of $y_{n,t+j}$ to a unit impulse u_{mt} is given by the $(n,m)^{th}$elements of the matrices Φ_j. Therefore, the elements of Φ_j represent the responses to u_t innovations. Since $\Phi_j \to 0$ as $j \to \infty$ for stationary processes, then the effect of an impulse vanishes over time, that is, it is transitory. However, the Impulse Response Function analysis is subject to a problematic assumption, that is, a shock occurs only in one variable at a time. Such assumption it is reasonable if and only if the shocks in different variables are independent. Whenever this feature is not satisfied, the error terms consist of all the influences and variables that are not directly included in the set of y variables (Lütkepohl, 2005). Correlation of the error terms may signal that a shock in one variable is likely to be followed by a shock in another variable. Therefore, setting all other residuals to zero could give a misleading picture of the dynamic relationship between the variables. To overcome this problem, obtaining uncorrelated residuals of different equation, it is possible to rewrite the VAR model with a decomposition of the white noise covariance matrix given by:

$$\Sigma_u = W \Sigma_\varepsilon W'$$

With Σ_ε being a diagonal matrix with positive diagonal elements and W a lower triangular matrix with unit diagonal. This decomposition is obtained from the Choleski decomposition, $\Sigma_u = PP'$ by defining a diagonal matrix D which has the same main diagonal as P and by specifying $W = PD^{-1}$ and $\Sigma_\varepsilon = DD'$, leading to the definition of a VAR model which reflect the actual system, without the possibility of having an instantaneous impact:

$$y_t = A_0^* y_t + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \varepsilon_t$$

Where $A_0^* \coloneqq I_K - A$[34].

A common problem with the orthogonalized impulse responses is the ordering of the variables, since it cannot be determined by statistical methods. In fact, the variables must be ordered in such a way that the first variable is the only one with a potential immediate impact on all other variables. The second may have an immediate impact on the last K – 2 components of y_t but not on y_{t1} and so on.

[33] Under stability assumption, a general VAR(p) process as in the (4.1) has a representation: $y_t = A(L)^{-1} u_t = \Phi(L) u_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j}$; This form of the process is called the Moving Average (MA) representation (or Wold MA representation), where y_t can expressed in terms of past and present error or innovation u_t.

[34] See Appendix C for further explanation.

4.5. *Applied methodology and VAR estimation*

Given the preliminary estimates obtained, together with the insights from the previous analysis, it is possible to proceed to the evaluation of the dynamic interdependence between the variables. Despite the fact that the time series can be considered integrated of order 0, and proven to be stationary, to obtain a more reliable results, the VAR models will focus on the changes in rig counts and the relative changes in oil prices. Since it has been observed that the contemporaneous correlations among variables it is very weak and very volatile (Figure 25), it is possible to exclude *a priori* the possible presence of contemporaneous effects between them (lag 0).

The considered lag structure is specified by considering both the cross-correlation analysis, showed in Figure 28-29, Table 1 and Figure 37, which is backed by economic reasoning to support it. Several will be the VAR models which will be considered, in order to obtain a more comprehensive framework and thus, useful results.

The most basic model to consider is characterized by a bivariate autoregressive model, with ΔSOR_t for the change in shale oil rig count and $\Delta NSOR_t$ for the change in non-shale oil rig count.

Model 1: standard VAR model

$$\begin{bmatrix} \boldsymbol{\Delta CSO_t} \\ \boldsymbol{\Delta CNSO_t} \end{bmatrix} = C + \sum_{j=1}^{p} \Phi_j \begin{bmatrix} \boldsymbol{\Delta CSO_{t-j}} \\ \boldsymbol{\Delta CNSO_{t-j}} \end{bmatrix} + \varepsilon_t$$

Where the optimal lag length is chosen using a maximum lag of 13 weeks, i.e. 3 months and 1 week. The model, despite its meaningless use, allows to define the optimal lag length by considering only the autoregressive terms of the time series. As shown in figure 49, both the AIC and the HQ suggest as optimal length a VAR (5). Since two of the three criteria considered above concur, a VAR of order 5 will be taken into consideration. The figure 50 displays the results. In this first model, it is possible to observe that the change in shale oil, (CSO), is affected by its own past in a stronger manner with respect to the change in non-shale oil rig count, (CNSO). In fact, it shows significant values at lag 1,2 and 4, being affected by the change in non-shale at lag 2, 4 and 5. Conversely, the change in non-shale oil is not explained by its own autoregressive term, while explained by the lags in the change in shale oil, at lag 1,2 and 5. Both series do not show evidences of serial dependence as shown by the multivariate

Portmanteau test, the last three rows, Q(6), Q(7) and Q(8) where it is reported the p-values of the test for lags 6, 7 and 8, respectively.

VAR Lag Order Selection Criteria
Endogenous variables: CSO CNSO

Lag	LogL	LR	FPE	AIC	SC	HQ
0	−3270.255	NA	5174.212	14.22720	14.24516	14.23427
1	−3183.687	172.0063	3613.574	13.86821	13.92209	13.88943
2	−3144.205	78.10662	3096.979	13.71393	13.80374*	13.74930
3	−3139.615	9.040747	3089.051	13.71137	13.83710	13.76088
4	−3129.481	19.87095	3007.775	13.68470	13.84636	13.74836
5	−3118.612	21.21820*	2919.299*	13.65483*	13.85242	13.73264*
6	−3115.988	5.100401	2936.847	13.66082	13.89432	13.75276
7	−3113.732	4.364847	2959.248	13.66840	13.93783	13.77449
8	−3112.639	2.105213	2996.951	13.68104	13.98639	13.80128
9	−3109.181	6.630365	3004.102	13.68339	14.02467	13.81778
10	−3104.927	8.118623	3000.892	13.68229	14.05949	13.83082
11	−3100.444	8.517131	2994.720	13.68019	14.09332	13.84287
12	−3098.996	2.740010	3028.267	13.69129	14.14033	13.86811
13	−3094.445	8.566648	3021.198	13.68889	14.17386	13.87986

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Figure 49: VAR Lag order selection criteria

The second model introduced is a restricted VAR model, represented as the following,

Model 2:

$$\begin{bmatrix} \boldsymbol{\Delta CSO_t} \\ \boldsymbol{\Delta CNSO_t} \end{bmatrix} = \boldsymbol{C} + \boldsymbol{\Phi_1} \begin{bmatrix} \boldsymbol{\Delta CSO_{t-1}} \\ \boldsymbol{\Delta CNSO_{t-1}} \end{bmatrix} + \boldsymbol{\Phi_m} \begin{bmatrix} \boldsymbol{\Delta CSO_{t-1|t-4}} \\ \boldsymbol{\Delta CNSO_{t-1|t-4}} \end{bmatrix} + \boldsymbol{\Phi_q} \begin{bmatrix} \boldsymbol{\Delta CSO_{t-1|t-13}} \\ \boldsymbol{\Delta CNSO_{t-1|t-13}} \end{bmatrix} + \boldsymbol{\varepsilon_t}$$

Where $y_t = [\begin{smallmatrix} \Delta CSO_t \\ \Delta CNSO_t \end{smallmatrix}]$ and $y_{t-1|\,t-j} = \sum_{i=1}^{j} y_{t-i}$. This model takes into consideration the sum of the lagged value for the month $\Phi_m\ y_{t-1|t-4}$ and the quarter $\Phi_q y_{t-1|t-13}$ in order to verify whether there is an impact on these fixed summed lags rather than all lags, as in the VAR(5). This model, despite its slightly lower adj. R-squared, it is more informative considering the economic rationale behind. In fact, it is assumed that the summed change in rig counts, both monthly and for the quarter, could be explanatory of the relationship between them. As observable in Figure 51, the change in shale oil is affected by both monthly sums while the change in non-shale is explained by the first lag of the shale oil, its monthly sum and the quarterly sum of the change in non-shale. In this sense, it is possible to observe that the change in non-shale oil present a slower response with respect to its autoregressive term, since it is affected by the quarter results rather than the monthly one[35]. Even though these two models might help understanding what the relationship between the variables is, a more useful and complete model need to consider other variables. An extension of the classic Vector AutoRegression is the Vector AutoRegression with eXogenous variables (VAR-X) that allows for a more extensive analysis.

A VAR model with exogenous variables is expressed as following:

$$Y_t = \nu_0 + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + B_1 X_{t-1} + \cdots + B_q X_{t-q} + U_t \tag{4.3}$$

Where $Y_t \in \mathbb{R}^k, X_t \in \mathbb{R}^m$ is a vector of exogenous variables, $\nu_0 \in \mathbb{R}^k$ is a vector of intercepts, the A_j's are $(K \times K)$ coefficient matrices, B_j's are $(K \times M)$ coefficient matrices and $U_t \in \mathbb{R}^k$ is the vector of errors. The general model can be referred as VAR-X(p,q) where p is the lag order of the autoregressive terms while q is the lag order for the exogenous variables. The equation (4.3) allows for a contemporaneous term (X_t) without any loss of generality[36]. This model allows to include in the regressions the effect of oil shocks, as in the following:

Model 3: Restricted VAR-X with oil-log returns

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{\Delta CSO_t} \\ \boldsymbol{\Delta CNSO_t} \end{bmatrix} = \boldsymbol{C} + \boldsymbol{\Phi_1} \begin{bmatrix} \boldsymbol{\Delta CSO_{t-1}} \\ \boldsymbol{\Delta CNSO_{t-1}} \end{bmatrix} + \boldsymbol{\Phi_m} \begin{bmatrix} \boldsymbol{\Delta CSO_{t-1|t-4}} \\ \boldsymbol{\Delta CNSO_{t-1|t-4}} \end{bmatrix} + \boldsymbol{\Phi_q} \begin{bmatrix} \boldsymbol{\Delta CSO_{t-1|t-13}} \\ \boldsymbol{\Delta CNSO_{t-1|t-13}} \end{bmatrix} \\
+ \boldsymbol{\Psi_1 COP_{t-1}} + \boldsymbol{\Psi_m COP_{t-1|t-4}} + \boldsymbol{\Psi_q COP_{t-1|t-13}} \\
+ \boldsymbol{\delta}[\boldsymbol{DGS10_{t-1}} + \boldsymbol{DTWEXBGS_{t-1}} + \boldsymbol{EFFR_{t-1}} + \boldsymbol{GSPC_{t-1}} + \boldsymbol{NFCI_{t-1}} \\
+ \boldsymbol{STLFSI2_{t-1}} + \boldsymbol{TEDRATE_{t-1}} + \boldsymbol{VIX_{t-1}}] + \boldsymbol{\varepsilon_t}
\end{aligned}$$

Where COP are the oil-log returns and the final variables in square brackets are a collection of economic and financial variables. This model considers the lagged sum of rig counts, monthly and quarterly, and the lagged sum for the oil-log returns, for the previous month and the previous quarter. This model allows for oil price to impact on rig counts, verifying whether it is as suggested by the economic theory. The figure 52 displays the results. Starting from the CSO, it shows a characterizing feature. In fact, this rig count it is positively affected by both the monthly sum of shale and non-shale rig count. What is interesting is that both the monthly and the quarterly sum of the oil shocks are affecting this series, with the latter having higher level of significance. An explanation could be the fact that the shale rig, having a faster life cycle and being more costly, are characterized by having higher sensitiveness to oil shocks. The

[35] The effect it is significant only at the 10% level of confidence.
[36] See Appendix C for the proof.

coefficients are positive meaning that an increase (a decrease) in the oil prices over the previous month/quarter leads to an increase (a decrease) in the shale oil rig count of the current week. Conversely, the non-shale rig count presents significant values for the first lag of the shale rig count as well as the first lag of the change in oil price. Moreover, this model highlights the delay in the response of the rig count, being affected by the quarter sum of both the oil shocks and the change in shale oil. However, the signs of the coefficients are inconsistent. All the control variables are not significant. The model shows absence of serial dependence as shown by the multivariate Portmanteau test at the bottom of the figure, where are reported the Q-stat for Q(2), Q(3) and Q(4) with the related p-values of the test for lags 2, 3 and 4, respectively. Moreover, by looking at the adjusted R-squared, it is possible to state that the oil-log returns together with the economic and financial variables increase the explanatory power of the model.

Despite the insights obtained by the previous model, it assumes symmetry in the response to oil-log returns. Therefore, to verify and detect the responsiveness of the rig count to the oil shocks while considering the sign of the oil shocks, an extension of the model is required.

Model 4: restricted VAR-X with positive and negative oil shocks

$$\begin{aligned}\begin{bmatrix}\Delta CSO_t\\ \Delta CNSO_t\end{bmatrix} = C + \Phi_1\begin{bmatrix}\Delta CSO_{t-1}\\ \Delta CNSO_{t-1}\end{bmatrix} + \Phi_m\begin{bmatrix}\Delta CSO_{t-1|t-4}\\ \Delta CNSO_{t-1|t-4}\end{bmatrix} + \Phi_q\begin{bmatrix}\Delta CSO_{t-1|t-13}\\ \Delta CNSO_{t-1|t-13}\end{bmatrix} + \Psi_1^+ COP_{t-1}^+ \\ + \Psi_m^+ COP_{t-1|t-4}^+ + \Psi_q^+ COP_{t-1|t-13}^+ + \Psi_1^- COP_{t-1}^- + \Psi_m^- COP_{t-1|t-4}^- \\ + \Psi_q^- COP_{t-1|t-13}^- + \delta[DGS10_{t-1} + DTWEXBGS_{t-1} + EFFR_{t-1} + GSPC_{t-1} \\ + NFCI_{t-1} + STLFSI2_{t-1} + TEDRATE_{t-1} + VIX_{t-1}] + \varepsilon_t\end{aligned}$$

Where $COP_{t-1|\,t-j}^- = \sum_{i=1}^{j} COP_{t-i} I(COP_{t-i} < 0)$ and $COP_{t-1|\,t-j}^+ = \sum_{i=1}^{j} COP_{t-i} I(COP_{t-i} > 0)$. This model breaks down the oil shocks, dividing them into positive and negative. This would allow to verify the importance of the lags and related sign. Figure 53 provides the results. Focusing on the change in shale oil, it is still observed the positive impact of both the monthly sum of the changes. Moreover, the series is impacted by the quarterly sum of the negative shocks, with a positive effect. This highlights two important feature of the series. First, the change in shale oil rig count shows an asymmetric response with respect to the oil shocks. The presence of the asymmetry is assumed by the absence of significant coefficients for positive shocks and by the presence of a significant coefficient for the negative ones. Moreover, the only significant coefficient with respect to the negative shocks, is the one which considers the sum of the previous 13 weeks. An interpretation of the coefficient could be that, being the shale oil rigs a proxy of the production, as the production increases then the oil price decreases because of an excess supply. Despite showing the asymmetry feature, the change in non-shale rigs instead presents a different behaviour. It is positively affected by both the previous week and the sum of the previous 13 weeks of the shale oil rigs change. However it is negatively affected by the negative oil shocks at lag 1, meaning that the larger the negative oil shocks, the larger the negative effect on the non-shale rigs, underscoring the importance of the size of oil revenues on traditional drilling activity. Despite the strong, negative sign for negative oil shocks at lag 1, the overall coefficient for the quarter sum of negative shocks is positive, confirming what has been assumed, and observed, so far that is, the quarterly negative oil shocks are affecting positively the non-shale rig count. Once again, an interpretation of this relationship could be thinking about the rig counts as a proxy for the production.

Despite the negative oil shocks effect is not always positive considering different lags, from the last two models it is visible the presence of asymmetry between positive and negative shocks. The positive effect of the oil returns could be explained by the fact that, despite the fall in oil price, the rigs in progress will continue to drill until the wells are completed because it is expensive to shut down a rig once it started due to installation costs and rent contracts. The absence of significance in the positive shocks’ coefficients could be explained by the fact that the process of drilling and extracting oil takes time, mostly if we consider the non-shale oil whose production process is slower. Therefore, even if the oil shocks are positive, this might have a longer delay considering the physical time required for the activation of new drills. As observed by Khalifa et al. (2017), adding new drills, considered an investment decision, may require financing which might not be immediate and might depend on the oil reserves collateral value. Thus, the time span required to see the effect of positive shocks might be longer.

All the control variables in the model are not significant. The model shows absence of serial dependence as shown by the multivariate Portmanteau test at the bottom of the figure, where are reported the Q-stat for Q(2), Q(3) and Q(4) with the related p-values of the test for lags 2, 3 and 4, respectively.

The last model that has been considered is an extension of the Model 4, at which it is added the interaction among variables. When there is an interaction term, the effect of one variable that form the interaction depends on the level of the other variable in the interaction. Therefore, this would allow to verify whether the lagged values related to rig counts are affected by the oil shocks’ sign.

Model 5: restricted VAR-X with positive and negative oil shocks and interaction among variables

$$\begin{bmatrix} \Delta CSO_t \\ \Delta CNSO_t \end{bmatrix} = C + \Phi_1 \begin{bmatrix} \Delta CSO_{t-1} \\ \Delta CNSO_{t-1} \end{bmatrix} + \Phi_m \begin{bmatrix} \Delta CSO_{t-1|t-4} \\ \Delta CNSO_{t-1|t-4} \end{bmatrix} + \Phi_q \begin{bmatrix} \Delta CSO_{t-1|t-13} \\ \Delta CNSO_{t-1|t-13} \end{bmatrix} + \Psi_1^+ COP_{t-1}^+$$

$$+ \Psi_m^+ COP_{t-1|t-4}^+ + \Psi_q^+ COP_{t-1|t-13}^+ + \Psi_1^- COP_{t-1}^- + \Psi_m^- COP_{t-1|t-4}^-$$

$$+ \Psi_q^- COP_{t-1|t-13}^- + \Theta_1^+ \begin{bmatrix} \Delta CSO_{t-1} \\ \Delta CNSO_{t-1} \end{bmatrix} \times [COP_{t-1}^+] + \Theta_1^- \begin{bmatrix} \Delta CSO_{t-1} \\ \Delta CNSO_{t-1} \end{bmatrix} \times [COP_{t-1}^-]$$

$$+ \Theta_m^+ \begin{bmatrix} \Delta CSO_{t-1|t-4} \\ \Delta CNSO_{t-1|t-4} \end{bmatrix} \times [COP_{t-1|t-4}^+] + \Theta_m^- \begin{bmatrix} \Delta CSO_{t-1|t-4} \\ \Delta CNSO_{t-1|t-4} \end{bmatrix} \times [COP_{t-1|t-4}^-]$$

$$+ \Theta_q^+ \begin{bmatrix} \Delta CSO_{t-1|t-13} \\ \Delta CNSO_{t-1|t-13} \end{bmatrix} \times [COP_{t-1|t-13}^+] + \Theta_q^- \begin{bmatrix} \Delta CSO_{t-1|t-4} \\ \Delta CNSO_{t-1|t-4} \end{bmatrix} \times [COP_{t-1|t-13}^-]$$

$$+ \delta[DGS10_{t-1} + DTWEXBGS_{t-1} + EFFR_{t-1} + GSPC_{t-1} + NFCI_{t-1}$$

$$+ STLFSI2_{t-1} + TEDRATE_{t-1} + VIX_{t-1}] + \varepsilon_t$$

Considering the model above, whose results are displayed in figure 54, the shale oil rig count shows different features with respect to the model 4. Despite the constant being still significant with more or less the same impact on the series, it is possible to see that the first autoregressive term is now significant with a negative impact on the series. Moreover, the sum of the previous month result is negligible while the previous quarter sum of oil returns is significant, displaying the same impact as in the model 4. With the introduction of interactions among the exogenous variables, the model shows an interesting dynamic. Considering the interaction between the first lag of the change in shale oil and the positive oil shocks at the first lag, the coefficient is significant. Therefore, the series display a positive response to its first lag, provided that the oil shocks are positive.

Focusing instead on the change in non-shale oil rigs, it displays coefficients similar to model 4 with respect to the first lag and the previous quarter sum of the change in shale oil. Moreover, the model highlights once again how the negative oil shocks are important, with a negative response at lag 1 while showing a positive impact if the previous quarter result is considered. Furthermore, both coefficients for the first lag of the rigs are significant, highlighting the asymmetric relationship with respect to negative oil shocks. In particular, considering the interaction between its autoregressive term and the lagged value of the negative oil shocks, it amplify the magnitude of the negative oil returns, underscoring once again the fundamental role played by the oil price for the determination of the rig counts number. Finally, the model shows the presence of asymmetric response with respect to the oil shocks, with the negative ones having higher impact on the change of the rig counts.

All the models presented have diagnosed in order to confirm their validity. The results for the last two, more specific, model, are reported in the figure 55 and figure 56. Both models display absence of residuals serial correlations, however the Jarque-Bera test reject the null hypothesis of multivariate normal residuals for both models.

To verify how the different models perform over time, the Figure 57 displays their rolling adjusted R-squared, with a 144-weeks window. For all models and for both series, the values oscillate over times. However, the introduction of more specific models increases the ability of explain the model's behavior. By starting from an autoregressive model, the overall adjusted R-squared increases as the oil-log returns are introduced. Moreover, the classification of those shocks between positive and negative increases the measure, mostly after the 2014 oil glut to later drop at the end of 2017, for both shale and non-shale rigs. Another effect of considering the oil shocks, is the increase in the measure whenever there is a fall in the oil price. Furthermore, the overall measure increases as the interaction among variables is considered, mostly for the non-shale oil rig. This suggests that the model might be more capable of describing the behaviour of both series after fall in oil price, as in 2014 and the recent drop due to the COVID outbreak.

The evolutions of the coefficients and their p-values are provided from Figure 58 to Figure 81, where the red line represents the 5% level of confidence while the green dotted line the 10% level of confidence. Considering the last model, by looking at the coefficients and the p-values is possible to obtain several information. Starting from the equation of the change in shale oil, the constant is significant during the 2014 oil glut, with a positive value. Moreover, it is observed once again the asymmetry through the significance of the sum of negative oil shocks for the previous 13 weeks, starting from 2013 until 2018. Similar result is found for the positive oil shocks, however with a lower impact. Considering their coefficients, is visible how the negative shocks impact positively the rig count while the positive in a negative manner[37]. Interesting is the significance of the control variables, found to be relevant during oil drops, in particular the DGS10 during the 2014 oil drop as well as the EFFR and the STLFSI2. These might indicate the importance of the funding for the oil industry. Finally, considering the interaction among variables, the negative oil shocks are found significant mostly during bearish

[37] These dynamics are visible in the model 4 too, confirming the presence of asymmetry with respect to the oil shocks.

oil market. By looking at change in non-shale oil equation, the p-values show that in general neither positive nor negative oil shocks seem to be relevant, not even during bearish oil market. This could be probably due to the longer delay affecting this type of technology. However, considering the evolution for the coefficients of the interaction among variables, the relationships between the change in shale oil with the sum of the previous 4 and 13 weeks negative oil shocks are relevant. Considering the financial covariates, the DGS10 and the STLFSI2 are found to be relevant, the former during the 2014 oil glut.

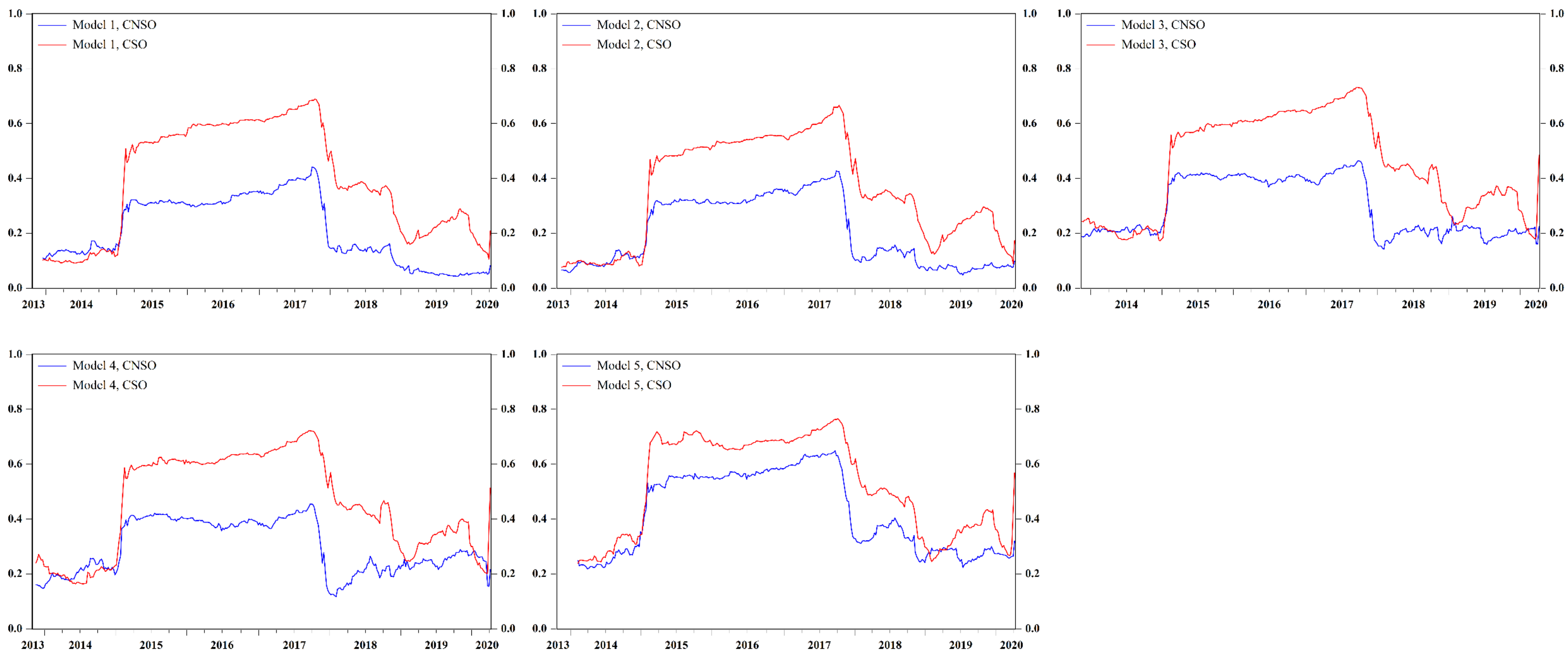

Figure 57: Rolling Adjusted R-squared for all models

Conclusion

The main aim of this work is to investigate the relationship between changes in oil price and changes in rig counts, considering the different nature of the rig. Understanding this relationship is of significant interest to analysts, investment and commercial banks, policy makers and oil companies among others. Starting from a descriptive analysis of the variables, it has been proven that the relationship between the oil and the rigs is not contemporaneous but lagged, in line with the analysed literature. This implies that changes in the rig counts are not the immediate effect of the changes in oil price but they are the results of the oil price changes up to three months before.

The use of the exceedance correlations methodology has proven the presence of an asymmetric response considering the different quantiles of the oil price, showing a stronger impact of oil returns below the median to the change in both series. In this sense, the non-shale rig count has proven to be more resilient to negative oil shocks, having a lower negative correlation considering different lags than the shale rigs. The asymmetry is then underscored by the less stable responses of the rigs to oil prices above the median, which become more visible as the lags considered increase. This could be translated in a different reaction of the rig counts, with a stronger impact during bearish oil market. The absence of significance for positive oil shocks' coefficients could find an explanation in the need of longer delay considering the physical time required for the activation of new drills. Therefore, the time span needed to see the effect of positive oil shocks might be longer.

The study then proposes five different VAR models, starting from a more general version made of 5 autoregressive terms to later consider the sum of the previous monthly and quarterly change. The latter model shows that the series seems to be affected by these sums. This might be due to the completion of DUC, which is a process that requires some time due to physical and economic reasons. Then the oil price is introduced to increase the effectiveness of the model and split by the sign. The model confirms both the presence of asymmetry, showing significant coefficients for the negative oil shocks, and the presence of a lagged relationship. An interesting finding is that, despite the previous analysis has shown the non-shale oil rig count as more resilient, the model display that this series is affected in a stronger manner to negative oil changes than the shale oil rigs. Moreover, the divergences of coefficients arose from the last model, where interactions among variables are introduced, can have several explanations. Among them, a reason for the presence of positive sign for both type of oil shocks could be given by the different response of the rig counts considering the length of time. Despite suffering a relevant decrease as *"short-term"* adjustment due to negative oil shocks of the previous week, considering the overall result of the previous quarter negative, the series react positively. This dynamic underscore the importance of revenues in the oil industry, mostly for the traditional one, and the potential presence of adjustment dynamics of the production related to different levels of oil price. A lower level of the oil price has multiple effects. Among them, the fluctuation of the oil price moves the separation line between the contingent resources (potentially recoverable depending on the condition of commercial markets and profitability of the investment), leading to a different estimation of the reserves. This might have direct consequences on the oil production and therefore on the rigs. However, to prove this, an analysis considering the weekly production should be considered to verify the speed of the adjustment of oil supply to price changes. The lack of data about the weekly production represents a limit for this study. Furthermore, despite the fact the cross-correlation displays the relationship as one-sided that is, is the oil price which affect the rig counts and not the other way around, it is required a deeper analysis which would considering the oil price as endogenous variable rather than weakly exogenous.

BIBLIOGRAPHY

Anderson, S. T., Kellogg, R., & Salant, S. W. (2014). *Hotelling under pressure* (No. w20280). National Bureau of Economic Research.

Apergis, N., Ewing, B. T., & Payne, J. E. (2016). A time series analysis of oil production, rig count and crude oil price: Evidence from six US oil producing regions. *Energy*, *97*, 339-349.

Baffes, J., Kose, M. A., Ohnsorge, F., & Stocker, M. (2015). The great plunge in oil prices: Causes, consequences, and policy responses. *Consequences, and Policy Responses (June 2015)*.

Bohorquez M.O, (2014). Hydrocarbon Discovery Potential in Colombian Basins: Creaming Curve Analysis. (Adapted from oral presentation given at AAPG 2014 European Regional Conference & Exhibition, Barcelona, Spain, May 13-15, 2014)

Caporin, M., Pelizzon, L., Ravazzolo, F., & Rigobon, R. (2018). Measuring sovereign contagion in Europe. *Journal of Financial Stability*, *34*, 150-181.

Carmalt, S. W. (2016). The economics of oil: a primer including geology, energy, economics, politics. Springer.

Dahl, C., & Duggan, T. E. (1998). Survey of price elasticities from economic exploration models of US oil and gas supply. *Journal of Energy Finance & Development*, *3*(2), 129-169.

Ellwanger, R., Sawatzky, B., & Zmitrowicz, K. (2017). Factors Behind the 2014 Oil Price Decline. *Bank of Canada Review*, *2017*(Autumn), 1-13.

Fattouh, B., Poudineh, R., & Sen, A. (2016). The dynamics of the revenue maximization–market share trade-off: Saudi Arabia’s oil policy in the 2014–15 price fall. *Oxford Review of Economic Policy*, *32*(2), 223-240.

Gülen, G., Browning, J., Ikonnikova, S., & Tinker, S. W. (2013). Well economics across ten tiers in low and high Btu (British thermal unit) areas, Barnett Shale, Texas. *Energy*, *60*, 302-315.

Hamilton, J. D. (1983). Oil and the macroeconomy since World War II. *Journal of political economy*, *91*(2), 228-248.

Henriques, I., & Sadorsky, P. (2011). The effect of oil price volatility on strategic investment. *Energy Economics*, *33*(1), 79-87.

Höök, M., Davidsson, S., Johansson, S., & Tang, X. (2014). Decline and depletion rates of oil production: a comprehensive investigation. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372(2006), 20120448.

Hosking, J. R. (1980). The multivariate portmanteau statistic. *Journal of the American Statistical Association*, *75*(371), 602-608. Hamilton, J. D. (2009). Understanding crude oil prices. *The Energy Journal*, *30*(2).

Howard, A. W., & Harp Jr, A. B. (2009). Oil and gas company valuations. *Business Valuation Review*, *28*(1), 30-35.

Hubbert, M. K. Nuclear Energy and the Fossil Fuels. 1956. Shell Development Company-Exploration and Poduction Research Division. p.10

Huntington, H. G. (1998). Crude oil prices and US economic performance: where does the asymmetry reside? *The Energy Journal*, *19*(4).

Inkpen, A. C., & Moffett, M. H. (2011). The global oil & gas industry: management, strategy & finance. PennWell Books.

Kaiser, M. J. (2010). Economic limit of field production in Louisiana. *Energy*, *35*(8), 3399-3416.

Kaiser, M. J. (2012). Haynesville shale play economic analysis. *Journal of Petroleum Science and Engineering*, *82*, 75-89.

Kellogg, R. (2014). The effect of uncertainty on investment: Evidence from Texas oil drilling. *American Economic Review*, *104*(6), 1698-1734.

Khalifa, A., Caporin, M., & Hammoudeh, S. (2017). The relationship between oil prices and rig counts: The importance of lags. *Energy Economics*, *63*, 213-226.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of econometrics*, *54*(1-3), 159-178.

Longin, F., & Solnik, B. (2001). Extreme correlation of international equity markets. *The journal of finance*, *56*(2), 649-676.

Lütkepohl H, (2011). *Vector Autoregressive Models*, Economics Working Papers ECO2011/30, European University Institute.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Mohn, K. (2008). Efforts and efficiency in oil exploration: a vector error-correction approach. *The Energy Journal*, *29*(4).

Mohn, K., & Osmundsen, P. (2011). Asymmetry and uncertainty in capital formation: an application to oil investment. *Applied Economics*, *43*(28), 4387-4401.

Mork, K. A. (1989). Oil and the macroeconomy when prices go up and down: an extension of Hamilton's results. *Journal of political Economy*, *97*(3), 740-744.

Mork, K. A., Olsen, O., & Mysen, H. T. (1994). Macroeconomic responses to oil price increases and decreases in seven OECD countries. *The Energy Journal*, *15*(4).

Osmundsen, P., Roll, K. H., & Tveterås, R. (2010). Exploration drilling productivity at the Norwegian shelf. *Journal of Petroleum Science and Engineering*, *73*(1-2), 122-128.

Osmundsen, P., Roll, K. H., & Tveteras, R. (2012). Drilling speed—The relevance of experience. *Energy Economics*, *34*(3), 786-794.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica: journal of the Econometric Society*, 1361-1401.

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

Pierce, D. A. (1977). Relationships—and the lack thereof—between economic time series, with special reference to money and interest rates. *Journal of the American Statistical Association*, *72*(357), 11-22.

Prest, B. C. (2018). Explanations for the 2014 oil price decline: Supply or demand?. *Energy Economics*, *74*, 63-75.

Ringlund, G. B., Rosendahl, K. E., & Skjerpen, T. (2008). Does oilrig activity react to oil price changes? An empirical investigation. *Energy Economics*, *30*(2), 371-396.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, 1-48.

Smith, J. L., & Lee, T. K. (2017). The price elasticity of US shale oil reserves. *Energy Economics*, *67*, 121-135.

SPE. 2011. PRMS Guidelines Nov 2011, 222 pp. Richardson, Texas USA: Society of Petroleum Engineers.

Speight, J. G. (2011). An introduction to petroleum technology, economics, and politics. John Wiley & Sons.

Toews, G., & Naumov, A. (2015). The relationship between oil price and costs in the oil industry. *The Energy Journal*, *36*(Adelman Special Issue).

Waheed, M., Alam, T., & Ghauri, S. P. (2006). Structural breaks and unit root: evidence from Pakistani macroeconomic time series. *Available at SSRN 963958*.

Zivot, E., & Andrews, D. W. K. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of business & economic statistics*, *vol.1, pp.251-270*

Zou, C. (2017). Unconventional petroleum geology. Elsevier.

REPORT:

BP Statistical Review of World Energy, 2019.

World Oil Outlook 2014. Organization of the Petroleum Exporting Countries (OPEC), 2018.7

SITOGRAPHY:

https://www.eia.gov/

https://fred.stlouisfed.org/

http://www.petroleum.co.uk/

APPENDIX A

In the economic literature it is common to assume that the evolution of economic variables in a specific period are the realizations of random variables. Therefore, time-series are generally considered to be generated by a stochastic process. To define the concept of stochastic process is required first to introduce two fundamental concepts.

Let $(\Omega, \mathcal{F}, P)$ be a *probability space*, where:

- Ω is a nonempty set of all elementary events, called sample space;
- $\mathcal{F}$ is a sigma-algebra of events or subsets of Ω, i.e. a family of subsets closed with respect to countable union and complement with respect to Ω;
- Pr is a probability measure defined for all members of $\mathcal{F}$.

A *random variable* y is a real valued function defined on Ω such that for each real number *c*, $A_c = \{\omega \ \in \Omega | y(\omega) \leq c\} \in \ \mathcal{F}$. A_c is an event for which the probability is defined in terms of Pr. A K-dimensional random vector or a K-dimensional vector of random variables is a function y from Ω into the K-dimensional Euclidean space $\mathbb{R}^K$, that is, y maps $\omega \in \Omega$ on $y(\omega) = (y_1(\omega), \dots, y_K(\omega))'$ such that for each $c = (c_1, \dots, c_K)' \in \mathbb{R}^K$,

$A_c = \{\omega | y_1(\omega) \leq c_1, \dots, y_K(\omega) \leq c_K\} \in \mathcal{F}$, where the function F: $\mathbb{R}^K \rightarrow [0,1]$ defined by F(c) = Pr(A_c) is the joint distribution function of y.

Suppose Z is some indexes set with at most countably may elements, a stochastic process is a real valued function

$y : Z \times \Omega \rightarrow \mathbb{R}$ such that for each fixed $t \in Z, y(t, \omega)$ is a random variable. The random variable corresponding to a fixed t is usually denoted by y_t.

Consequentially, a K-dimensional *vector* (or *multivariate*) *stochastic process* is a function

$y : Z \times \Omega \rightarrow \mathbb{R}^K$ where for each fixed $t \in Z$, $y(t, \omega)$ is a K-dimensional random vector. The notation y_t in this case corresponds to a random vector for a fixed $t \in Z$.

Finally, we can define the stochastic process as a family of real random variable $\boldsymbol{Y} = \{y_i(\omega); i \in T\}$, all defined on the same probability space $(\Omega, \mathcal{F}, P)$. The set T is called the index set of the process. If $T \subset \mathbb{Z}$, then the process is called a *discrete* stochastic process. If T is an interval of $\mathbb{R}$, then the process is called a *continuous* stochastic process.

The underlying stochastic process is said to have generated the time series and it is generally called data generation process (DGP).

APPENDIX B

The difference operator

The first difference operator Δ is defined as $\Delta = (1 - L)$ and is used to express the difference between two consecutive realizations of a time series.

$\Delta Y_t = (1 - L)Y_t = Y_t - LY_t = Y_t - Y_{t-1}$

More in general, the differentiation of order *j* is defined as:

$\Delta^j Y_t = (1 - L)^j Y_t,$ for $j = 0,1,2$ ...

The Lag operator

The lag operator, also called backward (shift) operator, is denoted by L and is an operator that shift the time index backward by one time unit: $LY_t = Y_{t-1}$

More in general, $L^j Y_t = Y_{t-j}$. If:

- *j* = 0, then: $L^0 Y_t = Y_t$
- *j > 0,* the series is shifted *j* periods backward: , $L^j Y_t = Y_{t-j}$
- *j < 0,* the series is shifted | *j* | periods forward: $L^j Y_t = Y_{t+|j|}$.

APPENDIX C

Assuming a VAR(p) process

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \qquad (1)$$

This process can be rewritten in order to obtain uncorrelated residuals of each equation. Decomposing the white noise covariance matrix as the following:

$$\Sigma_u = W \Sigma_\varepsilon W'$$

With Σ_ε being a diagonal matrix with positive diagonal elements and W a lower triangular matrix with unit diagonal. This decomposition is obtained from the Choleski decomposition, $\Sigma_u = PP'$ by defining a diagonal matrix D which has the same main diagonal as P and by specifying $W = PD^{-1}$ and $\Sigma_\varepsilon = DD'$.

Premultiplying $y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$ by $A \coloneqq W^{-1}$ is possible to obtain

$$Ay_t = A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \varepsilon_t, \qquad (2)$$

where $A_i^* \coloneqq AA_i, i = 1, \ldots, p$ and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Kt})' \coloneqq A_{ut}$ has a diagonal covariance matrix, $\Sigma_\varepsilon = E(\varepsilon_t \varepsilon_t') = AE(u_t u_t')A' = A\Sigma_u A'$.

Adding $(I_K - A)y_t$ to both sides of (2) we obtain

$$y_t = A_0^* y_t + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \varepsilon_t \qquad (3)$$

Where $A_0^* \coloneqq I_K - A$. Since W is lower triangular with unit diagonal, the same holds for A.

Therefore:

$$A_0^* = I_K - A = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ \beta_{21} & \cdots & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ \beta_{K1} & \beta_{K2} & \cdots & \beta_{K,K-1} & 0 \end{bmatrix}$$

Is a lower triangular matrix with zero diagonal. Considering the VAR(p) process express in the equation (3), the first equation contains no instantaneous $y's$ on the right-hand side. More in general, the k^{th} equation may contain $y_{1t}, \ldots, y_{k-1,t}$ and not $y_{kt}, \ldots, y_{Kt}$ on the right-hand side. Therefore the eq. (3) cannot have an instantaneous impact on y_{kt} for $k < s$.

VAR-X (*p*,*q*) process

Assuming:

$$Y_t = \nu_0 + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + B_1 X_{t-1} + \cdots + B_q X_{t-q} + U_t \qquad (4)$$

And

$$X_t = c_0 + C_1 X_{t-1} + \cdots + C_r X_{t-r} + V_t, \qquad (5)$$

which implies that Y_t does not Granger-cause X_t, that is a weak form of exogeneity.

If a contemporaneous X_t is included in the model (4), it becomes

$$Y_t = \nu_0 + A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + B_0 X_t + B_1 X_{t-1} + \cdots + B_q X_{t-q} + U_t \qquad (6)$$

If $B_0 \neq 0$, it may be possible that Y_t has an indirect impact on X_t via possible mutual dependence of U_t and V_t. If so, then (5) and (4) form a system of simultaneous equations.

To show that the absence of a contemporaneous X_t in (4) is characterized by no loss of generality, it is sufficient to rewrite (6) as a reduced VAR-X form by substituting (5) in (6) obtaining:

$$\begin{aligned} Y_t = \nu_0 &+ A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + \\ &+ B_0 (C_1 X_{t-1} + \cdots + C_r X_{t-r}) + B_1 X_{t-1} + \\ &+ \cdots + B_q X_{t-q} + U_t + B_0 V_t, \end{aligned}$$

Which is of the same form as (4).

VAR models estimates

Figure 50: VAR estimation, model 1[38]

	CSO	CNSO
C	-0.153445	-0.157420
	[-0.36864]	[-0.58205]
CSO_{t-1}	**0.166109**	**0.171710**
	[3.53633]	[5.62612]
CSO_{t-2}	**0.167960**	0.056383*
	[3.45773]	[1.78644]
CSO_{t-3}	0.048347	-0.020644
	[0.98556]	[-0.64766]
CSO_{t-4}	**0.152250**	0.024694
	[3.22758]	[0.80567]
CSO_{t-5}	0.046167	**0.077258**
	[0.97495]	[2.51096]
$CNSO_{t-1}$	0.100881	-0.005695
	[1.40865]	[-0.12239]
$CNSO_{t-2}$	**0.316151**	-0.005048
	[4.40643]	[-0.10829]
$CNSO_{t-3}$	-0.002578	-0.062345
	[-0.03535]	[-1.31554]
$CNSO_{t-4}$	0.126356*	-0.018826
	[1.73252]	[-0.39729]
$CNSO_{t-5}$	**0.211734**	0.072039
	[2.96215]	[1.55110]
Adj. R^2	0.399369	0.184558
Akaike AIC	7.264388	6.402055
Schwarz SC	7.361111	6.498779
	Q-stat	p-value
Q(6)	6.908674	0.1408
Q(7)	10.39188	0.2386
Q(8)	14.89908	0.2470

[38] The bold values denote the 5% statistically significant coefficients, the asterisks denote the 10%.
the associated t-statistics for each variable is reported in the square bracket. The first line in the bottom reports the adjusted R-squared for each of the two equations composing the VARX model.
Residuals do not show evidences of serial dependence as shown by the multivariate Portmanteau test, last three rows of the table where are reported the p-values of the test for their respective lags..

Figure 51: VAR estimation, model 2[39]

	CSO	CNSO
C	-0.139484	-0.182611
	[-0.32749]	[-0.67437]
CSO_{t-1}	0.039024	**0.152313**
	[0.66612]	[4.08943]
$CNSO_{t-1}$	-0.025970	0.029460
	[-0.31670]	[0.56508]
$CSO_{t-1\|t-4}$	**0.144357**	**0.038935**
	[5.63629]	[2.39105]
$CNSO_{t-1\|t-4}$	**0.139240**	-0.029505
	[3.15334]	[-1.05101]
$CSO_{t-1\|t-13}$	0.003370	0.012076
	[0.28035]	[1.58011]
$CNSO_{t-1\|t-13}$	-0.001081	-0.030185*
	[-0.03774]	[-1.65698]
Adj. R^2	0.371090	0.175803
Akaike AIC	7.306294	6.400465
Schwarz SC	7.367453	6.461624
	Q-stat	p-value
Q(2)	6.153798	0.1880
Q(3)	13.58909	0.0931
Q(4)	14.59980	0.2641

[39] The bold values denote the 5% statistically significant coefficients, the asterisks denote the 10%.
the associated t-statistics for each variable is reported in the square bracket. The first line in the bottom reports the adjusted R-squared for each of the two equations composing the VARX model.
Residuals do not show evidences of serial dependence as shown by the multivariate Portmanteau test, last three rows of the table where are reported the p-values of the test for their respective lags.

Figure 52: Var estimation, model 3[40]

	CSO	CNSO
C	0.261694	0.006783
	[0.62705]	[0.02458]
CSO_{t-1}	-0.026365	**0.133471**
	[-0.46593]	[3.56759]
$CNSO_{t-1}$	-0.058847	0.007834
	[-0.74648]	[0.15030]
COP_{t-1}	-9.756818	**-23.94408**
	[-0.80484]	[-2.98745]
$COP_{t-1\|t-4}$	11.82556*	3.316450
	[1.92017]	[0.81450]
$COP_{t-1\|t-13}$	**11.64567**	**6.016367**
	[3.37391]	[2.63636]
$CSO_{t-1\|t-4}$	**0.126383**	0.027266
	[4.92226]	[1.60617]
$CNSO_{t-1\|t-4}$	**0.133461**	-0.031978
	[3.13842]	[-1.13738]
$CSO_{t-1\|t-13}$	0.011755	**0.016681**
	[1.01481]	[2.17807]
$CNSO_{t-1\|t-13}$	0.011611	-0.026546
	[0.42350]	[-1.46446]
$DGS10_{t-1}$	-4.667517	-0.762339
	[-1.08884]	[-0.26898]
$DTWEXBGS_{t-1}$	-0.857650	-0.617542
	[-1.32440]	[-1.44237]
$EFFR_{t-1}$	10.61153	-2.005543
	[1.56756]	[-0.44810]
$GSPC_{t-1}$	-0.002787	-0.009732
	[-0.17625]	[-0.93078]
$NFCI_{t-1}$	-1.932637	-2.723071
	[-0.08791]	[-0.18734]
$STLFSI2_{t-1}$	1.556085	0.043013
	[0.66521]	[0.02781]
$TEDRATE_{t-1}$	-4.083643	-8.891717
	[-0.36114]	[-1.18935]
VIX_{t-1}	-0.114387	-0.140542
	[-0.50420]	[-0.93699]
Adj. R^2	0.453221	0.225093
Akaike AIC	7.225157	6.397606
Schwarz SC	7.382422	6.554871
	Q-stat	p-value
Q(2)	4.011804	0.4044
Q(3)	11.23453	0.1888
Q(4)	12.16858	0.4322

[40] The bold values denote the 5% statistically significant coefficients, the asterisks denote the 10%.
the associated t-statistics for each variable is reported in the square bracket. The first line in the bottom reports the adjusted R-squared for each of the two equations composing the VARX model.
Residuals do not show evidences of serial dependence as shown by the multivariate Portmanteau test, last three rows of the table where are reported the p-values of the test for their respective lags.

Figure 53: VAR estimation, model 4[41]

	CSO	CNSO
C	**2.640346**	0.561568
	[2.28023]	[0.73863]
CSO_{t-1}	-0.028245	**0.136001**
	[-0.50002]	[3.66690]
$CNSO_{t-1}$	-0.059045	0.012336
	[-0.75049]	[0.23880]
$CSO_{t-1\|t-4}$	**0.117911**	0.021383
	[4.54118]	[1.25424]
$CNSO_{t-1\|t-4}$	**0.131141**	-0.032244
	[3.05705]	[-1.14475]
$CSO_{t-1\|t-13}$	0.010317	**0.015036**
	[0.89029]	[1.97624]
$CNSO_{t-1\|t-13}$	0.013948	-0.020475
	[0.50240]	[-1.12322]
COP^{-}_{t-1}	-18.29804	**-59.18716**
	[-0.88322]	[-4.35106]
$COP^{-}_{t-1\|t-4}$	13.20326	9.572929
	[1.18593]	[1.30957]
$COP^{-}_{t-1\|t-13}$	**16.93279**	**8.945477**
	[3.08740]	[2.48412]
COP^{+}_{t-1}	-0.489417	15.96539
	[-0.02184]	[1.08506]
$COP^{+}_{t-1\|t-4}$	10.81917	-0.024621
	[0.90442]	[-0.00313]
$COP^{+}_{t-1\|t-13}$	4.255349	3.047157
	[0.86898]	[0.94771]

	CSO	CNSO
$DGS10_{t-1}$	-4.740281	-2.089790
	[-1.09110]	[-0.73260]
$DTWEXBGS_{t-1}$	-0.846232	-0.491939
	[-1.29916]	[-1.15024]
$EFFR_{t-1}$	10.87510	0.317649
	[1.57446]	[0.07004]
$GSPC_{t-1}$	-0.001532	-0.003769
	[-0.09559]	[-0.35807]
$NFCI_{t-1}$	1.398026	-2.108672
	[0.06316]	[-0.14510]
$STLFSI2_{t-1}$	1.400009	-0.476108
	[0.59687]	[-0.30914]
$TEDRATE_{t-1}$	-0.894073	-10.06039
	[-0.07743]	[-1.32697]
VIX_{t-1}	-0.123427	-0.152572
	[-0.54521]	[-1.02644]
Adj. R^2	0.436615	0.211225
Akaike AIC	7.222728	6.382995
Schwarz SC	7.397466	6.557734
	Q-stat	p-value
Q(2)	4.325660	0.3637
Q(3)	10.67939	0.2205
Q(4)	11.74532	0.4663

[41] The bold values denote the 5% statistically significant coefficients, the asterisks denote the 10%.
the associated t-statistics for each variable is reported in the square bracket. The first line in the bottom reports the adjusted R-squared for each of the two equations composing the VARX model. Residuals do not show evidences of serial dependence as shown by the multivariate Portmanteau test, last three rows of the table where are reported the p-values of the test for their respective lags.

Figure 54: VAR estimation, model 5[42]

	CSO	CNSO
C	2.077677*	0.761033
	[1.76154]	[0.99085]
CSO_{t-1}	$\mathbf{-0.200907}$	**0.162611**
	[-2.83096]	[3.51868]
CNSO_{t-1}	-0.028130	-0.098008
	[-0.24327]	[-1.30156]
$\text{CSO}_{t-1\|t-4}$	0.087306*	-0.016957
	[1.91383]	[-0.57083]
$\text{CNSO}_{t-1\|t-4}$	0.155266	-0.107712*
	[1.55756]	[-1.65930]
$\text{CSO}_{t-1\|t-13}$	0.014143	0.040425*
	[0.42555]	[1.86789]
$\text{CNSO}_{t-1\|t-13}$	0.020711	-0.057397
	[0.20501]	[-0.87247]
COP^{-}_{t-1}	-19.21277	$\mathbf{-48.83178}$
	[-0.90704]	[-3.54022]
$\text{COP}^{-}_{t-1\|t-4}$	4.279637	2.121674
	[0.35727]	[0.27199]
$\text{COP}^{-}_{t-1\|t-13}$	**16.92529**	**7.819732**
	[3.07412]	[2.18106]
COP^{+}_{t-1}	6.745171	13.39338
	[0.29541]	[0.90076]
$\text{COP}^{+}_{t-1\|t-4}$	9.352010	0.521790
	[0.77732]	[0.06660]
$\text{COP}^{+}_{t-1\|t-13}$	4.523510	-0.651821
	[0.90311]	[-0.19984]
$\text{CSO}_{t-1\|t-4} * COP^{-}_{t-1\|t-4}$	-0.328336	$\mathbf{-0.458119}$
	[-1.07634]	[-2.30621]
$\text{CSO}_{t-1\|t-4} * COP^{+}_{t-1\|t-4}$	0.358268	0.063077
	[0.94579]	[0.25571]
$\text{CNSO}_{t-1\|t-4} * COP^{-}_{t-1\|t-4}$	-0.506591	-0.556463
	[-0.69453]	[-1.17155]
$\text{CNSO}_{t-1\|t-4} * COP^{+}_{t-1\|t-4}$	-1.559435*	-0.203940
	[-1.71558]	[-0.34454]
$\text{CSO}_{t-1\|t-13} * COP^{-}_{t-1\|t-13}$	-0.053893	-0.093591
	[-0.57455]	[-1.53220]
$\text{CSO}_{t-1\|t-13} * COP^{+}_{t-1\|t-13}$	-0.068262	$\mathbf{-0.167638}$
	[-0.57418]	[-2.16536]

	CSO	CNSO
$\text{CNSO}_{t-1\|t-13} * COP^{-}_{t-1\|t-13}$	0.170409	0.115341
	[0.77161]	[0.80201]
$\text{CNSO}_{t-1\|t-13} * COP^{+}_{t-1\|t-13}$	0.157209	0.267650
	[0.42578]	[1.11318]
$\text{CSO}_{t-1} * COP^{-}_{t-1}$	-1.422508	**1.681862**
	[-1.19090]	[2.16223]
$\text{CSO}_{t-1} * COP^{+}_{t-1}$	**6.747638**	-0.504167
	[3.92086]	[-0.44988]
$\text{CNSO}_{t-1} * COP^{-}_{t-1}$	1.240640	-2.915932*
	[0.46219]	[-1.66818]
$\text{CNSO}_{t-1} * COP^{+}_{t-1}$	-3.340829	2.168390
	[-0.84059]	[0.83784]
DGS10_{t-1}	-6.231572	-3.686808
	[-1.43714]	[-1.30569]
DTWEXBGS_{t-1}	-0.884155	-0.400405
	[-1.36924]	[-0.95223]
EFFR_{t-1}	8.930646	1.819530
	[1.28358]	[0.40160]
GSPC_{t-1}	0.005746	-0.000189
	[0.34744]	[-0.01758]
NFCI_{t-1}	-6.926580	-16.19598
	[-0.31498]	[-1.13100]
STLFSI2_{t-1}	-0.020649	-0.706146
	[-0.00884]	[-0.46401]
TEDRATE_{t-1}	4.041245	-7.219257
	[0.35055]	[-0.96165]
VIX_{t-1}	0.023234	-0.092967
	[0.10018]	[-0.61555]
Adj. R^2	0.463826	0.262838
Akaike AIC	7.198870	6.340971
Schwarz SC	7.487189	6.629290
	Q-stat	p-value
Q(2)	2.86767	0.5784
Q(3)	12.60024	0.1236
Q(4)	14.45430	0.2674

[42] The bold values denote the 5% statistically significant coefficients, the asterisks denote the 10%.
the associated t-statistics for each variable is reported in the square bracket. The first line in the bottom reports the adjusted R-squared for each of the two equations composing the VARX model. Residuals do not show evidences of serial dependence as shown by the multivariate Portmanteau test, last three rows of the table where are reported the p-values of the test for their respective lags.

Figure 55: VAR model 4 diagnostic

Model 4

VAR Residual Serial Correlation LM Tests

Lag	LRE*stat	df	Prob.
1	15.07145	4	0.0046
2	4.948551	4	0.2926

VAR Residual Normality Tests
Orthogonalization: Cholesky (Lutkepohl)
Null Hypothesis: Residuals are multivariate normal

Component	Skewness	Chi-sq	df	Prob.*
1	0.037520	0.111916	1	0.7380
2	−0.253815	5.121564	1	0.0236
Joint		5.233480	2	0.0730

Component	Kurtosis	Chi-sq	df	Prob.
1	4.262059	31.65677	1	0.0000
2	4.019009	20.63777	1	0.0000
Joint		52.29454	2	0.0000

Component	Jarque-Bera	df	Prob.
1	31.76868	2	0.0000
2	25.75933	2	0.0000
Joint	57.52802	4	0.0000

*Approximate p-values do not account for coefficient estimation

Figure 56: VAR model 5 diagnostic

Model 5

VAR Residual Serial Correlation LM Tests

Lag	LRE*stat	df	Prob.
1	7.619506	4	0.1066
2	3.023528	4	0.5539

VAR Residual Normality Tests
Orthogonalization: Cholesky (Lutkepohl)
Null Hypothesis: Residuals are multivariate normal

Component	Skewness	Chi-sq	df	Prob.*
1	0.148066	1.742914	1	0.1868
2	−0.146997	1.717844	1	0.1900
Joint		3.460758	2	0.1772

Component	Kurtosis	Chi-sq	df	Prob.
1	3.868233	14.98234	1	0.0001
2	3.843290	14.13388	1	0.0002
Joint		29.11622	2	0.0000

Component	Jarque-Bera	df	Prob.
1	16.72526	2	0.0002
2	15.85172	2	0.0004
Joint	32.57698	4	0.0000

*Approximate p-values do not account for coefficient estimation

Figure 58: Rolling coefficients for Model 1, CSO equation

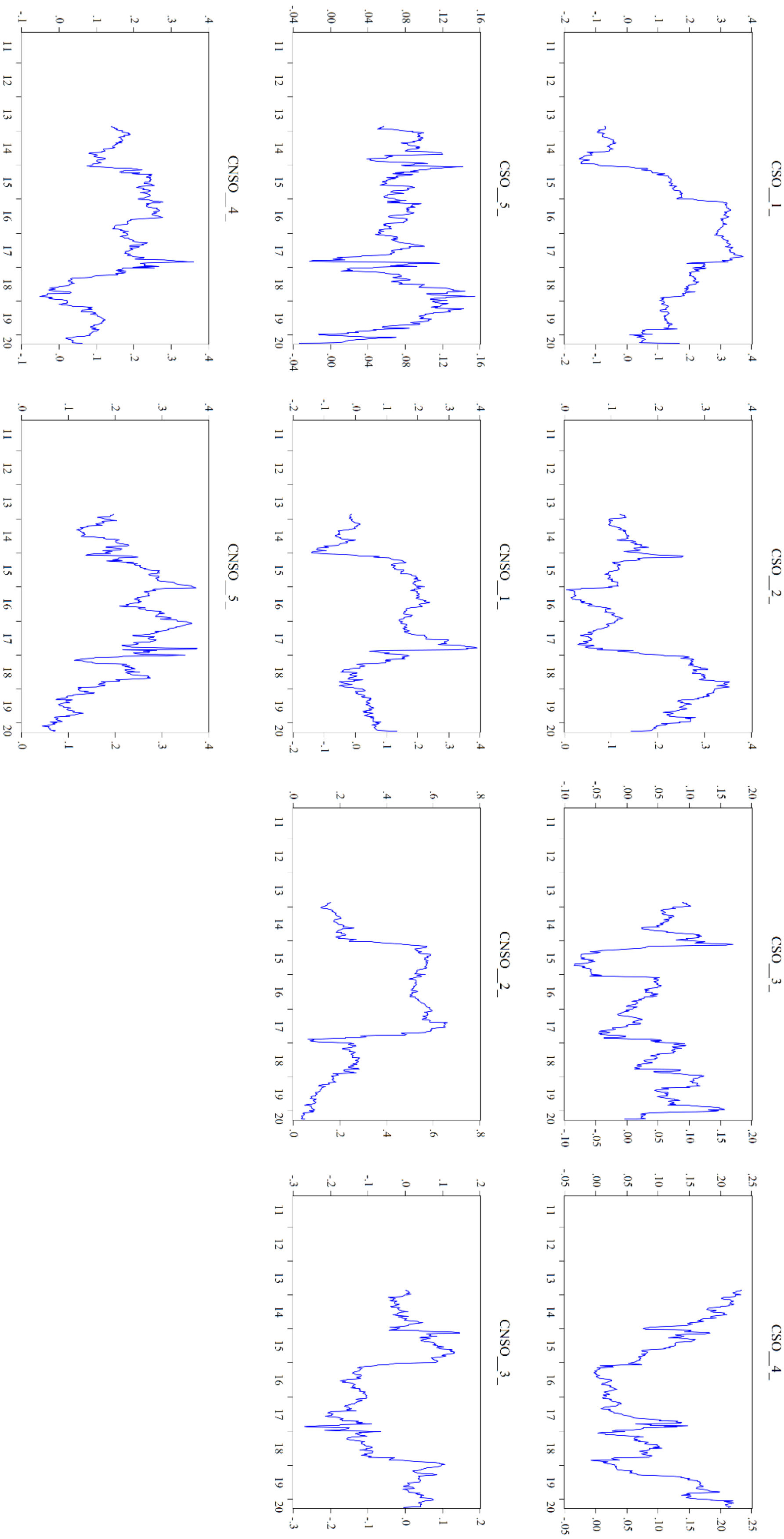

Figure 59: Rolling p-values for Model 1, CSO equation

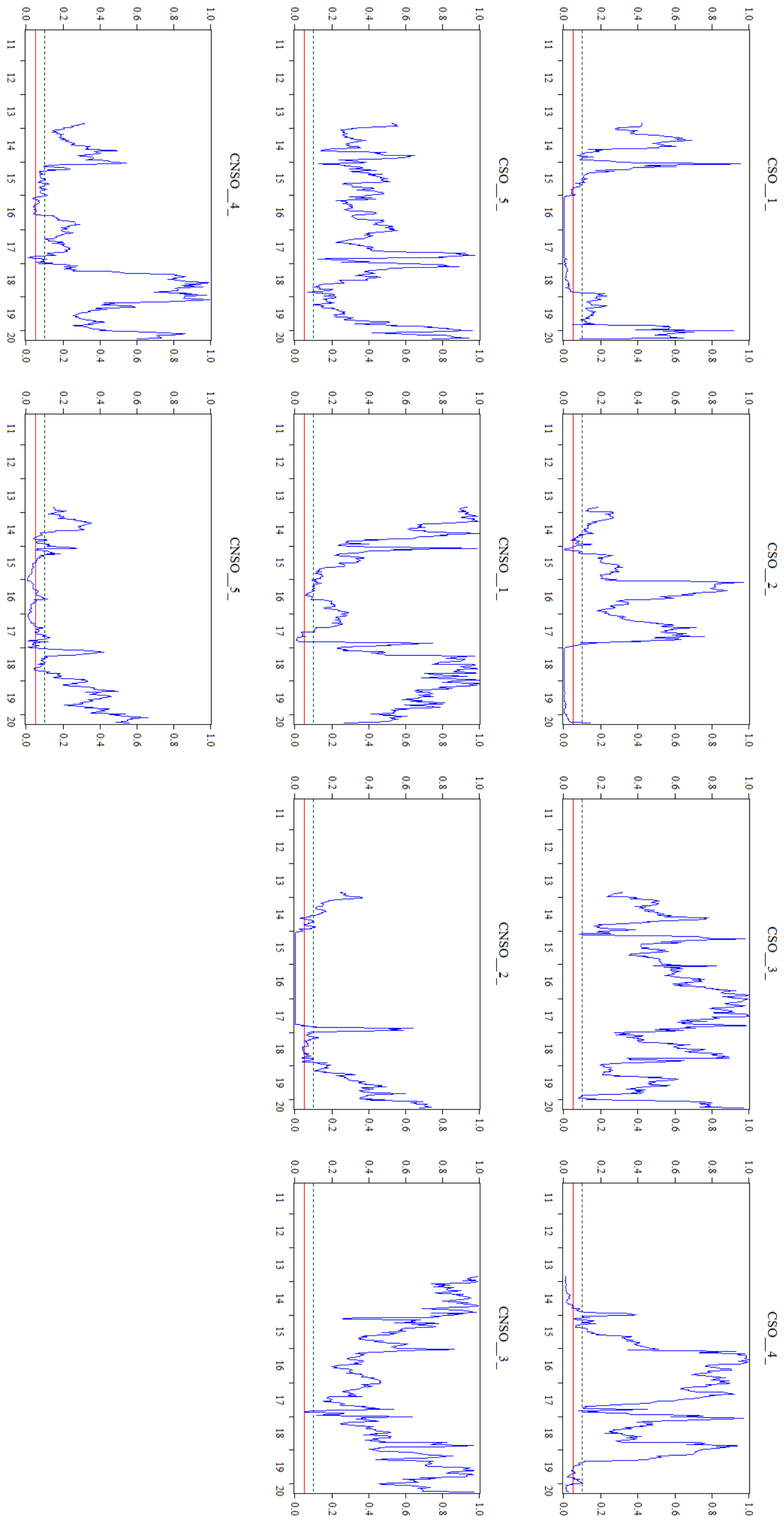

Figure 60: Rolling coefficients for Model 1, CNSO equation

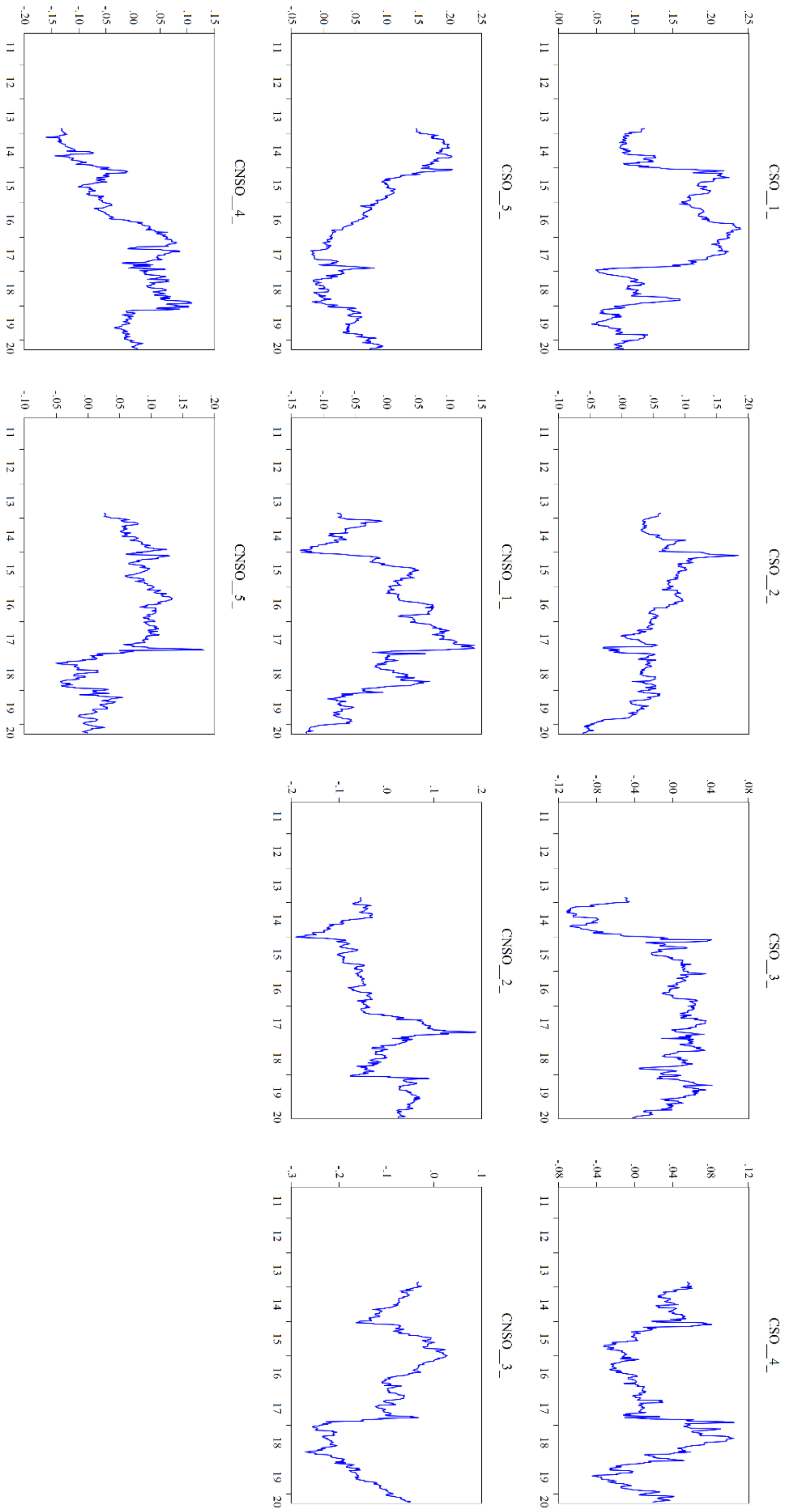

Figure 61: Rolling p-values for Model 1, CNSO equation

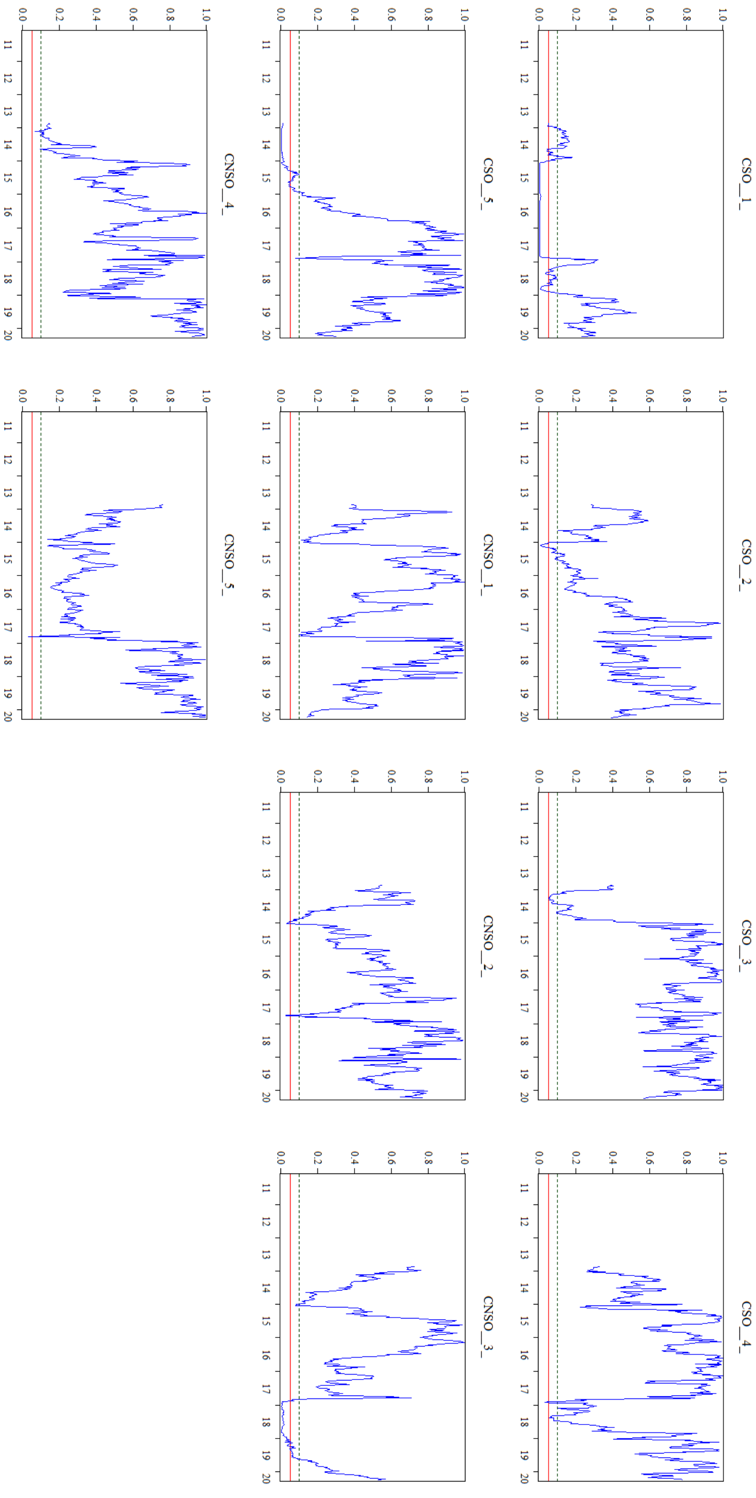

Figure 62: Rolling coefficients for Model 2, CSO equation

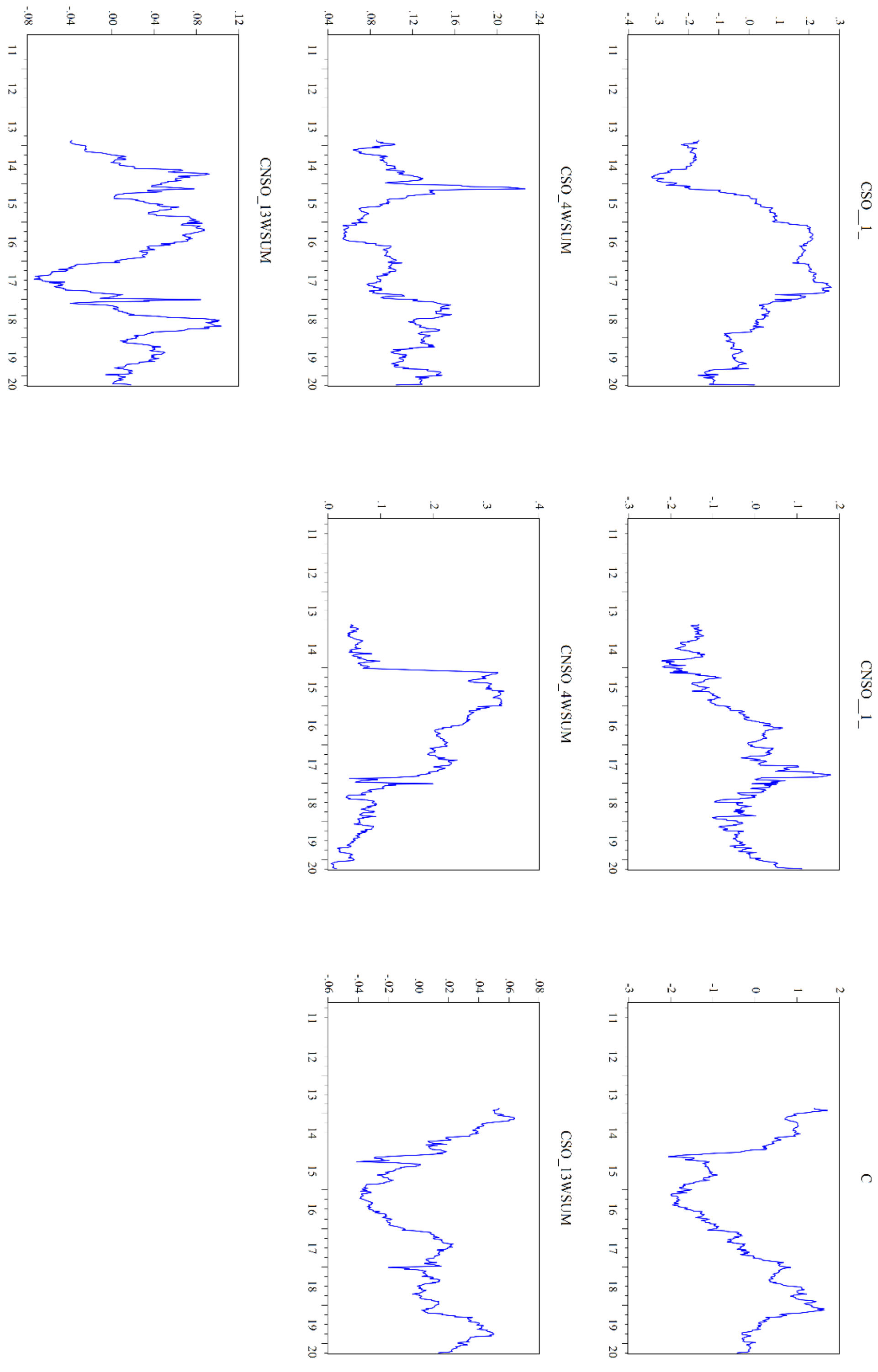

Figure 63: Rolling p-values for Model 2, CSO equation

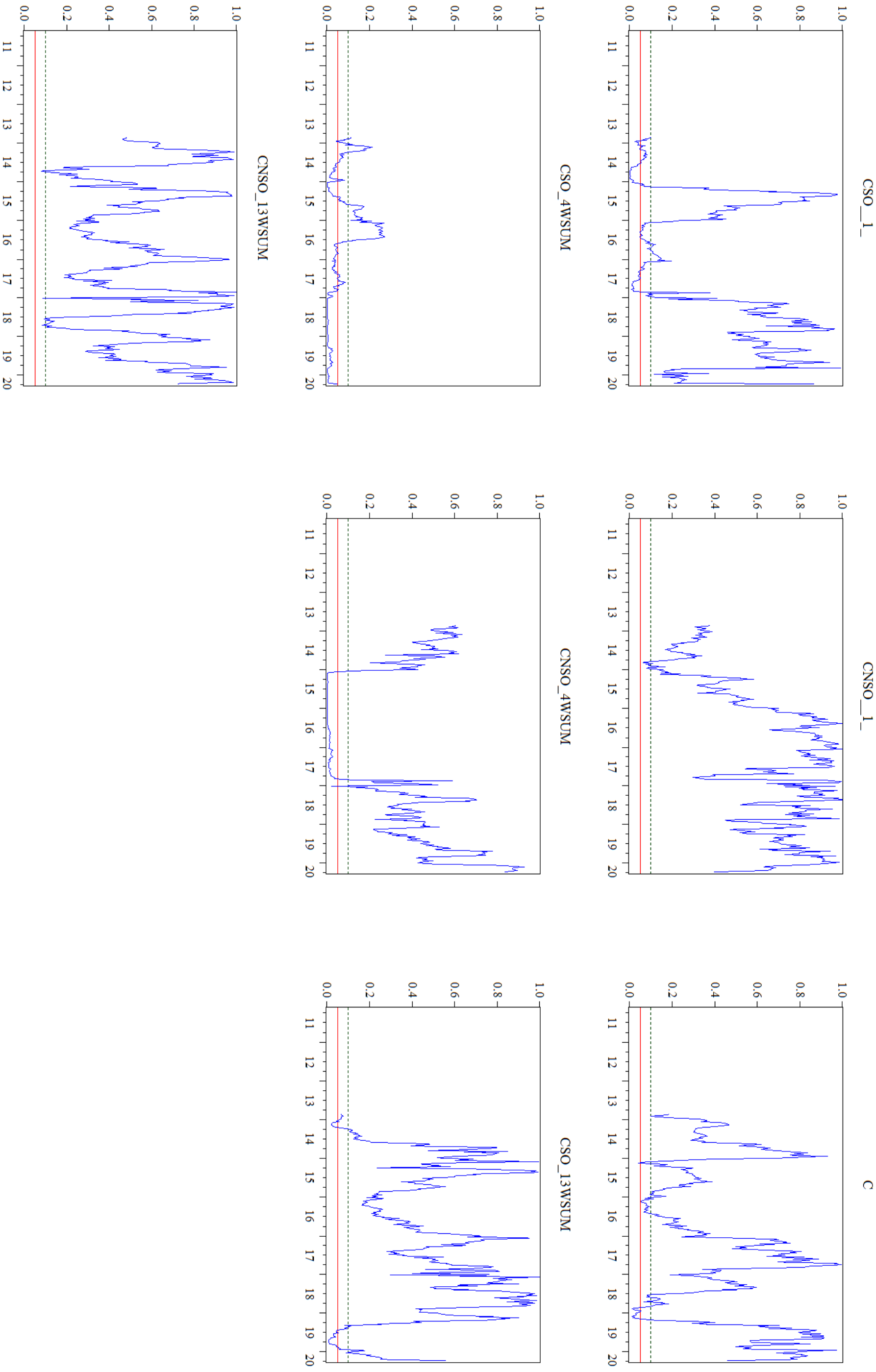

Figure 64: Rolling coefficients for Model 2, CNSO equation

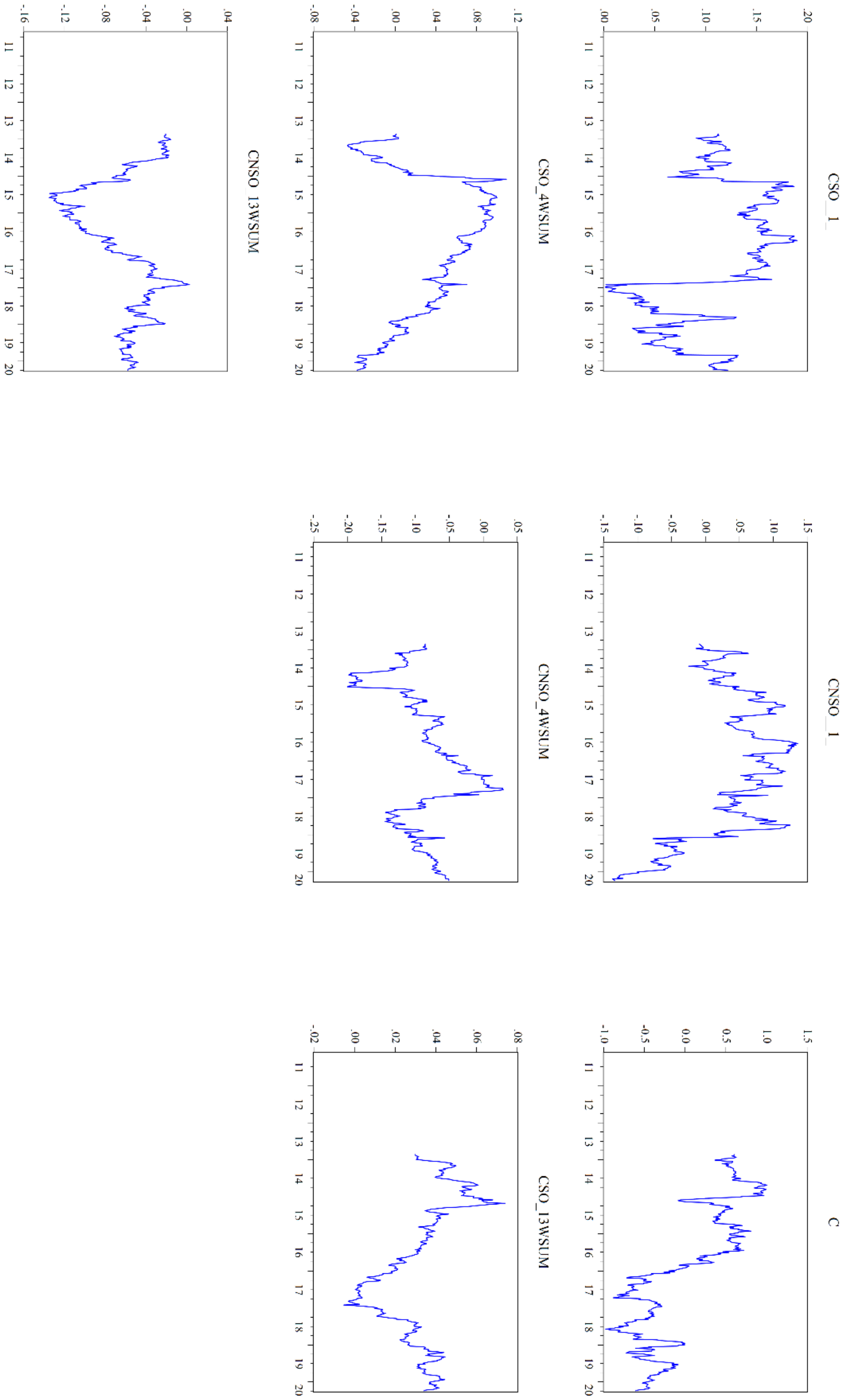

Figure 65: Rolling p-values for Model 2, CNSO equation

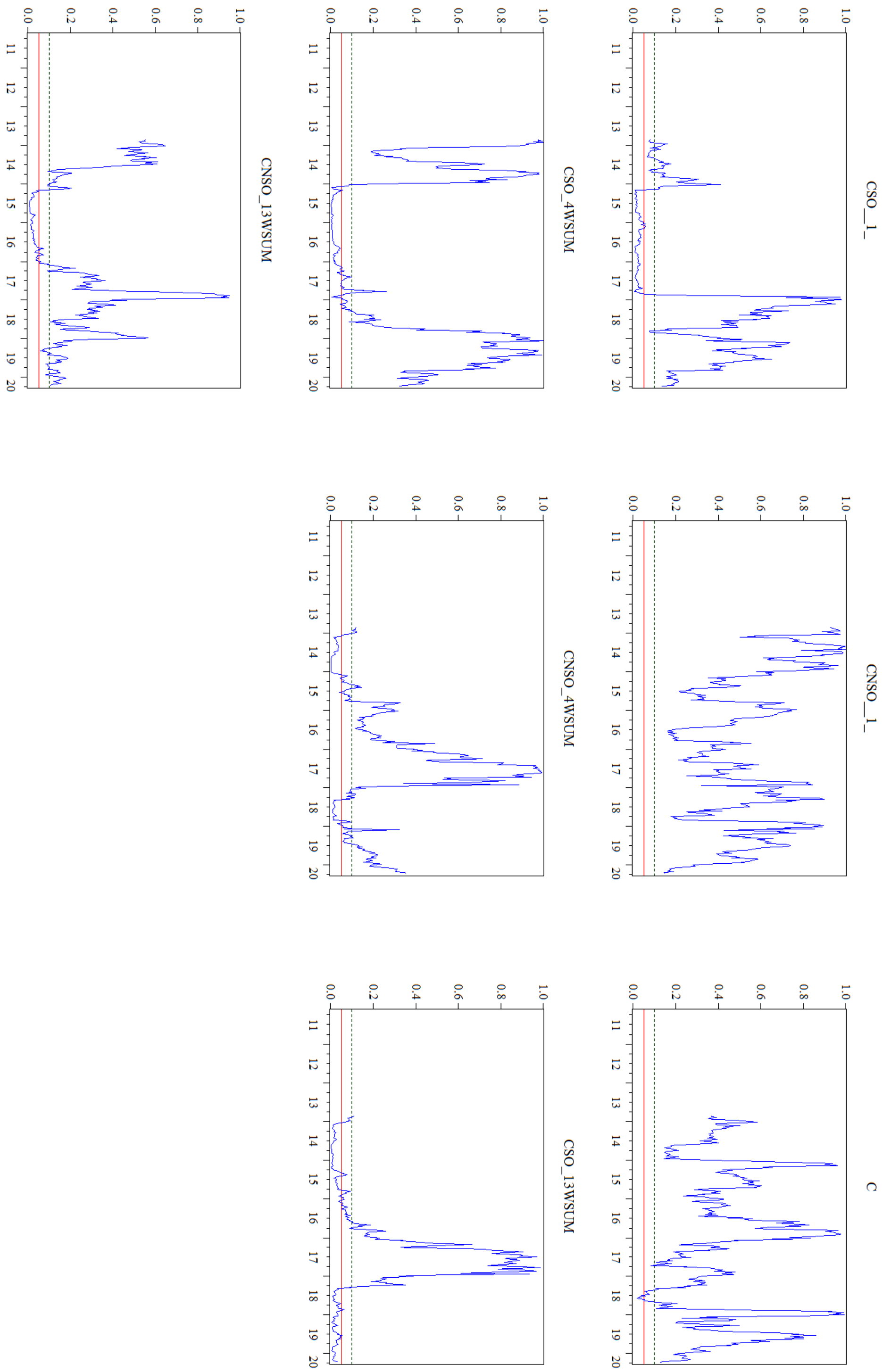

Figure 66: Rolling coefficients for Model 3, CSO equation

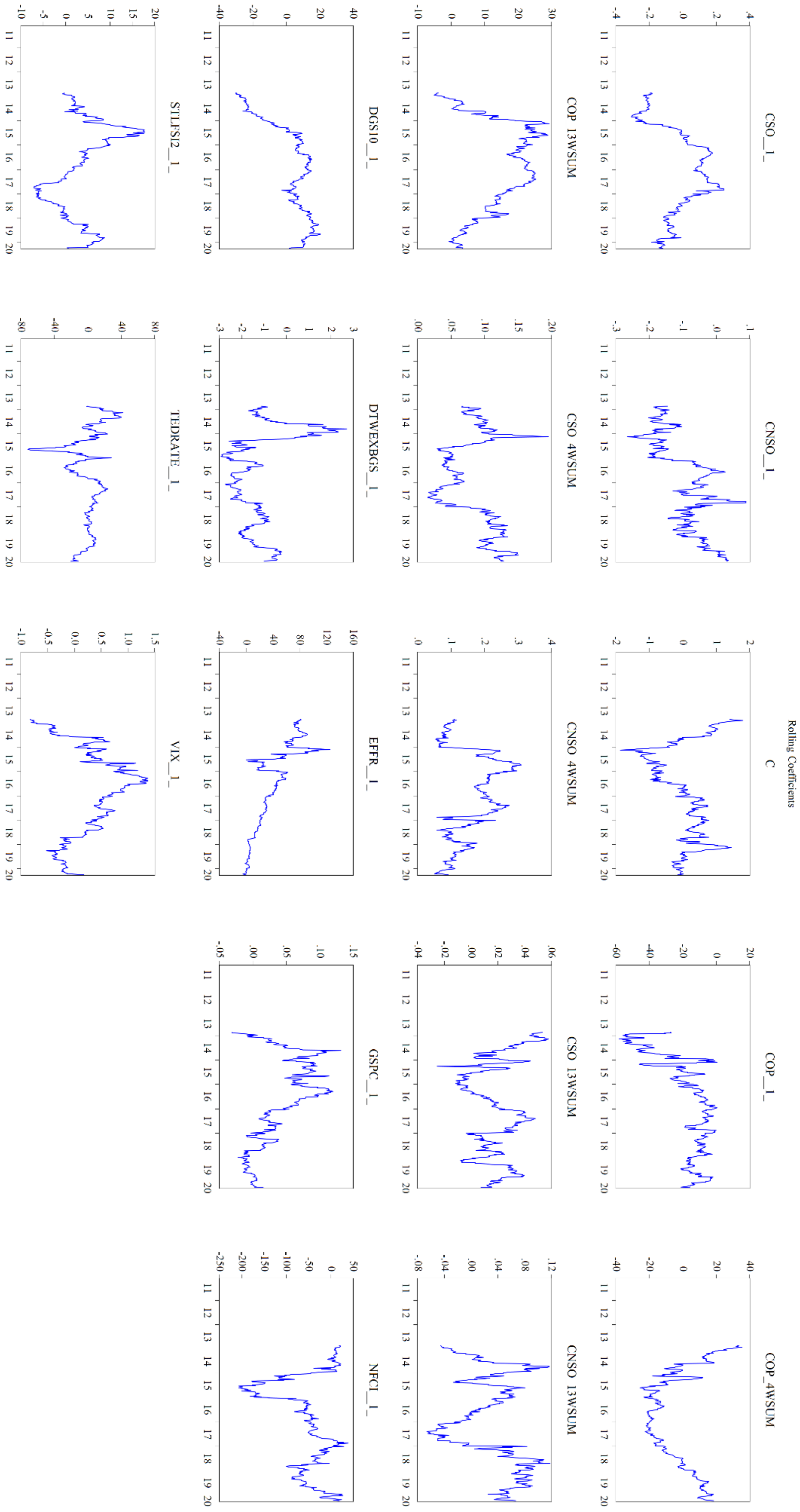

Figure 67: Rolling p-values for Model 3, CSO equation

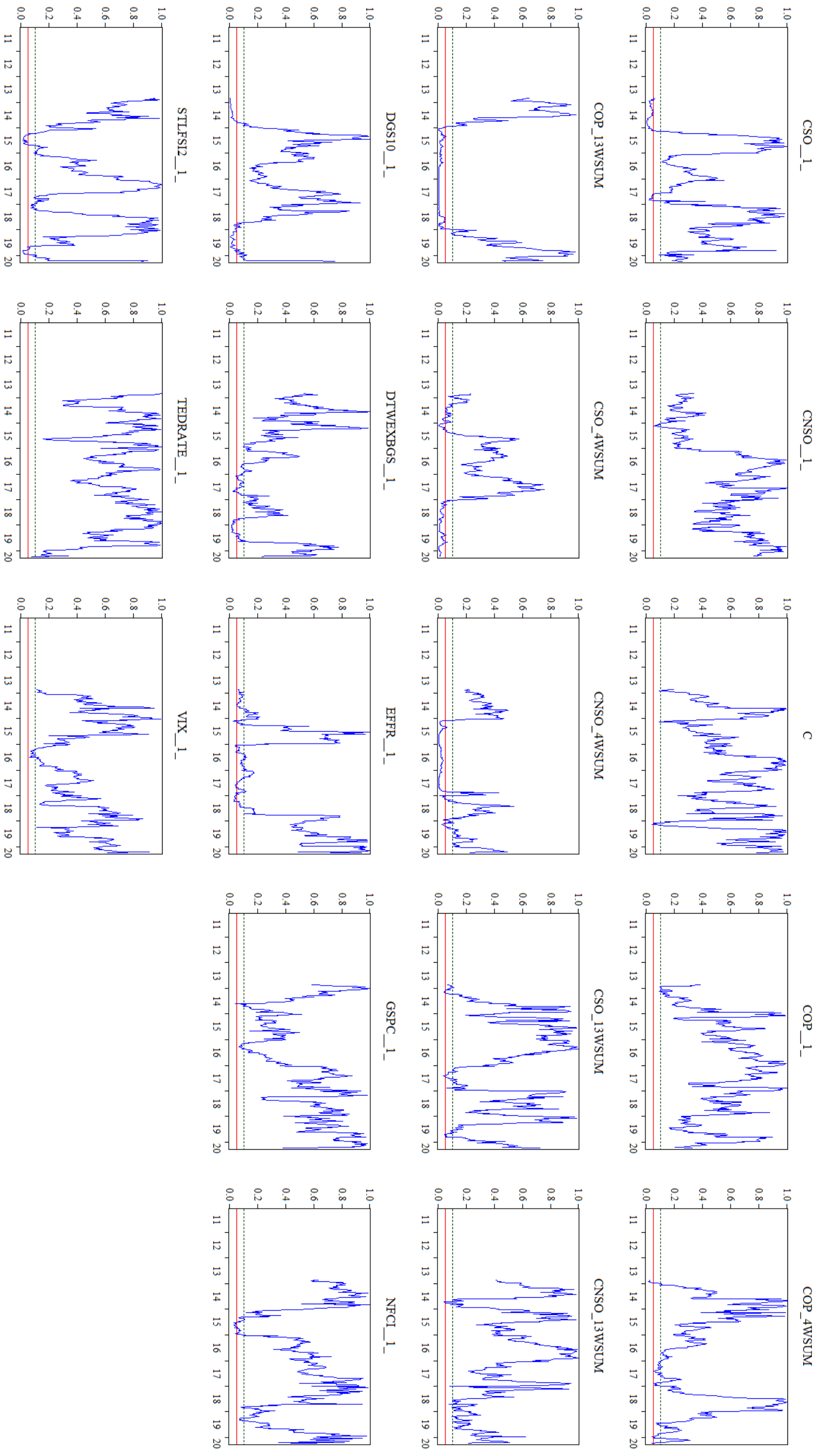

Figure 68: Rolling coefficients for Model 3, CNSO equation

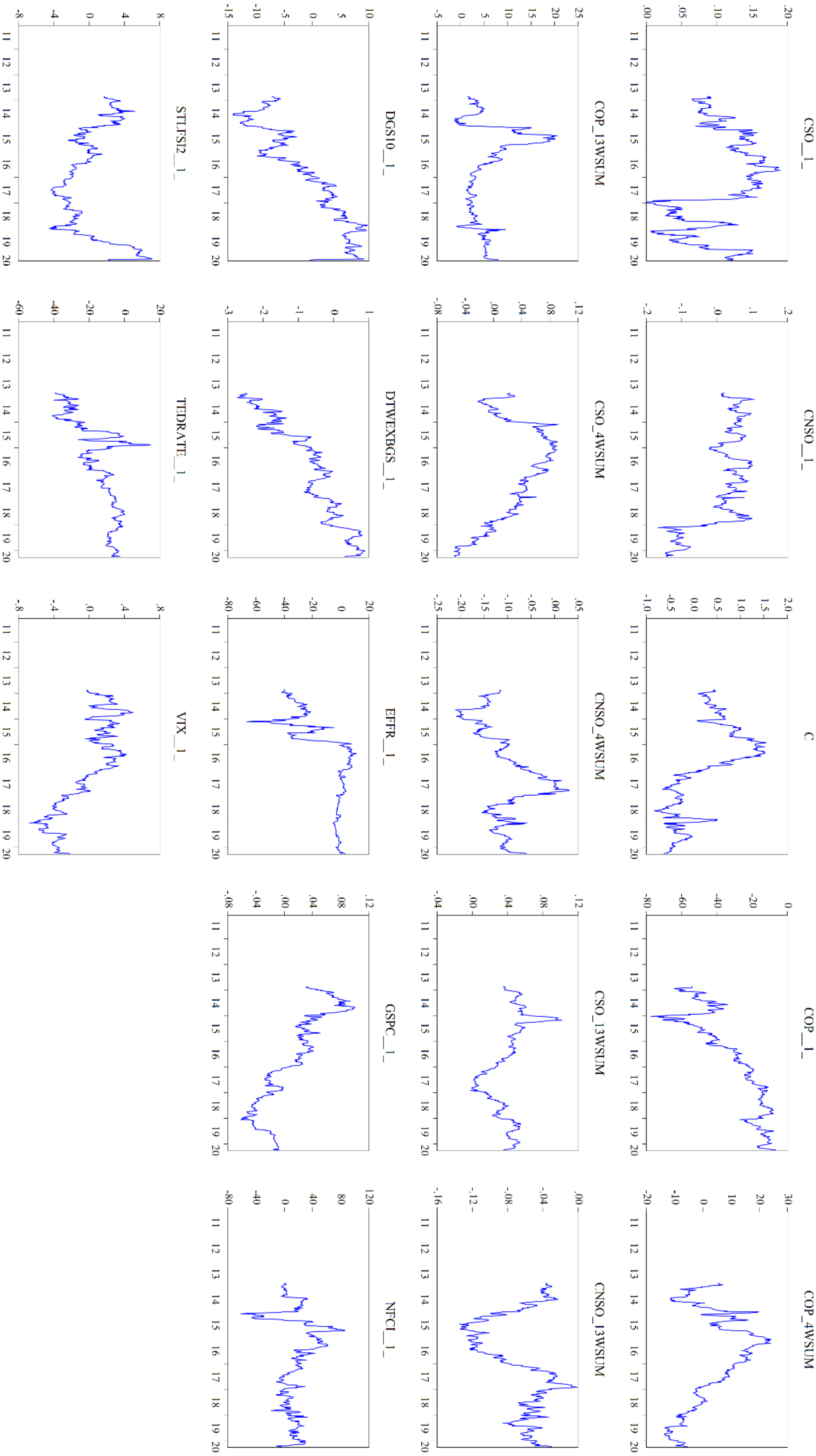

Figure 69: Rolling p-values for Model 3, CNSO equation

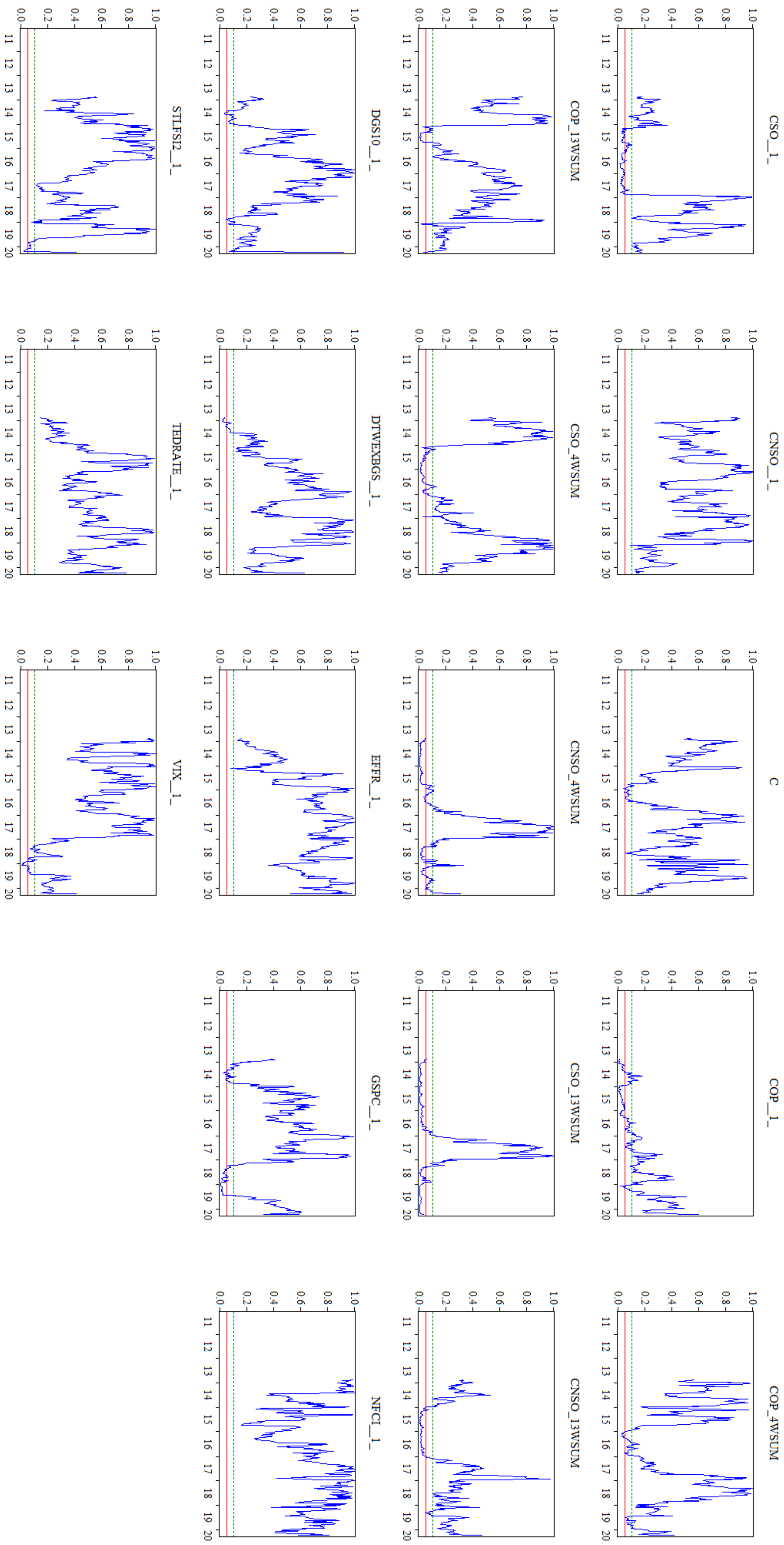

Figure 70: Rolling coefficients for Model 4, CSO equation

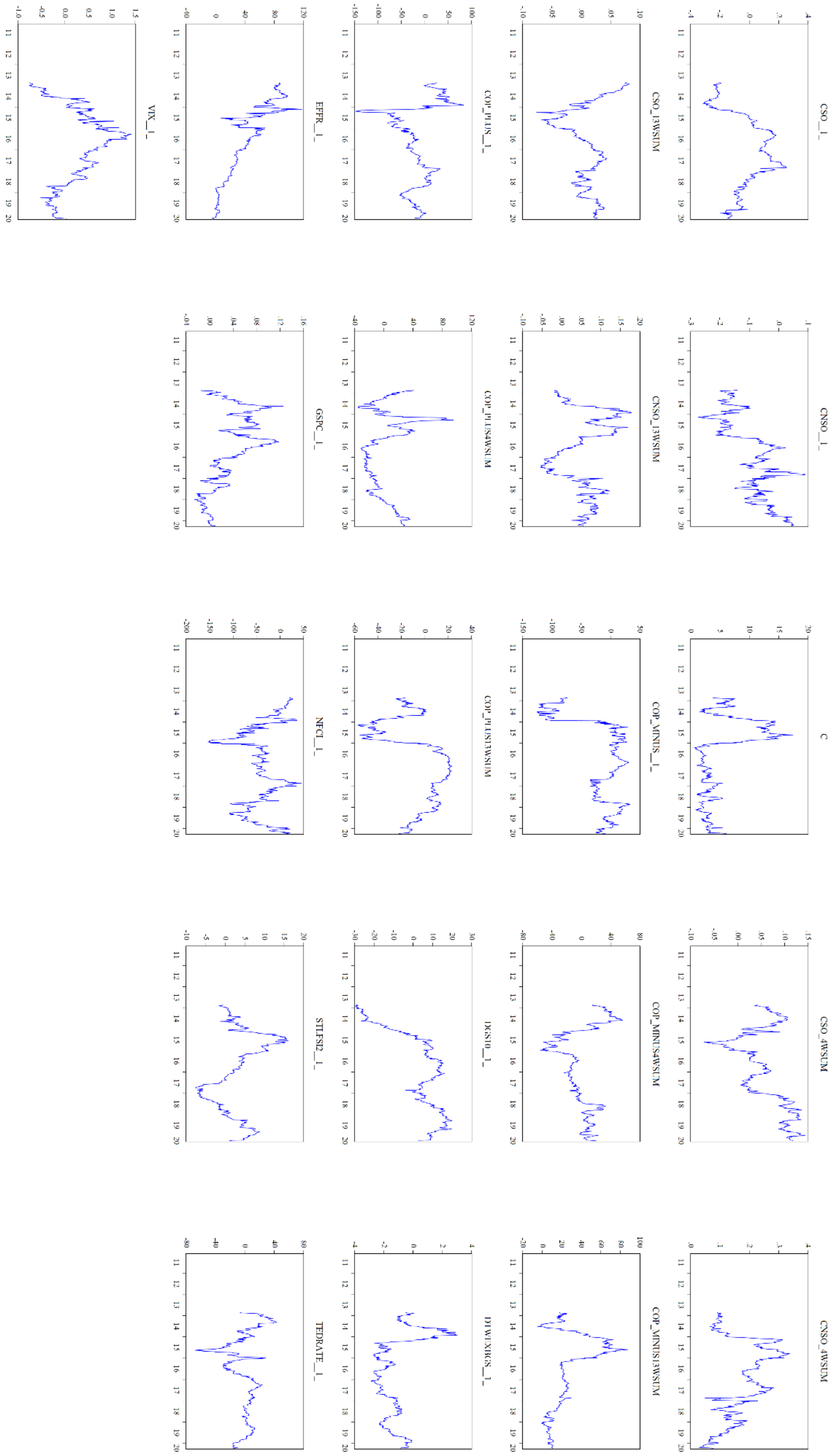

Figure 71: Rolling p-values for Model 4, CSO equation

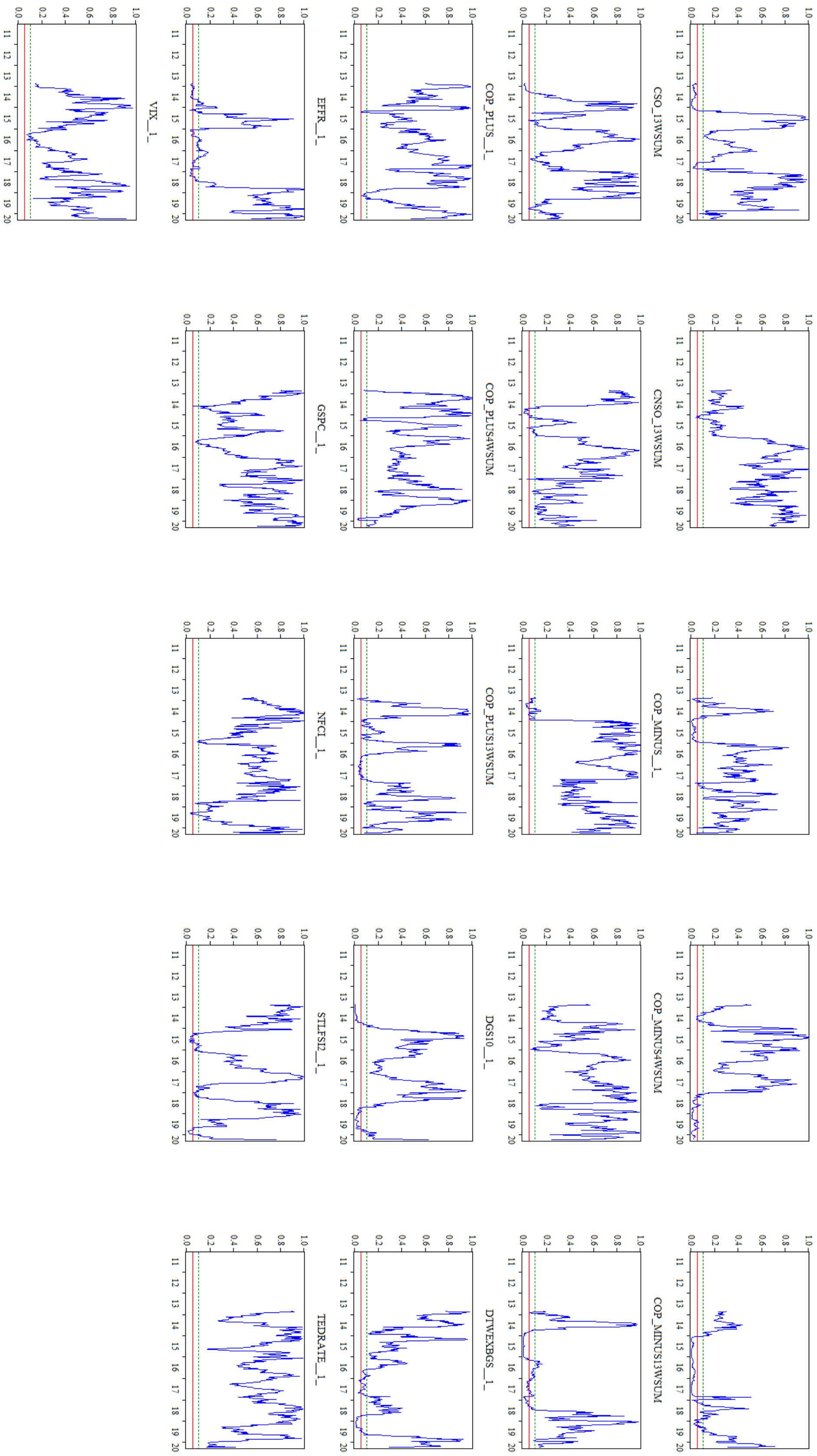

Figure 72: Rolling coefficients for Model 4, CNSO equation

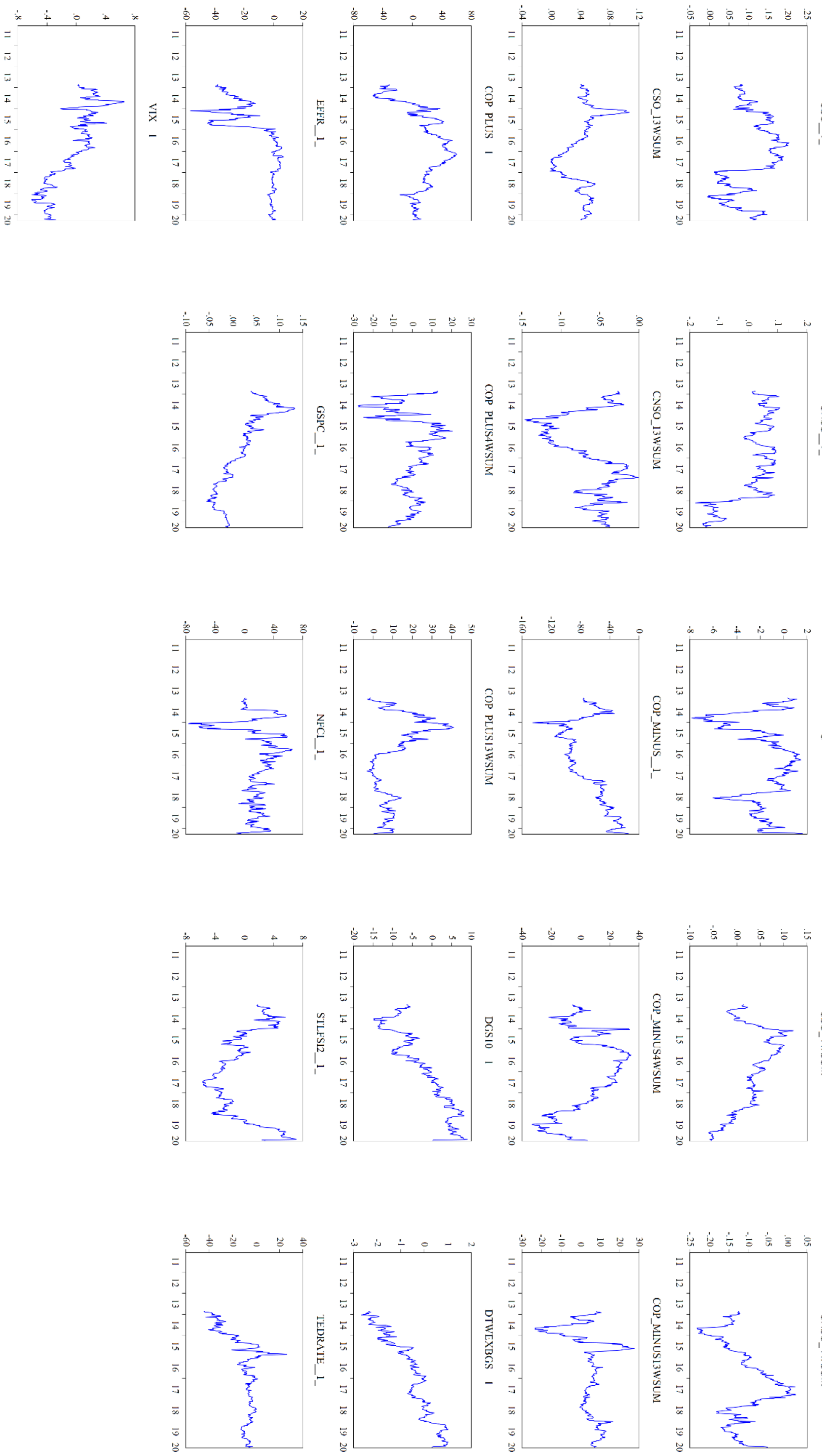

Figure 73: Rolling p-values for Model 4, CNSO equation

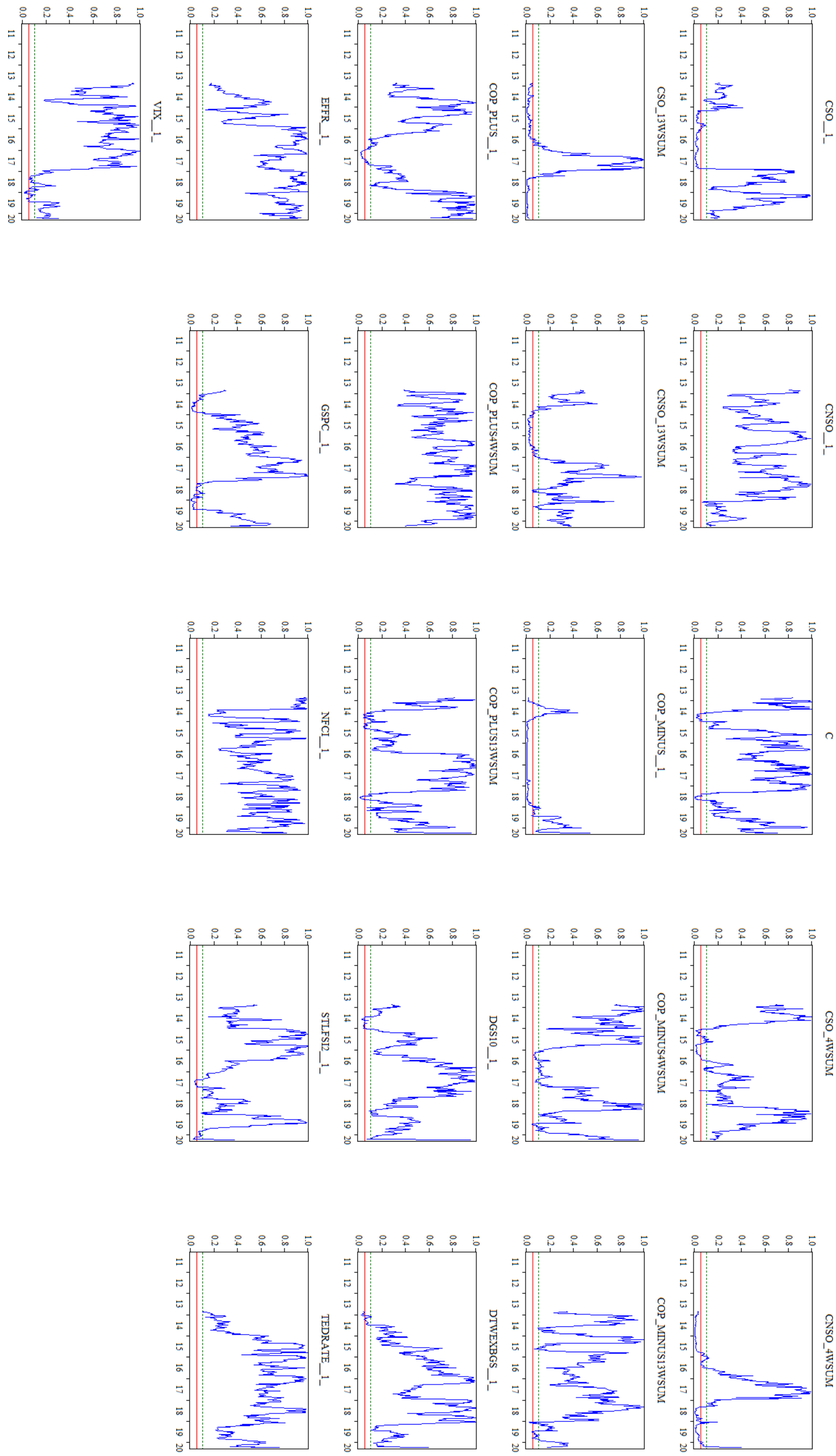

Figure 74: Rolling coefficients for Model 5, CSO equation

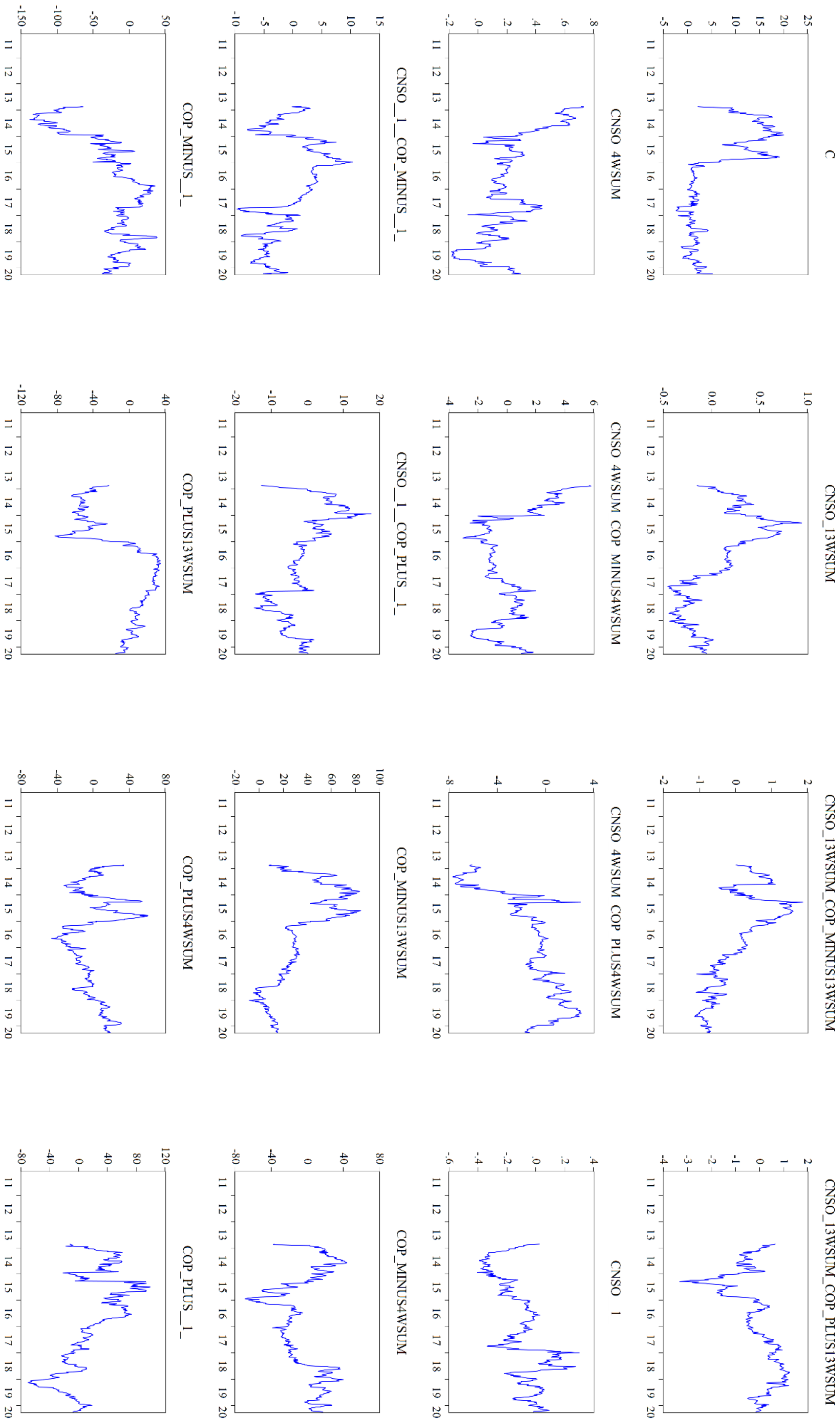

Figure 75: Rolling coefficients for Model 5, CSO equation

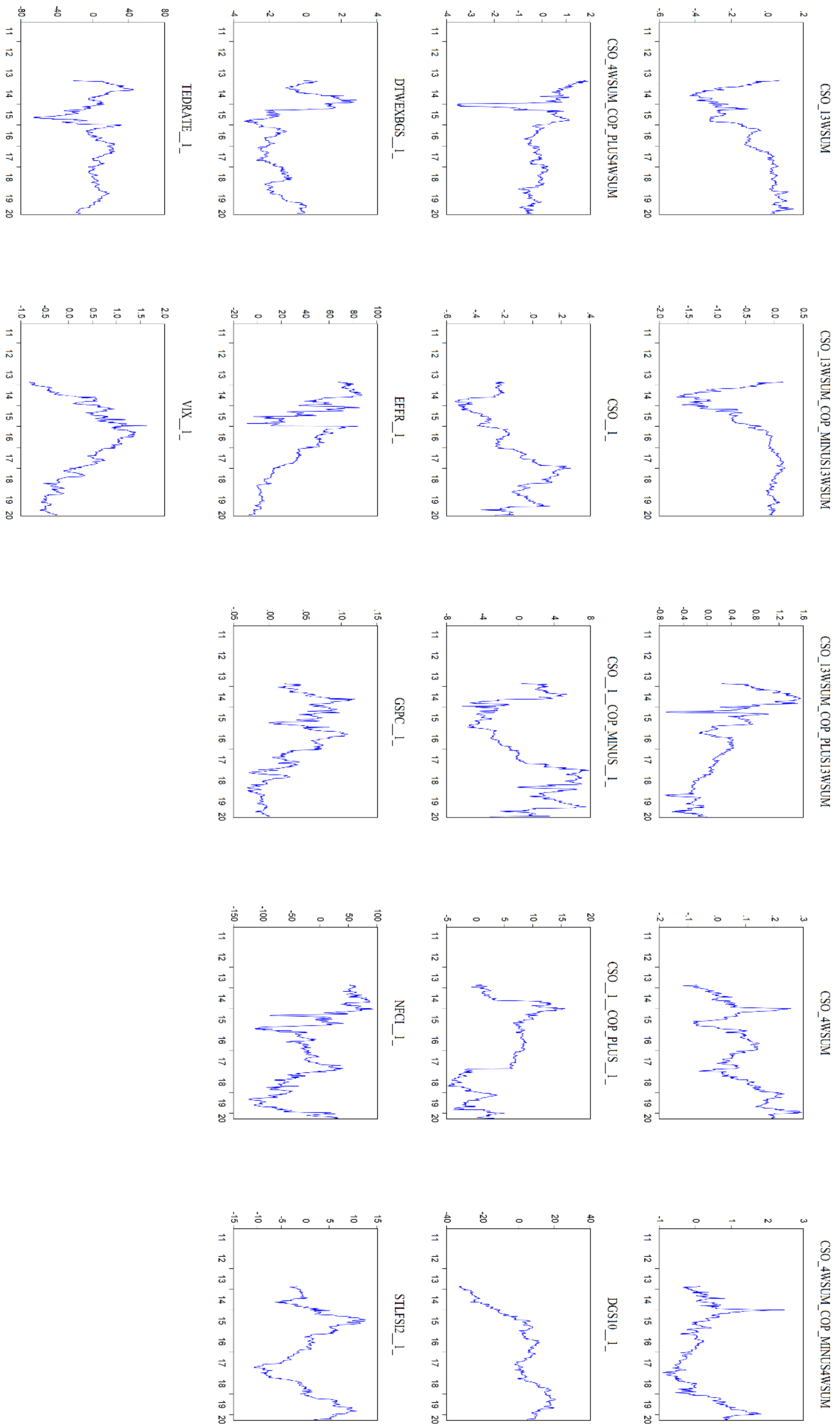

Figure 76: Rolling p-values for Model 5, CSO equation

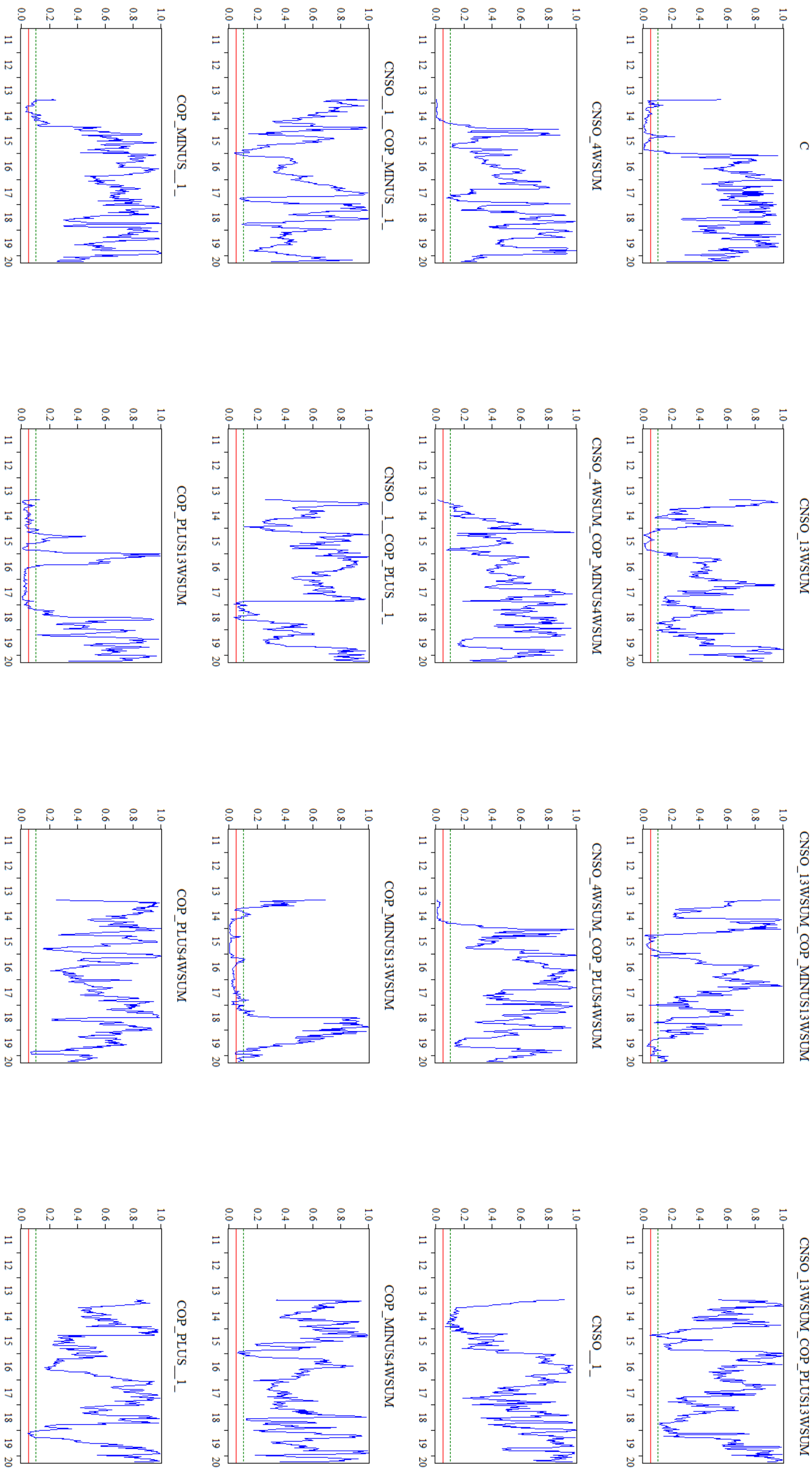

Figure 77: Rolling p-values for Model 5, CSO equation

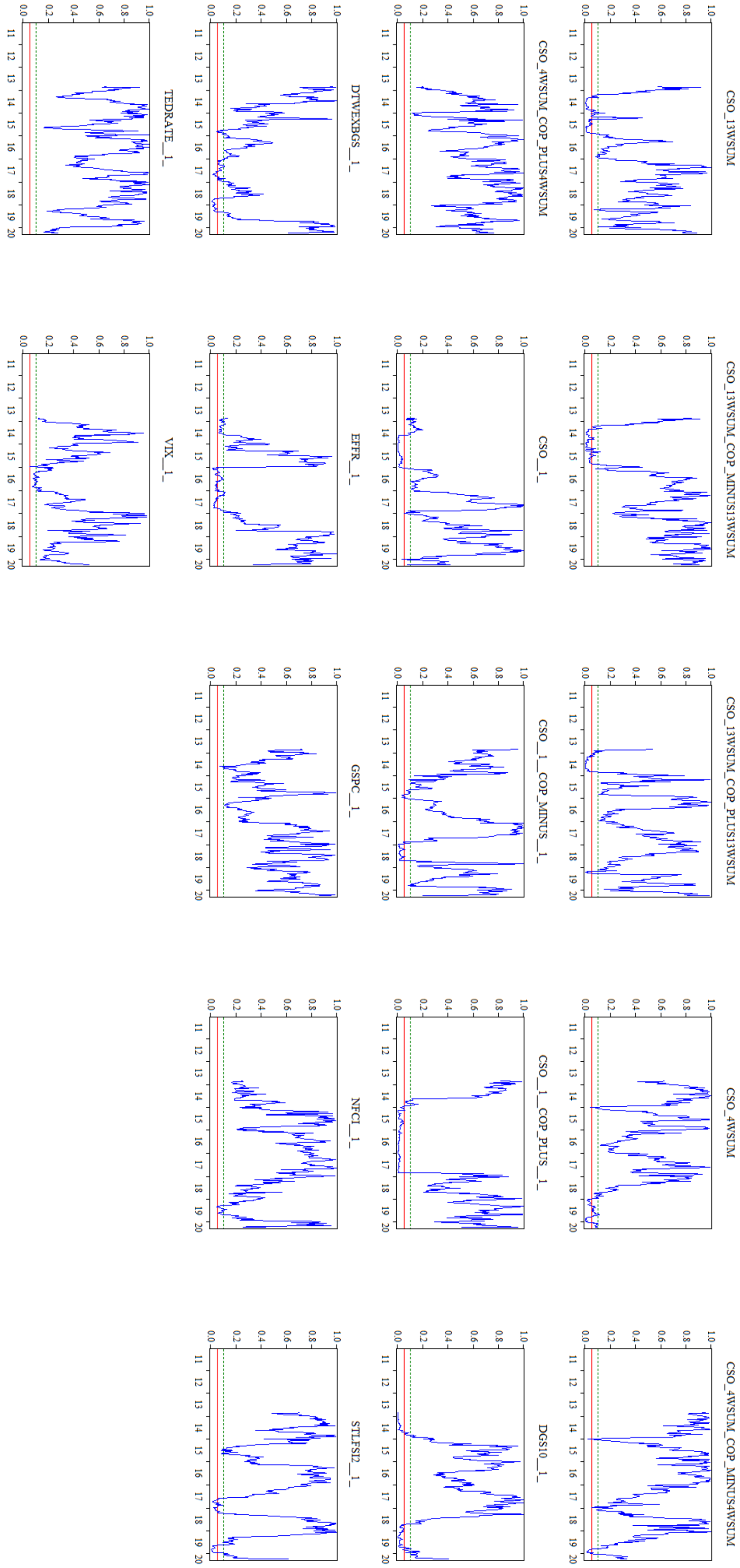

Figure 78: Rolling coefficients for Model 5, CNSO equation

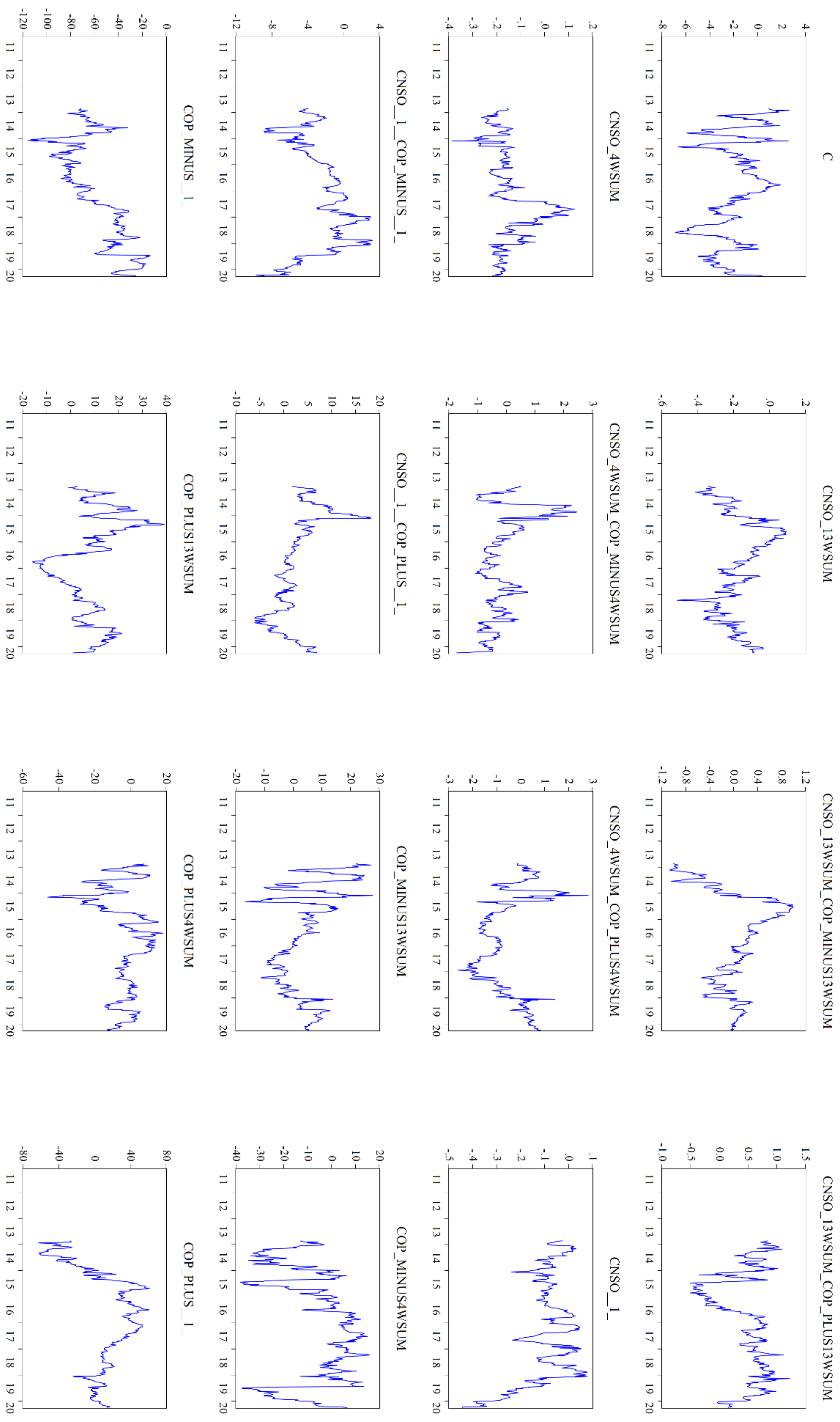

Figure 79: Rolling coefficients for Model 5, CNSO equation

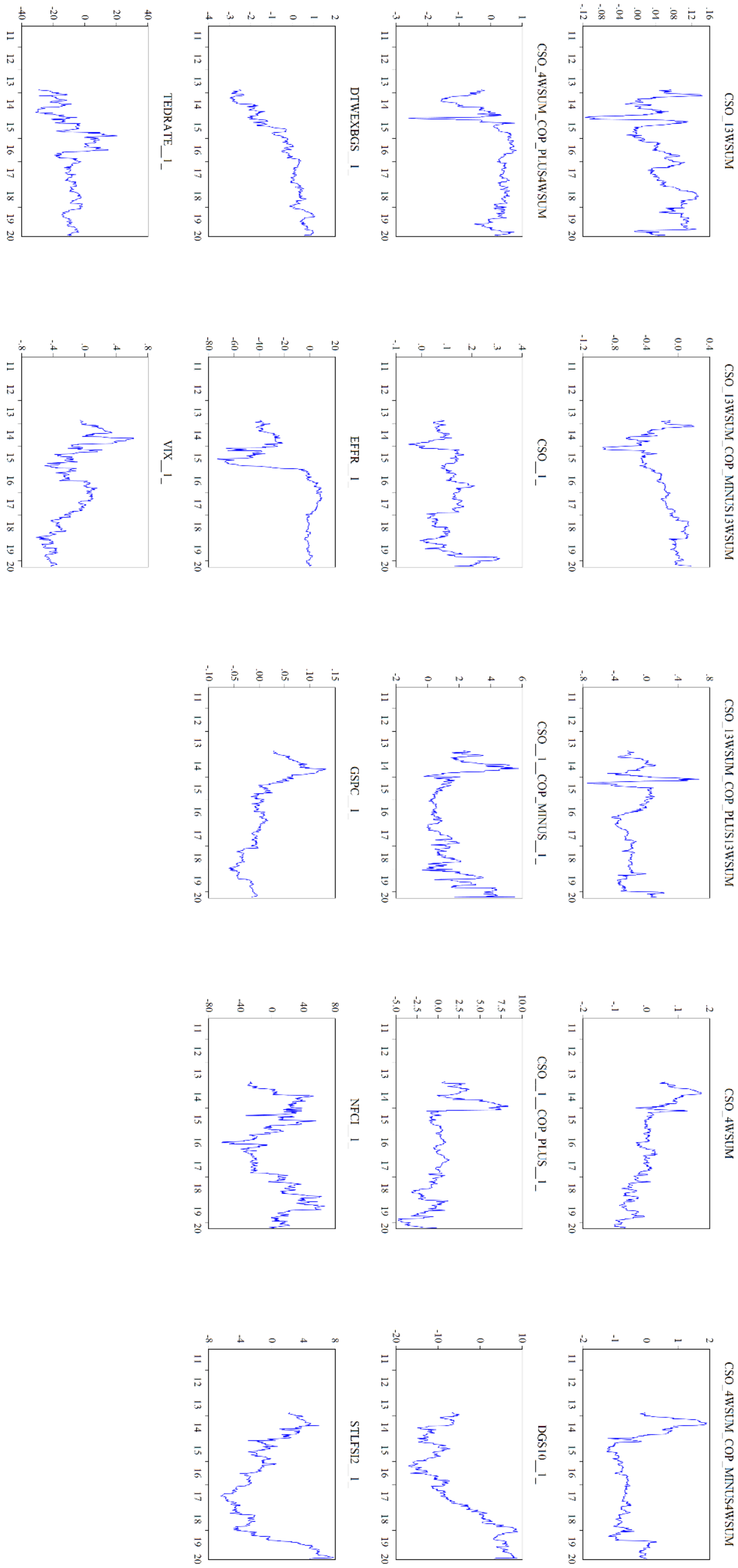

Figure 80: Rolling p-values for Model 5, CNSO equation

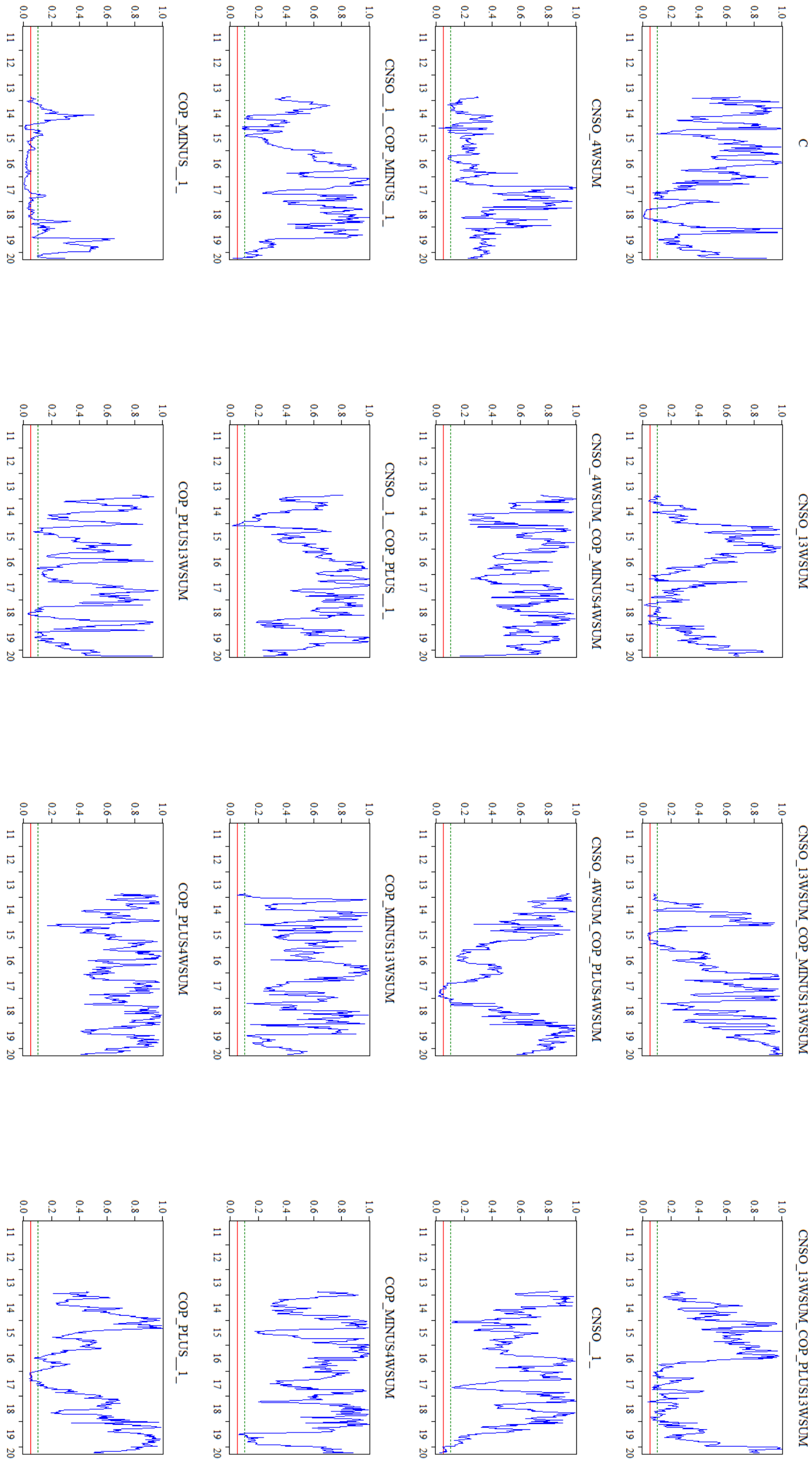

Figure 81: Rolling p-values for Model 5, CNSO equation