SPILLOVER EFFECTS BETWEEN OIL AND NATURAL GAS PRICES: NEW EVIDENCE FROM INCREASING SHARES OF NATURAL GAS IN THE ENERGY MIX

KONSTANTINIDIS CHRISTOS

SID: 3302120025

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Energy Systems

NOVEMBER 2013

THESSALONIKI-GREECE
SPILLOVER EFFECTS BETWEEN OIL AND NATURAL GAS PRICES: NEW EVIDENCE FROM INCREASING SHARES OF NATURAL GAS IN THE ENERGY MIX

KONSTANTINIDIS CHRISTOS

SID: 3302120025

Supervisor: Prof. Nikolaos Apergis

Supervising Committee Members: Prof. Dr.

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Energy Systems

NOVEMBER 2013

THESSALONIKI-GREECE
ACKNOWLEDGEMENTS

On the conclusion of my dissertation I would like to express my warmest thankfulness to my supervisor, Professor Nikolaos Apergis. With his continuous support and guidance and his cooperative spirit he encouraged me to work hard and achieve my research. Furthermore, my deep gratitude goes to a member of the academic staff of the International Hellenic University, Dr. Theologos Dergiades. He offered me continuous provision of advice regarding the econometric part of my dissertation.
ABSTRACT

Over the recent years, the oil markets had to confront with a big challenge coming from natural gas markets. The penetration of natural gas (NG) in energy mix on a worldwide basis has provoked differentiation in oil pricing and has changed the geopolitical balances. Nowadays, NG has prevailed in energy market. The globalization of NG market in conjunction with NG economic and technical characteristics reinforce its presence in energy markets. This study examines the econometric relationship between West Texas Intermediate (WTI) crude oil price and Henry Hub (HH) natural gas price in the US. The two fuels are highly correlated and present a strong connection. Yet, the considerable changes in the US energy market the last two decades and the tremendous growth of NG globally altered the field. The last researches on topic captured that the oil and gas prices in the US move independently from each other at least in the short-run. The main question is if the divorce between gas and oil prices is permanent. This research performs well-known econometric techniques in order to reveal the relationship between HH and WTI spot prices since 1997 until now. The results showed that HH and WTI have not a long-run cointegration equation and after an exogenous shock are not converged in an equilibrium point. Prior to these evidence three exogenous economic variables are checked in order to find their impact on the HH and WTI time series. The NG residential consumption and the storage level in the US did not reveal any impact on HH and WTI in contrast to Gross Domestic Product (GDP) that affects considerably the two fuels’ prices. The causality testing showed that the two series cause each other but only in the short-run under the absence of cointegration. Furthermore the volatility of two time series do not have impact on the HH and WTI price determination. Although the oil and gas in the US are strongly connected the first elements of a separate price movement is already a fact.

Keywords: natural gas, oil, Henry Hub, West Texas Intermediate, spot prices, causality, cointegration, volatility.
# CONTENTS

**ACKNOWLEDGEMENTS** ........................................................................................................................................... 3

**ABSTRACT** .......................................................................................................................................................... 4

1. INTRODUCTION .................................................................................................................................................. 7

2. LITERATURE ....................................................................................................................................................... 11

2.1 (Dis)Integration trends in energy markets ........................................................................................................ 11

2.1.1 Gas integration trend .................................................................................................................................. 11

2.1.2 Oil disintegration trend ............................................................................................................................ 12

2.2 Dominant players in gas market .................................................................................................................... 13

2.2.1 NOCs presence in oil sector .................................................................................................................... 14

2.2.2 NOCs presence in gas sector .................................................................................................................. 14

2.3 Oil and gas linkage .......................................................................................................................................... 15

2.3.1 The reasons of linkage ............................................................................................................................ 16

2.3.2 Economic factors that strengthen the oil-gas linkage ............................................................................... 18

2.4 Topic research ................................................................................................................................................. 19

2.4.1 European Union ...................................................................................................................................... 20

2.4.2 Pacific Asia .............................................................................................................................................. 21

2.4.3 North America ....................................................................................................................................... 21

3. METHODOLOGY .................................................................................................................................................. 28

3.1 Unit root and Stationarity tests ...................................................................................................................... 28

3.1.1 Augmented Dickey-Fuller (ADF) test ....................................................................................................... 28

3.1.2 Phillips-Perron (PP) test ........................................................................................................................ 30

3.1.3 Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test ........................................................................... 31

3.2 Cointegration tests .......................................................................................................................................... 32

3.2.1 Engle and Granger test ........................................................................................................................... 32

3.2.2 Phillips-Ouliaris test .................................................................................................................................. 34

3.2.3 Johansen cointegration test .................................................................................................................... 34

3.3 Vector Autoregression Model (VAR) ............................................................................................................. 35

3.4 Causality tests ................................................................................................................................................... 36

3.4.1 Toda-Yamamoto test-Linear causality ....................................................................................................... 36

3.4.2 Wald test-Linear causality ....................................................................................................................... 37

3.4.3 Breitung and Candelon test-Nonlinear causality .................................................................................... 38

3.5 ARCH and GARCH estimation .................................................................................................................... 40

3.5.1 GARCH ...................................................................................................................................................... 40
4. EMPIRICAL APPLICATION .............................................................................................. 43
  4.1 Data sources ............................................................................................................ 43
  4.2 Stationarity testing .................................................................................................. 48
  4.2.1 Endogenous variables ....................................................................................... 48
  4.2.2 Exogenous variables ......................................................................................... 50
  4.3 Cointegration testing .............................................................................................. 53
  4.4 Coefficients estimation .......................................................................................... 55
  4.4.1 Parameter estimates .......................................................................................... 56
  4.5 Causality testing ..................................................................................................... 59
  4.5.1 Wald test-Linear causality ................................................................................. 59
  4.5.2 Breitung and Candelon-Nonlinear causality ...................................................... 61
  4.6 Volatility effects testing ........................................................................................ 63
  4.6.1 GARCH (1,1) .................................................................................................. 63
  4.6.2 Causality testing ................................................................................................ 64
    4.6.2.1 Toda-Yamamoto-Linear causality .............................................................. 64
    4.6.2.2 Breitung and Candelon-Nonlinear causality .............................................. 66
5. CONCLUSION-DISCUSSION ......................................................................................... 70
REFERENCES .................................................................................................................. 76
1. INTRODUCTION

During the previous years the energy market has experienced considerable changes. The traditional energy carriers as coal and petroleum products hold their share in energy mix while new energy sources as the renewables increase their presence in energy markets. The fuel that presents the most significant change is the natural gas (NG). The NG has increased its share in energy mix and enjoys tremendous growth.

NG augmented presence in energy mix is explained by many reasons. Firstly, the rising demand mainly from developing countries is covered largely by NG. Additionally, the NG is considered as “clean energy” due to low carbon dioxide emissions in contrast to coal and oil. The Kyoto protocol and other global commitments that were adopted mainly from western countries acquire the use of less polluting fuels in order to achieve the corresponding environmental targets. Furthermore, the technological breakthrough in the US has increased the shale gas production considerably. The minimization of the gas production cost gives an impetus for further use of NG in energy consumption. Finally, the liquefied natural gas (LNG) trade growth that is taking place nowadays diversifies the possible energy sources for the importer countries. The geopolitical importance of a diversified energy mix and the energy security factors make the LNG a very important alternative fuel with high prospects of further rising.

In this framework it is very important to examine the linkage of NG with the traditional petroleum products. NG was shown in the energy market much later than the traditional oil products. The pricing of this new product was based on oil prices and this relationship has been continued until now. The reasons for the strong relationship vary.

To begin with, the “common nature” of oil and gas products that are found in the same wells strengthen the corresponding linkage. Thus any deviation in the production of one product affects the other one considerably. The changes in the affected product may be either in the production rates or they may have a direct impact on energy prices.
In addition, the strong correlation between the two fuels derives from the connection that is presented in their final use. Despite the fact that the substitutability between the two fuels has reduced regarding the previous years, many industrial and residential users have the ability to shift from one fuel to another. This option in the final use of energy carriers develops a strong connection between oil and gas.

Furthermore, the main reason for which oil and gas are highly correlated and do not move independently from each other is that NG supply contracts are based on the oil price. The state-owned energy companies cover the largest part of the NG supply globally. Due to the fact that NG production projects are very costly and the energy prices in a liberalized market present volatility the guarantee of the returns is essential for the energy companies. In order to assure the high level of returns the NG supply agreements are oil-indexed contracts that consist of stringent clauses. Also the fact that new, costly NG supply projects are under construction or are planned for the future exacerbate the problem and strengthen the oil and gas linkage.

All the above reasons explain sufficiently the strong connection between the two fuels and their corresponding co-movement. The majority of previous studies on topic captured the correlation of the two fuels and their similar price path and confirmed the potential linkage.

On the other hand, in previous years some studies revealed that the connection between the fuels is altering; they consider that the two fuels follow separate and independent ways in their price route, at least in the short-term. The price drivers of NG and oil are different and independent from each other and that donates a relative autonomy in NG price. This evidence corresponds to the US energy market that has different characteristics from the European and Asian ones. In European and Pacific Asia markets the connection is still very mighty and will continue to be.

The reasons for the altering in the oil and gas linkage are many. Nowadays, the substitutability between the two fuels has declined due to the economic and technical characteristics of the two energy fuels. The common presence of the fuels is encountered in domestic use for heating purposes and in double-fired boiler in the industrial sector and tends to disappear. On the other hand, oil is used mainly in the transport sector in contrast to the NG that is used in the industrial sector and in power generation where it substitutes coal and nuclear power. Thus the different use of the
two fuels leads to disparate price movements. In addition, the liberalized energy markets that have been developed create a price determination that corresponds to the supply and demand equilibrium in contrast to the oil-indexed NG agreements. These liquid markets provide a variety of spot and future contracts and lead to the oil and gas decoupling trend. Despite the fact that the liberalized and liquid markets in other continents have not the desirable results for many reasons, in the US they are the opposite. Finally, in the US the shale gas overproduction of the last years in conjunction with the insufficient infrastructure as the obsolete pipeline transmission system and export capacities created an oversupply of NG. These high rates of NG storage pushed down the price of NG which decoupled from oil price. The first evidence from an independent movement of NG is fact, at least in the short-run. Yet, the main question is whether the divorce between the fuels is permanent.

Therefore, the main goal of this study is to examine the relationship between NG and oil prices and their potential decoupling trend. Also the effects of the oil price on the NG price determination are investigated.

The market in which the study focuses on is the US energy market. Henry Hub (HH) and West Texas Intermediate (WTI) are chosen for the NG and oil prices, respectively. The fact that HH prices are available since 1997 reduces the analysis’ interval. The data are spot prices from HH and WTI in a quarterly basis from Q1 1997 to Q4 2013. Regarding with the time series the log level is used. A number of well-known econometric techniques are performed in order to examine the relationship of HH and WTI spot prices.

Firstly, the study examines the time series in order to detect their integration order. The unit root and stationarity testing is implemented with the Augmented Dickey-Fuller (ADF) test, Phillips Peron (PP) test and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test. The next step is the examination of long-run equilibrium between the WTI and HH. Three different cointegration tests as Engle and Granger, Phillips-Ouliaris, and Johansen tests are performed. According to the cointegration test there is no cointegration equation between the HH and WTI.

Therefore, a Vector Autoregression (VAR) is chosen in order to investigate the possibility of causality between the WTI and HH time series. The VAR consists of two endogenous variables (WTI and HH) and three exogenous ones as Gross
Domestic Product (GDP), NG storage and residential consumption in the US. Prior to testing for causality, the coefficients parameters estimates of exogenous variables are checked to reveal how statistical significant are in WTI and HH time series and the degree that affects them. Additionally the linear causality test is performed under a Wald test framework. The Wald test includes a set of restrictions detecting the potential causality between the HH and WTI. Except for the linear causality test the possibility of the nonlinear causality is examined. The nonlinear causality testing implements a Breitung-Candelon approach checking for causality in a frequency domain.

The most economic variables present volatility in their prices. The effect of the volatility in investment is important due to the fact that increases the corresponding risk. This study investigates the volatility effect of WTI on HH and vice versa. For that reason a Generalized Autoregression Conditional Heteroscedasticity (GARCH) model is performed. After the GARCH implementation the conditional variance of the model is the volatility of the time series. The new time series of volatility are tested to detect the possibility of causality between them. Furthermore the causal relationships among the HH and WTI volatility time series and the log levels of HH and WTI are checked. The existence of causality in the volatility framework requires testing for linear and nonlinear causality. The linear causality is performed with the Toda-Yamamoto test while the nonlinear test following the Breitung-Candelon approach.

The remainder of this dissertation is structured as follows. Chapter 2 consists of the theoretical framework of the oil and gas market and their strong linkage in conjunction with the existing literature on the subject. Chapter 3 describes the methodology that is performed. Chapter 4 provides the data and presents the empirical application and the corresponding results of the econometric analysis. Finally, Chapter 5 summarizes the findings and presents a discussion about them.
2. LITERATURE

2.1 (Dis)Integration trends in energy markets

Over the recent years, oil and gas markets have been in a development process. The market environment is more competitive and challenging and the balances have changed. Also the market is more volatile as new technological breakthroughs and geopolitical shifts affect the stability of previous years. The gas sector is evolving tremendously under the unconventional expansion and the trade in liquefied natural gas (LNG). The US shale gas production has changed the field as the US are evolving into a natural gas (NG) exporter and reduce their dependence over the traditional exporter countries. Also the growth of National Oil Companies (NOCs) and their progress in the extraction and supply efficiency make them dominant players in the market. Oil and gas markets follow different paths in their operational models. The oil sector appears a tendency to disintegration in contrast with the gas sector that follows higher vertical integration. The vertical integration up to now was the dominant model but since all the above changes have occurred, it is not the prevailing model. Different models and different players in the market compose a complex market environment.

2.1.1 Gas integration trend

Nowadays the gas market is under an evolving process and the differences in the sector regarding with the previous decades are tremendous. The US shale gas development and the rise of unconventional globally create a new framework in the gas supply. The LNG trade growth diverses the energy sources for importer countries and adds more flexibility in the energy market. The new dominant players develop partnerships in order to eliminate their lack in technology expertise and to compete the traditional players with success. Due to high capital intensity in natural gas, partnerships between small and big players are shaped in the market and stock shares are acquired from companies unrelated with the energy market. The new opportunities that are created and the strong prospects that follow the natural gas future, inspire a vast amount of investments in the gas sector. New, non-traditional players such as
banks, utility and car industry companies and hedge funds hold shares in unconventional upstream businesses in natural gas and especially in LNG projects. Also National Oil Companies invest in new projects and in corresponding infrastructure in order to assure a part of earnings that are coming from the tremendous growth of the LNG and natural gas development. These projects are capital intensive due to the fact that innovative technologies and supply chain advanced improvements lead to the growth of the projects. The Research and Development (R&P) in Exploration and Production (E&P) companies in the upstream activities and the upgrade of supply chain efficiency are high-cost processes and the partnerships, in the profitability framework, are necessary. All the above lead the companies to shift from the obsolete business models to new, which correspond in the altering market conditions. As a result, a higher vertical integration is developed in the sector. The prospects of high profitability in gas sector lead non-traditional companies to invest in both upstream and downstream projects with unrelated activities regarding their current ones. This fact results in vertical integration in the gas sector. On the other hand, the dominant players in the market explore the opportunity that emanates from the market conditions and increase their integration level. Finally, the dominant players which control a vast fleet can attain cost reducing and involve into the spot trading markets to explore arbitrage opportunities from cost inefficiencies and differences in markets.

2.1.2 Oil disintegration trend

The petroleum products contain an important share in energy mix globally. Despite the fact that the oil market is dominant in energy field, its evolution is very different from the natural gas market development. The oil sector does not reflect high growth prospects like the gas sector. In the US has been observed a reduction in energy consumption (probably owing to economic crisis and fuel competition) in conjunction with the high crude oil prices. The result was a decline in the refining margins. Also the production in upstream activities rose considerably. That leads many companies to redefine their strategy to downstream activities. Major companies as Exxon Mobil and Chevron acquired profits from sales in downstream operations.
This issue is very controversial in our days. It is not generally accepted that disintegration strategy is efficient and moves in the right way, but the integrated model value is characterized from uncertainty. However, the appropriate and profitable model depends on the magnitude of the company and its portfolio. The supermajors that have large economies of scale and a diversified portfolio with global upstream activities can follow an integrated model with success. The NOCs are the appropriate companies to implement an integration model. The rising demand in their countries and the low pressure for returns allows them to integrate their activities. Finally the tremendous growth in petrochemical activities mainly in East Asia creates strong tendency for partnerships between NOCs and International Oil Companies (IOCs).

To sum up, the oil sector is removed from the vertical integration as the market conditions are not the appropriate ones mainly for medium-size companies. On the other hand, the gas sector evolves rapidly with shale gas production and LNG trade growth. The non-traditional companies move upstream to generate profits and strong partnerships are developed due to high capital intensity in the unconventionals, LNG and offshore gas projects. The supermajors which enjoy high economies of scale and globally diversified portfolio in both the gas and oil sector implement vertical integration models.

2.2 Dominant players in gas market

The evolution of natural gas market in the previous years was tremendous and its share in energy mix is considerable. Predictions about the future energy markets place the natural gas in the dominant position among the fuels. This sharp rise of gas creates opportunities for businesses which are involved in upstream and downstream activities in the sector. The leading players in gas market are the IOCs and the NOCs.

The International Oil Companies are known as “big sisters”. They are multinational companies with great profitability rates that possess mainly upstream activities and are headquartered in western countries (especially in the US and the UK). Before the rise of OPEC (Organization of the Petroleum Exporting Countries) the “big sisters” were a monopoly in energy markets.
The main competitors of the IOCs are the NOCs. The OPEC union was the first attempt of the state-owned companies to control the energy markets and to exploit their national reserves for geopolitical and profitability reasons. The state-owned companies are involved in all levels of the value chain from downstream to upstream projects and invest in various gas projects globally.

The high production and exploration costs in gas market lead companies to partnerships and co-operation. In this way companies diversify their portfolio and reduce the investment risk.

The state-owned companies sell their natural gas through long term contracts (take-or-pay agreements) with oil indexed prices. This happens in order to assure high returns corresponding to costly pipelines, offshore and LNG projects. That creates a strong economic relationship between producer and importer counties but also it reflects overreliance of the latter and geopolitical gains for the former.

2.2.1 NOCs presence in oil sector

The NOCs are already the main competitor to IOCs. But their strategy focuses on enhancing and stabilizing the NOCs position in the market in next years. For that reason NOCs proceed in offensive acquisitions especially in undeveloped fields in order to lock future profits.

Primarily NOCs development focused on oil sector for many reasons. The higher demand and consumption for oil than gas made oil more attractive to NOCs. Also the NOCs intended to secure oil supplies because of their importance in their countries’ economies. Finally the premium of oil prices was higher than gas prices.

2.2.2 NOCs presence in gas sector

The NOCs due to their advance in technological knowledge and the increased demand from developing countries has acquired a competitive position in the gas market. This tendency is shown from the NOCs vast amount of investments especially
in LNG, shale gas and offshore projects. In the shale gas field Asian NOCs invest in North America production in order to develop technology expertise. In addition, in the East Africa there are new offshore projects that attract many funds from NOCs. The natural gas sector is a challenge for NOCs due to the different market conditions than oil market, their lack of technology knowledge in production and exploration and the costly and risky projects as well.

The NOCs present different development in oil and gas sectors. In the short-term NOCs are dominant players in conventional oil production and expand their investments in unconventionals. The long-term will be characterized from large investments in LNG, offshore and shale gas projects.

2.3 Oil and gas linkage

In the previous years oil and gas were considered as alternative energy fuels. But technology development and the necessity for cost-efficient use of energy lead oil and gas to different sectors depending on their technical and economic characteristics. Nowadays, oil and gas are seen as complementary energy carries. Oil is mostly used in the transport sector and it has a little share in power generation. Yet, natural gas (NG) has a strong contribution to electricity production where it substitutes nuclear power and coal. Nowadays, the different use of the two fuels would be expected to weaken the strong linkage between them. But this is not the case and that is reflected in similar price paths between the two fuels. Also the liberalized and liquid spot markets that have developed especially in the US and the UK did not change the linkage. The reasons for the linkage resistance are many.

The most important reason is that the National Oil Companies base the long-term gas agreement on oil prices. That is a common practice (maybe it is fair) which secures a high level of returns in order to finance the costly pipelines and offshore projects and the new infrastructure that is necessary for the efficient gas supply chain. In the EU due to the planned or under construction pipeline and infrastructure projects from Russia and private consortiums (South Stream, North Stream, TAP) the linkage between oil and gas will continue to be very strong. For example the UK that has regulated a liberalized gas market is strongly connected with continental gas because
of the new Interconnector pipeline. Moreover, in the US despite the high level of production and the liquid spot markets as Henry Hub, the link is still mighty. The reasons for oil and gas linkage are many and are presented analytically below.

2.3.1 The reasons of linkage

The reasons for the strong connection between oil and gas prices vary. This linkage affects all markets from non-liberalized to liberalized. The reasons for the linkage are:

➢ The substitutability between oil and gas has declined in the previous years but many end-users have the ability to shift from one fuel to another. This is common for heating purposes in residential buildings and in power generation as well. But very strong evidence for this approach is the industrial sector in which there are double-fired boilers which are shifted between oil and gas depending on their current prices.

➢ Oil and gas reserves are not always separated. Many times gas and oil products are discovered together and a distinction process is necessary to extract the fuels. If the prices between the fuels were independent from each other and there are considerable differences, then the producers would focus on extracting the fuel with the more attractive price. The result for this would be reflected on supply and demand disequilibrium and in price volatility in gas and petroleum markets.

➢ When the natural gas was first appeared in Europe as alternative fuel to the dominant of the market (petroleum and coal) did not have a market value. Its worth based on other fuels that had already used. So the natural gas contracts were based on oil prices as the power generation linked with oil or coal price. Frequently a specific crude oil was the price index for gas contracts.

➢ As it was referred above the large amount of investments which are necessary for exploration and production in gas projects lead the state-owned gas producers to oil-indexed gas prices to ensure a high level of returns. In addition the joint ventures and banks that finance projects require long term agreements for gas selling in order to secure the appropriate cash flows that
respond to the debt. In this way the gas price is attractive for investors, banks and state-owned businesses at the expense of consumers.

- The liquefied natural gas (LNG) sharp rise in recent years has changed the gas market conditions. The LNG trade connects different regions and leads the natural gas market to a higher level of integration. On the other hand, the oil-indexed LNG agreements tie the gas price with the crude oil price. Thus the oil-gas linkage was enhanced from the LNG market development.

The EU is under a liberalization transition process. The intention of the commission is a more competitive and liquid market in which gas prices will accrue from supply and demand equilibrium. On the other hand, Russia, Norway and Netherlands supply the 60% of gas consumption in Europe. This supply is oil-indexed under long-term contracts and has direct effect on gas pricing. Despite the re-regulated European energy market, the influence of the long-term agreements is catalytic and the linkage remains very strong.

Historically, in the US the gas and oil prices were correlated for many and different reasons. Yet, during the last years has been observed a reduction of the linkage and the strong connection. This is a very controversial issue between researches. Between November 2010 and November 2011 the oil barrel increased by 11% and at the same time the gas price fell by 1.3%. This is an evidence of the different price path of the two fuels. Some analysts insist that the evolving market conditions of the gas sector lead to the price differentiation. There are three main reasons on which this argument is based.

Firstly, the sale gas revolution made the US depend less on gas imports and approach the target of a net exporter country. The shale gas production developed oversupply in the market and pushed down the gas price independently from the current oil price. Besides, the vast amount of resources both in Europe and Asia presage decoupling of prices in near future.

Secondly, the globalization of natural gas market is already a fact. Traditionally the gas market was fractured in different regions in contrast to the globally traded oil. But in previous years that has changed and the main contributor to
this is the LNG trade. The rise of LNG trade between continents leads to an integration of the gas market and in the convergence of the gas prices worldwide. Under a more integrated gas sector the gas price will gradually decouple from oil.

Finally, the big companies’ intention to invest in all parts of the gas value chain from downstream to upstream and to hold stakes in different projects in the sector establishes the vertical integration model as the dominant in sector. These changes alter the strong connection between oil and gas prices.

The first indications of a decoupling process between natural gas and oil prices are a fact. Many times in previous years the prices coupled together after a decoupling period but this time seems to be permanent. Actually, the correlation of the two fuels exists but with less intensity than before. The new tendency in energy markets is the natural gas independent price path.

2.3.2 Economic factors that strengthen the oil-gas linkage

There are economic factors such as supply and demand that enhance the oil and gas linkage. Increases in crude oil prices affect the gas price path in many ways.

In the supply side the possible scenarios are:

1. A crude oil price rise that comes from crude oil demand increase, is possible to lead to natural gas production cost increment and finally to gas price increase. The crude oil and natural gas producers use similar resources, as drilling rigs and labor. If the crude price increases then it will be a motivation for further production activities by the projects operators. That leads to an increase of relevant factors and the gas production cost rises. Thus the final gas price is augmented.

2. An increase in crude oil price that comes from crude oil demand rise is possible to increase the produced natural gas as co-product of crude oil and will lead to natural gas price decrease. The natural gas is found in two forms. The non-associated gas that is not found with large quantities of oil in the same reservoir. On the other hand, the associated gas is found with significant oil quantities in the same...
reservoir as free gas (associated) or as gas in solution with oil (dissolved).

In the demand side we can identify the followings:

1. An increase in crude oil price leads to the substitution from crude oil to natural gas. Thus the natural gas demand increases and pushes upward the gas prices. The substitution from one fuel to another is possible due to competition of fuels in power generation. Also in industrial sector, many companies have dual-fired boilers that give them a shift choice.

2.4 Topic research

Historically gas price did not follow an independent path. The connection with oil products was very strong and that seems to continue up to now. However the conditions of this relationship have changed and in some geographical areas the oil and gas prices seem to move independently from each other for a long period without a trend convergence in their long-run term.

The gas market presents entirely different characteristics from the oil market which is a global market. Petroleum products are traded globally between different regions and that is compatible with their specific properties. Oil is a fuel with high energy density that is reflected in low cost shipments. On the other hand the relatively low energy density of the natural gas and the costly pipelines and LNG projects remove the gas sector from a worldwide integration. These high costs decrease the arbitrage opportunities and develop separated gas markets.

The most important studies that reflect the connection between NG and oil prices are presented below. Furthermore some indicative studies that illustrate the NG integration in different gas market are included.
2.4.1 European Union

The European Union is a big gas consumer. The main suppliers of Europe are Russia, Norway and the Netherlands. Also Algeria is an important source of gas especially for the north countries as Italy and Spain. These supply agreements are long term contracts based on oil prices that usually include ‘take or pay’ clauses.

On the other hand, the UK gas sector is characterized from a liquid spot market. The most important trading hub in the UK is the National Balancing Point (NBP). The NBP is a spot-traded gas market that is established from 1990s and is used as an indicator for Continental Europe gas prices. In addition, a trading hub with paramount importance in Continental gas markets is the Title Transfer Facility (TTF) in the Netherlands. The TTF is a virtual trading gas market identical to NBP.

Furthermore the new interconnector between Zeebrugge and Bacton linking the continental Europe with the UK and augment the level of gas market integration in Europe. The gas market in Europe is developing rapidly. The environmental commitments of the European countries and the CO2 emissions targets increase the share of natural gas in European energy mix due to the fact that the natural gas is considered as 'clean energy'.

In previous years many researchers examined the gas market integration in Europe and the results are very interesting and useful. A part of the most important studies has been chosen in order to value their results.

In Europe, the German gas market is the largest one and it is supplied by the biggest producer countries as Norway, the Netherlands and Russia. Thus Germany is considered as indicator for the EU gas market. Asche (2002) examined the European gas market integration based on German gas market. The results showed that there is an integrated German gas market and therefore an integrated European gas market. Siliverstovs (2005) revealed a high degree of integration in European gas market over the period from 1990s to 2004. A different approach was presented by Robinson (2007) who investigated the gas market integration of Europe over the 1980s and 1990s and found mixed results depending on the country that was examined and the test was implemented.
Panagiotidis and Rutledge (2007) investigated the relationship between the Brent crude oil and the UK gas prices after the market liberalization over the interval 1996-2003. The findings showed that gas prices did not decoupled from the oil price and are affected considerably from changes in oil price path. Also the same results appeared from Asche (2006) who explored that Brent oil and the UK gas prices are cointegrated before and after the critical point of the Interconnector opening.

2.4.2 Pacific Asia

The Pacific-Asia area is a large LNG importer. The main importers are Japan, China and South Korea. Japan is an island country and the LNG share in its energy mix is considerable. The lack of pipeline connection with supplier countries leads the energy policy in Japan to turn to LNG imports. Also the recent nuclear accident in Fukushima will strengthen the presence of natural gas in country at the expense of nuclear power. Furthermore, China and South Korea are developing countries which have augmented energy needs. Thus in order to build a diversified energy mix and to reduce the C0₂ emissions these countries have increased the LNG imports. The major suppliers in Asia are Australia, Malaysia, Indonesia and Qatar and the agreements are oil-indexed long-term contracts. The indexation benchmark is the Japanese Crude Cocktail (JCC) that consists of different crude oils. Siliverstovs (2005) investigated the gas market integration of Pacific-Asia. Also the research examined the international integration of gas market and their relationship with oil price. The results showed that LNG price in Japan is integrated with Brent crude oil and European gas in contrast with the US market that seems to move independently from Japanese LNG price.

2.4.3 North America

The North American gas market presents entirely different elements from European and Asian gas markets. The US gas sector is characterized from gas market liberalization. There are many private production companies, big suppliers that feed the consumers in industrial, residential and electricity generation sector and local
distribution companies for local transmission purposes. Furthermore, supply and demand economic factors are crucial contributors of gas price in contrast with the European prices.

There are many gas trading hubs in different regions. These trading hubs provide liquid prices and a variety of contracts. The most important trading hub is the Henry Hub (HH) that is considered as the benchmark price of gas in the US. In addition, very important trading hubs are Topock and Transco Zone 6 which are regional trading nodes in the US transmission system in West and East respectively. Finally, the gas transmission system consists of a vast amount of pipelines grids that are encountered in trading hubs.

The natural gas market in this area is under considerable development. Nowadays the shale gas boom drives the gas sector. The technological breakthrough in shale gas extraction procedure gave a significant impetus in gas production. As a result, shale gas production has reduced the dependence of US on gas imports and developed a complete and liquid domestic gas market.

Also the LNG trade rise linked the US gas markets with different consumer and producer regions and enhanced the integration level among the global gas markets. The LNG agreements are oil-indexed and increase the co-integration of oil and gas markets. Although the export capacity of the US cannot correspond to the gas oversupply, the progress in the LNG sector is significant.

As the new pioneer technology in unconventionals and shale gas extraction processes developed an overproduction in the gas sector, at the same time the high volatility of previous years in gas prices discouraged the investors to finance new gas projects. Thus, this oversupply was not absorbed mainly due to inefficient infrastructure and limited transmission and export capacity. The high gas inventories rates of the US provide independency in gas price from oil price path, at least in short-term.

Oil and gas prices were traditionally linked for many reasons that listed above. That seems to change in accordance with previous studies in North America gas sector at least in the short-run. The main object of previous studies on the subject and
the research topic of this study is to find out if the gas and oil prices in US have decoupled and are removed from the long-run equilibrium on a permanent trend.

There are many studies that investigated the gas and oil price correlation in N. America. Serletis and Herbert (1999) examined the price similarities between fuel oil, natural gas and power generation in North America. The mid-Atlantic area of Pennsylvania, Maryland, New Jersey and Delaware with high energy consumption were chosen for the research. The energy carriers in the area are connected considerably. First of all large industrial users have the choice to shift between natural gas and fuel oil depending on their prices. Also fuel oil and natural gas are the peaking sources in electricity generation. All the above common characteristics of the fuels enhance the likelihood of finding a strong relationship between these energy carriers. The data of the research are daily from 25 October 1996 to 21 November 1997 on the Henry Hub and Transco Zone 6 natural gas spot prices. The results of this research revealed that there are shared trends between Transco Zone 6 and Henry Hub gas prices and the price of fuel oil. This means that US natural gas prices are cointegrated with US fuel oil prices. Thus the arbitraging mechanisms operate effectively into the data period.

At the same year, Barcella (1999) investigated the relationship between petroleum products and natural gas prices in the US. The research in the US electric power sector showed a considerable inter-fuel competition and that illustrated in the high correlation and cointegration between crude oil or refined products and gas prices.

The market integration in natural gas and crude oil markets and the integration between them in the US were searched from Bachmeier and Griffin (2006). The study resulted in strong integration between the US oil and gas prices. At the same year Villar and Joutz (2006) searched the econometric relationship between Henry Hub (HH) natural gas prices and West Texas Intermediate (WTI) oil prices in US. The analysis showed a cointegrating relationship between the two fuel prices. The data covered the period from January 1989 to December 2005. The spot prices of HH and WTI on a monthly basis were selected for the analysis due to the fact that monthly prices based on the short-run movements that are captured quite well on this basis. The results confirmed the expected stable relationship between natural gas and oil
price despite the decoupling trend that appeared at times. Also it was proved that despite the strong impact of WTI prices on natural gas price the opposite is negligible.

On the other hand, there are studies with different results about the oil and gas relationship in the US market. Silverstovs (2005) examined the degree of natural gas market integration in three different regions: Europe, North America and Japan. Also it is explored the relationship between the three gas markets and their linkage to the oil price. The data was monthly prices for imported gas in Europe and the US (LNG and pipeline gas) and LNG monthly price in Japan. In the US it is chosen the Henry Hub spot price. In the Continental Europe the Brent Crude oil is chosen in order to examine the interaction between oil and global gas. The observations cover the period from November 1993 through March 2004.

The results of the study are very interesting. In North America it is observed integration in natural gas price probably from changes in regulation and gas to gas competition. Also in Continental Europe there is a high level of gas market integration. In addition, Japanese and European gas markets are highly integrated. On the other hand the gas markets across the Atlantic are not integrated. The US gas price moves independently from the other gas markets due to the limited arbitrage opportunities across the Atlantic. Furthermore, there is strong cointegration evidence between natural gas and Brent price in Europe. Finally the most useful evidence is the lack of cointegration between the US gas price and the Brent oil price in contrast with previous studies and the corresponding cointegration in other markets between gas and oil.

In the US gas markets there are many exogenous facts that affect the gas price path. Brown and Yücel (2008) investigated the US natural gas price drivers. But the difference from the previous research on the specific topic is that it implements very important additional factors that affect the gas price path. These factors are seasonality, weather, storage and shut in production. Although the crude oil and natural gas price have a strong relationship, it is affected by the additional factors, as well. If these factors are taken into account, the crude oil price explains quite well the natural gas movements. The data for the relationship between natural gas and oil prices cover the period from 7 January 1994 to 14 July 2006. But the additional factors as seasonality, weather, natural gas storage and disruptions on natural production
shorten the analysis’ interval from 13 June 1997 to 14 July 2006. The observations are weekly prices of the Henry Hub and West Texas Intermediate. The weather and seasonality factors are implemented with heating degree days, deviations from normal heating degree days, cooling degree days, and deviations from normal cooling degree days. For the storage effect in gas prices the study examines the storage differential that is the difference between the storage in a given week and the average for that week over the past five years. The shut in production refers to Gulf of Mexico production site. The study implements an error correction model taking into account the additional factors that were referred above, apart from the crude oil prices and explains sufficiently the gas price path. It is shown that the gas prices are related to crude oil prices and adjust to deviations of the latter. The relationship is stable in the long run and presents complex short-run dynamics.

Additionally, in the same year Brown and Yücel (2008) in a different approach examined the integration level of the US natural gas market. The North American natural gas market consists of regional interstate natural gas markets. In the mid-1990s the gas consumption increased mainly due to the gas used in power generation. Also the seasonal variation in demand was covered by storage and not by production and net imports. That leads to price volatility. As a result, pipeline companies were not guaranteed regarding their returns and reduced their investments in new pipeline projects. Thus the pipeline system was pushed to capacity and the arbitrage opportunities decreased considerably. All the above affected the gas market integration level in the US. The study implement causality tests between the Henry Hub and gas regional nodes and electricity regional interchange nodes on the US electricity grids as well. The data of the analysis are daily prices from 3 February 1997 to 17 January 2007. The prices are collected from the Henry Hub, Transco Zone 6 and Topock gas markets. The Topock and Transco Zone 6 are regional trading nodes in the US transmission system in West and East respectively. In the electricity sector the nodes that are used for the analysis are the PJM and the Paolo Verde in East and West respectively.

The causality tests between the Henry Hub and the regional gas markets revealed that movements in the regional market prices lead to corresponding movements in Henry Hub prices. Also movements in Henry Hub prices direct the two regional markets price path. Additionally, in regional level, in both the East and West
Topock and Transco Zone 6 prices direct those of Palo Verde and PJM respectively. The causality is bidirectional and that means the electricity prices movements in both East and West direct the natural gas prices in the corresponding area. That is logical due to the fact that gas is the primary fuel in power generation. Finally the Henry Hub is not correlated with the regional electricity prices. The independent influence that is exerted from electricity prices on regional gas prices probably is not arbitraged to HH. That reveals limited arbitrage between regional gas markets which is caused by delivery constrains. The factors that direct the regional gas prices are independent of the HH price drivers. The key finding is that the US market is characterized by limited arbitrage due to inefficient capacity and bottlenecks in natural gas transmission system. All the above showed that the integration in US gas market has decreased.

A different approach to the previous studies depicted Victor Lux Tonn and McCarthy (2010). This study explored the relationship between natural gas and oil futures prices. The researchers are using the wavelet analysis and the results are very important. The wavelet analysis presents the correlation between oil and gas prices in time domain and in frequency domain as well. The data of the analysis are the daily closing prices of the near month futures contracts for natural gas and light sweet crude from 3 April 1990 to 2 October 2007. The wavelet analysis showed that only if the frequencies from the two series are within limited ranges, the gas and oil prices are highly correlated. When the frequency periods of futures oil prices are between specific ranges, the oil and gas price movements are synchronized. Thus the prices of natural gas and oil futures have high covariance at high frequencies but considerably less at low frequencies.

Finally, Peter Erdos (2012) examined if the oil and gas price decouple within the analysis’ interval. The findings showed that the 2009 is a milestone for the correlation among the fuels’ prices. After 2009 the natural gas price seems to move independently from the crude oil price in the US, unlike the previous years. The data consist of Henry Hub and West Texas Intermediate weekly spot prices in the US. In the UK the National Balancing Point (NBP) is chosen. The period of the analysis starts from January 1994 and ends in December 2011 but the NBP price is only available since 1997. The econometric technique of this study implements a vector error correction model (VECM) to examine the relationship between oil and gas prices but takes into account the exogenous shocks that affect the natural gas price.
Similarly with the Brown and Yücel (2008) the exogenous variables are: Natural gas storage deviations from the average (five-year), deviations of heating and cooling degree days from seasonal norms and natural gas shut-ins production in the Gulf of Mexico.

The results can be divided into two periods. Before 2009, in both the US and the UK the crude oil price directs the natural gas price and the latter adjusted to the former after exogenous shocks that decoupled temporarily from their equilibrium. Furthermore, in both the UK and the US markets, the gas price reverts to equilibrium with crude oil individually. Also deviations from equilibrium in the US market leads to similar deviations in the UK market and the opposite. Finally after an exogenous shock, the US and the UK gas prices decoupled from each other for a period of 20 weeks on average and converged mainly due to the arbitrage conditions across the Atlantic and the crude oil contribution to mediation. After 2009, the US natural gas has decoupled from the European natural gas price and from crude oil price unlike the UK gas that remained cointegrated with the crude oil price. The oversupply of gas from the shale gas boom in the US and the limited export capacity lead the gas to an independent movement from crude oil price. The arbitrage across Atlantic decreased considerably, resulted in a division of the co-movement of the US and the UK gas price path. Finally the results from the study are very important due to the fact that for the first time has been revealed a permanent decoupling trend of the US gas from European gas and crude oil price.
3. METHODOLOGY

3.1 Unit root and Stationarity tests

Stationarity is a very important task in empirical analysis. The detection or not of stationarity determines the process will be applied. In the field of econometrics the variables are split into two general categories: The stationary time series and the non-stationary time series.

In stationary time series the mean and the variance are constant and in conjunction with the autocovariances they are independent of time. The stationary time series have a deterministic trend with constant increments. As a result any random shock that changes the variable or the error term value at a specific time period tends to be dissolved and any lasting effects on the time series are eliminated. These time series are weakly stationary or covariance stationary.

In non-stationary time series the variance and autocovariance are time-varying. In this way the sharp fluctuations of time series after an exogenous shock have permanent effects on them. The non-stationary time series have a stochastic trend (unit root) which is a trend with random increments. Also their time-varying nature is incompatible with the classical assumptions of linear regression such as multivariate normality, homoskedasticity and autocorrelation. Thus the regression that consists of non-stationary time series is considered as spurious with a very high $R^2$ that explains only the similar rising trend among the variables over time.

In order to develop a reliable regression it is necessary to find the order of integration of the non-stationary time series. The integration order shows how many times the non-stationary series need to be differentiated in order to be transformed into stationary series.

3.1.1 Augmented Dickey-Fuller (ADF) test

In order to implement a Dickey-Fuller (DF) test an AR process is necessary. The AR is the Autoregression model in which the explanatory variables are
lags of the dependent variable. In standard Dickey-Fuller test the appropriate model is the AR (1) that is shown below:

\[ Y_t = \kappa Y_{t-1} + \nu + \lambda t + \epsilon_t \]

where \( \kappa, \lambda \) are parameters to be estimated and \( \epsilon_t \) is white noise. The equation of the dependent variable may consist of a constant or a constant and time trend. Also there are tests without constant but are not used in practice. In the above equation the constant and trend situation is applied. The value of \( \kappa \) is related to the autocorrelation function behavior and to stationarity concept. If \( |\kappa| \geq 1 \), then \( Y_t \) time series is non-stationary. On the other hand, if \( |\kappa| < 1 \), then time series is stationary.

By subtracting the \( Y_{t-1} \) from both sides of equation, the model is modified into the following form:

\[ \Delta Y_t = \mu Y_{t-1} + \nu + \lambda t + \epsilon_t \]

where \( \mu = \kappa - 1 \). The null hypothesis denoted with \( H_0 \) and the alternative with \( H_1 \) and are presented analytically below:

\[ H_0 : \mu = 0 \]
\[ H_1 : \mu < 0 \]

In null hypothesis \( H_0 \), \( \mu = 0 \) and consequently \( \kappa = 1 \). As a result the time series is non-stationary. On the other hand, in the alternative hypothesis \( H_1 \), \( \mu < 0 \) and consequently \( \kappa < 1 \).

Thus the time series is stationary. The test-statistic which is used to assess the null hypothesis is defined as:

\[ test \ statistic = \frac{\hat{\mu}}{SE(\hat{\mu})} \]

where \( \hat{\mu} \) is the estimate of \( \mu \) and the \( SE(\hat{\mu}) \) is the coefficient standard error.
In Dickey-Fuller test approach an AR(1) model is used. However, it is possible that the lags of Y which are included as explanatory variables in the right-hand side of AR equation may be more than 1. In that case the unit root test is not valid because the assumption of $\varepsilon_t$ as white noise is disrupted. In order to surpass the problem an AR (p) model is performed, where p is the number of Y lag terms. The modification of the simple Dickey-Fuller test is the Augmented Dickey-Fuller (ADF) test. The equation of $\Delta Y$ is transformed into:

$$\Delta Y_t = \mu Y_{t-1} + \sum_{i=1}^{p} a_i \Delta Y_{t-i} + \nu + \lambda t + \varepsilon_t$$

The null hypothesis and the test statistic are applied in the same way as in Dickey-Fuller approach.

### 3.1.2 Phillips-Perron (PP) test

Phillips and Perron (1988) introduced a unit root test that is widely used in econometric analysis. The difference from the ADF lies in the way it deals with the serial correlation and heteroskedasticity in the errors. The ADF includes a number of lags in first differences of $Y_t$ in order to encounter the serial correlation problem. The Phillips-Perron (PP) test in a different approach uses a non-parametric method in which the test statistic is modified and corrects the heteroskedasticity and any serial correlation of the errors in the regression. In PP, the asymptotic distribution of test statistic is not affected by the serial correlation. Also the modified test statistic has the same asymptotic distribution with the ADF unit root test.

The PP test involves the Dickey-Fuller test equation:

$$\Delta Y_t = \mu Y_{t-1} + \nu + \lambda t + \varepsilon_t$$

where the $\varepsilon_t$ is I (0) and it is possible to be heteroscedastic. For that reason the test estimates the following equation:

$$Y_t = \pi Y_{t-1} + \nu + \lambda t + \varepsilon_t$$
The PP test may consist of a constant or a constant and a trend or neither of the two in the equation form. In the above equation a constant and a trend situation has chosen.

Under the PP test the test statistic is transformed into the following form:

\[
\tilde{t}_\mu = t_\mu \left( \frac{\gamma_0}{f_0} \right)^{1/2} - \frac{T(f_0 - \gamma_0)}{2f_0^{1/2}s} \left( se \left( \hat{\mu} \right) \right)
\]

where \( \hat{\mu} \) is the estimate and \( t_\mu \) is test statistic of \( \mu \), \( s \) is the standard error of the test regression and \( se \left( \hat{\mu} \right) \) is the coefficient standard error. Additionally \( f_0 \) is the estimator of the residual spectrum at zero frequency. Finally \( \gamma_0 \) is an estimate of the error variance in Dickey-Fuller equation of \( \Delta Y_t \) (\( \gamma_0 \) is equal to \( (T - \kappa)s^2 / T \) where \( k \) is the number of regressors).

The main advantage of the Phillips Perron test over the Augment Dickey-Fuller test is that the PP is robust to general forms of heteroskedasticity in the errors \( \epsilon_t \). Another advantage of PP is that the determination of the lag length is not necessary.

3.1.3 Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test

In the previous tests, ADF and PP, the null hypothesis is the non-stationarity of the time series. Under the Kwiatkowski, Phillips, Schmidt and Shin (1992) test, in a different approach, the null hypothesis determines the stationarity of the series. The KPSS is based on the residuals from the OLS regression of \( Y_t \) on the exogenous variables \( X_t \):

\[
Y_t = X_t' \delta + u_t
\]

where \( X_t \) are optional exogenous regressors (constant or constant and trend).

The test statistic is the Lagrange Multiplier (LM) and is defined as:
\[ LM = \frac{\sum S(t)^2}{T^2 f_0} \]

where \( f_0 \) is an estimator of the residual spectrum at zero frequency and \( S(t) = \sum_{\pi=1}^{\infty} u_{\pi} \)

is a sum function that is based on the residuals \( u_{\pi} = \hat{Y}_t - X_t \hat{\delta}(0) \).

### 3.2 Cointegration tests

The most economic variables contain a unit root and are integrated of order 1, \( I(1) \). Engle and Granger (1987) proposed the theory that if there is a linear combination of two non-stationary, \( I(1) \) time series which is stationary, \( I(0) \) then there is a long run equilibrium between them. The long run equilibrium of two time series denotes that despite the decoupling trend (different ways) that the variables may present in short-term, the deviation from equilibrium will not be permanent. The time series convergence occurs after the departure and that reflects their long run relationship. The cointegrated time series present cointegration equations that determine the long run relationship. The existence of cointegration between time series is very important. In case that there is cointegration and it is not examined, then the analysis is deprived of long information and its reliability is reducing. If the variables are not cointegrated the causality between them is examined in a Vector Autoregression (VAR) framework. On the other hand, if the variables are cointegrated the Vector Error Correction Model (VECM) that contains the cointegration equation is performed.

#### 3.2.1 Engle and Granger test

The Engle and Granger test is based on the residuals of the regression between the time series. Assuming the following equation:

\[
Y_t = \delta D_t + \varphi X_t + u_t
\]
where \( D_t \) are the deterministic trend regressors. The first step is the estimation of the long run relationship by OLS in cointegration regression. The cointegration regression is estimated as:

\[
\hat{Y}_t = \delta D_t + \varphi X_t
\]

In addition the disequilibrium errors are defined as:

\[
\hat{u}_t = Y_t - \hat{Y}_t = Y_t - \delta D_t - \varphi X_t
\]

These residuals are used in order to test the null hypothesis of no cointegration equation. Engle and Granger (1987) introduced the Augmented Dickey-Fuller to check the existence of a unit root. If the residuals are stationary, the null hypothesis is rejected and there is cointegration and long run relationship between the time series. If the residuals are not stationary (have a unit root), the null hypothesis is not rejected and there is no cointegration equation between the time series.

The first order autoregressive process follows this equation:

\[
\hat{u}_t = \kappa \hat{u}_{t-1} + v_t
\]

where \( v_t \) is white noise. Finally the autoregression process with \( p \) lags is defined as:

\[
\Delta \hat{u}_t = (\kappa - 1) \hat{u}_{t-1} + \sum_{t=1}^{p} a_l \Delta \hat{u}_{t-l} + v_t
\]

Engle and Granger cointegration test despite its usefulness has some important limitations: Firstly the parameter’s estimates are biased when the sample is finite and this weakness is very crucial. Furthermore in long run parameters no testing is possible. Finally, the existence of more than one cointegration relationship generates problems.
3.2.2 Phillips-Ouliaris (PO) test

The Engle and Granger test presents serious weaknesses and limitations. In order to avoid these problems many techniques have been developed. Phillips and Ouliaris (1988) proposed a cointegration test similar to Engle and Granger that is based on the residuals that derive from the OLS estimation of the cointegrating equation of $Y_t$. This test is based on the non–parametric Phillips-Perron methodology, in contrast to the Engle and Granger test in which the parametric augment Dickey-Fuller approach is used. The estimation of $\kappa$ is based on the following regression equation:

$$\Delta \hat{u}_t = (\kappa - 1) \hat{u}_{t-1} + w_t$$

The test examines the stationarity in the residuals to detect the existence of cointegration relationship between the time series. The null hypothesis determines that there is not cointegration and it is rejected if the residuals are stationary.

Although the Engle and Granger procedure is easy to be implemented, the results present differences regarding with the determination of the dependent variable. In Phillips-Ouliaris test whichever variable is chosen to be the dependent the results are exactly the same (invariant to normalization). On the other hand, Phillips-Ouliaris test, present an important disadvantage that it can estimate only a single cointegration equation. This problem is solved with the use of Johansen cointegration test.

3.2.3 Johansen cointegration test

In Johansen cointegration test more than one cointegration relationships are permitted, in contrast to the Engle and Granger and Phillips Peron test. Johansen (1995) proposed a procedure based on a Vector Autoregression (VAR) model. The VAR of order $K$ can be written as:

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{j=1}^{K-1} \Gamma_j \Delta Y_{t-j} + AX_t + \epsilon_t$$
where \( Y_t \) is a \( n \times 1 \) vector of variables that are integrated of order one, \( I(1) \), \( e_t \) is a \( n \times 1 \) vector of innovations, \( X_t \) is the vector of deterministic variables and \( \Pi \) a coefficient matrix. Relatively to the \( \Pi \) rank size we can identify three different cases:

- If \( \Pi \) equal to zero there is not cointegration.
- If \( \Pi \) has a full rank then all \( Y_t \) must be stationary.
- If \( \Pi \) has reduced rank, less than full, but no zero, there is cointegration.

The third case is the most interesting due to the fact that cointegration exists between the variables. In this case the rank of \( \Pi \) has reduced compared to \( Y_t \) vector and \( r < n \), where \( r \) is the cointegration rank or the number of cointegrating relations. Also \( \Pi \) can be written as \( \Pi = \alpha \beta' \) and is \( I(0) \), where \( \alpha \) and \( \beta \) are two \( n \times r \) matrices. The Johansen test estimates the \( \Pi \) matrix from an unrestricted VAR and test whether we can reject the restrictions implied by the reduced rank of \( \Pi \).

### 3.3 Vector Autoregression Model (VAR)

The Vector Autoregression model (VAR) is proposed by Sims (1980) in order to deal with the assumption of interactions between the time series. VAR models are a generalization of univariate autoregression models (AR) and are used to capture the interdependencies among time series. Also a VAR is characterized by its convenience for estimation and forecasting. A VAR model is a system of equations in which the evolution of each variable is expressed as a function of its own lags and the lagged values of the rest variables. Assuming the VAR model of order \( d \) (number of lags) and \( n \) endogenous variables:

\[
Y_t = C_0 + \sum_{j=1}^{d} \sum_{i=0}^{n} A_{ij} Y_{t-j} + \varepsilon_t
\]

where \( Y_t = \{Y_{t1},...,Y_{tn}\} \) is the vector of endogenous variables, \( C_0 \) is the vector of intercepts, \( A_i \) is a \( (n \times n) \times d \) matrix of autoregressive coefficients and \( \varepsilon_t = \{\varepsilon_{t1},...,\varepsilon_{tn}\} \) is the vector of error terms. The coefficients of matrix \( A_i \) are estimated by Ordinary Least Squares (OLS).
The bivariate VAR model with k lags has the following form:

\[ Y_{1,t} = C_{1,0} + A_{1,1} Y_{1,t-1} + \ldots + A_{1,k} Y_{1,t-k} + B_{1,1} Y_{2,t-1} + \ldots + B_{1,k} Y_{2,t-k} + X_{1,t}' \theta + \epsilon_{1,t} \]

\[ Y_{2,t} = C_{2,0} + A_{2,1} Y_{2,t-1} + \ldots + A_{2,k} Y_{2,t-k} + B_{2,1} Y_{1,t-1} + \ldots + B_{2,k} Y_{1,t-k} + X_{2,t}' \theta + \epsilon_{2,t} \]

where \( X_t \) is the vector of exogenous variables.

In order to use a VAR model the time series should be stationary and I (0). In case that the series are integrated of order d, I (d), some modifications are necessary. If the series are not stationary, I (d) and are not cointegrated as well, the time series should be differentiated d times before run the regression. If the time series are cointegrated an extension of VAR model is used. The new model contains an error correction term and is the Vector Error Correction Model (VECM). The VECM is not part of this study and it is not referred further.

### 3.4 Causality tests

#### 3.4.1 Toda-Yamamoto test—Linear causality

Toda and Yamamoto (1995) based on the Granger causality developed a different approach. The alternative procedure does not take into consideration the possibility of long-run relationship among the variables or their integration order. That means the unit root and stationarity tests can be omitted and the cointegration testing as well. In this modified version of Granger causality an augmented level Vector Autoregression model is performed. Toda and Yamamoto causality test consists of the following equations:

\[ Y_t = \alpha + \sum_{i=1}^{h \times d} \beta_{t,i} + \sum_{j=1}^{k \times d} y_{j,t} X_{t-j} + u_{yt} \]

\[ X_t = \alpha + \sum_{i=1}^{h \times d} \theta_{t,i} + \sum_{j=1}^{k \times d} \delta_{j,t} Y_{t-j} + u_{xt} \]

where \( h \) and \( k \) the optimal lag length of \( Y_t \) and \( X_t \), d is the maximal order of integration order of the system variables and \( u_{yt} \), \( u_{xt} \) the error terms that are assumed as white noise.
The first step of the procedure is to determine the maximal order $d$. After that a VAR model is performed with $\kappa + d$ lags. Finally, the null and alternative hypothesis includes a set of restrictions in order to detect the potential causal relationship between the variables.

### 3.4.2 Wald test - Linear causality

The Wald Test is based on an unrestricted regression. In addition, it implements some restrictions. If a restriction is true then it should be “approximately” satisfied by the unrestricted model. In that case the LS estimator is asymptotically normally distributed:

$$
\sqrt{n}\left(\hat{\beta} - \beta_0\right) \overset{d}{\longrightarrow} N\left(0, \sigma^2 Q_X^{-1}\right)
$$

Assuming a linear regression model:

$$
Y = X \beta + \varepsilon
$$

where $Y$ and $\varepsilon$ are $n$ vectors and $\beta$ is a $\kappa$ vector of parameters to be estimated.

The linear restriction for this model is defined as:

$$
H_0 : R \beta_0 = r
$$

where $R$ is a $q \times k$ matrix ($q$ is the number of restrictions), $r$ is a $q$ vector and $H_0$ is the null hypothesis. The LS estimator is transformed into:

$$
\sqrt{n}\left( R \hat{\beta} - r \right) \overset{d}{\longrightarrow} N\left(0, \sigma^2_Q R Q_X^{-1} R'\right)
$$

The consistent estimator of $Q_X^{-1}$ is substituted by $\left(\frac{X'X}{n}\right)^{-1}$ and the Wald statistic is equal to:

$$
W = \left( R \hat{\beta} - r \right)' \left( \sigma^2 Q R X' \right)^{-1} \left( R \hat{\beta} - r \right)
$$
Under the null hypothesis, W is asymptotically distributed as $\chi^2(q)$. In the Wald test, the restrictions are tested without having to estimate the restricted model.

### 3.4.3 Breitung and Candelon test-Nonlinear causality

Granger (1969), Geweke (1982) and Hosoya (1991) proposed frequency-domain causality measures and test procedures. Yao and Hosoya (2000) proposed a Wald Test in order to examine causality at given frequency which included non-linear restrictions on the autoregressive parameters and improved this procedure with the delta method based on numerical derivatives. Based on this approach Breitung and Candelon implemented a simple empirical test to capture prediction’s reliability at given frequencies. This test is performed also to cointegrated systems. The causality testing at frequency domain is based on decomposition of spectral density. The test procedure implements a bivariate VAR based on linear restrictions on the VAR parameters. This approach can be extended in order to be used in higher dimensional systems. Finally this procedure detects the existence of short-run and long run predictability.

If $z_t = [x_t, y_t]$ is a two–dimensional vector at $t=1, \ldots, T$ and it has a VAR representation, it can be expressed as:

$$\Theta(L)z_t = \epsilon_t$$

where $\Theta(L)$ is a $2 \times 2$ lag polynomial with $L^k z_t = z_{t-k}$. If the system is stationary, the MA representation is defined as:

$$z_t = \Phi(L)\epsilon_t = \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ \Phi_{21}(L) & \Phi_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} = \Psi(L)\epsilon_t = \begin{bmatrix} \Psi_{11}(L) & \Psi_{12}(L) \\ \Psi_{21}(L) & \Psi_{22}(L) \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix}$$

where $\Theta(L)^{-1} = \Phi(L)$, $\Psi(L) = \Phi(L)G^{-1}$ and $\eta_t = G\epsilon_t$. If the spectral density of $x_t$ is defined as:

$$f_x(\omega) = \frac{1}{2\pi} \left\{ |\Psi_{11}(e^{-i\omega})|^2 + |\Psi_{12}(e^{-i\omega})|^2 \right\}$$
then the measure of causality between the time series has the following form:

\[
M_{y \rightarrow x}(\omega) = \log \left[ \frac{2\pi f_c(\omega)}{\left| \Psi_{11}(e^{-i\omega}) \right|^2} \right] = \log \left[ 1 + \frac{\left| \Psi_{12}(e^{-i\omega}) \right|^2}{\left| \Psi_{11}(e^{-i\omega}) \right|^2} \right]
\]

In case that the \( \left| \Psi_{12}(e^{-i\omega}) \right| = 0 \), y does not cause x at frequency \( \omega \). If the time series are non-stationary and cointegrated the procedure is performed with some modifications and the causality has the following form:

\[
M_{y \rightarrow x}(\omega) = \log \left[ 1 + \frac{\left| \Psi_{12}(e^{-i\omega}) \right|^2}{\left| \Psi_{11}(e^{-i\omega}) \right|^2} \right]
\]

Besides, the measure can be applied in higher-dimensional systems.

The null hypothesis of the test procedure is that y does not cause x at frequency \( \omega \) and is defined as:

\[
M_{y \rightarrow x}(\omega) = 0
\]

Yao and Hosoya (2000) proposed the replacing of \( \left| \Psi_{11}(e^{-i\omega}) \right|^2 \) and \( \left| \Psi_{12}(e^{-i\omega}) \right|^2 \) with estimates from the fitted VAR in conjunction with the delta method. Due to the fact that \( \left| \Psi_{12}(e^{-i\omega}) \right| \) is a complicated nonlinear restriction and the derivative is difficult to be estimated, Yao and Hosoya (2000) used a numerical differentiation.

It is obvious that \( M_{y \rightarrow x}(\omega) = 0 \) if \( \left| \Psi_{12}(e^{-i\omega}) \right| = 0 \). Using a number of modifications the set of linear restrictions that are necessary for the procedure are:

\[
\sum_{k=1}^{p} \theta_{12,\delta} \cos(k\omega) = 0
\]

\[
\sum_{k=1}^{p} \theta_{12,\delta} \sin(k\omega) = 0
\]

where \( \theta_{12,\delta} \) is the (1,2) element of \( \Theta_{\delta} \), and \( \Theta(L) \) is a function of \( \Psi(L) \).
If we let $a_j = \theta_{1j}$ and $\beta_j = \theta_{2j}$, the VAR equation of $x_t$ is transformed into:

$$x_t = a_1 x_{t-1} + ... + a_p x_{t-p} + \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + \epsilon_t$$

The null hypothesis $M_{y \to x}(\omega) = 0$ is equivalent to:

$$H_0 : R(\omega) \beta = 0$$

Where $R(\omega) = \begin{bmatrix} \cos(\omega) & \cos(2\omega) & ... & \cos(p\omega) \\ \sin(\omega) & \sin(2\omega) & ... & \sin(p\omega) \end{bmatrix}$ and $\beta = [\beta_1, ..., \beta_p]'$.

Finally the F-statistic is distributed as $F(2, T-2p)$ for $\omega \in (0, \pi)$.

The Breitung-Candelon approach can be applied also in higher dimensional systems and when time series are I (1) and cointegrated with some basic modifications. This study implements the test with two time series and in a not cointegration framework.

### 3.5 ARCH and GARCH estimation

#### 3.5.1 GARCH

Assuming the following equation of $Y_t$:

$$Y_t = X_t ' \beta + \epsilon_t$$

where the $X_t$ is a function of exogenous variables and $\epsilon_t$ is the error term $\epsilon_t : N(0, \sigma^2)$. In this case the variance of the series is constant and there is homoscedasticity as $\text{var}(\epsilon_t) = \sigma^2$. In case that the variance is not constant the sample is heteroscedastic. If the existence of heteroscedasticity is ignored the results of the model will be unreliable and the errors estimates will be incorrect. In order to deal with the volatility in time series Engle (1982) proposed a stationary nonlinear model for $Y_t$. In Autoregressive Conditional Heteroscedasticity model (ARCH) the conditional variance of $Y_t$ evolves with an autoregressive process and depends on the
immediately previous squared error term. The conditional variance equation in ARCH (1) process is defined as:

\[ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 \]

The general form with q terms is ARCH (q) and is equal to:

\[ \sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \ldots + \alpha_q \varepsilon_{t-q}^2 \]

Although the ARCH model deals with volatility quite well it presents some limitations, though. The most important of those are:

- The difficulty in the number of lags (q) selection
- In order to capture all the dependence in the conditional variance the number of q should be very large.
- Some constraints mainly non-negativity may be violated.

Bollerslev (1986) and Taylor (1986) independently from each other generalized the ARCH in order to surpass the limitations of the model and to make it more realistic. The result was the Generalized Autoregressive Conditional Heteroscedasticity (GARCH). The simplest model form, the GARCH (1, 1) is defined as:

\[ Y_t = X_t' \beta + \varepsilon_t \]
\[ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \]

The first equation is the mean equation with \( X_t \) as function of exogenous variables and \( \varepsilon_t \) is the error term. The second equation is the conditional variance equation that consists of three components:

- \( \omega \) is a constant term
- \( \varepsilon_{t-1}^2 \) is the lag of squared error term that measures the volatility in previous period and is considered as the ARCH component of the conditional variance equation (the second term in parenthesis in the GARCH (1,1)).
• \( \sigma^2_{t-1} \) is the forecast variance of the last period and is considered as the GARCH component of the conditional variance equation (the first term in the parenthesis in the GARCH(1,1)).

In the same way the GARCH (p, q) model is defined with q ARCH and p GARCH components. In this study the simple model of GARCH (1, 1) will be used.
4. EMPIRICAL APPLICATION

4.1 Data sources

The data correspond to the NG and oil spot prices in the US energy market. The Henry Hub price is chosen for the NG. Henry Hub is a trading hub in natural gas transmission system in Erath, Louisiana in South America. The settlement prices at the Henry Hub (HH) are used as benchmarks for the entire North American natural gas market. On the other hand, the West Texas Intermediate (WTI) is the benchmark crude oil for the N. America and the oil prices for the analysis are based on this. The data consists of quarterly observations of HH and WTI. The interval of the analysis is limited due to the fact that the HH price is available since 1997. Therefore, the data cover the period from Q1 1997 to Q2 2013. Furthermore the exogenous variables that are used are the US Gross Domestic Product (GDP), NG Residential Consumption in the US and the NG Storage level. Similar to HH and WTI prices the observations of the three exogenous variables are on a quarterly basis.

In order to implement the econometric techniques on the above time series some modifications are necessary. The HH and WTI prices correspond to real prices as their nominal ones were divided by the Consumer Price Index (CPI). Prior to this modification the monthly prices of HH and WTI was transformed into quarterly prices as the average price of three successive months. In the same way the residential consumption and storage are quarterly observations that derive from the average of three successive months. Finally the GDP is coming from the nominal price divided by the CPI and is the real GDP. GDP nominal price was acquired directly on the quarterly basis.

All the above time series are not used in their levels. The natural logarithms of the time series are chosen for the econometric analysis. The logarithmic transformations address the potential heteroskedasticity in time series and remove the scale effects on them. In addition, with the natural logarithms the parameter estimates can be interpreted as constant elasticities. Finally the differentiated logarithms of the time series represent their growth rates or the returns of the time series. Thus returns \((z) = \ln z_t - \ln z_{t-1}\).
A number of figures are presented below in order to capture the evolution of the time series in the analysis’ interval and the correlation between HH and WTI. Figures include the observations in log levels.

**Figure 1:** WTI real prices (Logarithm)

**Figure 2:** WTI real prices returns
Figure 3: HH real prices (Logarithm)

Figure 4: HH real prices returns
Figure 5: WTI-HH real prices (Logarithms)

Figure 6: WTI-HH real prices returns
**Figure 7:** NG Storage level (Logarithm)

**Figure 8:** Real GDP (Logarithm)
4.2 Stationarity testing

4.2.1 Endogenous variables

In order to detect the integration order in our variables many different tests are applied. The unit root and stationarity tests that are performed are the following: The augmented Dickey-Fuller test (ADF), Phillips-Peron test (PP) and the Kwiatkowski–Phillips–Schmidt–Shin test (KPSS). For the test implementation the log level of the West Texas Intermediate and Henry Hub is used. The results from the 3 tests showed that WTI is not stationary at its level but it is stationary at first difference. Thus the WTI is integrated of order 1, I (1). On the other hand the HH presents mixed results in unit root tests. Some tests show that HH is integrated of order 1 and others show that HH is integrated of order 0. That is a common phenomenon in econometric analysis and there are many reasons for this, such as the number of observations, type of observations, etc. From the theoretical background of the study and its purposes we will assume that HH is integrated of order 1, I (1). Also *, **, *** denote rejection of the null hypothesis at the 10%, 5% and 1% respectively. The tables that are presented below illustrate the unit root tests results analytically.
Table 1: ADF unit root test results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Log(level)</th>
<th>First difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no trend</td>
<td>trend</td>
</tr>
<tr>
<td></td>
<td>t-Stat.(k)</td>
<td>t-Stat.(k)</td>
</tr>
<tr>
<td>WTI</td>
<td>1.365(0)</td>
<td>-3.284(0)*</td>
</tr>
<tr>
<td>HH</td>
<td>-2.914(1)**</td>
<td>-2.886(1)</td>
</tr>
</tbody>
</table>

Notes: k represents the selected lag length. The lag length selection based on Schwartz information criterion with $k_{min}=0$ and $k_{max}=10$. *,**,*** denote rejection of the null hypothesis at the 10%, 5% and 1% significance level, respectively.

The ADF test defines as null hypothesis the existence of a unit root. The rejection of the null hypothesis is the stationarity of time series. The results showed that for WTI we reject the null hypothesis at first difference. Thus the WTI is I(1). For the HH we fail to reject the null hypothesis at its log level for 1% but we reject the null at 5% and 10% significance levels. Also we reject the null hypothesis at first difference. Thus for 1% significance level, the HH is I(1).

Table 2: Phillips-Perron unit root test results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Log(level)</th>
<th>First difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no trend</td>
<td>trend</td>
</tr>
<tr>
<td></td>
<td>t-Stat.(b)</td>
<td>t-Stat.(b)</td>
</tr>
<tr>
<td>WTI</td>
<td>-1.445(1)</td>
<td>-3.435(2)*</td>
</tr>
<tr>
<td>HH</td>
<td>-2.045(4)</td>
<td>-2.005(4)</td>
</tr>
</tbody>
</table>

Notes: b represents the selected bandwidth. The Bartlett Kennel spectral estimation method with Newey-West procedure for the bandwidth selection is used for the test. *,**,*** denote rejection of the null hypothesis at the 10%, 5% and 1% significance level, respectively.
The Phillips-Peron test similar to the ADF test has the null hypothesis of a unit root. That means the variable is not stationary. The results showed that for WTI we reject the null hypothesis at first difference and WTI is integrated of order 1, I(1). Also for the HH we reject the null hypothesis at first difference and HH is I(1) as well.

Table 3: KPSS stationarity test results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Log(level)</th>
<th>First difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no trend</td>
<td>trend</td>
</tr>
<tr>
<td></td>
<td>LM-Statistic(b)</td>
<td>LM-Statistic(b)</td>
</tr>
<tr>
<td>WTI</td>
<td>0.909(6)***</td>
<td>0.098(5)</td>
</tr>
<tr>
<td>HH</td>
<td>0.229(6)</td>
<td>0.229(6)***</td>
</tr>
</tbody>
</table>

Notes: b represents the selected bandwidth. The Bartlett Kennel spectral estimation method with Newey-West procedure for the bandwidth selection is used for the test. *, **, *** denote rejection of the null hypothesis at the 10%, 5% and 1% significance level, respectively.

In Kwiatkowski–Phillips–Schmidt–Shin test (KPSS), in contrast to the previous tests, the null hypothesis is the stationarity of the time series. The results showed that for WTI we reject the null hypothesis only at its level. Thus the WTI is I(1). Also in HH the null hypothesis of stationarity is not rejected in both its level and at first difference and it is integrated of order 0, I(0).

4.2.2 Exogenous variables

In order to implement a VAR model the exogenous variables should be examined for stationarity similar to the endogenous variables occasion. In a regression equation all the variables that are included should be integrated at the same order. As a result, after testing for the endogenous variables’ stationarity the integration order of the exogenous variables is necessary to be researched. The tests that will be
performed are the well-known Augmented Dickey-Fuller test, Phillips-Peron test and Kwiatkowski–Phillips–Schmidt–Shin test. The approach is identical to the previous part of WTI and HH time series and it is presented analytically below.

### Table 4: ADF unit root test results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Log(level)</th>
<th>First difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no trend</td>
<td>trend</td>
</tr>
<tr>
<td></td>
<td>t-Stat.(k)</td>
<td>t-Stat.(k)</td>
</tr>
<tr>
<td>GDP</td>
<td>-1.691(1)</td>
<td>-1.705(1)</td>
</tr>
<tr>
<td>Storage</td>
<td>-0.336(7)</td>
<td>-5.276(4)***</td>
</tr>
<tr>
<td>R.Consumption</td>
<td>-2.871(4)*</td>
<td>-3.225(4)*</td>
</tr>
</tbody>
</table>

Notes: k represents the selected lag length. The lag length selection based on Schwartz information criterion with $k_{\min}=0$ and $k_{\max}=10$. *, **, *** denote rejection of the null hypothesis at the 10%, 5% and 1% significance level, respectively.

The ADF test in the exogenous variables showed that in GDP the null hypothesis is rejected at first difference and the variable is integrated of order 1 or is $I(1)$. The storage presents similar results and it is integrated of order 1, $I(1)$ as the null hypothesis is rejected at first difference. On the other hand, the residential consumption (r.consumption) in log level is stationary at 10% significance level and it is stationary at first difference as well. Thus the r.consumption is $I(1)$ at 1% significance level.
Table 5: Phillips-Perron unit root test results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Log(level)</th>
<th>First difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no trend</td>
<td>Trend</td>
</tr>
<tr>
<td></td>
<td>t-Stat(b)</td>
<td>t-Stat(b)</td>
</tr>
<tr>
<td>GDP</td>
<td>-1.499(4)</td>
<td>-1.506(4)</td>
</tr>
<tr>
<td></td>
<td>-5.263(1)**</td>
<td>-5.270(1)**</td>
</tr>
<tr>
<td>Storage</td>
<td>-7.013(25)**</td>
<td>-7.934(18)**</td>
</tr>
<tr>
<td></td>
<td>-15.251(12)**</td>
<td>-14.793(12)**</td>
</tr>
<tr>
<td></td>
<td>-30.370(11)**</td>
<td>-29.845(11)**</td>
</tr>
</tbody>
</table>

Notes: b represents the selected bandwidth. The Bartlett Kennel spectral estimation method with Newey-West procedure for the bandwidth selection is used for the test. *, **, *** denote rejection of the null hypothesis at the 10%, 5% and 1% significance level, respectively.

The Phillips–Peron test presents different results from the ADF test. In GDP variable the null hypothesis is rejected at first difference and it is integrated of order 1, I(1). In contrast to GDP, the storage and the r.consumption are integrated of order 0, I(0) as the null hypothesis is rejected at their log level.

Table 6: KPSS stationarity test results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Log(level)</th>
<th>First difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no trend</td>
<td>trend</td>
</tr>
<tr>
<td></td>
<td>LM-Statistic(b)</td>
<td>LM-Statistic(b)</td>
</tr>
<tr>
<td>GDP</td>
<td>0.612(6)**</td>
<td>0.207(6)**</td>
</tr>
<tr>
<td></td>
<td>0.133(4)</td>
<td>0.078(4)</td>
</tr>
<tr>
<td>Storage</td>
<td>1.160(3)**</td>
<td>0.127(21)</td>
</tr>
<tr>
<td></td>
<td>0.149(12)</td>
<td>0.100(12)</td>
</tr>
<tr>
<td>R.Consumption</td>
<td>0.207(12)</td>
<td>0.096(12)</td>
</tr>
<tr>
<td></td>
<td>0.357(11)</td>
<td>0.325(11)**</td>
</tr>
</tbody>
</table>

Notes: b represents the selected bandwidth. The Bartlett Kennel spectral estimation method with Newey-West procedure for the bandwidth selection is used for the test. *, **, *** denote rejection of the null hypothesis at the 10%, 5% and 1% significance level, respectively.
The KPSS in contrast to the previous tests defines as null hypothesis the stationarity of time series. In KPSS test the GDP is integrated of order 1, I(1) as the null hypothesis is rejected at its log level. In addition, the storage presents the same result as the GDP and it is integrated if order 1, I(1). Finally the r.consumption is integrated of order 0, I(0) as the null hypothesis is not rejected in both its log level and at first difference.

The majority of unit root and stationarity tests revealed that the exogenous variables are integrated of order 1. The evidence of variables integrated of order 0 is presented in some cases but it is ignored due to the theoretical background of the study. The most possible reason is the small sample of observations. Therefore, it is overpassed and the exogenous variables are considered as I(1).

4.3 Cointegration testing

In the previous part we determined the integration order of the HH and WTI time series. The results of the unit root and stationarity testing showed that the two variables are stationary at first difference or are integrated of order 1, I(1). But the linear combination of two (or more) non stationary time series may be stationary according to Engle and Granger (1987). In this study three different cointegration tests are implemented in order to detect an equilibrium relationship between the HH and WTI time series. Engle and Granger, Phillips-Ouliaris, and Johansen cointegration tests were chosen for the analysis. The cointegration testing between two variables showed that there is no cointegration equation and the variables do not have a long-run equilibrium relationship.

In Engle and Granger cointegration test the null hypothesis is that the time series are not cointegrated. The P-values of the t-statistic and z-statistic define the final result. In order to reject the null hypothesis, the P-values should be lower than the significance levels. The automatic lags specification based on Schwartz criterion. The results revealed that P-values are much higher than all significance levels. Thus we fail to reject the null hypothesis and the time series are not cointegrated.
Table 7: Engle and Granger Cointegration test results

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Tau-statistic</th>
<th>Probability</th>
<th>z-statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogHH</td>
<td>-2.635</td>
<td>0.237</td>
<td>-14.869</td>
<td>0.124</td>
</tr>
<tr>
<td>LogWTI</td>
<td>-1.217</td>
<td>0.855</td>
<td>-3.532</td>
<td>0.840</td>
</tr>
</tbody>
</table>

Notes: The lag length selection based on the Schwartz information criterion.

Under the Phillips-Ouliaris cointegration test the null hypothesis is that the time series are not cointegrated. Similar to the Engle and Granger cointegration test the P-values of the t-statistic and z-statistic define the rejection or not of the null hypothesis. If the P-value is lower than the significance levels, then we reject the null hypothesis. The results showed that the P-values are higher than any important significance level and we fail to reject the null hypothesis. Thus the time series of HH and WTI are not cointegrated.

Table 8: Phillips-Ouliaris Cointegration test results

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Tau-statistic</th>
<th>Probability</th>
<th>z-statistic</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogHH</td>
<td>-1.954</td>
<td>0.555</td>
<td>-7.878</td>
<td>0.483</td>
</tr>
<tr>
<td>LogWTI</td>
<td>-1.050</td>
<td>0.893</td>
<td>-2.709</td>
<td>0.895</td>
</tr>
</tbody>
</table>

Johansen cointegration test is performed in a VAR framework in order to detect the existence of cointegration relationships. A very important factor of the Johansen cointegration test is the optimal lag selection. This study implements the Akaike information criterion (AIC) that determines the optimal lag length. According to AIC the optimal lag selection is 3 lags. Furthermore the null hypothesis is that the series are not cointegrated. If the trace statistic is higher than critical value at 5% significance level, we reject the null hypothesis. The final results showed that the trace statistic is lower than the critical value at 5% and we fail to reject the null hypothesis. Thus the time series are not cointegrated. Finally it is well known that between two variables that are tested for cointegration only one cointegration equation can exist and the second row of Johansen test in the next table can be omitted.
Table 9: Johansen Cointegration test results

<table>
<thead>
<tr>
<th>Number of cointegration Equations</th>
<th>Trace statistic</th>
<th>0.05 Critical value</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0: r=0, H1: r=1</td>
<td>7.016</td>
<td>15.495</td>
<td>0.576</td>
</tr>
<tr>
<td>H0: r &lt;1, H1: r=2</td>
<td>0.360</td>
<td>3.841</td>
<td>0.548</td>
</tr>
</tbody>
</table>

Notes: The analysis performs a VAR model approach. The lag length selection is based on the Akaike’s information criterion (AIC).

The three different tests that were implemented showed exactly the same result that the series are not cointegrated. That reveals the absence of cointegration equation between HH and WTI time series. The time series have not an equilibrium relationship in the long-term and any deviation from their short-run co-movement seems to be permanent. The time series probably will follow different ways and move independently from each other in the long-term. The independent price path of HH and WTI reflects that the time series are affected from different factors that determine their route. In addition the next step in empirical analysis, the causality testing, is correlated with the cointegration results. In case that the time series are cointegrated the appropriate procedure is the implementation of a vector error correction model (VECM) in order to detect the existence of causality between the time series. On the other hand the absence of a cointegration relationship between the time series leads to a VAR approach. The results of this study showed the lack of cointegration between HH and WTI and a VAR model is performed in order to detect the existence of causality that runs from HH to WTI and vice versa.

4.4 Coefficients estimation

The HH and WTI are affected from many factors that determine their price path. These factors may be different for the two time series. In previous studies on topic many different factors were used that were assumed as contributors in the HH and WTI prices. Some of them showed to be highly correlated with the two time series and determine their evolution in a high degree. On the other hand, many tested factors revealed that their contribution in HH and WTI price are negligible and non-
significant. This research based on the framework of previous studies examines the possibility of well-known economic variables to affect the HH and WTI prices. A VAR approach is used due to the fact that there is no cointegration equation between the variables. In VAR model we will consider as endogenous variables the HH and WTI time series. On the other hand, the economic variables that will be tested for their impact on the endogenous variables are the Gross Domestic Product (GDP) of the US, the US natural gas storage and the residential consumption (r.consumption) of natural gas in the US. These variables are considered as exogenous and are involved into the VAR model. The VAR model is bivariate and consists of two equations in which the dependent variables are the endogenous ones and they depend on previous observations of endogenous variables plus the exogenous ones. Also the model includes exogenous regressors as the constant term. The next step in this study is the assessment of the coefficients. The level of each coefficient in conjunction with its P-value will show how important are the exogenous variables in the HH and WTI evolution path.

4.4.1 Parameter estimates

The parameter estimates for the exogenous variables revealed very important facts. The findings of the analysis present mixed results. Some variables seem to be very important price drivers of the HH and WTI. On the other hand, there are variables that are not significant and their effect in endogenous variables is negligible.

The estimations are taking place in log level. The difference from the analysis in level is the interpretation of the results. In log level the coefficient estimation has the meaning that if the exogenous variable increased by one percent, it is expected the dependent variable to be increased by $\beta$ percent assuming the $\beta$ as the parameter estimation of the specific coefficient. In the above case the sign of the coefficient was positive and the increase of the exogenous variable leads to the increase of the dependent variable. But the sign of the coefficient may be negative and an increase in exogenous variable leads to a decrease in the dependent variable and vice versa. Additionally in order to interpret the parameter estimates of the exogenous variables as was expressed above, all the other independent variables should be constant.
The exogenous variable of residential consumption showed similar results in both HH and WTI regressions. The coefficient of residential consumption is not statistically significant due to the fact that the P-value in HH and WTI regressions is much higher than all significance levels. Thus the residential consumption does not affect the HH and WTI prices.

In contrast to residential consumption, the Gross Domestic Product seems to be very important in HH and WTI price determination. In WTI regression an increase in GDP by one percent leads to an increase in WTI price by 4.23%. In addition, in HH regression the corresponding increase in the dependent variable from one percent GDP increase is 3.22%. The P-value in both HH and WTI regressions is much lower than all significance levels and the GDP is significant at 1%. That means the GDP affects the HH and WTI price considerably. In addition, GDP is one of the main price drivers of HH and WTI and its contribution to their evolution is very crucial.

Finally, the third exogenous variable in VAR model the storage presents mixed results. In HH regression the P-value of the storage coefficient is much higher than all the significance levels and it is not statistically significant. On the other hand, in WTI regression the storage is statistically significant at 10% significant level. The impact of storage in WTI price is -0.39%. That means if the storage change by one percent, then the WTI will change by 0.39% but in the opposite direction due to the negative sign. Thus at 10% significance level the impact of storage in WTI is important. The results are presented in the table below.

<table>
<thead>
<tr>
<th>Exogenous Variables</th>
<th>LogWTI Coefficient</th>
<th>LogWTI P-value</th>
<th>LogHH Coefficient</th>
<th>LogHH P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>4.2539</td>
<td>0.0000</td>
<td>3.2238</td>
<td>0.0029</td>
</tr>
<tr>
<td>Storage</td>
<td>-0.3944</td>
<td>0.0819</td>
<td>-0.2431</td>
<td>0.3435</td>
</tr>
<tr>
<td>R.Consumption</td>
<td>-0.0452</td>
<td>0.1498</td>
<td>0.0208</td>
<td>0.5596</td>
</tr>
</tbody>
</table>

Notes: The analysis performs a VAR model approach. The lag length selection is based on the Akaike’s information criterion (AIC).
The fact that the natural gas storage variable does not affect the HH price but affects the WTI price seems to be strange. The explanations may be many. Firstly the small size of the observations’ sample is possible to create this difference in the results. In addition, when a VAR is deprived of important exogenous variables that affect the endogenous ones or include exogenous variables that are not statistically significant the results may be altered and their reliability is reducing. In order to overpass this problem and to acquire more accurate results a new VAR is applied. The new VAR consists of the two endogenous and two exogenous variables. The difference from the previous VAR is that the residential consumption is excluded from the model.

The results from the new VAR showed that a part of the parameter estimates remains the same but generally there are important differences. The GDP presents similar results in the new model as the P-values are the same regarding with the previous model and the changes in the coefficient magnitude are minimal and almost negligible. On the other hand, the storage revealed that it is not statistically significant in both the WTI and HH price in all the significance levels. Thus, the results in new VAR approach illustrate a more accurate picture of the HH and WTI price determination and seem to be more logical. The final results are presented analytically below.

**Table 11: Parameters estimates results**

<table>
<thead>
<tr>
<th>Exogenous Variables</th>
<th>LogWTI</th>
<th>LogHH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>P-value</td>
</tr>
<tr>
<td>GDP</td>
<td>4.2864</td>
<td>0.0000</td>
</tr>
<tr>
<td>Storage</td>
<td>-0.2779</td>
<td>0.1924</td>
</tr>
</tbody>
</table>

Notes: The analysis performs a VAR model approach. The lag length selection is based on the Akaike’s information criterion (AIC).

In conclusion, the parameter estimates showed that the GDP variable is an important contributor to WTI and HH price and affects them considerably. On the other hand the residential consumption does not affect the HH and WTI and it does
not have a significant impact on them. Finally the storage presents mixed results in VAR approach with three exogenous variables. After the exclusion of the residential consumption in new VAR model the storage is not significant in both the HH and WTI price.

4.5 Causality testing

After the assessment of the coefficients, a causality testing is implemented. In order to examine the possible existence of linear causal relationship between HH and WTI the process that is applied is a Wald test. The Wald test as has been referred in the methodology part of the study, involves some restrictions in order to capture the potential causality. The results from the Wald test procedure are presented analytically below.

4.5.1 Wald test-Linear causality

In the previous part the coefficients of exogenous variables were estimated and their contribution in WTI and HH price path was revealed. In order to detect the possibility of causality between the two endogenous variables a Wald test is implemented. The Wald test defines as null hypothesis a number of linear restrictions.

In the first equation of VAR model the WTI is the dependent variable. The null hypothesis determines that the coefficients of HH at this equation as equal to zero. In case that the P-value of the test is lower than the significance levels the null hypothesis is rejected and previous prices of HH affects the WTI price.

In the second equation of VAR model the HH is the dependent variable. The Wald test sets a number of coefficients equal to zero as the null hypothesis. The coefficients are these of WTI previous prices that affect the HH price. If the P-value is lower than the significance levels the null hypothesis is rejected and previous prices of WTI affects the HH price. The table below presents the results analytically.
Table 12: Wald test results

<table>
<thead>
<tr>
<th>Wald test</th>
<th>Dependent variable: LogWTI</th>
<th>Dependent variable: LogHH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Causality direction: LogHH → LogWTI</strong></td>
<td><strong>Causality direction: LogWTI → LogHH</strong></td>
<td></td>
</tr>
<tr>
<td>Restrictions</td>
<td>P-value</td>
<td>Restrictions</td>
</tr>
<tr>
<td>C(4)=C(5)=C(6)=0</td>
<td>0.0046</td>
<td>C(11)=C(12)=C(13)=0</td>
</tr>
</tbody>
</table>

The C(4), C(5) and C(6) are the coefficients of previous prices of HH or the coefficients of HH lags in WTI regression. In addition, C(11), C(12) and C(13) are the coefficients of previous prices of WTI in HH regression.

The results showed that the P-value in both the WTI and HH occasion is much lower than all the significance levels. The null hypothesis is rejected and the coefficients are not equal to zero. The conclusion is that previous prices of HH and WTI affect the price of WTI and HH respectively. There is causality that runs from HH to WTI in all significance levels and vice versa. The Wald test showed that there is bidirectional causality between the two variables in all significance levels.

The Wald test was implemented exactly in the same way as in case of the three exogenous variables, after the removal of the residential consumption from the VAR model. The findings were similar to the initial VAR model. The differences are infinitesimal and are changes in the fourth decimal of P-value in both variables. Thus the P-values are much lower than all the significance levels. The causality evidence is very strong and the changes without the residential consumption variable in VAR process are negligible. The small differences are shown below.
Table 13: Wald test results

<table>
<thead>
<tr>
<th>Wald test</th>
<th>Dependent variable: LogWTI</th>
<th>Dependent variable: LogHH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Causality direction:</strong> LogHH→LogWTI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restrictions</td>
<td>P-value</td>
<td>Restrictions</td>
</tr>
<tr>
<td>C(4)=C(5)=C(6)=0</td>
<td>0.0093</td>
<td>C(11)=C(12)=C(13)=0</td>
</tr>
</tbody>
</table>

The C(4), C(5) and C(6) are the coefficients of the HH previous prices in WTI regression and C(10), C(11) and C(12) are the coefficients of WTI previous prices in HH regression.

4.5.2 Breitung and Candelon-Nonlinear causality

In order to examine the possibility of linear causality between the HH and WTI variables the Wald test was performed. The test showed that causality runs from HH to WTI and vice versa due to the fact that the null hypothesis of no causality between the variables was rejected in all the significance levels. Thus there is bidirectional causality at 1% between HH and WTI time series.

Additionally, the causality among variables may exist also in a nonlinear form. The examination of nonlinear causality requires a very different approach from the one that has already been implemented. As it is referred in the methodology part a very important approach is the Breitung and Candelon test that allows checking the possibility of causality in a nonlinear framework. Besides, the existence or not of linear causality does not affect the result of the nonlinear causality test and they are independent from each other.

The Breitung and Candelon approach using a specific critical value examines the possibility of causality. When derive statistics are above the critical value there is causality. If derive statistics are below the critical value there is no causality. Furthermore in case that there is causality the frequencies define the kind of causality. If causality appears in low frequencies that is long-run causality. On the other hand,
causality in high frequencies corresponds to the short-run causality. In graphs below the causality between HH and WTI is illustrated.

Figure 8: WTI prices cause HH prices

Figure 9: HH prices cause WTI prices
In the first graph the possibility of causality that runs from HH to WTI is examined. The derive statistics are below the critical values in all frequencies. Thus there is no nonlinear causality that runs from HH to WTI. Therefore, the HH price does not affect the WTI price in nonlinear framework. In contrast to the previous case the testing for the opposite causality direction (from WTI to HH) showed that there is causality as the derive statistics are above the critical value in specific frequencies. In graph the function is illustrated to overpass the critical values from 0.9 frequency level until the end of 3.14 frequency level. That means there is nonlinear causality that runs from WTI to HH. The level of frequency in which the derive statistics overpass the critical value showed that the kind of causality is mean causality. In conclusion, there is nonlinear mean causality that runs from WTI to HH. HH depends on WTI and has not an independent price path as the contribution of the latter to the former is significant.

4.6 Volatility effects testing

4.6.1 GARCH (1,1)

According to the methodology part, any time series is possible to present heteroskedasticity. That means the variance of the time series is not constant. In order to deal with the heteroskedasticity this study performs a GARCH model that is a generalized Autoregression Conditional Heteroskedasticity model (ARCH). The GARCH (1,1) is chosen for the analysis. This model consists of one GARCH component and one error term or ARCH component in the conditional variance equation.

After the GARCH implementation the conditional variance of the time series is the uncertainty of them. The uncertainty of the time series illustrates how volatile are the specific ones. Using the volatility of HH and WTI the potential causality between them is examined in order to detect the possibility of the HH volatility to cause the corresponding of WTI and vice versa. Also the potential causality that runs from HH and WTI volatility to the log level of the time series is investigated. Therefore, any possible correlation of the HH and WTI volatility time series with each
other and with the log level of the WTI and HH is detected. Finally, all the above tests for causality will depict the impact of the volatility in the HH and WTI time series.

4.6.2 Causality testing

Similar to the causality approach in the log level of the HH and WTI the procedure consists of two tests. The first one is the causality test of Toda-Yamamoto that detects the potential linear causality without investigating the integration order or the existence of cointegration among the variables. Toda-Yamamoto implements a VAR model where the optimal lag length is based on the AIC information criterion and the test defines as lag length the optimal lag plus one. The nonlinear causality is examined with the Breitung-Candelon test.

4.6.2.1 Toda-Yamamoto-Linear causality

As it has been referred in methodology the Toda-Yamamoto causality test examines the possibility of linear causality. Three different cases of causality were tested in this study. Firstly the causality between the volatility time series of HH and WTI. In addition, the test checked the potential causality that runs from HH volatility to WTI. The last and very crucial implementation of the test is the causality that runs from WTI volatility to HH time series. All the above Toda-Yamamoto causality tests use a Wald test approach that examines a set of restrictions in order to confirm or not the existence of causality. The results of the tests are presented below.

<table>
<thead>
<tr>
<th>Table 14: Toda-Yamamoto causality test results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Toda-Yamamoto test</strong></td>
</tr>
<tr>
<td><strong>Dependent variable: WTI volatility</strong></td>
</tr>
<tr>
<td>Causality direction: HH volatility → WTI volatility</td>
</tr>
<tr>
<td>C(6)=0</td>
</tr>
<tr>
<td><strong>Dependent variable: HH volatility</strong></td>
</tr>
<tr>
<td>Causality direction: WTI volatility → HH volatility</td>
</tr>
<tr>
<td>C(3)=0</td>
</tr>
</tbody>
</table>
In the first equation of Toda-Yamamoto the volatility of WTI is the dependent variable and the $C(6)$ is the coefficient of previous prices of HH volatility. The P-value is lower than 5% and the null hypothesis of no causality at this level is rejected. That means linear causality runs from HH volatility to WTI volatility at 5% significance level.

In the second case, in the HH volatility regression equation $C(3)$ is the coefficient of previous prices of WTI volatility time series. The P-value is much higher than all the significance levels and the null hypothesis is not rejected. Thus there is not causality that runs from WTI volatility to HH volatility time series.

The results of causality test between the HH and WTI volatility showed that causality runs only from HH volatility to WTI volatility at 5% significance level. In the following table the results from the causality test between the volatility of HH and WTI and their log levels are shown.

**Table 15**: Toda-Yamamoto causality test results

<table>
<thead>
<tr>
<th>Toda-Yamamoto test</th>
<th>Dependent variable: LogWTI</th>
<th>Causality direction: HH volatility $\rightarrow$ LogWTI</th>
<th>Dependent variable: LogHH</th>
<th>Causality direction: WTI volatility $\rightarrow$ LogHH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restrictions</td>
<td>P-value</td>
<td>Restrictions</td>
<td>P-value</td>
<td></td>
</tr>
<tr>
<td>$C(5)$=$C(6)$=$C(7)$=0</td>
<td>0.2624</td>
<td>$C(5)$=$C(6)$=$C(7)$=0</td>
<td>0.2902</td>
<td></td>
</tr>
</tbody>
</table>

Under the first equation the dependent variable is the WTI and the causing variable is the volatility of HH. The P-value is much higher than all the significance levels and the null hypothesis is not rejected. The restrictions $C(5), C(6)$ and $C(7)$ are obviously the previous prices of HH volatility time series. Therefore, there is not causality that runs from HH volatility to WTI time series.
Similar to the previous case in the regression equation of HH the null hypothesis is not rejected and the WTI volatility does not cause the HH time series. In conclusion, we can identify that the HH and WTI volatility time series do not cause the WTI and HH time series, respectively.

4.6.2.2 Breitung and Candelon—Nonlinear causality

The nonlinear causality test is performed under a Breitung-Candelon approach. The existence of nonlinear causality is examined similar to the linear causality test. That means in the first step the causality between the volatility time series of WTI and HH is checked. Additionally, causality is examined between the WTI and HH volatility time series and the HH and WTI the time series, respectively. The Breitung-Candelon test detects the possibility of causality in a frequency domain.

The test between the volatility time series of HH and WTI revealed that the volatility of the WTI does not cause the volatility of the HH as the derive statistics are below the critical value in all frequencies. On the other hand, the volatility of the HH causes the corresponding of WTI as the derive statistics are above the critical values in all frequencies. Therefore the causality that exists has the characteristics of short-run and long-run causality as well.
Figure 10: WTI volatility causes HH volatility

Figure 11: HH volatility causes WTI volatility
The causality test between the HH volatility and the WTI time series showed that the derive statistics of the test are below the critical value and therefore there is no causality that runs from HH volatility to WTI time series. Additionally, similar to the previous test, there is not causality that runs from WTI volatility to the HH time series in all frequencies. In conclusion, the findings revealed that there is not nonlinear causality that runs from the HH and WTI volatility to the WTI and HH time series, respectively.

Figure 12: HH volatility causes WTI prices
Figure 13: WTI volatility causes HH prices
5. CONCLUSION-DISCUSSION

This study examines the relationship between NG and oil prices in the US. These fuels were traditionally linked and do not move independently from each other. The HH prices formation does not correspond to market fundamentals but are strongly correlated to WTI prices. The majority of the studies in literature suggest the existence of a strong relationship between the two time series. Many times after an exogenous shock the NG and oil prices were decoupled temporarily but in long-term they converged in their equilibrium point.

On the other hand, during the last years some researchers propose the theory that NG and oil prices are affected from entirely different factors and follow separate price paths. In the US the oil and NG prices seem to move independently from each other in contrast to the European and Asian energy markets where the oil and gas linkage is still very mighty. The evidence for this decoupling trend in the US is very strong and can be explained from technological and economic changes. Although the shale gas technological breakthrough in last years lead to an overproduction of NG, the transmission system inefficiencies and the low export capacity developed bottlenecks in the transmission system of NG and pushed down the NG price removing it from the oil price. Furthermore, the reduce of substitutability between oil and NG fuels in industrial boilers and their use for different purposes and sectors remove the fuels prices from a co-movement.

The econometric techniques that were used in the research investigated the potential decoupling trend between NG and oil prices in the US. In order to deal with the time series integration order three different unit root and stationarity tests were performed. The augmented Dickey-Fuller test (ADF), Phillips-Peron test (PP) and the Kwiatkowski–Phillips–Schmidt–Shin test (KPSS) are chosen. The results showed that the endogenous variables of HH and WTI are integrated of order one. Also the exogenous variables of Gross Domestic Product (GDP), NG residential consumption and NG storage are integrated of order one. Despite the fact that some tests showed that time series are integrated of order zero that was minority and it was ignored due to the study’s theoretical background and its purposes. In addition, the type of observations and the small sample may create this phenomenon.
After the determination of time series’ integration order three different cointegration tests are performed in order to detect the existence of a common stochastic trend between HH and WTI time series. The tests that were used are the Engle and Granger, Phillips-Ouliaris, and Johansen cointegration tests. In case that cointegration exists there is long-run relationship with occasional decoupling between HH and WTI. All the above tests conclude that there is no cointegration equation between HH and WTI. Therefore, HH and WTI do not have a long-run equilibrium point. If the possibility of cointegration be ignored the relationship between the time series can be misspecified and the parameters estimates’ reliability is reducing.

One of the main reasons for the potential permanent decoupling trend between HH and WTI that is examined is that the contributors of HH and WTI price formation seem to be different. In order to reveal the effect of other economic variables in HH and WTI price many potential contributor variables are checked in literature. This study investigated the impact of GDP, NG residential consumption and NG storage on the WTI and HH prices. For that reason a VAR model is chosen as the appropriate technique due to lack of cointegration that did not allow using a VECM model. The VAR model consists of two endogenous variables, HH and WTI and three exogenous ones that are referred above. The parameters estimates showed the magnitude and the significance of the exogenous variables’ coefficients.

In conclusion the GDP is shown as the main contributor in HH and WTI price evolution in contrast to the residential consumption that does not have any impact on them. The NG storage seems to be significant only on WTI price but not on HH price, something that is not expected. The implementation of a new VAR after the exclusion of the residential consumption augments the reliability of the estimates and presents some differences. The NG storage is not significant on HH and WTI time series in contrast to the GDP that remains very important contributor on them.

Regarding with the effect of HH on WTI and vice versa, the causal relationship between them is examined. The existence of causality is investigated in linear framework and in nonlinear as well. The linear causality implemented a Wald test that revealed bidirectional causality between HH and WTI time series in all significant levels. On the other hand, the nonlinear testing takes place under a Breitung-Candelon approach. The results captured that nonlinear causality runs from
WTI to HH time series. The kind of this specific causality is considered as mean causality. In absence of cointegration equation between the variables the possibility of long-run causality is excluded due to the fact that there is no long-run equilibrium point. The causality that is investigated in both the linear and nonlinear framework corresponds to short-run causality.

In order to deal with the heteroskedasticity in time series a GARCH (1, 1) model is performed. The conditional variance of GARCH model corresponds to time series volatility. After the determination of the HH and WTI volatility the possible causal relationship between them is examined in both the linear and nonlinear framework. Similar to the causality testing in log levels of HH and WTI the linear causality is tested with the Toda-Yamamoto test and the nonlinear causality uses a Breitung-Candelon approach. The results showed that linear causality runs from HH volatility to WTI volatility at 5% significance level but there is no causality in the opposite direction. The nonlinear testing captured only the existence of causality from HH volatility to the WTI volatility time series. This causal relationship has the characteristics of short-run and long-run causality.

The next step in the study is the examination of causal relationship between the HH and WTI volatility time series and the WTI and HH time series respectively. Under the same procedure for the linear and nonlinear causality the Toda-Yamamoto and the Breitung-Candelon approach were used. The results from the linear testing identified that the HH and WTI volatility time series do not cause the WTI and HH time series, respectively. Similar to the linear case the findings of nonlinear testing revealed that there is no nonlinear causality that runs from the HH and WTI volatility to the WTI and HH time series, respectively. In the table below the overall results from all the causality testing are depicted analytically.
The three exogenous variables, US GDP, NG residential consumption and storage in the US were chosen in order to check their possible impact on the HH and WTI time series as they were considered as potential contributors in HH and WTI price route.

GDP is the gross domestic product in the US and corresponds to the productivity and the magnitude of the US economy. In case that the real GDP that is used has an upward trend the US economy enjoys high growth rates. The purchasing power of consumer augments and the logical result is that the NG and oil demand will increase. This demand rise will lead to a price rise. On the other hand, a decrease of GDP will lead to opposite results. Thus, HH and WTI are possible to be affected from changes in GDP. The findings revealed that the GDP is a very crucial contributor to HH and WTI prices and affects them considerably verifying the theoretical estimation.

NG residential consumption corresponds to residential consumer demand for NG. Furthermore it reflects the residential consumers’ preference for NG over other competitive fuels. Residential consumers are a large part of the overall NG consumers in the US and their share in NG consumption is not negligible. The preference for NG at the expense of other fuels due to environmental awareness or economic reasons is possible to affect the NG and oil prices and to contribute to their evolution. The findings showed that the residential consumption is not a significant contributor to
HH and WTI price path and its impact on them is negligible. Thus the theoretical estimation was not verified by empirical results.

NG storage in the US has a very important influence on HH and WTI prices. A number of factors as overproduction of shale gas, low export capacity, peak demand, shut-ins in production sites and frequent natural disasters fluctuate the NG storage level and that correspond to implications on HH and WTI price levels. According to Brown and Yücel (2008), especially in the US gas market the seasonal variation in demand depends on storage inventories rather than in production changes and that leads to more volatile prices. The difficult part in storage handling is that its influence is depicted under conditions. The empirical results showed that the NG storage is not significant on HH and WTI prices evolution in contrast to the theoretical background. This specific combination of samples’ magnitude and the form of NG storage time series that is used in this study is not the appropriate to illustrate the force of gas storage on HH and WTI price series. In conclusion, the NG storage is possible to affect considerably the gas and oil prices in the US but this influence was not reflected in the empirical results.

The results of cointegration testing are very important and confirm the findings of the last studies on topic. HH and WTI do not share a common trend and are not linked with a cointegration equation. Therefore, after an exogenous shock the HH and WTI prices present a decoupling trend that seems to be permanent in absence of a long-run equilibrium point that would lead to prices convergence. The lack of long-run relationship between HH and WTI depicts the independent price path of them as they move separately from each other in the long term.

Causality runs from HH to WTI and vice versa in linear approach that corresponds to short-run causality. Furthermore WTI causes the HH in nonlinear framework and the relationship is considered as mean causality. The causal relationship between HH and WTI was expected as they affect each other considerably. They are not independent and they follow a co-movement in the short-run.

The uncertainty of HH and WTI corresponds to volatility of these time series. The volatile nature of HH and WTI is possible to influence their price evolution. The results showed that only HH volatility cause the corresponding of WTI and the
volatility time series do not affect the log levels of them. This conclusion is very important as it illustrates that the uncertainty of HH and WTI has no impact on the WTI and HH prices respectively.

The general picture of the cointegration and causality testing in both the log levels and the volatility of time series revealed a strong evidence of permanent decoupling trend between HH and WTI probably due to the fact that are influenced by different factors. The overproduction of shale gas and the inefficiencies in the gas transmission system remove the oil and gas prices from each other in the long-run. In addition, the gas markets under the globalization of natural gas will set their prices based on other gas markets prices in contrast to current oil-indexed gas prices. Furthermore, the vertical integration of the gas sector and the economic activities of companies that correspond to the entire gas value chain is possible to decouple the gas prices from the oil ones. Although the correlation between the two fuels is still mighty the linkage in pricing formation is decreasing. In addition, the volatility effects that were tested depict that an exogenous shock in one fuel price does not correspond to the other. The absence of long-run equilibrium point and the unaffected time series from prices volatility seem to develop a permanent divorce for NG and oil prices in the US.

All the above findings converge to the results of previous studies in last years that revealed strong evidence for permanent decoupling trend between HH and WTI and independent price routes. Although the oil and gas in the US are strongly connected the first elements of a separate price movement is already a fact.
REFERENCES


