Universidade de Évora - Escola de Ciências Sociais

Mestrado em Gestão

Área de especialização | Finanças

Dissertação

Os determinantes do preço do petróleo crude e o papel da especulação financeira

Filipe Miguel de Mira Ferreira Marques Cachapa

Orientador(es) | Andreia Dionísio
Fernanda Peixe

Évora 2019

Universidade de Évora - Escola de Ciências Sociais

Mestrado em Gestão

Área de especialização | Finanças

Dissertação

Os determinantes do preço do petróleo crude e o papel da especulação financeira

Filipe Miguel de Mira Ferreira Marques Cachapa

Orientador(es) | Andreia Dionísio

Fernanda Peixe

Évora 2019

A dissertação foi objeto de apreciação e discussão pública pelo seguinte júri nomeado pelo Diretor da Escola de Ciências Sociais:

- Presidente | Marta da Conceição Cruz Silvério (Universidade de Évora)
- Vogal | Jacinto António Setúbal Vidigal da Silva (Universidade de Évora)
- Vogal-orientador | Fernanda Peixe (Universidade de Évora)

Évora 2019

Resumo

O preço do barril de petróleo crude representa uma série temporal de elevada volatilidade, que tem sido alvo de estudo por parte de diversos autores e investigadores.

O estudo presente nesta dissertação de mestrado tem como objetivos: encontrar variáveis económicas e financeiras que influenciem significativamente o preço do barril de petróleo e ajudem a explicar as variações observáveis nessa série temporal; explorar o papel da especulação financeira relativamente à *commodity* em questão.

Para tal, recorreu-se a técnicas de modelação financeira que permitiram analisar a influência de variáveis fundamentadas na literatura. Obteve-se um modelo VECM relativo às variáveis que revelaram poder ter uma relação de longo prazo com a variável dependente, e ainda um modelo ECM onde foram incluídas outras variáveis.

As variáveis que se revelaram significativas foram o Consumo de Gás, o Consumo de Carvão, o Consumo de Petróleo e os PIB's dos EUA e China.

Palavras-chave:

Determinantes; Preços do Petróleo; Especulação; Modelo VECM; Modelo ECM.

Abstract

The Determinants of Crude Oil Prices and the Role of Financial Speculation

The price of the crude oil barrel represents a timeseries of high volatility, which has been studied by many authors and investigators.

The research contained in this master's dissertation has the following objectives: to reach economic and financial variables which significantly influence the oil barrel price and help explain that series' observable variations; explore the role of financial speculation regarding the commodity in question.

In order to do that, financial modeling techniques that allowed to analyze the influence of literature-based variables were applied. A VECM model regarding the variables that seemed to present a long-term relation with the dependent variable was obtained, and also an ECM model where other variables were included.

The variables that showed a significant effect were the World Gas Consumption, the World Coal Consumption, the World Oil Consumption and the USA's and China's GDP's.

Keywords:

Determinants; Crude Oil Prices; Speculation; VECM model; ECM model;

Agradecimentos

Devo começar por agradecer às professoras Andreia Dionísio e Fernanda Peixe, pela orientação prestada, e pela paciência e disponibilidade que demonstraram ao longo de todo este processo de elaboração da dissertação.

Tenho também de deixar o meu agradecimento a todos os professores que me transmitiram conhecimentos e forneceram bases para poder elaborar este trabalho, quer os professores do mestrado em Gestão, quer os professores da licenciatura em Matemática Aplicada à Economia e Gestão.

Agradeço à minha família, e em especial aos meus pais e irmãos, que me ajudaram a tornar na pessoa determinada e ambiciosa que sou, e que me permitiram ter o percurso académico que até este momento tenho vindo a desenvolver.

Agradeço também aos meus amigos que, de uma forma ou de outra, me apoiam e inspiram a levar a cabo os objetivos que traço de forma clara e confiante, como foi o caso da elaboração desta dissertação.

Por fim, ainda que não menos relevante, agradeço à minha namorada Teresa, pela constante paciência e conselhos, e sobretudo, por ser o meu grande apoio ao longo desta etapa.

Contents

List of Appendices

List of Figures

List of Tables

List of Abbreviations

ADF – Augmented Dickey-Fuller

AIC – Akaike's Information Criterion

ARCH – Autoregressive Conditional Heteroskedasticity

BP – British Petroleum

DF – Dickey-Fuller

ECM – Error Correction Model

FRED - Federal Reserve Bank of St. Louis' Federal Reserve Economic Data

FPE – Final Prediction Error

GDP – Gross Domestic Product

HQIC – Hannan and Quinn's Information Criterion

OECD – Organization for Economic Cooperation and Development

OLS – Ordinary Least Squares

OPEC - Organization of Petroleum Exporting Countries

RESET – Ramsey's Regression Equation Specification Error Test

SBIC – Schwarz's Bayesian Information Criterion

USA – United States of America

VAR – Vector Autoregressive

VECM – Vector Error Correction Model

1 – Introduction

1.1 – Theme Justification

Crude oil is an energy source with limited capacity when it comes to its availability and obtainability. Nowadays, it is the main source of primary energy used by man, and therefore it plays an increasingly crucial role on our day-to-day life. The dependency that human beings have developed for this fossil fuel has become undeniable, and the fact that crude oil is both a limited and a not-renewable resource, which is being consumed at a progressively increased rhythm, may certainly have substantial consequences on economy and finance on a global scale.

Taking these facts under consideration, there are multiple questions regarding crude oil that arise and are interesting to analyze from an economical and financial point of view, such as the determination of its price. It is easy for one to realize that the price of crude oil barrels changes every day, at every hour and every minute. So, which are, after all, the determinants that account for those variations?

The goal of the elaboration of the research contained in the present dissertation is to try to understand which are the main variables that influence crude oil prices and how they explain the high volatility of those prices.

Additionally, we take into account the role of financial speculation on crude oil prices, since this is one the most thought-provoking – yet difficult to model – topics that can be explored in this context.

1.2 – Problem Formulation, Objectives and Hypotheses

The fact that crude oil prices constitute an extremely volatile time series leads to multiple authors and researchers being interested in studying, analyzing and trying to develop techniques that allow them to conclude which are the main determinants that influence the values that can be observed in the series throughout time.

Since this is a time series whose values exert a great impact upon the development and evolution of economy and financial markets on a global scale, therefore influencing and conditioning the value of many other series and variables, it seems quite clear that determining the main variables that explain its evolution is a problem for which contributions are highly relevant.

Even though the goals of the research that sustains this dissertation do not consist on producing prediction techniques, by determining the key factors that influence the crude oil barrel prices we have more information that can be used to anticipate the future values of the series.

Thus, the problem in question lays on the urge for determining the main factors of the crude oil prices time series, for the determination of those variables may allow us to better understand the series itself as well as the way that markets may anticipate the changes that crude oil prices may suffer at different time points.

The main objective of the present research is to understand which are the determinants of crude oil prices. In order to fulfill that main objective, three specific objectives have also been defined, which are:

➢ Determining the nature of the influence of the significant variables on crude oil prices;

➢ Determining the extension of the impact caused by those variables;

➢ Analyzing the role of financial speculation on the crude oil prices' determination.

The hypotheses that were taken under consideration so that we could aim at accomplishing the already mentioned objectives were the following:

➢ Hypothesis #1: Higher values of Worldwide Oil Reserves lead to lower Crude Oil Prices.

➢ Hypothesis #2: Higher values of Worldwide Oil Production lead to lower Crude Oil Prices;

➢ Hypothesis #3: Higher values of Worldwide Oil Consumption lead to higher Crude Oil Prices.

➢ Hypothesis #4: Higher values of Worldwide Refining Capacity lead to lower Crude Oil Prices.

➢ Hypothesis #5: Higher values of Worldwide Gas Consumption lead to lower Crude Oil Prices.

➢ Hypothesis #6: Higher values of Worldwide Coal Consumption lead to lower Crude Oil Prices.

➢ Hypothesis #7: Higher values of the USA's Gross Domestic Product lead to higher Crude Oil Prices.

➢ Hypothesis #8: Higher values of China's Gross Domestic Product lead to higher Crude Oil Prices.

➢ Hypothesis #9: Periods of apparently higher financial speculation activity in the markets can lead to both increases or decreases in Crude Oil Prices.

It is important to mention that this last hypothesis, regarding financial speculation, will be approached slightly aside from the others, since it concerns a question that is much more difficult to measure on a quantitative level.

1.3 – Approach

With the goal of understanding which are the most significant determinants of crude oil prices, data regarding the prices and its possible determinants was extracted from the BP Statistical Review of World Energy and the Federal Reserve Bank of St. Louis' Federal Reserve Economic Data (FRED). The time period taken into consideration was the period between 1980 and 2017 and the explanatory variables were divided into three different categories according to their nature.

Since the variables used consisted on time series, several tests were applied in order to check if they were stationary. For the non-stationary variables, a Vector Error Correction Model (VECM) was estimated, and long-term relationships were obtained and interpreted.

An Error Correction Model (ECM) was then estimated, assuming the long run relationships already obtained, and adding trend stationary variables as well as a dummy variable accounting for the years of presumably higher financial speculation.

1.4 – Structure of the Dissertation

This dissertation has its structure organized into five sections, being that each of those sections has its own sub-sections, where the topics on which the section focuses are explored with more details and specifications.

Section 1, the present section, is an introductory chapter where the choice of the theme and its financial and economic importance is justified. The objectives and hypotheses that sustain the research are also mentioned is section 1, as well as a brief description of the methods used.

Section 2 consists on a review of the most relevant existing literature on the theme, mainly when it comes to the determinants and methodologies that the authors most usually choose to include in their models to attempt to reach conclusions regarding the determination of crude oil prices. The role of financial speculation and the way that authors have tried to approach it is also mentioned.

Section 3 describes the methodology used in this research. Information regarding the data obtained is provided and the variables used are defined. The tests and methodologies that allowed one to test the previously defined hypothesis and reach the objectives of the research are also explained in this section.

Section 4 is the section where the results obtained are presented and discussed.

As for section 5, it is the section where the conclusions that were reached based on the results obtained are exposed. Because this is the final section of the dissertation, it also contains a few considerations regarding future studies and what could have been approached differently in the present research.

2 – Literature Review

2.1 – Background

The high volatility presented by crude oil prices throughout the years constitutes a topic that has been explored and approached by numerous researchers, especially because of the substantial importance that this commodity has on financial markets worldwide.

Many authors and researchers have tried to give out their contribute to this topic by offering different perspectives and approaches according to their visions and positions regarding what they consider to be the main causes that can help explain the unstable and constantly changing behavior of the crude oil prices.

According to Jager (2008), the best way to address this question is by breaking it into smaller and more easily identifiable factors. Jager (2008, p.2) states that "The oil market leads to a complex system of interactions not accessible as a complete network. The complexity has to be reduced by phenomenological or mathematical arguments. It is an effective strategy to start with a more detailed and therefore complex mathematical model and to reduce it to a sub model describing essential features of the system".

This statement can be seen as a premise to any research study that regards this topic, for only by identifying the factors that determine and influence crude oil prices we can become closer to studying and understanding the behaviors and evolution of the prices themselves.

However, the importance behind the identification of those factors goes far beyond the capacity to analyze and understand the evolution of the crude oil prices' time series itself, because it can also have implications on how policy makers think and act when it comes to decision making processes.

Liu, Wang, Wu & Wu (2016, p. 363) stress out the importance of this particular aspect, by stating that "The identification of price determinants not only helps to explain the origins of volatile oil prices but also has important policy implications. For example, if the major determinant of oil price is speculation, regulators and supervisors could weaken the negative impacts of volatile oil prices on the economy by implementing strict policies on the derivative markets. If the price of oil is determined by demand and supply fundamentals, then policy makers can do little to affect prices but only focus on promoting energy conservation or developing alternative energy sources."

Therefore, the identification of the determinants of crude oil prices is, indeed, a crucial step towards understanding its movements and the market itself. With that being settled, the obvious follow-up question that arises lies on comprehending which factors should be taken under consideration.

Even though the strategy suggested by Jager (2008) states that it is beneficial to start with more complex models and work them down into simpler and more effective ones, it is still important to know ahead which determinants should be submitted to significance tests, and which should not even be included in the models whatsoever.

By taking a look at the multiple determinants that authors and researchers have been including in their models so far, it might become easier to recognize which ones actually must be tested and taken under consideration in this study.

The rest of this literature review section will be organized according to the following sub-sections: Determinants of Crude Oil Prices; Methodologies Used; The Role of Financial Speculation[1].

2.2 – Determinants of Crude Oil Prices

While many authors mostly choose to include exogenous macroeconomic variables in their models, in order not only to test their significance when it comes to the direct influence they have on crude oil prices but also to analyze how they behave when the oil prices go up and/or down, there are others who claim that that kind of approach is not well defined and can lead to erroneous conclusions.

An example of that is the work of Kilian (2008), where the author starts by explaining that there is no sense in trying to analyze the impact that crude oil price changes provoke on macroeconomic variables. Kilian (2008) supports this position by stating that the cause and effect relation's direction is not well defined, and also that the observable variations in crude oil prices concern questions that are related with demand and supply shocks, and

[1] References to financial speculation as a variable and methodologies regarding it will be made purely in the last subsection of this literature review, and not in the "Determinants of Crude Oil Prices" or the "Methodologies Used" subsections.

those sorts of questions may vary quite a lot, depending on their origins and consequences.

What Kilian (2008) actually does is that he 'splits' the crude oil price's determinants into three different categories: shocks related with the physical availability of oil (oil supply shocks); shocks to the current demand for crude oil driven by fluctuations in the global business cycle (aggregate demand shocks); and, finally, shocks driven by shifts in the so called 'precautionary demand for oil' (precautionary demand shocks[2]).

The main conclusions that Kilian (2008, pp.19-20) reached have allowed him to formulate that the more usual macroeconomic models, which are built based on the premise that crude oil prices are exogenous variables, may be erroneous, as previously mentioned. He also leaves some recommendations by stating that the models that consider crude oil prices to be endogenous should focus on the demand side of the oil market, instead of mainly focusing on supply-wise aspects. Based on his analysis, the author also states that each of the demand and supply shocks in the oil market is pre-determined.

Kilian's (2008) approach to this topic is quite innovative[3], yet his work was no exception when it comes to the methodology used in the choosing of the variables. As a matter of fact, a significant part of the authors and researchers who approach this topic, including those who take macroeconomic factors under consideration, choose to separate the determinants included in their models into different categories.

Badr and El-Khadrawi (2016) are another example of researchers who have chosen to build up on that procedure, since these authors have separated the determinants taken under consideration into three different groups. In this case, the determinants were grouped according to: supply related factors; demand related factors; factors associated with the behavior of financial market participants.

As for the variables contained in the model that Badr and El-Khadrawi (2016) worked with, they were the following: China's gross domestic product growth; the Organization for Economic Cooperation and Development (OECD) countries' gross domestic product

2 It is important to mention that this last category concerns shocks that have to do with a demand that arises as a consequence of the oil demanding agents' uncertainty about shortfalls of expected supply relative to expected demand.

3 This statement will become clearer in the next sub-section of the literature review section (2.3 – The Methodologies Used).

growth; days of forward consumption of the OECD and China crude oil stocks; worldwide refining capacity; capacity utilization by the Organization of Petroleum Exporting Countries (OPEC); average world temperature; worldwide oil stocks; worldwide oil reserves; nominal effective exchange rate of the U.S. Dollar.

The results that Badr and El-Khadrawi (2016) reached seem to indicate that the growth of China's and the OECD countries' gross domestic products influences the crude oil prices in a significant and robust way. They also reached conclusions regarding the fact that the alterations verified in the worldwide refining capacity seem to be significant for the determination of the crude oil prices.

By now, it should be clear that, as previously said, something that characterizes most research papers in this area lies on the fact that the authors adopt the same technique of separating the variables into different categories. However, just by contemplating the multiple variables that were used and the conclusions that were reached in the latter mentioned paper, it is easy to see that there is a large range of approaches to this topic. While Kilian (2008) focuses a lot more on understanding how the supply and demand related factors affect the crude oil prices, Badr and El-Khadrawi (2016) tend to go with a much more pragmatic and macroeconomic approach to the problem.

Another variable gathering approach that is worth being mentioned in this literature review is the approach of King, Deng and Metz (2012), who focus a lot more on the role of political and mediatic events and their influence on the movements of crude oil prices, although they also dedicated a full section of the paper to the role of financial trading, and another one to the impact caused by OPEC decisions.

King, Deng and Metz (2012, p. 51), found that political events and economic news have a significant effect on crude oil prices. By contrast, they find limited evidence that financial trading by noncommercial traders (or "speculators") had a significant effect on oil prices, and they find that that significant effect did not occur during the price run-up period that occasioned the most public debate on this topic.

Although their approach differs from the previous ones, King, Deng and Metz (2012) also decided to use the strategy of separating the variables into different categories – in this case: political events and governmental decisions; economic factors; and natural events that have the potential to temporarily affect the price of oil.

Once that it has already been mentioned which are some of the variables that researchers choose to include in the models and how most of those researchers choose to aggregate them, it is now important to move on to another aspect that is worth analyzing, which is how all these variables express themselves when it comes to the way they influence crude oil prices. This constitutes an imperative point of the researches regarding this topic, because it is important to look not only at the determinants themselves and whether or not they are significant, but also at the nature of their possible influence, which can consist on a positive or a negative impact.

The following table (Table 1) summarizes which variables have mostly been chosen by some authors to include in their research papers, and what was the nature of each variable's influence, both regarding the authors' expectations and the actual observable impact:

Table 1 – The Most Commonly Used Variables in the Literature.

Title	Author(s)	Publication Date	Variables Used	Expected Impact	Actual Impact
"Disentangling the Determinants of Real Oil Prices"	Li Liu, Yudong Wang, Chongfeng Wu & Wengfeng Wu	2016	Oil Supply Shocks	-	-
			USA's Oil Demand Shocks	+	+
			China's Oil Demand Shocks	+	+
			Precautionary Demand Shocks	+	+
			Derivative Market Speculation Shocks	+	+
"The Determinants of Crude Oil Prices"	Osama Badr & Ahmed F. El-Khadrawi	2016	China's Gross Domestic Product Growth	+	+
			OECD's Gross Domestic Product Growth	+	+
			Days of Forward Consumption of the OECD and China Crude Oil Stocks	-	-
			Total Refining Capacity Worldwide	-	-
			OPEC's Capacity Utilization	+	+
			Average World Temperature	+	Non Significant
			Worldwide Oil Stocks	-	-
			Worldwide Oil Reserves	-	-
			Nominal Effective Exchange Rate of the USD (NEER)	-	-
"Assessing the Factors Behind Oil Prices Changes"	Stéphane Dées, Audrey Gasteuil, Robert Kaufmann & Michael Mann	2008	OPEC's Capacity Utilization	+	+
			Difference Between the Fourth Month Contract for WTI and the Near Month Contract for WTI	+	+
			USA's Refinary Utilization Rate	-	-

			Days of Forward Consumption of OECD Crude Oil Stocks	-	-

Source: Own Elaboration

2.3 – Methodologies Used

When analyzing and reflecting on the several approaches that authors and researchers have taken throughout the years in order to explore a certain topic, focusing purely on the factors that those authors have chosen to study might not be absolutely clarifying. It is equally important to contemplate and examine the way that those variables have been worked and the way they have been used by the researchers, in order to reach solid results and conclusions. In fact, it is possible that two or more research papers on the same topic take different courses, even if the variables used were the same, since their modeling might still differ from one another.

This is even more so when it comes to research works regarding the topic in question - that is, the oil market and the way the prices change and evolve – since this is a subject that allows authors to go through many different routes. Thus, one could never truly review the existing literature on this topic without referring to the different methodologies on which authors have relied so far.

Some of the methodologies used to approach this subject over the years have, in part, already been mentioned in this literature review. References regarding the kind of variables chosen and even the way that authors have chosen to aggregate them have been made in the previous sub-section (2.2 – Determinants of Crude Oil Prices), and those choices do constitute methodologies.

However, in the present sub-section, we focus more on what distinguishes each research papers' approaches when it comes to the models used and the way the data was both processed and tested.

Starting with Badr and El-Kadrawi's (2016, p. 6) methodology, in order to estimate their model, the authors used Engle and Granger's (1987) procedure, claiming that it is an appropriate way to do it due to the fact that it allows researchers to estimate both long-run and short-run relationships between crude oil prices and the considered explanatory variables. The authors then explain that, even if the time series data is non-stationary, there is still a possibility to generate a linear combination from the data in question, and if that linear combination is stationary, then the time series are cointegrated.

This observation is particularly important, because, as the researchers continue to describe, the goal of the cointegration test is to find out whether this combination is stationary or integrated of order 0 by applying an Augmented Dickey-Fuller[4] (ADF) type-test on the residuals, which is called the Engle and Granger (1987) test. The test that was just mentioned allows the authors to conclude whether there is a cointegration relation among the variables – if the null hypothesis is rejected – which can be interpreted as the long-term equilibrium relationship among the variables of the model.

Badr and El-Kadrawi's (2016) found that all of the variables used were integrated of the same order, I(1), because none of them was stationary in the first place, but they all revealed stationarity in their first differences, thus leading to an estimation of an Error Correction Model (ECM) via Ordinary Least Squares (OLS). According to the researchers, this model "reflects the short-term relationship by entering the residual lagged for one period".

As for the results and the interpretation of the methods used, the results of the previously mentioned regression indicate that, because there is a presence of a cointegration relationship in the model used, it "(…) can be interpreted as an equation for the long-term determinants of oil price".

The approach of Killian (2008) led him to develop a monthly index of global real economic activity, which was explicitly designed to capture shifts in the demand for industrial commodities driven by the global business cycle and had its base on dry cargo single voyage ocean freight rates. As Killian (2008, p.3) describes, "While it is evident from informal evidence that demand shocks play an important role in the crude oil market, the problem of quantifying these demand shocks so far has proved elusive. One difficulty is that there are no readily available indices that capture shifts in the demand for industrial commodities driven by the global business cycle." This quotation justifies the need to develop the previously mentioned index, which was then used as one of several variables contained in a VAR model based on monthly data.

[4] The ADF test was first developed by Dickey and Fuller (1979).

However, the development of the monthly index of global real economic activity was not the only point that deserved an innovative approach from the author when it comes to the methodology used. Another aspect that researchers - including Killian (2008, p.3) – have to face while modeling the prices of crude oil prices based on its determinants has to do with the fact that "(…) the expectation shifts underlying the precautionary demand shock are not observable". Thus, the author had to find a way to overcome this issue, so Killian (2008, p.4) allowed "(…) a structural dynamic simultaneous equations model to pin down the oil-market specific component of demand as the residual."

As for the estimation of the previously mentioned VAR model, Killian (2008, p.10) explains that "The reduced form VAR model is consistently estimated by the least-squares method. The resulting estimates are used to construct the structural VAR representation of the model."

Liu et al. (2016, pp. 364) also gave their contribution to the existing literature when it comes to the methodology that they have assumed. Their paper tries to evaluate not only the effects of economic and financial fundamentals on the prices of crude oil, and not only the possible effects of financial speculation, but both these effects combined.

The authors also found that the existing literature on the topic did not really take sign restrictions techniques as a common practice throughout the researches, and that fact resulted in an approach that contrasts from most studies on crude oil prices. Specifically, as the authors explain, "(…) we use a sign restriction approach that restricts the impulse responses according to the standard economic theories. This approach makes more sense than the traditional short-term exclusive restrictions because it allows for the contemporaneous reaction between two variables. This advantage of sign restriction is extremely useful for disentangling oil price shocks."

Thus, Liu et al.'s (2016, pp. 364) approach is innovative, since "(…) the exogenous restriction cannot address the contemporaneous responses between oil prices and speculating positions". Plus, as the authors themselves claim: "To the best of our knowledge, sign restriction is applied in very few studies on the identification of oil price shocks except in some notable papers".

A component that has been targeted in the methodologies of several authors and used by them in the models that support their studies is financial speculation. However, given that this is such a specific factor – and one that requires special attention in the methods used to measure it on a quantitively way too – it will be approached with more detail in the next sub-section of this literature review (2.4 – The Role of Financial Speculation).

2.4 – The Role of Financial Speculation

Over the years, but especially since the early 2000's, financial speculation and its role on commodity markets such as the crude oil market have become topics increasingly more debated and analyzed by researchers who try to explain how this phenomenon works and what influence does it have on the determination of commodity prices. The increase in attention given by authors to this particular aspect matches the seemingly greater influence exerted by financial speculation. Basically, researchers are noticing this influence more and more because the influence itself seems to be increasing since the turning of the century, although some authors claim that this impact has cooled down a bit over the last years.

But before getting into too much analysis, it is important to clarify what financial speculation is, and how does it work. Alquist & Gervais (2011, pp.2) define financial speculation as "(…) a firm holding a net position, either long or short, in the expectation of earning a positive return, and not because it is a commercial user of oil". On a larger definition, one can say that financial speculation in the oil market can be seen as the act of financial and economical agents 'betting' on the increase/decrease of crude oil prices, in the hope that those bets end up influencing the price movements in order for them to match the agents' expectations, thus leading to favorable scenarios when it comes to the possibility of generating earns.

As for the difficulty of taking financial speculation's influence under consideration and including it in the models used to try to explain crude oil changes, it has to do with the fact that this particular factor is extremely difficult to measure, given that its influence is not quantifiable in an exact sense. What this means is that financial speculation, opposing to variables such as the world oil reserves, USA's gross domestic product, etc., cannot be directly included in a model as a numeric variable, because its influence on crude oil prices, while seemingly existent, is not directly measurable. That being said, it becomes clear that a way to solve this issue must be found if we truly intend to include financial speculation in a model that tries to explain the behaviors of crude oil prices.

An alternative that quickly comes to mind when trying to overcome this issue consists on finding another variable - one that represents financial speculation and reflects its influence on crude oil prices, while still being measurable and quantifiable. As long as this alternative variable presents the same kind of behavior as speculation when it comes to provoking/reacting to changes in crude oil prices, it could be included in the model instead of financial speculation itself. This technique would allow the researcher to extrapolate how financial speculation influences crude oil prices, and then take conclusions based on that.

A critical stage of this procedure would be to select a variable that reflects speculation's behavior and influence on crude oil prices in the most reliable way possible. Killian and Lee (2014, p.73) state that "An alternative view is that speculation may also be conducted by oil producers who have the option of leaving oil below the ground in anticipation of rising prices (…). An accumulation of below-ground inventories by oil producers in anticipation of rising prices would be equivalent to a reduction in flow supply. In short, flow supply shocks and speculative supply shocks are observationally equivalent." These statements could already serve as guidelines as to which are some of the variables that could be taken under consideration as possible replacers of financial speculation in a model that sets itself to include the determinants that help explain the changes in crude oil prices. Following Kilian and Lee's (2014) logic, if flow supply shocks and speculative supply shocks are equivalent – at least when it comes to the observation of their behavior – then by including variables that relate to flow supply shocks, such as the world refining capacity or the worldwide production, we would already be taking under consideration financial speculation's influence on the real prices of oil.

Kilian clearly believes that oil inventories and supply shocks hold the key to including a speculative component in the models used to study the evolution of crude oil prices, since the previously mentioned one is not the only paper where this author tries to relate these questions. The work of Killian and Murphy (2014) also follows this approach, as the authors suggest that most of the relevant data regarding financial speculation is already contained in the inventory data.

There are, however, other authors who have chosen to represent the speculative component of their models in different ways. Liang and Liu (2017, p.2) are an example of that, as the authors "propose the non-commercial traders' net long position of light sweet crude oil in NYMEX as the proxy to capture the speculative behavior in the futures market".

While it is a vastly used approach, basing the effects of financial speculation on the crude oil futures' market might not be the best solution. As Oliveira (2017, p.8) describes, "there is no way in isolating the speculative component in net open positions, unless one separates commercial from non-commercial positions. This is so because noncommercial traders are foreseen as the speculators in the market, [while] commercial traders are related to hedging positions to protect their demand for oil". Oliveira (2017, p.5) also points out that the use of futures can be "disadvantageous due to the fact that crude oil futures only came into play in the 1980's", which is a downside since it restricts the size of the data samples used.

Regardless of the way that authors choose to represent the influence of financial speculation in the models used to analyze and evaluate the evolution of oil prices, one thing that is certain is that the importance of including this component and its impact in the analysis is being recognized by some authors, while disputed by others. Authors such as Cifarelli and Paladino (2010) have found evidences that there is in fact a relation between speculators' actions in the oil market and the changes of prices that occur. Those evidences are referenced by Zhang (2013, pp. 395), who states that Cifarelli and Paladino's (2010) results "(…) validate the existence of positive feedback in crude oil market and identify a significant role played by speculation in the oil market", outlining that that role lies on the fact that "(…) speculatively driven high prices can persist for a considerable time before fundamentals bring them down to fairer values".

The already mentioned Killian and Lee (2014, p.85) paper is one on which the authors have also found evidence that speculation does indeed affect real oil prices. The above paper finds evidence of speculation "(…) driving up the real price of oil in the physical market for crude oil in 1979 after the Iranian Revolution, in 1990 near the time of the invasion of Kuwait, in 2002 in the months leading up to the 2003 Iraq War, in early 2011 during the Libyan crisis and in early 2012 during the Iranian crisis. A common feature of all these episodes of speculative pressures is that they reflect concerns about the stability of oil supplies (…). We also found evidence that speculation may lower the real price of oil. We identified several episodes in which a reduction in speculative demand contributed to lower oil prices."

On the other hand, there are researchers, such as Hamilton (2009), who have studied and analyzed the role of financial speculation in commodity markets and come to the conclusion that its influence might not be that relevant. This author argues that the influence on crude oil prices that other researchers blame on financial speculation should be imputed to demand/supply shocks.

Hamilton (2009, pp.14-15) describes an hypothetical case based on the premise that speculation could, in fact, lead to increases in crude oil prices, and comes to the conclusion that there is no way that the increase of the price could come from a speculative component, inferring that "ultimately there are physical producers of crude oil and physical consumers of gasoline, and insofar as the activities of either have any response at all to the price, incentives for consumption would be reduced and incentives for production increased whenever the price of crude oil is driven up. For this reason, an ongoing speculative price bubble would have to result in continuous inventory accumulation, or else be ratified by cuts in production."

The author explains that the continuous inventory accumulation hypothesis must be seen as "(…) clearly unsustainable", and therefore could not be the correct outcome. Thus, the outcome must be that the speculative price bubble would be ratified by cuts in production, and so "(…) one might make the case that the supply cuts rather than the speculation itself has been the ultimate cause of the price increase."

Whichever is the case - that is, whether financial speculation does or not play a significant role in the crude oil market and in the determination of its prices – it is only by taking this component under consideration that we can establish a coherent and well defined model that attempts to conclude which really are some of the determinants of crude oil prices. The following sections of this study will certainly help clarifying these questions, and whether financial speculation, as well as the other variables that were mentioned throughout this literature review, ultimately do impact crude oil prices in a significant way.

3 – Methodology

3.1 – Hypotheses and Objectives of the Research

In this section we detail the hypotheses and objectives that have already been established in section 1 – Introduction. A clear definition of the hypotheses and objectives will determine which methods are the most appropriate to adopt in order to reach results and conclusions that are both valid and significant.

Prior to the definition of the hypotheses and both the main and the specific objectives of the research, it is important to define what is the research question of the study. The research question of this dissertation is the following: *Which are the economic and financial variables that most directly influence crude oil prices?*

This research sets its main objective precisely at understanding which are those variables, and which can be ruled out for not revealing themselves as relevant to the determination of crude oil prices. As for the specific objectives of the study, those have been defined as the following:

➢ Determining the nature of the influence of the significant variables on crude oil prices;

➢ Determining the extension of the impact caused by those variables;

➢ Analyzing the role of financial speculation on the crude oil prices' determination.

With the goal of achieving the previously mentioned objectives, several hypotheses were set:

➢ Hypothesis #1: Higher values of Worldwide Oil Reserves lead to lower Crude Oil Prices.

The results reached by Badr & El-Khadrawi (2016) go towards this hypothesis when it comes to the coefficient that these authors obtained for their variable of OPEC reserves, which was negative. This means that there seems to be a relation between OPEC reserves' levels and crude oil price on which the increase of OPEC reserves leads to decreases in crude oil prices. Based on these results, the same thought can be applied when it comes to World Reserves values, that is, that higher values of World Reserves should lead to lower prices in the oil market.

One possible argument that could support this hypothesis is the fact that when there are higher values of crude oil reserves worldwide, the urge to distribute them into the markets arises, and in order to ensure that that happens, the price comes down, so that the demand can also arise and meet the supply provided as a result of the high amounts of crude oil contained in those reserves.

➢ Hypothesis #2: Higher values of Worldwide Oil Production lead to lower Crude Oil Prices.

This hypothesis also holds its justification on the basic economic and financial notion that, when supply levels rise, the price must decrease so that the demand levels can meet with supply and an equilibrium point can once again be established throughout time.

Kilian (2008, pp.10) mentions that productors have "historically tended to restrict supply in order to prop up the price of oil.", which is a statement from which we can deduce that, when worldwide production reaches higher values, it is expected that its capacity to supply the crude oil market is higher, and that would push down crude oil prices.

➢ Hypothesis #3: Higher values of Worldwide Oil Consumption lead to higher Crude Oil Prices.

As for this hypothesis, its validation lays on the same logic as the previous ones, only in an inverse perspective.

If worldwide consumption for crude oil raises, that means that the demand levels will be growing, which leads to an increase in the observable price until, once more, an equilibrium point is reached, where demand and supply meet each other.

It is actually quite intuitive to think that, if a certain product is presenting higher levels of demand, there is an opportunity for its price to rise, because certainly some of the many parts that are looking to acquire it will be willing to pay more for it.

➢ Hypothesis #4: Higher values of Worldwide Refining Capacity lead to lower Crude Oil Prices.

World refining capacity reflects the capacity to produce crude oil barrels that refineries worldwide have, on a daily basis. Therefore, and taking into consideration some of the explanations that were given to support previously presented hypotheses, the logic that sustains this hypothesis is based on the assumption than a higher capacity of oil production leads to higher levels of supply, and therefore to a reduction in price.

Badr & El-Khadrawi (2016) also stand by this logic, and as a matter of fact, these authors have reached results that clearly indicate that “Refining capacity has a negative effect on the long-run level of oil prices.”, which basically means that higher values of world refining capacity lead to lower crude oil prices, which is hypothesis #4 in itself.

➢ Hypothesis #5: Higher values of Worldwide Gas Consumption lead to lower Crude Oil Prices.

This hypothesis is based on the natural assumption that, if financial agents are presenting higher values of demand for a different kind of energy source – gas, in this case – then their demand levels for crude oil will be lower.

Because demand for crude oil decreases in this situation, in order for demand to raise again, the prices must diminish, so that the agents are once again attracted into searching for crude oil in the financial markets.

In other words, it is expected that crude oil and gas are substitute goods, which means that a higher level of consumption/demand for one of them should lead to lower levels of consumption/demand for the other, which leads to lower prices in this second good, according to the explanation that was just given.

➢ Hypothesis #6: Higher values of Worldwide Coal Consumption lead to lower Crude Oil Prices.

The same explanation that was just given for Hypothesis #5 can also be used for Hypothesis #6.

Because coal also is an alternative to crude oil, it only seems logic to assume that crude oil prices’ behavior regarding higher levels of worldwide coal consumption should develop according to the substitute goods theory. Zamani (2016, pp.802) even stated that “(…) a high level of substitutability between coal and oil began many years ago (…)”, and that fact “(…) may lead to relationships between them”, which supports how this hypothesis is defined.

➢ Hypothesis #7: Higher values of the USA's Gross Domestic Product lead to higher Crude Oil Prices.

Higher values of a country's Gross Domestic Product usually indicate higher economic development and, of course, a higher capacity for the financial agents to be active in the markets.

With that being said, it is only logical to assume that higher levels of the world's biggest economy's Gross Domestic Product should lead to higher activity in the crude oil market, namely in the demand side of the market, thus leading to a rise of the prices.

➢ Hypothesis #8: Higher values of China's Gross Domestic Product lead to higher Crude Oil Prices.

The same logic presented in Hypothesis # 7 can be applied here.

Even though China is not yet the country with the world's highest Gross Domestic Product values[5], it is one of the fastest expanding economies in the world. Therefore, one can also assume that as China's Gross Domestic Product grows, higher levels of crude oil demand will appear as a result, thus leading to higher values of the commodity's price;

➢ Hypothesis #9: Periods of apparently higher financial speculation activity in the markets can lead to both increases or decreases in Crude Oil Prices.

From all the nine hypotheses presented in this study, this is certainly the one that would be both the most interesting one to confirm and the most difficult one to demonstrate on a quantitative level.

It is established based on the assumption that, because speculating agents hold their positions in the market until the right time comes and they can exchange their positions and get their earns, crude oil prices could be influenced to either raise or diminish depending on whether those agents decided to buy or sell first.

Explaining this argument more widely, what happens is that market agents can either decide to define their selling price first[6], or their buying price first.

[5] USA is actually the country with the highest Gross Domestic Product values worldwide, followed by non-other than China itself.
[6] This concept is called a short sale, which is a widely used practice that consists on a certain market player selling something in the market before he even owns it.

In the first case, the agents would have influenced the market in a way so that the price would decrease[7], and later they would decide to buy, thus earning the difference between that buying value and the higher selling value that they had already defined at first. In the second scenario, speculators decide to buy first, which would lead to increases in the prices[8], and later on they would sell their position once the prices would reach higher levels, interesting enough for those agents to take the profit they were after.

So, what this hypothesis defines is that, depending on the original decision of the speculators to either adopt a buying or selling position in the market at first, crude oil prices can either rise or descend. Plus, on the later moment, on which speculating agents decide to reverse their position - that is, when they decide to sell/buy after having bought/sold first – they are also conditioning the prices in the market, since multiple speculators would reverse their position at around the same moment as well.

3.2 – Variables Used

The variables used in this research were chosen taking under consideration the analysis that was previously made in the Literature Review section, regarding the variables that most authors who study oil prices chose to include in their models.

In this sub-section, the variables used will be presented and described as they were before the application of any tests or estimations, except for the graphical observation of the series themselves.

The dependent variable of this study is:

➢ **oilprice**: Crude Oil Prices. The values of this variable are expressed in U.S. Dollars.

The explanatory variables can be divided into three groups, which are the following: 1) Variables closely related to crude oil prices; 2) Macroeconomic variables; 3) A dummy variable concerning the time period from which the impact of financial speculation on crude oil prices may be more noticeable;

---

[7] This happens since multiple speculators would have implemented selling positions at around the same time.
[8] This happens since multiple speculators would have implemented buying positions at around the same time.

The first group of explanatory variables consists of six variables:

➢ **worldres**: Worldwide Crude Oil Reserves, in billions of barrels;

➢ **worldprod**: Worldwide Production of Crude Oil, in millions of tons;

➢ **worldcons**: Worldwide Consumption of Crude Oil, in millions of tons;

➢ **worldrefcap**: Worldwide Crude Oil Refining Capacity, in thousands of barrels per day.

➢ **worldgascons:** Worldwide Consumption of Natural Gas, in millions of tons;

➢ **worldcoalcons**: Worldwide Consumption of Coal, in millions of tons;

The second group of explanatory variables, the group of macroeconomic variables, consists of two variables:

➢ **usaGDP**: United States of America's Gross Domestic Product, in billions of U.S. Dollars;

➢ **chinaGDP**: China's Gross Domestic Product, in billions of U.S. Dollars.

It is important to justify that Gross Domestic Products from other countries could have also been considered and used in this research. However, the choice of including these two Gross Domestic Product variables relies on the fact that both the USA and China are the two largest and most influential economies worldwide, being that China is not only one of the biggest economic and financial potencies, but also the most boosting economy of the last two decades.

The third group of variables consists of a single variable:

➢ **precrises**: Variable concerning the time period from which the impact of financial speculation on crude oil prices seems to be more noticeable, which is the time period of build-up until the financial crises of 2008. This is a dummy variable which assumes the value "1" if the observation belongs to the time period from 1998 to 2008, and the value "0" otherwise (from 1980 until 1997 or from 2009 until 2017);

Having explained which explanatory variables were included in the study and how they were separated according to their characteristics, it only makes sense that, before getting into detail regarding the tests and estimations to which the variables were submitted, one presents the *a priori* expectations of the nature of the impact that those variables might exert on crude oil prices. The following table summarizes the expected signs for each of the explanatory variables used in the present study:

Table 2 – Expected Signal for Each Explanatory Variable.

Explanatory Variable	Expected Signal
World Oil Reserves	-
World Oil Production	-
World Oil Consumption	+
World Refining Capacity	-
World Gas Consumption	-
World Coal Consumption	-
USA's GDP	+
China's GDP	+
Financial Speculation Dummy	+/-

Source: Own Elaboration

As it can be observed, the expected signals for each explanatory variable are defined according to the previously presented hypotheses of the research, which had already establish which behavior we should expect from each of the variables regarding their possible influence on the dependent variable and its values.

3.3 – Data

The databases from which the sample used in the present study was extracted were the BP Statistical Review of World Energy 2018 and the Federal Reserve Bank of St. Louis' Federal Reserve Economic Data (FRED). The BP Statistical Review of World Energy is a statistical report regarding all types of energy related data, which is released by BP on an annual basis. BP ensures that all the statistics presented in the BP Statistical Review of World Energy 2018 are taken from governments and published data. As for the Federal Reserve Economic Data (FRED), it is a database that relies on 87 different sources and contains data concerning over 500000 economic time series, which is being managed and sustained by the Federal Reserve Bank of St. Louis.

The BP Statistical Review of World Energy 2018 was used to gather data for the variables that are related with crude oil. The Federal Reserve Economic Data (FRED) was used to obtain data for the macroeconomic variables used in the research, which basically consist on Gross Domestic Products from two countries.

Because there were variables which had more observations than others when it comes to the data available, the period chosen for the overall sample used in the research concerns the time period between 1980 and 2017, since 1980 was the furthest period for which there were observations available for all the variables.

All the data collected consists on time series, and the periodicity of the observations is annual. There is a total of 38 observations per variable used.

For a first impression of the data, we compute some descriptive statistics, which are presented in Table 3:

Table 3 – Descriptive Statistics of the Selected Variables.

Variable	Mean	Std. Deviation	Min	Max
Crude Oil Prices	59.126	31.039	19.120	121.240
World Oil Reserves	1238.156	323.163	683.500	1702.430
World Oil Production	3528.235	503.215	2762.890	4387.140
World Oil Consumption	3543.341	524.666	2761.080	4469.680
World Refining Capacity	82506.480	8663.773	72421.900	98139.410
World Gas Consumption	2071.469	586.334	1224.280	3155.970
World Coal Consumption	2688.628	713.546	1793.340	3865.260
USA's GDP	10022.650	5050.346	2857.310	19485.390
China's GDP	2953.718	3773.148	191.150	12237.700

Source: Own Elaboration

After taking a first look at Table 3, the first thing that we notice is the wide range of the Crude Oil Prices series, since its values over a time period of 38 years vary from 19.12 USD to 121.24 USD per barrel. This fact suggests that this time series could be a very challenging and a very interesting one to analyze and try to explain.

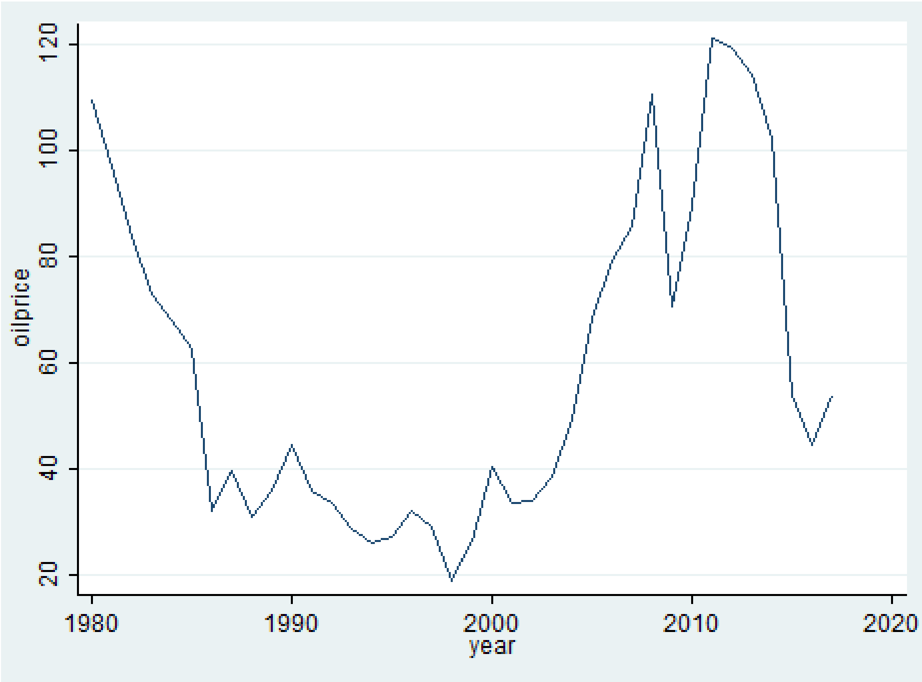
Looking at the maximum and minimum values of the World Consumption of Crude Oil and comparing them to the respective values of the World Consumption of Gas and the World Consumption of Coal, we can see that the World Consumption of Crude oil has a smaller amplitude of variation over the years of analysis, with the maximum value being 61.88% higher than the minimum. The World Consumption of Gas has varied 157.78% between its highest and lowest values, and the World Consumption of Coal has varied 115.53%, thus meaning that when it comes to worldwide consumption, crude oil could actually change less than the other two series of energy sources included in the research.

Another interesting aspect of the descriptive statistics that were obtained is how much the USA's Gross Domestic Product and China's Gross Domestic Product have grown over the years. In both cases, the minimum value corresponds to the first year of observations (1980), and the highest value corresponds to the latest year (2017), being that during that time-period, there were few years on which the values of each of these countries´ Gross Domestic Product have decreased.

With the purpose of better interpreting the results that will be presented in the next section (4 – Results Analysis), and also taking in consideration the graphical analysis of each time-series, all the variables[9] were converted into logarithms, so that one could analyze the elasticities between them[10]. The graphs that were obtained for each series before applying the logarithmic transformation to them - and that ultimately led to the decision of applying that transformation - were the following ones:

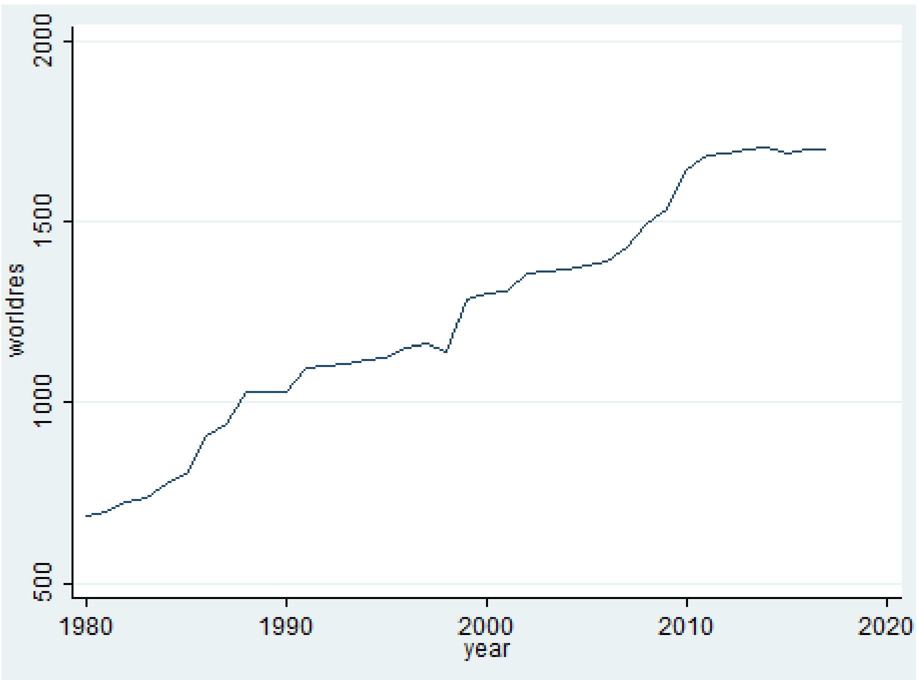
9 The only exception was, of course, the dummy variable..
10 This included the dependent variable.

Fig. 1 – Crude Oil Prices' Time-Series Graph.



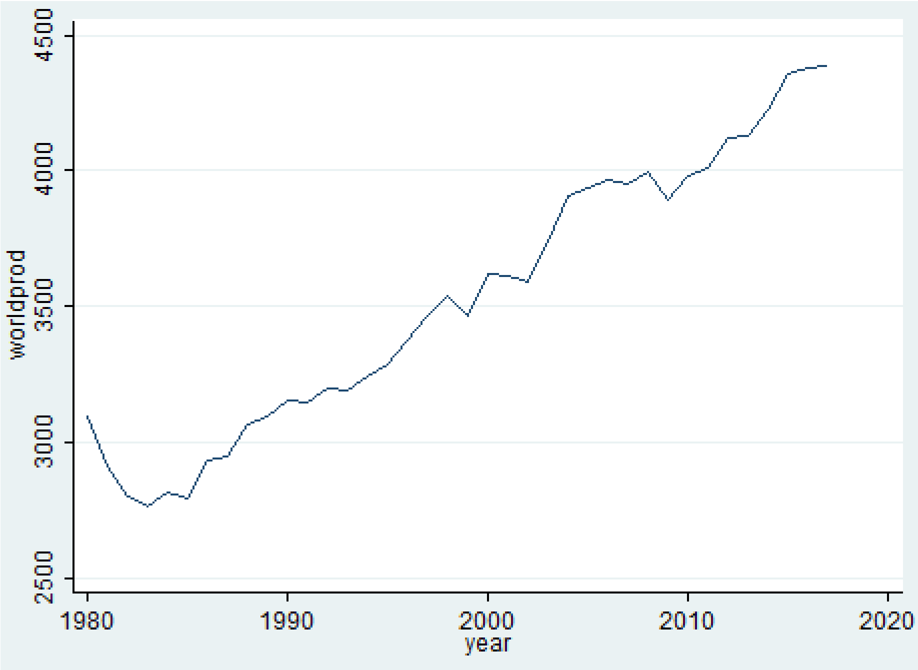
Source: Own Elaboration

Fig. 2 – Worldwide Oil Reserves' Time-Series Graph.



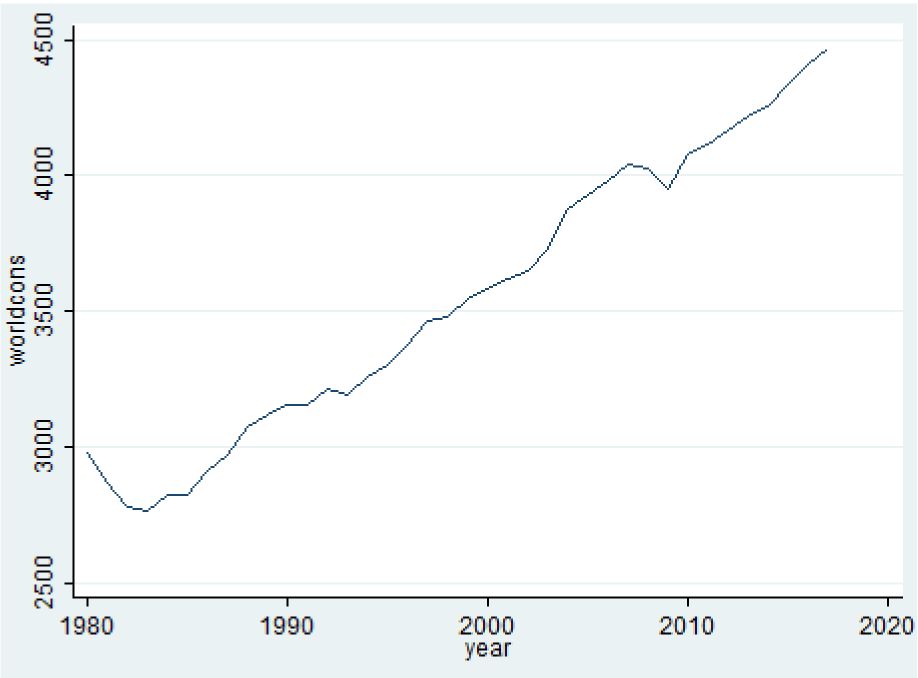
Source: Own Elaboration

Fig. 3 – Worldwide Oil Production's Time-Series Graph.



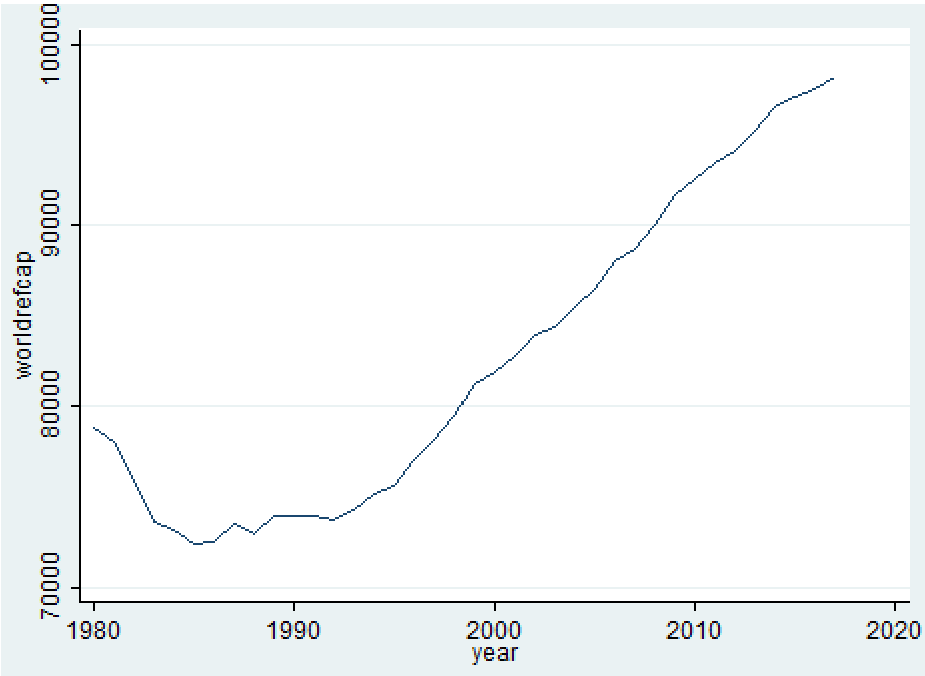
Source: Own Elaboration

Fig. 4 – Worldwide Oil Consumption's Time-Series Graph.



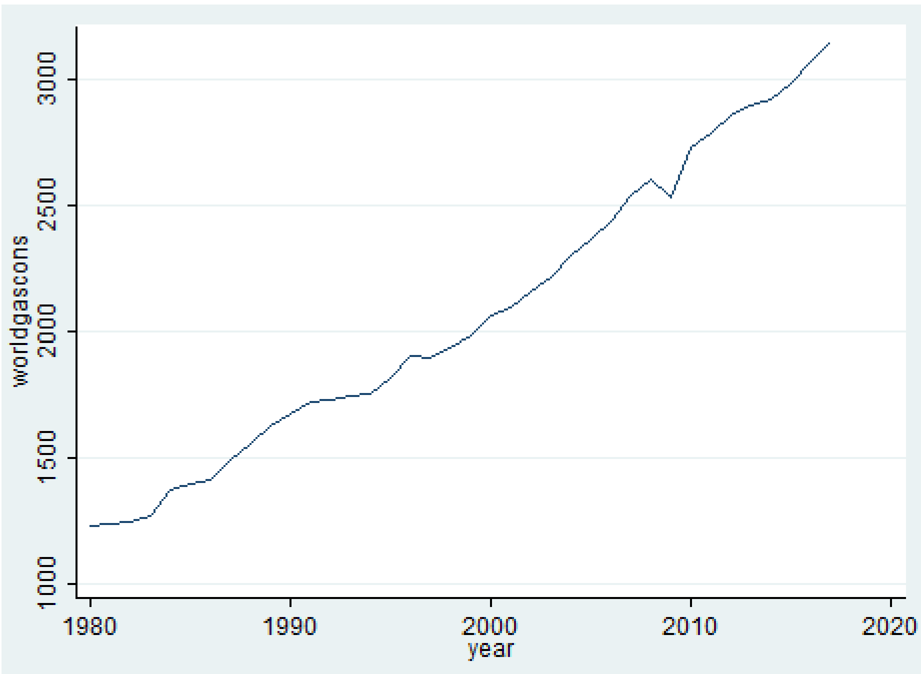
Source: Own Elaboration

Fig. 5 – Worldwide Oil Refining Capacity's Time-Series Graph.



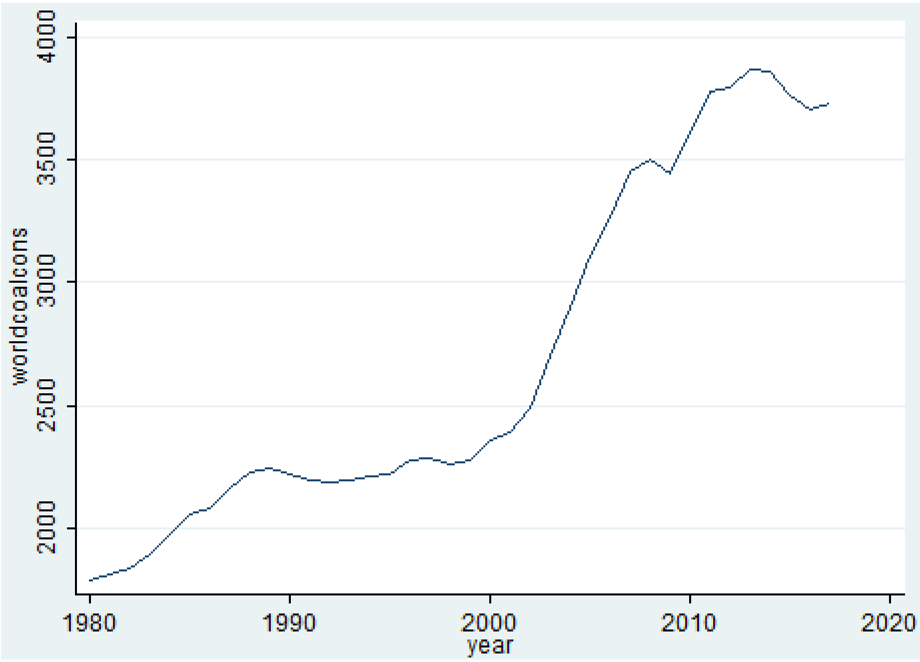
Source: Own Elaboration

Fig. 6 – Worldwide Gas Consumption's Time-Series Graph.



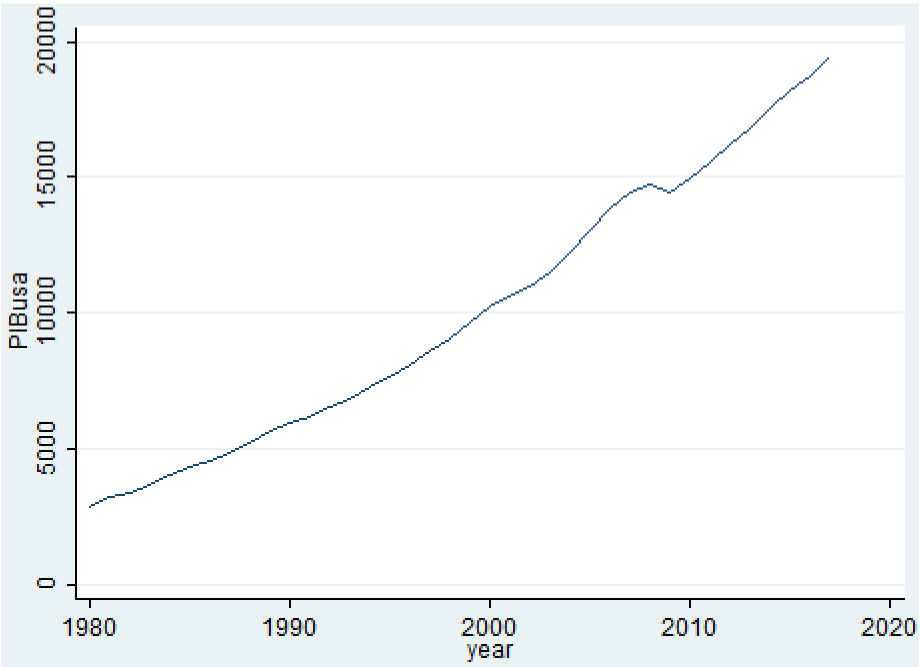
Source: Own Elaboration

Fig. 7 – Worldwide Coal Consumption's Time-Series Graph.



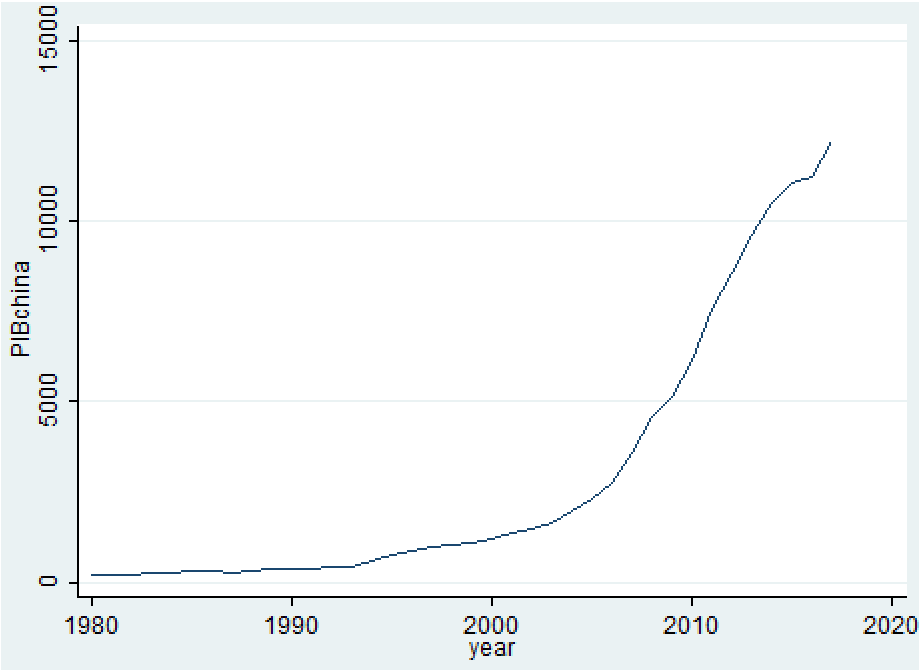
Source: Own Elaboration

Fig. 8 – USA's Gross Domestic Product's Time-Series Graph.



Source: Own Elaboration

Fig. 9 – China's Gross Domestic Product's Time-Series Graph.



Source: Own Elaboration

As for the graphs regarding the time-series with the logarithmic transformation applied to them, they are present in the Appendix.

3.4 – The Model

All the variables that were mentioned in the sub-section 3.2 – Variables Used had to be submitted to a few tests before being included in a model that could allow us to draw conclusions regarding their influence and its significance on the determination of crude oil prices.

The first of those procedures consisted on applying to each series a test that could allow us to conclude whether the series presented stationarity or not. Although there is more than one test that can be applied in order to reach conclusions regarding the stationarity of the series, the choice relied on the most widely applied test in this context, the Augmented Dickey-Fuller test, which had already been mentioned in section 2 – Literature Review.

In order to use the ADF test, we must choose the number of lags (k) to include in the ADF regression. Since we have annual data, we decided to set the initial number of lags included in the ADF test to k = 3. That number was further reduced if the last lag was nonsignificant (using the t-statistic).

Both the DF and the ADF statistics test the null hypothesis that a unit root is present in the time series, which would indicate that the series is nonstationary against the alternative of stationarity.

However, if the conclusion that results from the DF/ADF tests is that the series is not stationary, then one must apply the DF/ADF once more, this time to the first differences of the series. The goal of this second test is to reach conclusions regarding the order of integration of the original series. If the series of first differences is stationary or integrated of order zero, I(0), then the original series is integrated of order one or I(1). If the series of the first differences is nonstationary, then we must apply yet another test to check whether the series of the second differences is stationary, which would then mean that the original series is I(2), and so on.

After applying the unit root tests, and assuming that we find I(1) variables, we investigate any long-term relations among those I(1) variables using the Johansen (1991) cointegration tests[11]. These tests, the trace test and the maximum eigenvalue test, allow us to verify if there is indeed one or even more cointegration vectors among the variables.

The aim of these procedures was to verify if it would be possible to obtain a long-term model that could reveal the nature of the possible long-term equilibrium relations between the non-stationary series. Both the equilibrium relations and the short-term dynamics are estimated in one step, using a Vector Error-Correction Model (VECM). In order to test Johansen´s assumptions, a test for autocorrelation and a test for normality were performed after the VECM estimation.

After the VECM estimation, we obtain the cointegrating vectors, which give us the long-term determinants of oil price, and compare the results with our hypothesis.

The first equation of the VECM gives us already the short-term dynamics for the oil price variation. However, we decided to further build a single-equation dynamic model, based on the first equation of the VECM, but including more information, and excluding nonsignificant variables. This single equation ECM model includes: the dependent

[11] These tests were first developed and applied by Soren Johansen in his 1991 paper "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models".

variable (changes in crude oil prices), lags of the I(0) variables, lagged differences of the I(1) variables, the lagged equilibrium errors (from the cointegrating vectors previously estimated) and the dummy variable related to the period of supposedly higher financial speculation.

The dummy variable was included in order to check if the fact that the observations belonged to the 10 year time period that anticipated the big financial crises of 2008 - which was a period on which a speculative bubble surely could have formed – could have affected crude oil prices.

The final version of the ECM model left out the lagged variables that were not statistically significant, provided that their exclusion does not result in autocorrelation.

The final model was then tested for functional form misspecification using Ramsey's (1969) RESET test, for autocorrelation with the Breusch-Godfrey (1978) test, for heteroskedasticity - the Breusch-Pagan (1979) test - and for autoregressive conditional heteroskedasticity (ARCH effect[12]).

[12] This test was firstly applied by Engle (1982)

4 – Results Analysis

The first results we must analyze are the results of the stationarity tests. As was previously mentioned in the Methodology section, the initial number of lags considered in the ADF tests was k = 3. It was then reduced to k = 2 lags if the third lag was not statistically significant, according to the t-test, and k = 1 lag if the second lag was also nonsignificant. If none of the lags appeared to be significant, as happened in some cases, then the number of lags used was k = 0, which means that the unit root test used was the original Dickey-Fuller (DF) test.

The inclusion of a trend on the ADF/DF regressions was decided based on the graphical analysis of the logarithmic series, shown in the Appendix. However, when it comes to the nonstationary series, a second ADF/DF test had to be applied to the series in first differences, which have no trend.

The results obtained from the application of the ADF/DF tests are shown in Table 4:

Table 4 – ADF/DF Tests.

Variable	Number of Lags	Trend	Test Statistic	ADF Test Conclusion	I(0)/I(1)
Log(Crude Oil Prices) ΔLog(Crude Oil Prices)	0 0	Yes No	-2.232 -5.701***	Nonstationary Stationary	I(1)
Log(World Oil Reserves) ΔLog(World Oil Reserves)	0 0	Yes No	-1.603 -5.766***	Nonstationary Stationary	I(1)
Log(World Oil Production)	2	Yes	-3.441*	Trend Stationary	I(0)
Log(World Oil Consumption)	2	Yes	-3.516*	Trend Stationary	I(0)
Log(World Refining Capacity)	0	Yes	-5.932***	Trend Stationary	I(0)
Log(World Gas Consumption) ΔLog(World Gas Consumption)	0 0	Yes No	-2.753 -6.784***	Nonstationary Stationary	I(1)
Log(World Coal Consumption) ΔLog(World Coal Consumption)	3 0	Yes No	-2.777 -2.796*	Nonstationary Stationary	I(1)
Log(USA's GDP) ΔLog(USA's GDP)	0 0	Yes No	-1.930 -4.321***	Nonstationary Stationary	I(1)
Log(China's GDP) ΔLog(China's GDP)	1 0	Yes No	-2.164 -3.651**	Nonstationary Stationary	I(1)

Source: Own Elaboration. The '*', '**' and '***' indicate that the statistic was significant at a level of 10%, 5% and 1%, respectively.

The results of the ADF/DF tests in Table 4 show that the World Oil Production, the World Oil Consumption and the World Refining Capacity logarithmic series can be considered trend-stationary, although for the first two the result holds only at a 10% level. On the other hand, the series of the logarithms of Crude Oil Prices, World Oil Reserves, World Gas Consumption, World Coal Consumption, USA's Gross Domestic Product and China's Gross Domestic Product seem to be nonstationary.

Table 4 also shows the results of the ADF for the first differenced series, when the series in levels were nonstationary. All first differenced series were found stationary. Thus, we can confirm that the correspondent series in levels were I(1), not I(2).

For those I(1) series, if there is, at least, one linear combination among them that is I(0), the series are said to be cointegrated, and that linear combination can be interpreted as a long-term relation. The search for cointegration was done using Johansen´s approach, which starts with a VAR model. The order of the VAR needs to be chosen, and we use the Final Prediction Error (FPE), the Akaike's information criterion (AIC), the Hannan and Quinn information criterion (HQIC) and the Schwarz's Bayesian information criterion (SBIC) with that purpose. Table 5 contains the results regarding those four criteria and the VAR's order:

Table 5 – VECM Order Selection Criteria.

Lags	FPE	AIC	HQIC	SBIC
0	3.1e-12	-9.47046	-9.37842	-9.20383
1	8.0e-18	-22.3688	-21.7245	-20.5024 *
2	4.9e-18 *	-23.0566	-21.8601 *	-19.5904
3	6.1e-18	-23.3978 *	-21.649	-18.3318

Source: Own Elaboration

In order to know how many lags should be considered, and consequently, what the order of the VECM is, one must look at how many lags minimize each of the criteria, and then choose the number of lags based on how many minimize the most criteria. With 0 lags, none of the criteria are minimized, while with 1 and 3 lags, only one criterion is minimized (the SBIC and the AIC, respectively). Therefore, by analyzing Table 5, it is possible to conclude that the order of the VECM should be 2, because that is the number of lags that minimizes most criteria (both the FPE and the HQIC).

Having chosen the order of the VAR, the Johansen tests can be applied. The results are presented in Table 6:

Table 6 – Johansen Tests.

Rank	Trace Test	Maximum Eigenvalue Test
0	104.107***	35.136
1	68.971**	28.519
2	40.452	22.296
3	18.156	9.335
4	8.821	7.099

Source: Own Elaboration

From Table 6 we notice that the Trace Test strongly rejects the null hypothesis of no cointegration, with a significance level of 1%, meaning that there is at least one cointegration vector. However, the null hypothesis of the rank = 1 is also rejected against rank ≥ 1, at a significance level of 5%, but rank = 2 is not rejected, not even at a 10% level. Thus, there is evidence that there are two cointegration vectors among the variables.

The Maximum Eigenvalue Test does not agree with the Trace Test, test (no cointegration was found using the former), but, given the strong result of the latter, we proceed to assume two cointegrating vectors and estimate the VECM model. The estimates of those cointegrating vectors are presented in Tables 7 and 8:

Table 7 – The First Cointegration Vector.

Variable	**Coefficient**	**Std. Deviation**	**Test Statistic**	**Impact**
Log(Crude Oil Prices)	1	-	-	-
Log(World Gas Consumption)	39.415	8.932	4.410^{***}	Negative
Log(World Coal Consumption)	-10.004	2.816	-3.550^{***}	Positive
Log(USA's GDP)	-9.272	3.401	-2.730^{***}	Positive
Log(China's GDP)	-2.111	0.586	-3.600^{***}	Positive
Constant	-128.795	-	-	-

Source: Own Elaboration

Table 8 –The Second Cointegration Vector.

Variable:	**Coefficient**	**Std. Deviation**	**Test Statistic**	**Impact**
Log(World Oil Reserves)	1	-	-	-
Log(World Gas Consumption)	3.975	1.969	2.020^{**}	Negative
Log(World Coal Consumption)	-0.266	0.621	0.430	Non Significant
Log(USA's GDP)	-1.445	0.750	-1.930^{*}	Positive
Log(China's GDP)	-0.240	0.129	-1.860^{*}	Positive
Constant	-3,975	-	-	-

Source: Own Elaboration

Tests for normality and no autocorrelation in the VECM model were also applied, and the results are present in Table 9. The Jarque-Bera (1980) test does not reject the null hypothesis of normality, given that the p-value of the test was 0.144. The null hypothesis of no autocorrelation is also not rejected, using the Breusch-Godfrey (1978), since the p-value obtained was 0.549:

Table 9 –Tests to the VECM Model.

Null Hypothesis	LM Statistic	p-value
Normality	17.144	0.144
No Autocorrelation	34.309	0.549

Source: Own Elaboration

Therefore, no evidence was found of autocorrelation or absence of normality in the VECM model. Since these results correspond to the assumptions of the Johansen´s model, we have further motivation to trust our results.

$$logoilprice^{*} = 128.795 - 39.415logworldgascons + 10.004logworldcoalcons + 9.272logusaGDP + 2.111logchinaGDP + \quad (1)$$

The results of equation (1) give us the long-term determinants of Crude Oil Prices. Those are: World Gas Consumption, World Coal Consumption, USA’s Gross Domestic Product and China’s Gross Domestic Product, which are all significant at a 1% level, as shown in Table 7.

The results in (1) show that, *ceteris paribus*[13], the World Gas Consumption has a negative impact on the Crude Oil Prices on a long-term relation, since this explanatory variable’s coefficient is negative. This result goes according to what was pre-established in the hypotheses of the research, since higher values of World Gas Consumption should lead to lower values of consumption in the oil market, therefore causing a decrease in its price. As for this coefficient’s value, the results seem to indicate that an increase of the World Gas Consumption in 1% leads to a decrease of 39.415% on Crude Oil Prices.

As for the coefficient of the logarithm of World Coal Consumption, it is positive, unlike what was expected. What this result indicates is that higher values of World Coal Consumption lead to higher values of Crude Oil Prices in the long run, meaning that these two variables act as complementary goods, instead of substitute goods as would be expectable. Zamani (2016, pp.801)[14] claims that “a large share of the total delivered price of coal is transportation cost”, from which we can argue that the higher the levels of coal

13 The ceteris paribus condition is assumed in all the analyses of individual impacts that follow.

14 Even though this quotation goes to support the positive long-term relation obtained between the logarithm of World Coal Consumption and the logarithm of Crude Oil Prices, Zamani (2016) actually defends that these two goods are substitute goods.

consumption, the higher its price will be, and due to the fact that most of its price is conducted by the cost of transportation, an increase in the price of oil is also expected, because higher costs of transportation in the coal market mean that more oil is being consumed to provide that transportation, which ultimately leads to increases in Crude Oil Prices. Thus, although the signal of the logarithm of World Coal Consumption's coefficient was the opposite of what was expected, there seems to be a reasonable explanation to account for it. Looking at its value, it indicates that an increase of 1% in the World Coal Consumption leads to an increase of 10.004% on Crude Oil Prices.

Observing the logarithms of the two Gross Domestic Product variables in (1), we can see that both variables presented positive coefficients, meaning that higher values of both the USA's Gross Domestic Product and China's Gross Domestic Product lead to higher values of Crude Oil Prices, as expected. The USA's Gross Domestic Product has a greater impact on Crude Oil Prices, since an increase of 1% leads to an increase of 9.272 % on oil prices, while the impact of China's Gross Domestic Product is 2.111%.

$$\begin{aligned} logworldres^* = {} & 21.031 - 3.975 logworldgascons + 0.266 logworldcoalcons \\ & + \; 1.446 logusaGDP + 0.240 logchinaGDP \end{aligned} \tag{2}$$

The second cointegration vector and its equation (2) do not hold as much interest as the first ones, since they do not represent equilibrium relations regarding Crude Oil Prices. In the second equation, the dependent variable is the logarithm of World Oil Reserves, and the logarithm of World Gas Consumption have shown significance at a level of 5% with a negative coefficient, while the logarithms of the USA and China's Gross Domestic Products are significant at a level of 10%, and presented a positive relation with the dependent variable.

Johansen' approach allow us to obtain the equilibrium relations, shown above, together with the short-term dynamics, expressed in the VECM model. The first equation of the VECM could be our dynamic model for the oil prices. However, we choose to estimate a single-equation ECM model with more information, such as the I(0) variables that were left out of the cointegration analysis, and a dummy variable that could allow for price changes due to speculation. The residuals from the cointegration relations are included, so that the error correction mechanism is essentially the same. Those residuals, or estimated equilibrium errors, are defined by equations (3) and (4).

$$\hat{u}_t = logoilprice_t - logoilprice_t^* \quad (3)$$

$$\hat{v}_t = logworldres_t - logworldres_t^* \quad (4)$$

Where $logoilprice^*$ and $logworldres^*$ have been defined in equations (1) and (2).

The ECM model was estimated for the change in logarithm of the Crude Oil Prices, that is, the growth in Crude Oil Prices, as the dependent variable. The right-hand side variables included: the lagged dependent variable; the I(0) variables, contemporaneous and lagged; the change in the I(1) variables, also contemporaneous and lagged; the lagged residuals from the equilibrium relations, $\hat{u}_{t-1}$ and $\hat{v}_{t-1}$; a linear trend that accounts for the trend-stationarity of the I(0) variables in levels; and the dummy variable related to the period of supposedly higher financial speculation. A parsimonious model was then built by taking out the less statistically significant variables, provided that the model would still be dynamically complete (no autocorrelation).

The results of the final ECM model are shown in Table 10:

Table 10 –The Final ECM Model for the Growth in Crude Oil Prices.

Variable	Coefficient	Std. Deviation	p-value	Impact
ΔLog(World Gas Consumption)$_{t-1}$	3.554	2.026	0.091*	Positive
ΔLog(World Coal Consumption)$_t$ ΔLog(World Coal Consumption)$_{t-1}$	-5.062 7.537	1.889 1.749	0.012** 0.000***	Positive
Log(World Oil Consumption)$_{t-1}$	5.708	2.536	0.033**	Positive
Financial Speculation Dummy	0.083	0.103	0.430	Non Significant
$\hat{u}_{t-1}$	-0.208	0.081	0.016**	Negative
$\hat{v}_{t-1}$	1.358	0.513	0.013**	Positive
t	-0.082	0.034	0.021**	Negative

F-Test Statistic	5.040***
F-Test p-value	0.001***
R-Squared	0.599
Adjusted R-Squared	0.480

Source: Own Elaboration

$$\Delta logoilprice_t = -44.849 - 0.082t + 3.554\Delta logworldgascons_{t-1} - 5.062\Delta logworldcoalcons_t + 7.537\Delta logworldcoalcons_{t-1} + 5.708 logworldoilcons_t - 0.208\hat{u}_{t-1} + 1.358\hat{v}_{t-1} \quad (5)$$

Tests to the functional form of the model, for autocorrelation, for heteroskedasticity and for the possible presence of autoregressive conditional heteroskedasticity were applied to the model and the results can be seen in Table 11. Ramsey's RESET test presented a p-value of 0.190, which means that the null hypothesis that the functional form of the model is correct is not rejected at the usual significance levels. The Breusch-Godfrey test for Autocorrelation indicates that the null hypothesis that there is no autocorrelation in the model should not be rejected, given that its p-value is 0.177. As for the Breusch-Pagan test for heteroskedasticity, the p-value obtained is 0.577, thus, the null hypothesis that there is no heteroskedasticity in the model is not rejected. Finally, the LM test for ARCH effects also has a p-value that does not reject its null hypothesis, being that the value is 0.689 and therefore there seems to be no evidence of ARCH effects in the model.

Table 11 –Tests to the Final ECM Model.

Test	Test Statistic	p-value
Ramsey's RESET Test	1.720	0.190
Breusch-Godfrey Test for Autocorrelation	1.826	0.177
Breusch-Pagan Test for Heteroskedasticity	0.310	0.577
LM Test for ARCH Effects	0.160	0.689

Source: Own Elaboration

Hence, there are statistical evidences to state that the model's functional form is correct, there is no autocorrelation, there is no heteroskedasticity and there are no ARCH effects in it.

Looking at Table 10 and equation (5), we can start by saying that the model presents global significance, since the p-value of the F-Test for global significance is 0.001, and therefore is significant at levels of 10%,5% and 1%. Another interesting statistic that must be interpreted is the R-Squared value, which measures the proportion of the variation in the dependent variable that can be explained by the explanatory variables included in the model. In this case, it is possible to claim that the explanatory variables contained in the final ECM model explain 59,90% of the dependent variable's variations.

The growth in World Gas Consumption, the growth in World Coal Consumption, the growth in World Coal Consumption with one time lag, the growth in World Oil Consumption, the residuals from the VECM model with on time lag and the time component, all revealed themselves to be significant.

The growth in World Gas Consumption is significant, even though only at a significance level of 10%. This variable's positive signal in the ECM model contradicts the negative signal that was obtained in the cointegration model's first equation. The ECM model's results seem to indicate that there is a positive relation between the World Gas Consumption and Crude Oil Prices, meaning that these two commodities act as complementary goods, since higher levels of consumption in the gas market seem to coexist with higher levels of consumption in the oil market, thus leading to increases in Crude Oil Prices. According to Table 10, an increase of 1 percentage point in last year's growth of World Gas Consumption leads to an increase of 3.554 points in this year's growth of the prices of oil.

The same thing can be said about the results of the World Coal Consumption in the ECM model, since this variable's coefficient is -5.062 in moment 't', and 7.537 when there is a one time-period lag. Since the positive impact's absolute value is higher than the negative impact's absolute value, we can state that the overall impact of the World Coal Consumption on Crude Oil Prices is positive, just as the World Gas Consumption's impact. The difference here is that this result goes according to the long-term relationship that was obtained in the VECM model, where the impact of the World Consumption of Coal had already revealed itself to be positive. In the ECM model, the logarithm of World Coal Consumption and the logarithm of World Coal Consumption with one time-period lag were significant at significance levels of 5% and 1%, respectively.

The logarithm of World Oil Consumption, the only trend-stationary variable that was included in the final ECM model, was significant at a level of 5%. Its impact on Crude Oil Prices is positive, according to the results, and that means that higher values of World Oil Consumption can be associated with higher values of Crude Oil Prices, which is in accordance with the hypothesis that was defined regarding this explanatory variable. An increase of 1 percentage point in the of World Oil Consumption leads to an increase of

5.708 percentage points in the growth of Crude Oil Prices, according to the results obtained.

As was already explained, not all the variables that were taken into consideration for the ECM model ended up being included in it, since some of them were highly non-significant. The Financial Speculation Dummy variable's p-value of 0.430 clearly indicates that this variable was also one of the non-significant ones. However, the dummy variable was still included in the final ECM model, so that results regarding financial speculation could be approached and commented anyway. The non-significance of this variable in the ECM model can be interpreted in several ways: it can be said that, perhaps the time-period that was chosen does not truly represent a period of higher financial speculation, or; it can be said that while that time period does indeed reflect a period of presumably higher financial speculation activity, that activity simply did not have a significant impact on the determination of crude oil prices.

The coefficient of the lagged equilibrium error u is -0.208, negative and significant. Being negative, it gives the model the interpretation of an ECM – it means that the equilibrium error of the last period leads to a change of opposite sign in the prices, thus, bringing the variable back to its equilibrium value. The adjustment rate is 0.208 per year.

As for the interpretation of $\hat{v}_{t-1}$, it does not hold great interest, since it resulted from the second cointegration vector of the VECM model, which represents an equilibrium relation between variables other than Crude Oil Prices. Still, this variable was included in the ECM model and presented significance at a level of 5%.

Finally, looking at the results regarding the time component, 't', that was also included in the ECM, it can be said that there is significance at a level of 5%. The linear trend coefficient tells us that the growth rate of oil prices, holding all other things equal, decreased in this time period approximately 0.082 per year.

5 – Conclusions

The main objective of this research is understanding which economic and financial variables are the ones that most directly influence crude oil prices and the way that this time series behaves.

In order to do that, several hypotheses were established so that it would be possible to analyze how the variables that were included in the research and turned out to be significant would impact the dependent variable. An unit root and cointegration analysis was performed, in the context of a VECM model, based on data obtained from the BP Statistical Review of World Energy 2018 and the Federal Reserve Bank of St. Louis' Federal Reserve Economic Data databases, regarding annual observations of 38 time-periods for each variable, from the periods between 1980 and 2017.

From all the modelling and estimation techniques that were applied throughout this dissertation, it is possible to conclude that the variables that have a significant long-term relation with Crude Oil Prices are World Gas Consumption, World Coal Consumption, USA's Gross Domestic Product and China's Gross Domestic Product. Our results revealed: positive long-term elasticities between Crude Oil Prices and World Coal Consumption, USA's Gross Domestic Product and China's Gross Domestic Product, respectively. On the other hand, a negative long-term elasticity was found between Crude Oil Prices and World Gas Consumption.

These conclusions agree with these explanatory variables' respective hypotheses, except for the conclusion regarding the World Coal Consumption, which contradicted the pre-established hypothesis that higher values of World Coal Consumption should lead to lower Crude Oil Prices. However, this unexpected sign also appeared in the literature, and may be related to transportation costs, as discussed.

Although the long-term analysis is the most important as far as the oil prices determinants are concerned, we decided to further estimate a single-equation dynamic ECM model, which allowed us more freedom in choosing the final set of variables to include as explanatory variables for the growth of oil prices. This kind of model is more useful for prediction purposes. The variables that were significant in the ECM model were the growth of World Gas Consumption, the growth of World Coal Consumption, the growth of World Oil Consumption, the 'equilibrium errors' that resulted from the cointegration relations, and the time trend, 't'. The results of this model revealed that higher growth rates of Crude Oil Prices tend to be associated with higher growth rates of: World Gas Consumption; World Coal Consumption; World Oil Consumption; as time goes by, the rate of growth of Crude Oil Prices tends to decrease around 0.082 per year holding all other things equal; the 10 year time period that anticipated the big financial crises of 2008, on which higher speculative action could have been developed, did not influence Crude Oil Prices in a significant manner. These conclusions went according to the hypotheses only when it comes to the series of Crude Oil Consumption. The conclusion regarding World Gas Consumption contradicted not only its hypothesis, but also the result obtained in the VECM model, while the conclusion regarding World Coal Consumption contradicted its hypothesis but was in accordance with the VECM model's results. However, we must take into account that the hypotheses were formulated for the levels of the variables, and in the ECM model we have the growth rates. As for the dummy variable, we can conclude that either the time period that was chosen does not truly represent a period of higher financial speculation, or perhaps that speculative activity simply did not have a significant impact on the determination of crude oil prices.

It is believed that the results and conclusions that were reached with this research give a contribution to the existing literature on the topic and can be helpful in clarifying other authors on which variables they should ponder when analyzing the crude oil prices series.

Overall, the objectives of the research have been reached, although it would have been interesting to obtain more definite results and conclusions regarding the role of financial speculation. Another limitation of the research lies on the sample size, which could have been bigger even with annual observations, if there was more information of easy access regarding some of the time-series utilized.

As for future researches/developments on this topic, perhaps approaching the same objectives of this research while recurring to monthly or daily data could allow other researchers and authors to obtain different results, which could then be compared and analyzed alongside with the results that were reached in this dissertation. Approaching the quantification of financial speculation's influence on crude oil prices in different ways could also give future contributions to the topic.

References

AL-FATTAH, S., GUCKWA, M., HUANG, Z., HUNTTINGTON, H. & NOURI, A. (2013): “Oil Markets and Price Movements: A Survey of Determinants” *SSRN Electronic Journal.* pp. 5-7, 16-23 & 64-66

ALQUIST, R. & GERVAIS, O. (2011): “The Role of Financial Speculation in Driving the Price of Crude Oil” *The Energy Journal*. Vol. 34

BADR, O., EL-KHADRAWI, A. (2016): “The Determinants of Crude Oil Prices”. pp. 1-7

BEKIROS, S. & DIKS, C. (2008): “The Relationship Between Crude Oil Spot and Future Prices; Cointegration, Linear and Nonlinear Causality”. *CeNDEF.* pp. 2-7 & 19-21.

BERA, A. & JARQUE, M. (1980): "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals". *Economics Letters*. Vol. 6, pp. 255–259.

BREUSCH, T. (1978): “Testing for Autocorrelation in Dynamic Linear Models”. Australian Economic Papers. Vol. 17, pp. 334-355.

BREUSCH, T. & PAGAN, A. (1979): “A Simple Test for Heteroskedasticity and Random Coefficient Variation” Econometrica. Vol.47, pp.1287-1294.

CIFARELLI, G. & PALADINO, G. (2010): “Oil Price Dynamics and Speculation: A Multivariate Financial Approach“. *Energy Economics.* Vol. 32, pp. 363-372.

DEES, S., GASTEUIL, A., KAUFMANN, R. & MANN, M. (2008): “Assessing the Factor Behind Oil Prices Changes”. *Energy Economics*. Vol. 30, pp. 2609-2622.

DENG, A., KING, K. & METZ, D. (2012): “An Econometric Analysis of Oil Price Movements: The Role of Political Events and Economic News, Financial Trading and Market Fundamentals” *Bates White*. pp.1-6 & 24-28.

DICKEY, D. & FULLER, W. (1979): “Distribution of the Estimators for Autoregressive Time Series with a Unit Root” *Journal of the American Statistical Association*. Vol.74, pp. 427–431

ENGLE, R. & GRANGER, C. (1987): “Co-Integration and Error Correction: Representation, Estimation, and Testing” *Econometrica*. Vol. 55, pp. 251-276.

FATTOUH, B. (2008): "The Dynamics of Crude Oil Prices Differentials" *Oxford Institute for Energy Studies.* pp. 2-20

FERREIRA, P. (2016): "Princípios de Econometria" *Rei dos Livros.* pp. 193-298.

FREY, G., MANERA, M., MARKANDYA, A. & SCARPA, E. (2009): "Econometric Models For Oil Prices Forecasting: A Critical Survey". *CESifo Forum 2012*. Vol.1, pp. 29-49.

GODFREY, L. (1978): "Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables" *Econometrica.* Vol.46, pp. 1293 – 1301.

HAMILTON, J. (2009): "Understanding Crude Oil Prices". *The Energy Journal.* Vol.30, pp. 179-206.

JAGER, S. (2008): "Nonlinear and Stochastic Dynamical Systems Modeling Price Dynamics – Aspects of Financial Economics in Oil Markets" *Econbiz.* pp. 1-9 & 86

JOHANSEN, S. (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models" *Econometrica.* Vol. 59, pp. 1551-1580

KILLIAN, L. (2008): "Not All Oil Price Shocks are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market" *Journal of American Economic Review.* Vol. 99, pp. 1053-1069

KILIAN, L. & LEE, T. (2014): "Quantifying the Speculative Component in the Real Price of Oil: The Role of Global Oil Inventories" *Journal of International Money and Finance.* Vol.42, pp. 71–87.

KILIAN, L. & MURPHY, D. (2014): "The Role of Inventories and Speculative Trading in The Global Market for Crude Oil." *Journal of Applied Econonometrics.* Vol. 29, pp. 454–478.

LIU, C. & LIANG, C. (2017): "The Effect of Speculation in Futures Market on Oil Price" *2017 IAEE International Conference Singapore.*

LIU, L., WANG, Y., WU, C. & WU, W. (2016): "Disentangling the Determinants of Real Oil Prices" *Energy Economics.* Vol 56, pp 363-373.

OLIVEIRA, V. (2017): “Does Speculation Affect Oil Price Volatility?” *ISCAC Master’s Thesis.*

RAMSEY, J. (1969): “Tests for Specification Errors in Classical Linear Least Squares Regression Analysis” *Journal of the Royal Statistical Society, Series B.* Vol.31, pp. 350-371.

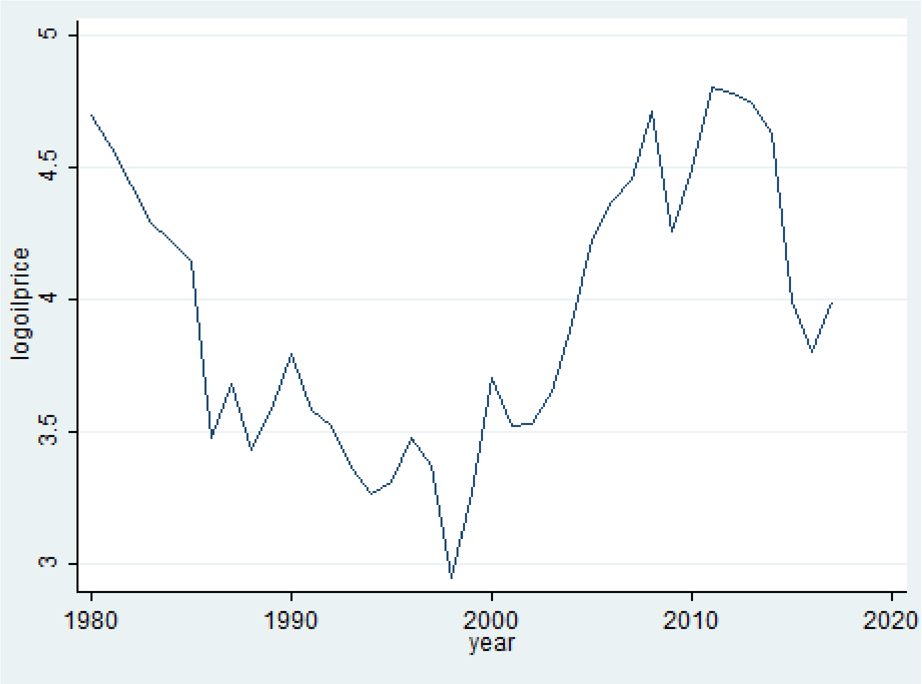
REITZ, S., SLOPEK, U. (2008): “Nonlinear Oil Prices Dynamics – A Tale of Heterogeneous Speculators?” *Deutsche Bundesbank.* pp.1-13.

ZAMANI, N. (2016): “The Relationship Between Crude Oil and Coal Markets: A New Approach” *International Journal of Energy Economics and Policy*. pp. 801-805.

ZHANG, Y. (2013): “Speculative Trading and WTI Crude Oil Futures Price Movement: An Empirical Analysis” *Applied Energy.* Vol.107, pp. 394-402.

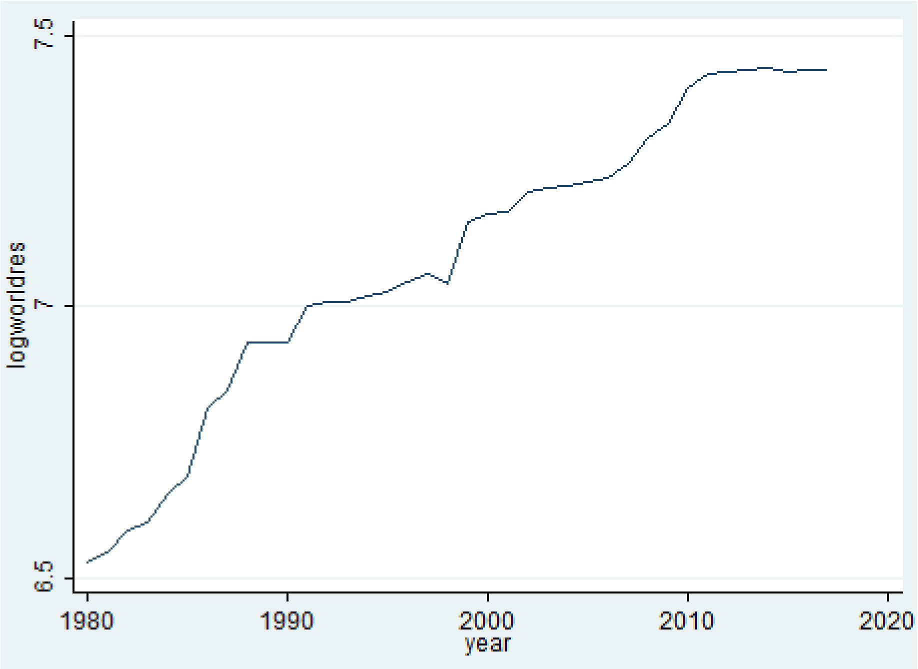
APPENDIX

Fig. 10 – Logarithm of Crude Oil Prices' Time-Series Graph.



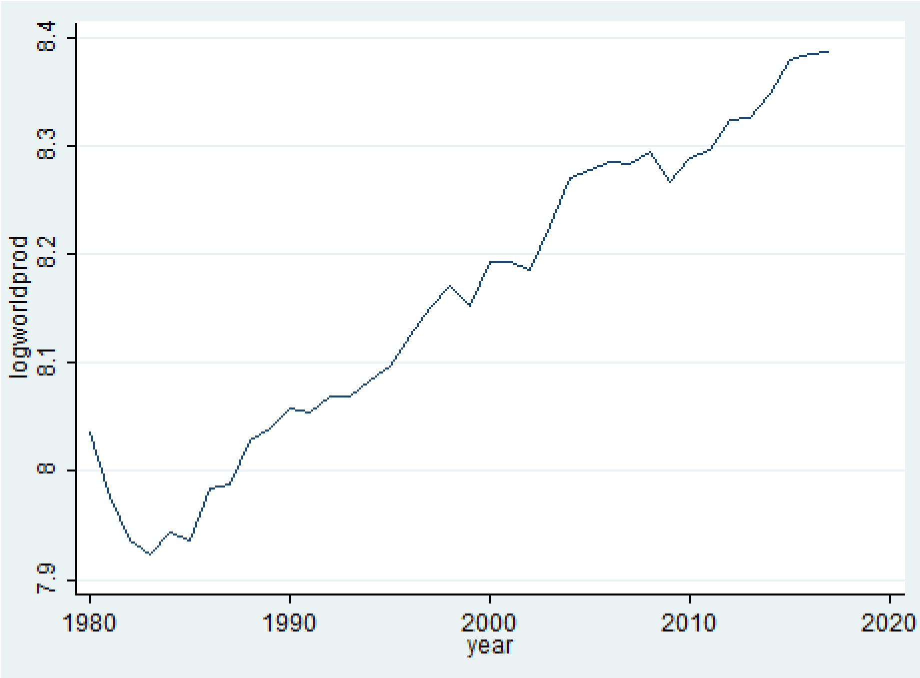
Source: Own Elaboration

Fig. 11 – Logarithm of Worldwide Oil Reserves' Time-Series Graph.



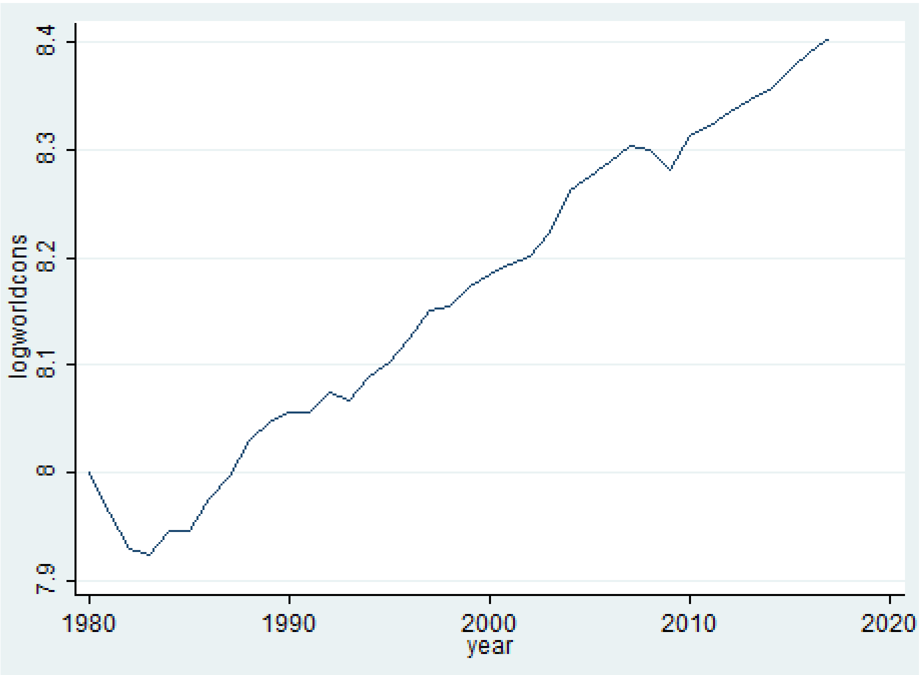
Source: Own Elaboration

Fig. 12 – Logarithm of Worldwide Oil Production's Time-Series Graph.



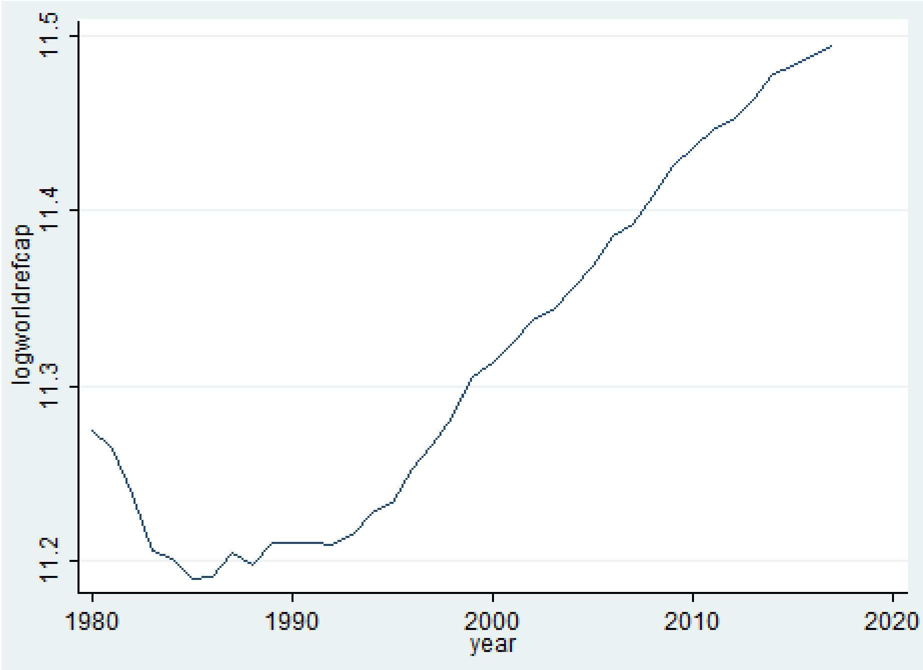
Source: Own Elaboration

Fig. 13 – Logarithm of Worldwide Oil Consumption's Time-Series Graph.



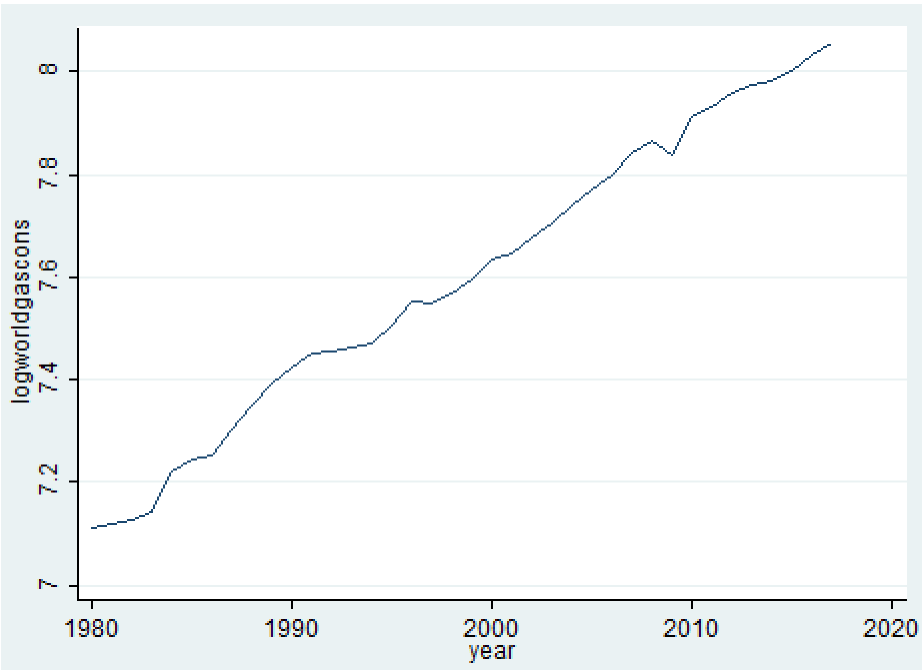
Source: Own Elaboration

Fig. 14 – Logarithm of Worldwide Oil Refining Capacity’s Time-Series Graph.



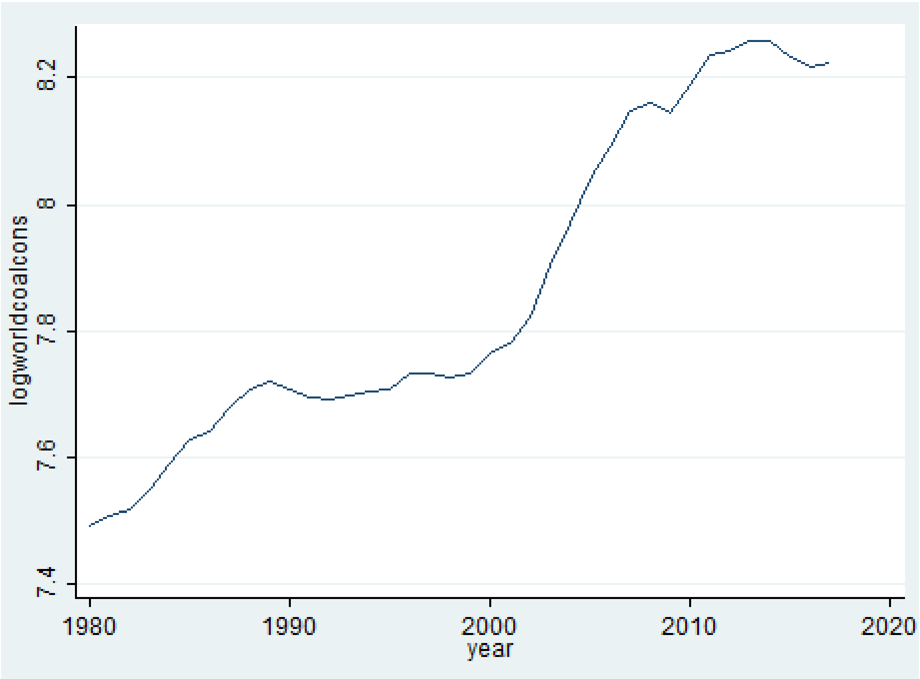
Source: Own Elaboration

Fig. 15 – Logarithm of Worldwide Gas Consumption’s Time-Series Graph.



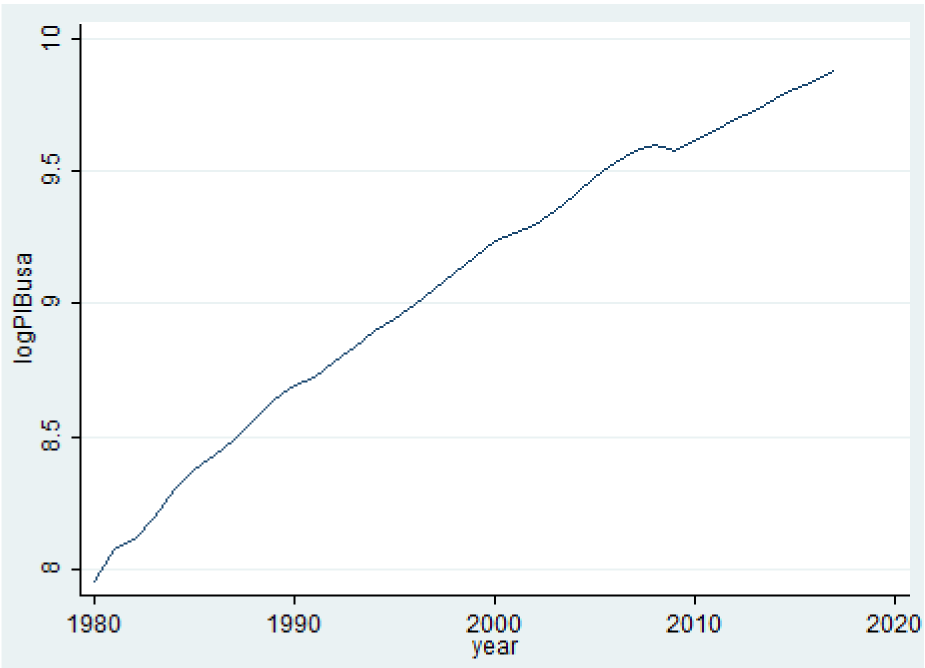
Source: Own Elaboration

Fig. 16 – Logarithm of Worldwide Coal Consumption's Time-Series Graph.



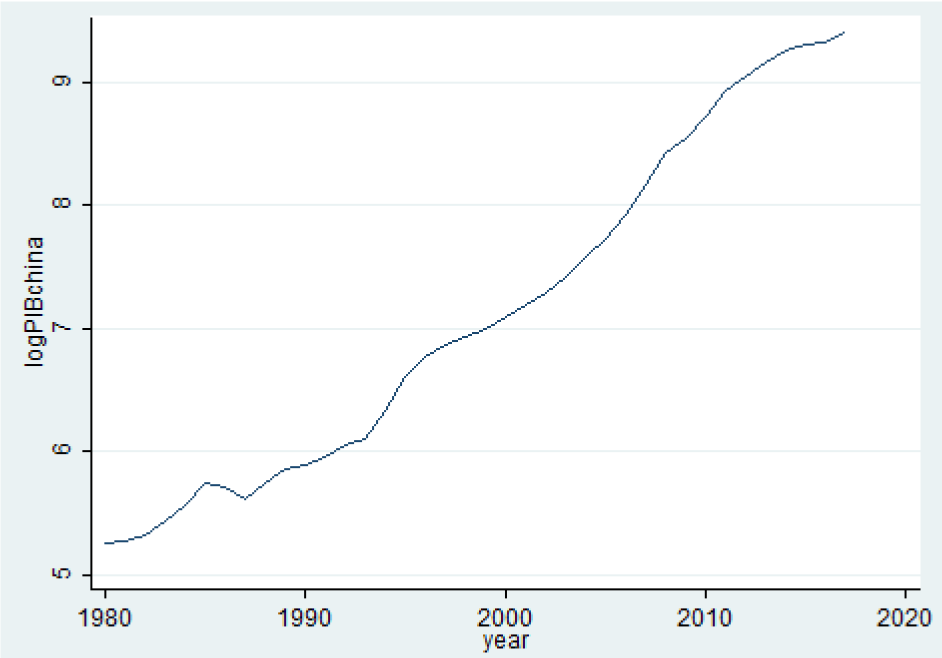
Source: Own Elaboration

Fig. 17 – Logarithm of USA's Gross Domestic Product's Time-Series Graph.



Source: Own Elaboration

Fig. 18 – Logarithm of China's Gross Domestic Product's Time-Series Graph.



Source: Own Elaboration