Erasmus University Rotterdam

MSc in Maritime Economics and Logistics

2015/2016

Empirical analysis of Natural Gas price: Linkage to Oil indexation

by

Puneet Mishra

Acknowledgments

With this last and final project, my incredible journey at MEL is about to finish. In few days I will be holding my master's degree (I still don't believe that I did it!!) and will be ready to face new challenges. One year in MEL, with an incredible group of enthusiastic mates, was great learning experience. Last year, at this very day I was excited but at the same time nervous too; it's been eight years since I sat in a classroom or opened a textbook. But time flies and here I'm writing my last piece of work at MEL. Now when I look back, I realise that I have learnt a lot from generous professors and vibrant classmates. From long and tiring group assignments to relaxing barbeque on a sunny day with friends or from exhausted group study for the ML-3 exam to crazy nights at Hamburg trip, it was an exceptional year with memories to cherish all my life.

I would like to take this opportunity to show my gratitude to few people, without them my journey would have never been so easy. First and foremost, I would like to thank my mother. She backed me in all my decisions and showed faith in me. My father who has been an inspiration for me and who always motivated me to strive harder. My brother, younger but more sensible and supportive than me. A true companion and my greatest strength. Without my family, I would never be able to chase my dreams. Thank you so much for your unconditional love and support.

Honestly speaking, I'm a terrible writer, and before MEL, I have never written anything before MEL. These last couple of months I have tried my level best to write something interesting and relevant, and it would have been not possible without the unprecedented guidance by my supervisor, Ted Welten. I would like to take this opportunity to thank you, professor, for being patient with me and in directing me in the best possible way.

I would like to thank MEL office, especially Renee and Martha for their unparalleled support and guidance throughout the year. My class of "Happy MEL", thank you guys for encouraging me in every possible way and giving me the countless beautiful memories.

My second family of five friends. Thank you guys for tolerating me, supporting me and pushing me towards the best. Last but definitely not the least, I would like to thank Katherine for loving me and supporting me in every possible way. I know at times I get stressed, but MEL would have never been possible without you.

Abstract

The price relationship between crude oil and natural gas is a debatable topic and there exists various theories arguing over the extent of linkage between natural gas with crude oil prices. However, overarching discussion indicates that the crude oil prices lead the natural gas prices. This also explain oil indexation of gas price being cited as the prime reason why crude oil prices concomitantly led to fall of natural gas prices. This research attempt to examine the long-run relationship between Natural gas and Crude oil prices using the historical data from January 1999 to June 2016. In this research author argue that though natural gas prices and crude oil prices share a fundamental long-term relationship, natural gas prices are susceptible to factors like weather, seasonality, inventory level, and disturbance in supply too.

The research uses the Conditional Error Correction model in Vector Error Correction model environment to analyse the natural gas and crude oil price relationship. There are two significant outcome of the model. First, the results illustrate that until mid of 2008, up to 21.89% volatility in natural gas prices are captured by the contemporaneous change in crude oil prices, after 2008, crude oil prices capture only 16.25% volatility in natural gas prices. The research quantifies that prior 2008, the error correction term is 10.4%, which implies that the natural gas price adjusts itself and correct 50% of the deviation in 6.29 weeks, also known as the half-life of error. But after 2008, the correction speed is reduced to 5.49% with half-life increased to 12.37 weeks.

This result has two important implications. First, the natural gas-oil price relationship is becoming week. And second, the external factors apart from oil prices are playing a significant role in natural gas prices. The reason for this week linkage could be the recent development in the natural gas market in the US, with shale gas revolution and the US becoming a natural gas exporter. It will be interesting to watch the natural gas market in near future and analysing the linkage of oil and natural gas prices.

Contents

List of Tables

List of Figures

List of Abbreviations

HH	Henry Hub prices
WTI	West Texas Intermediate
Ln	Logged e
HDD	Heating Degree Days
CDD	Cooling Degree Days
HDDEV	Deviation from normal Heating Degree days
CDDDEV	Deviation from normal Cooling Degree days
STORDIFF	Storage level of natural gas differential with 5-year average
SHUTIN	Shut down of natural gas production in Gulf of Mexico
VECM	Vector Error Correction Model
ECM	Error Correction Model
MMBtu	One Million British Thermal units
Bbl	Barrel

1. Introduction

Recent news like "Next Stop for U.S. Natural Gas Is 20-Year Low Amid Warm Weather" (Bloomberg, 2016b), "HOW LOW CAN GAS PRICES GO?" (Rakesh Upadhayay, 2016) or "Has Natural Gas Hit Bottom?" (Bloomberg, 2016a), draw the attention of the world. When analysing a commodity, it is important to know that the particular commodity price dependent on other commodity price fluctuation. There were several academic studies to find and establish the relationship of natural gas price with its nearest substitute, oil. The studies follow different methodologies, but all the studies result in a conclusion that natural gas price and crude oil price are co-integrated.

This study focused on investigating the long-term relationship between Henry-Hub natural gas prices and the WTI crude oil prices.
Theoretically, both the commodities share a stable link through supply and demand phenomena. As both the commodities are the primary source of energy. On the demand side, the growing energy demand and an option of fuel switching at power plants link the two commodities. Out of many reasons, the energy equivalent through the burner-tip mechanism of pricing link the two commodities. From wellhead to burner-tip, the price structure has many layers including the extraction cost and transportation cost. Comparing the energy sources by energy equivalent leads to competition between the two commodities.

On the supply side of the equation, there are again two sub-section. One where the crude oil and natural gas both are extracted as coproducts. The increase in the price of crude oil will lead to increase production of crude oil and result in increased production of natural gas. Undoubtedly, tagging the two commodities as complements in production. On the other hand, where the individual commodity is extracted from designated reservoirs or wells, the two commodities are linked through asset allocation. The companies involved in research and development (R&D) decide on which commodity they should focus. This decision is mainly driven by the prices of commodities. If the price of crude oil increases, the asset allocation in extracting crude oil grow and this lead to scarcity in research and development of natural gas. As a result, the supply of natural gas decline. So, two commodities share a rival relationship in production.

In general, it could be argued that there exists a long-term relationship between natural gas and crude oil but the commodities could be complement or rival.

Until late in 2008, natural gas prices were following crude oil prices. Even in the sudden fall of prices in mid of 2008, natural gas prices precisely follow crude oil prices. But in early 2009, where crude oil prices recover, the natural gas prices continuous to drop. It was that moment when the two commodities show the divergent trend. The two series move in a different direction and during this period the ratio between the two prices exponentially rise and reaches as high as 55-to-1.

This phenomenon of decoupling of natural gas price from oil price is not new in the natural gas market. In the early 1990s, the excess supply of natural gas in market kept the price low, as compared to prevailing oil price. At the beginning of 2000's, something opposite happened, the prices of natural were higher than the anticipated price. This random movement of natural gas price always brought the discussion to the table about the decoupling of natural gas price from oil price.

In February 2014, the market was arguing that will natural gas hit $15/MMBtu mark or not? It was sure that price would rise, the argument was on how fast it will happen. But after June 2014, the natural gas market saw one of the worst phases regarding price. The price never recovers and touch rock bottom with 1.60$/MMBtu. So, where all the predictions went wrong? Does the market forecaster consider in their forecast the hidden variable, oil price? Maybe the natural gas market had bright future, but oil price drags it down as oil price started to fall in June 2014. These incidents draw an attention of analyst in the business and the energy community.

The advancement in the Natural gas production process, horizontal drilling technology, accelerated the rate of production of natural gas through shale reservoirs. The natural gas production observes impressive rise in the share of shale gas share, where shale gas production increased by 540% since 2009, the total production of natural gas grew by 35%. Apparently, shale gas shakes the pricing structure of natural gas and diverge the oil and gas price series.

But when shale gas technology came into the picture, the world was facing global economic recession. So, it is hard to segregate the impact of global recession and shale gas and point out the cause of the disparity. Apparently, in past, the natural gas price and oil price time series do not move together. So, the question arises that do cointegration theory established by the academic studies are falling apart? Or they still hold but need further adjustment due to recent development.

During the analysis of prices from January 1999 to June 2008, there was no significant evidence found of a divergent trend of the two commodities due to a recession of 2001. But it can be argued that the recession of 2001 was barely a recession. Therefore, in this research, the long-term relationship between crude oil and natural gas price is analysed from 1999 to June 2016. Moreover, the segmented time frame, January 1999 to June 2008 and July 2008 to June 2016 are also examined to draw a comparison.

1.1 Research objective

This study is aimed to analyse the relation between the two-time series, natural gas price and oil price series and to explain the volatility in natural gas price through oil price and other variables. The research, therefore, aims to determine the cointegration relationship between the natural gas and oil price. Additionally, the study will seek to identify and will explain the variables which have a significant impact on natural gas prices. Moreover, the research also focuses on the period when the two commodities decoupled from the long-term relationship. The scope of this study is thus twofold. First, it focuses on natural gas price relation with the oil price, cointegration between natural gas and crude oil prices using the factors which can explain as carefully as possible the past trends in natural gas price. And second, the period is segmented in two different time-frame to analyse the decoupling phenomena of the two commodities.

As such, the central research question that this study aims to answer is:

" Focusing on cointegration between natural gas and crude oil price What is the long-term relationship between natural gas price and crude oil price?"

The main reason behind this research question is that as the market is very volatile and natural gas price fluctuation lead to grave concern in the energy sector. If the relationship of natural gas price and oil price can be explained by empirical analysis and natural gas price can be predicted depending on prevailing oil price than the market can anticipate and take appropriate action according to the situation.

To sufficiently answer this central research question, some sub-research questions must be answered:

Sub-research question 1: "What are the basic fundamental rules used in past to link natural gas and crude oil prices?"

Sub-research question 2: "What are the conditional factors/variables that show a significant trend with natural gas prices and are relevant in modelling the natural gas and crude oil price relationship?"

Sub-research question 3: "Which methodology to be used to analyse the relationship between the natural gas price and oil price time series?"

Sub-research question 4: "How significant and temporal natural gas and crude oil prices decouple?"

In the research, we use time series analysis of two-time series, natural gas price and oil price to analyse the relationship between the prices of the two commodities. The objective is to create an evolving model for a natural gas price which can capture and explain the natural gas price volatility. Also, the aim of research to analyse the how the relationship between natural gas and oil price evolved in last few decades.

1.2 Relevance

Natural gas is the purest form of an energy source as compared to its closest substitute coal and oil. With increased environmental concern in the world, natural gas shows a promising future in the energy sector. The natural gas extraction and storage requires a massive capital investment and the falling natural gas price bring a serious concern for the natural gas market.

A commodity price can fluctuate due to demand instability, supply uncertainty, nearest substitute price volatility and other exogenous variables. The current situation regarding the natural gas price is not promising. Oil is the closest substitute for the natural gas and oil market is in turmoil. Though the demand for natural gas is increasing with steady and slow pace, the stand of Australia in Pacific Basin and the US in Atlantic basin as a major supplier with the recently high supply of deliverable natural gas, flood the market with an excess supply of natural gas.

With falling oil price, it is essential to analyse the two commodities relationship. Though the natural gas price and crude oil price are cointegrated, supported by substantial academic research, it is crucial for the market analyst, traders and other service providers in a supply chain of natural gas from well to the end user to have some price stability. For the operator's perspective, it is required to know the price for arbitrage cargo. With natural gas service providers like FSRU companies' perspective, it is necessary to know the natural gas price as this is profoundly linked with LNG market.

With falling oil price, there is a general perception that oil prices are dragging the natural gas prices, but with the past data, it is not easy to justify. There were instances when the oil prices and natural gas prices move in the opposite direction. Oil is a global commodity, and its prices very much depend on the global upsets. In past, oil prices anchored the relationship between natural gas and oil prices. In present situation where there is a high environmental concern, the world wants to explore natural gas as an energy source, but with potential consequences of high volatility and recent development, it is important to analyse the linkage of natural gas price with oil price.

With falling oil and gas prices, now is the time to examine the trends of natural gas price time series and its cointegration with oil price time series in last few years and compare those with the past academic studies. The detailed analyses are required to understand the market dynamics and the role of various variables in quantifying the price fluctuation.

1.3 Research design and Methodology

In this thesis, both quantitative and qualitative methods are used to reach to a substantial conclusion on how to analyse the natural gas price structure, while checking the cointegration of natural gas and crude oil price series cointegration. In line with this, first, a VECM model will be used to identify and characterise the cointegration relationship between natural gas and crude oil price time series. The change in price for both natural gas and crude oil will also be modelled as a function of past change in both natural gas and crude oil price. This change will explain the inherent volatility left after particular shock. Also, as we assume that change in the price of one will change the price of another commodity (natural gas and crude oil), the incorporated "change in price" in the model will help us to understand and visualise how the change in the price of one commodity affect the price of another commodity. This effect of the history of a natural gas price time series and that of crude oil on natural gas price and vice versa.
After this, we will use Error correction mechanism (ECM) of VECM model. We have seen in past that the natural gas price and oil price move in the same direction but at times they move in opposite direction. The two-time series have deviated from their relationship. The rate with which they return to their relationship is important to know. For this we use ECM part of VECM model, to analyse and measure the speed at which the natural gas time series return to existing cointegration relationship after deviating from it because of some external exogenous variables or to a matter shocks.

So, in our model, there will be three components. First, the cointegration relationship between natural gas price and crude oil price. Second, the exogenous variable which leads to sudden shocks and these shocks persist for a long time and causing deviation in natural gas price and third, the rate with which natural gas price return to the long run relationship with crude oil. We are assuming that the crude oil price is driving the natural gas price and not vice versa. For this, while using ECM in the case of natural gas, we modify ECM and analyse previous change in natural gas price effect on the change in natural gas price. In this case, ECM model will behave like VAR model, and it is conditional ECM. The ECM model analyses the rate with which natural gas return to the relationship which is generated by VECM model.

Later we investigate our model with the segregated time models, where we examine how significant and temporal is the decoupling of natural gas prices.

1.4 Thesis Structure

Chapter: 2, gives the overview of the natural gas market since early 1930's and how the natural gas develops as a commodity.
Chapter: 3, contains the qualitative and quantitative description and analyses of few famous rules of thumbs used in past. The discussion in this section answer the Sub-research question: 1.
Chapter: 4, includes the relevant studies carried out in past and explain various methodologies used in past to analyse the relationship between the natural gas price and oil price.
Chapter: 5, describes the various parameters which affect natural gas prices apart from oil prices. The section contains the analyses of the parameters and their impact on natural gas price in past. This section answers the sub-research question:2 and prepares the variables for our model.
Chapter: 6, includes the quantitative methodology used in research along with the description of VECM and conditional ECM. The chapter explains the fundamental characteristics of time series. Also, the details of data used along with the reasons to use these data are provided in this section. This section answers the sub-research question:3.
Chapter: 7, elaborate the results of the model. Also, the section incorporates the analysis of results.
Chapter: 8, summarise the research and provide the conclusion. After analysing the results, this section answers the main research question. Also, limitation of this investigation and scope for further research is discussed in this section.

2. Development of Natural gas

In early 1900's, natural gas was treated as a by-product of crude oil in extraction. Crude oil transported through pipelines whereas there were no dedicated pipelines for natural gas. So, natural gas was not transported or stored, and if there is no local usage, the natural gas is either flared or vented from the rigs. In the late 1930s, the first long distance truck pipelines were laid to transport natural gas from, offshore natural gas basin in Houston in Texas to central Indiana.

The increasing transportation of natural gas through long-distance intrastate pipelines raised series regulatory concern. There was a tremendous need to establish a governing body to regulate these pipelines.
In the last, due to grave concern regarding the monopoly power of interstate pipelines and the industrial assortment, the federal government decided to step in and provide a regulatory framework for interstate pipelines.

The federal government intervened directly in the Natural Gas Act of 1938 for the passage of natural gas transported through interstate pipelines. The primary objective of the Act of 1938 was to protect customers against unreasonable pricing of natural gas due to the monopoly of interstate pipelines. Under this Act, Federal Power Commission issues a “certificate of public convenience and necessity” to the companies, and once the company have a certificate then only it can make the sale of natural gas across states. National Gas Act (NGA) specified that no new pipeline should be constructed to target a market which is already served by another pipeline. As the primary focus of NGA was on the transportation of natural gas, the sale from interstate pipelines were also regulated through NGA. At the time of constructing NGA, the focus was to provide natural gas at "just and reasonable" rate, but it missed out to precisely regulate the price at the wellhead. This lead to an unregulated grey patch of sale in a well-regulated natural gas market.

The companies in natural gas market argued that this act exempts production and gathering of natural gas, and they have no obligation on the sale of gas at wellhead price. The argument directly challenges the primary idea of this Act, to counter monopoly and save consumer interest against unreasonable high price, as producers can charge wellhead prices based on market and not on the actual cost of production. But later in 1954, a famous case "Phillips Petroleum Company v. Wisconsin” (347 U.S. 672 (1954)) supreme court ruling extended the federal price control over the wellhead gas.

From 1954 to 1960, Federal Power Commission (FPC) attempted to deal with individual companies. The rate at which each producer can sell natural gas was determined on a different level depending on their service standards. But this process turned out to be unfeasible as the number of different producer and rate cases were huge and this lead to the severe problem of backlogging. Facing this issue, in 1960 FPC decided to set the rate of interstate pipeline natural gas by regions. The commission divides the natural gas market into five areas. FPC was looking for the reasonable rate to establish a regional ceiling on natural gas price. Till the time it can determine "just & reasonable" rate, the FPC sets an interim cap rate. This interim rate depends on the average contract price paid during 1955-1960.

However, this process took a lot of time and by 1970, FPC was able to set the price of only 2 out of 5 regions. The regional price for stagnated and prices were same as ten years ago. So, the price laid down for the two areas were not able to cover "just as reasonable" rate targeted by FPC. The main reason was that there were many wells in one region with different production costs. So, by 1974, FPC realise it is next to impossible to set a reasonable regional rate and FPC set a national price ceiling. The new price cap was almost double of the prices of natural gas in 1960's, but still, it was not able to catch the various production cost of natural gas in the region. In the end, FPC accepted the fact that- first, it is not possible to set regional pricing. Second, it was not feasible to set the price depending upon the cost of service and third that new price also does not cover production cost for all individual well. But this system of price control remains until the passage of Natural Gas Policy Act (NGPA) in 1978. In 1983, US government ended federal regulation of wellhead natural gas prices. With time, natural gas market stabilised and consumer prices decreased with more discovery of natural gas. Consumers are switching from coal and petroleum product to natural gas as the source of natural gas for domestic usage.

As the market evolves due to complication in setting "just & reasonable" rates and prices depending upon service level, the natural gas market allowed to adjust itself. The deregulated market of natural gas allowed the forces like seasonality, weather, and inventory level to play a significant role in the natural gas prices. The removal of constraints and the free movement of the natural gas market on its own helped in the better analysis of natural gas linkage with other commodities.

In recent data published by EIA, on Net Generation by Energy Source: Total (All Sectors), 2006-April 2016, natural gas account for approximately 35% of total electricity generation. With all of its uses, natural gas accounts for 33% of total energy requirement of US (U.S. Energy Information Administration, 2015).

3. Natural gas and crude oil price relationship

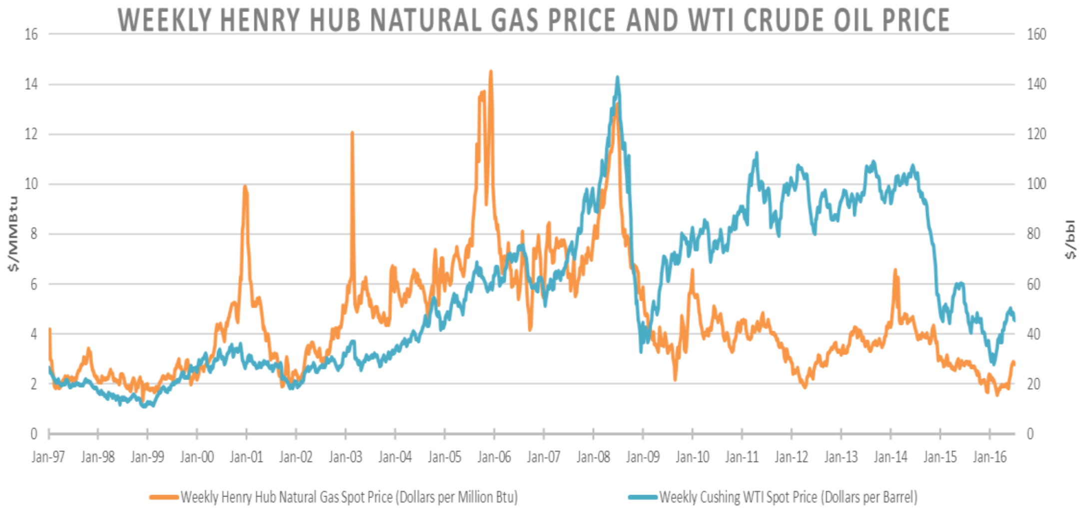
Theoretically, natural gas prices and crude oil prices are anchored together with a long-term relationship. In past, various researchers examined the relationship between the natural gas and different fuels like crude, residual fuel, distillate fuels. According to Medlock, the energy consumers and providers are very much interested in analysing the long-term relationship. The energy market traders are interested to know if there is a fundamental relationship exist between natural gas and crude oil like "rule of thumb". Also, even if there is a relationship, the important question is that is there any tendency of commodities, experienced a deviation from the relationship, to return to the fundamental relationship in a course of time? In past, various studies are done over the relationship of natural gas with other commodities but with different analysis methods.

Energy sources exhibit typical relationship of substitution. This relationship is the driver to many methodologies used in past to analyse the natural gas price with various commodities. Authors like Medlock argued that in energy and power plants, fuel switching between residual fuel and natural gas plays a significant role in the relationship between the natural gas and crude oil. They use a methodology based on fuel substitution. In the paper written by Medlock, author model a relationship between natural gas and crude oil but use the price of residual fuel as an intermediate to link the prince of natural gas and crude oil.

Authors like (Villar & Joutz, 2006) and (Brown & Yücel, 2008) used the price of natural gas and crude oil to analysed the relationship between the two-time series. Both the papers, though, checked for the various rule of thumbs but model the relationship between natural gas and crude oil directly. (Panagiotidis & Rutledge, 2007) used the UK wholesale gas price and Brent oil price to examine the long run relationship between the natural gas and crude oil price. The author also examined the effect of natural gas market development. There are various methodologies used to analysed the relationship between natural gas and crude oil.

The historical data revealed that the natural gas price and crude oil price follow the same trend. The two commodities exhibit a similar pattern. But there were a couple of instants in past when these commodities show a random relative movement.

Figure:1 is a graphical representation of the Henry Hub natural gas and WTI crude oil price series from April'2001 to July'2016. The Henry-Hub natural gas prices and WTI crude oil prices data are taken from the Bloomberg terminal database from Erasmus University. The Henry Hub prices are in dollars per million Btu and WTI crude oil prices are in dollar per barrel. As from the figure:1 it is clear that there were instantiates when the two commodities break their fundamental relationship and move in opposite direction. In 2001, the crude oil prices were almost stable with prices fluctuating between 28$/bbl and 20$/bbl whereas, the natural gas prices fall from almost 6$/MMBtu to less than 2$/MMBtu.



Note: The Henry Hub natural gas price and WTI crude oil price series is a weekly measured in dollar per MMBtu and dollar per barrel respectively, obtained from the Bloomberg Data terminal at Erasmus University.

Figure 1: Weekly Henry-Hub natural gas price and WTI crude oil price

Source: Author using data

During 2002-2004, WTI crude oil prices were increasing with slow, steady pace, whereas the natural gas prices observed a sudden and steep rise in early 2003. But this increase in price was sudden shock only, and prices fall back within a month. But the natural gas prices were more volatile and price series, as seen from the graph as well, were not stable. In 2005, the two commodities exhibited the opposite trend with natural gas prices rises to the all-time high of 15$/MMBtu, the WTI crude oil prices were stable at around 60$/bbl. The period of 2007 - 2008, seemed to be the time when the prices are again in the relationship as they were moving together. In 2008, both natural gas and crude oil prices fall as there was economic crisis and world economies were crawling. But after 2008, the relationship again changed, with natural gas prices started to fall and crude oil prices began to rise. Crude oil market saw continuous upward trend whereas natural gas market found its rock bottom with prices falls to the all-time low of 1.83$/MMBtu. There were several instances in 2011, 2012 and 2014 when commodity prices of natural gas and crude oil exhibit different behaviour.

To evaluate the two commodities, in the figure:2, we plotted the price ratio of two series over a period where price ratio is taken weekly between Henry Hub natural gas price and WTI crude oil price. The average ratio over a year is also plotted in the figure:2. It is clear that the ratio between the prices of two commodities is not constant and evolved over the time. The average ratio from 2001 to 2007 was around 10 and from the period of 2008 till 2010 it was between 10 and 20. During these periods, there were few weeks when there were sudden shocks, which leads to random rise and fall in two commodities price ratio. In the year of 2011, the ratio jump to 25 and within few months it reached to value more than 35. After 2012, there is the continuous drop in price ratio and for the last couple of years, the ratio is around 20.

The varying ratio over a period of 16 years clearly indicates that the Henry Hub natural gas prices and WTI crude oil prices, though exhibit fundamental relationship, the relationship is not stable and changes over the time.

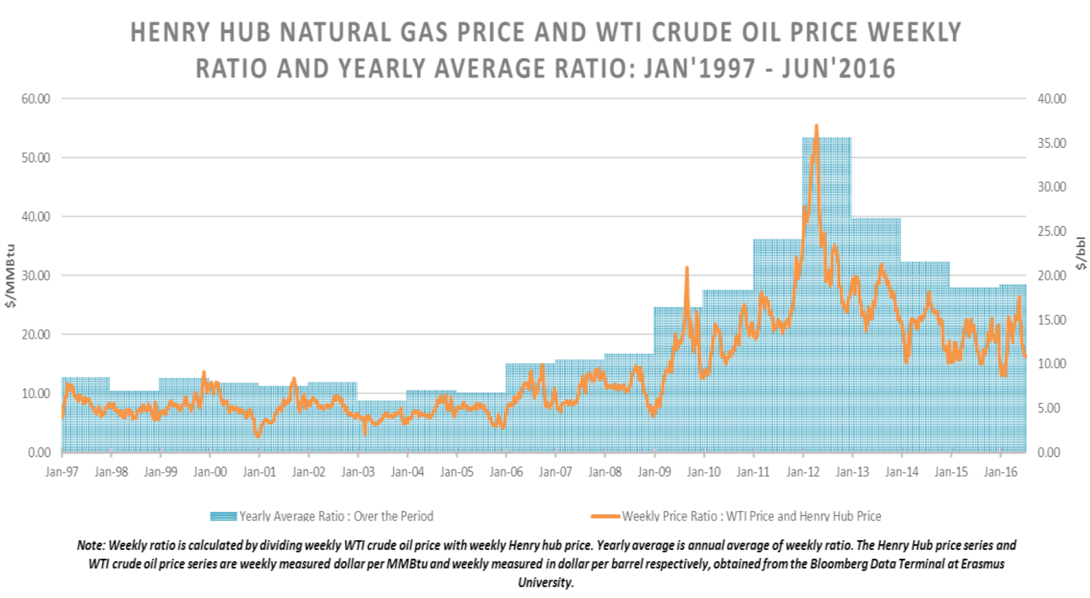


Figure 2: Natural gas price at Henry-Hub and WTI crude oil price weekly ratio and annual average

Source: Author using data

The price of natural gas is volume weighted average in $/mmBtu for delivery at Henry Hub, Louisiana. The local supply and demand of North America have a high impact on these prices. Also, the price of natural gas can be fluctuated by regional supply and demand, globally. So, the price of natural gas serves North America marketplace, but also influenced by other different local supply and demand, which move significantly apart at times. Whereas, the crude oil price is the arithmetic average in $/bbl. For West Texas Intermediate (WTI) crude oil traded at Cushing, Oklahoma.

Though WTI is a particular type of crude produced in this region, it is the benchmark for the global crude oil price. The global supply and demand, have a significant impact on WTI price. WTI signifies the global crude oil market with other crude oil indexes like Blend and Japan crude cocktail (JCC). So, the two-time series, natural gas and crude oil are depending on the different set of marketplaces, and this makes the relationship more complicated. There is always discussion about the general rule for natural gas prices and crude oil prices. Based on the usage in energy and power sector, the two commodities are expected to follow a rule of thumbs.

3.1 "Rule of thumb"

For many years there exist a rule of thumb between the price of natural gas and crude oil, it is called 10-to-1 rule as seen in the figure:3. According to this rule, the price of crude oil is ten times the price of natural gas. This rule is examined and discussed in past in various studies. As mentioned in (Brown & Yücel, 2008), this rule is not a derived state but rather observed from the data. The price series of natural gas and crude oil in 1990's exhibit and follow this rule.

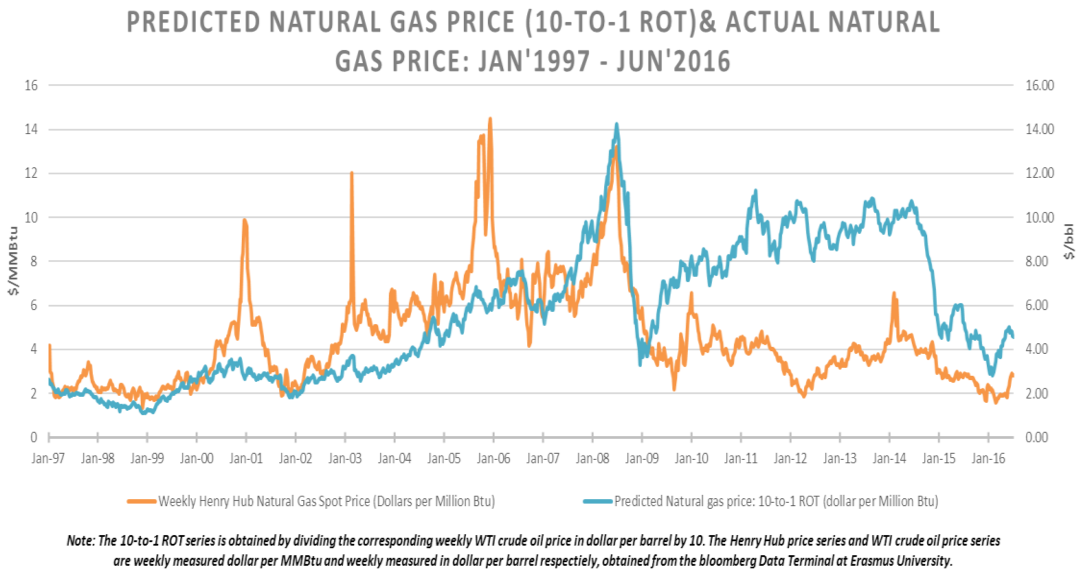


Figure 3: Comparison between Predicted price using 10-to-1 rule and Actual Natural gas price at Henry-Hub

Source: Author using data

The origin of this rule was simply from the data. Since 1997, data shows that the WTI crude oil price and Natural gas price have price ratio of approximately 10:1. As seen from the figure:3, the 10:1 ratio hold till 2009. Apparently, during this period the long run relationship between WTI crude oil price and natural gas price shows a relationship which has a characteristic of 10:1 ratio in price. But in later years, after 2009. This rule does not hold, and price ratio increased far above 10:1. The figure, illustrates the predicted price of natural gas for the corresponding WTI price using 10:1 rule. The 10:1 rule clearly explain the price relationship of two commodities from 1997 till 2009. But after that, the prices of natural gas fell far below than predicted price using the 10:1 rule. In last six years, it is clear that the 10:1 rule of thumb is not a rule which can predict the price relationship between natural gas and crude oil.

From early 2000, the general perception of oil and gas industry regarding the price relationship was shaken due to the inability of the 10-to-1 rule of thumbs to explain the price relationship of natural gas and crude oil. Many argued that as both commodities are energy carriers, the energy content of commodities should be the deciding factor for price relationship. The energy content of a barrel of crude is equivalent to the energy content of 5.825 MMBtu of natural gas. This lead to the rise of another rule of thumb, 6-to-1, where natural gas prices are expected to be onesixth crude oil prices at any given time. Therefore, according to this rule, a price of $60/bbl. of crude oil predict natural gas price to be $10/MMBtu.

Figure:4 illustrates the predicted prices of natural gas for the corresponding WTI crude oil prices using 6:1 ratio. From the figure, it is clear that this rule of thumb, to a much extent, successfully predict the price of natural gas in the interval of 1999 to 2006. During 1998-2005, the average ratio between the two commodities prices was 7.6 with a range of 2.68 to 13.68. After 2005, the prices of crude oil rises continuously, whereas the prices of natural gas were a fluctuating. Since 2005, the 6-to-1 "rule of thumb" never able to capture the price relationship of natural gas and crude oil and shows no promising future as a rule of thumbs.

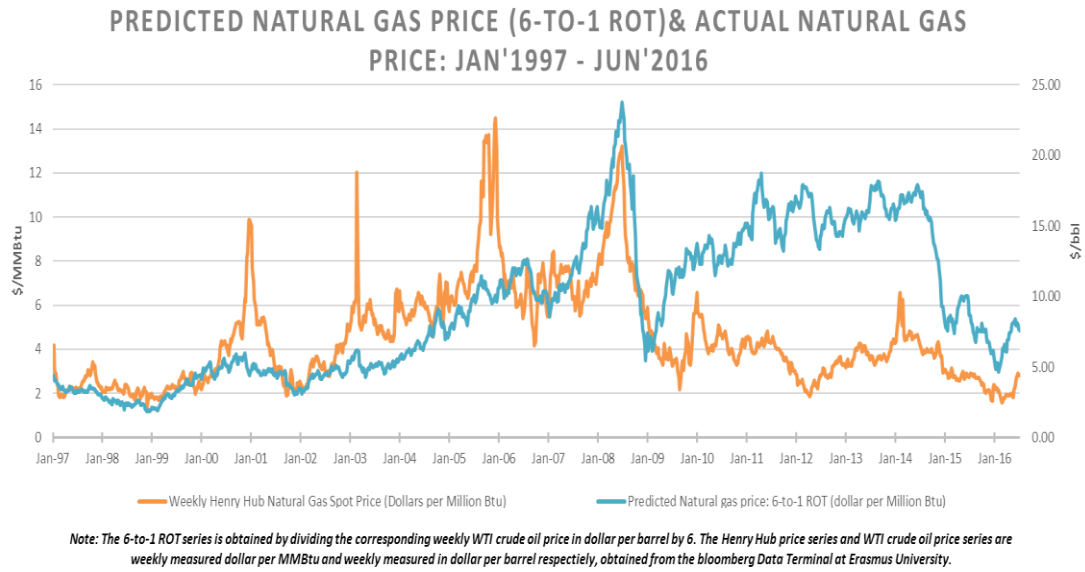


Figure 4: Comparison between Predicted price using 6-to-1 rule and Actual Natural gas price at Henry-Hub

Source: Author using data

Clearly, from the data, the ratio rule of 10-to-1 and 6-to-1 are incapable of capturing the relationship between the natural gas and crude oil and do not qualify as a "Rule of Thumb". In past, several authors challenge this assumption that the energy equivalent of commodities can give the best possible relationship between the prices. (Adelman & Watkins, 1996) and (Smith, 2004) argued that the cost of production, exploration, transportation and usage is different for both the commodities. Also, the two commodities serve diverse portfolio. So, the claims were made that price relationship does not solely depend on the energy equivalent and other factors significantly affect the price relationship. And historical data sufficiently supports these theories. Apparently, the actual prices are showing no or very less resemble to "rule of thumb".

3.2 Burner Tip party rule

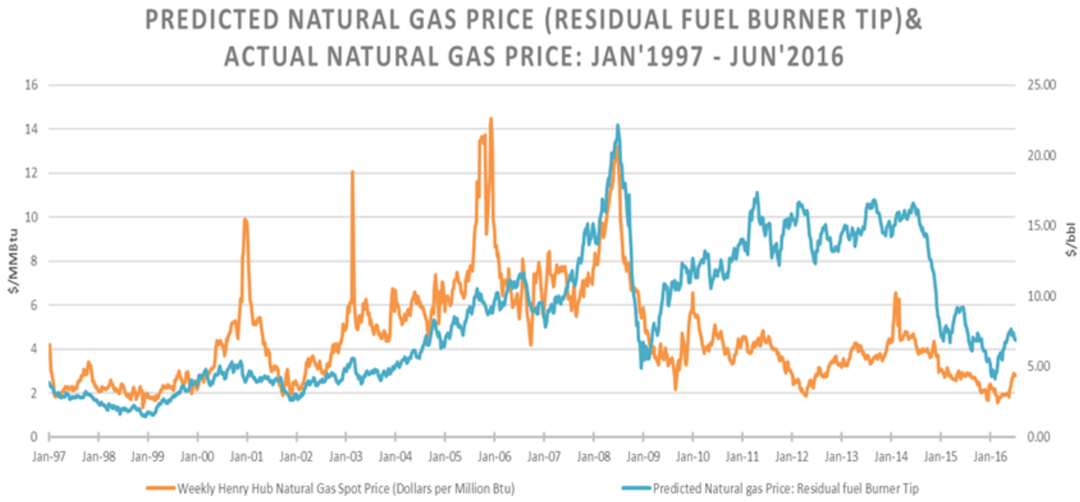
The theory of using intermediate commodity to link the natural gas and crude oil also prevailed in past. Authors clubbed the concept of fuel substitution and the energy equivalent and model the relationship between natural gas and crude oil price. (Brown & Yücel, 2008) also, used this concept and model the relationship between natural gas price and WTI crude oil price. In this paper, the author focused on the two different commodities residual fuel and distillate fuel and their competition with natural gas. In both the cases, the transportation cost includes in the calculation which did not appear in the 6-to-1 rule of thumb. The marginal difference between the transportation cost of two commodities, residual fuel and natural gas in one case and distillate fuel and natural gas in another case is used. Both the cases then translate to model the relationship between the natural gas and WTI crude oil price. Burner time is the process where the energy source is burnt to generate heat. The burning of the source of energy is marked as an end of a process which includes exploration, extraction, pumping, transportation. and delivery.

The burner tip parity rule is based on the concept that when there is the possibility of substitution between energy sources, substitution provides price competitiveness at the point of usage of energy source. In this case, for usage of fuel at burner time, the substitution between natural gas and residual fuel provide a competitive price structure. In this rule of burner tip parity, the price of crude oil is converted in petroleum product price and then using energy equivalent and marginal differential of transportation factor, relate the price at the major trading hub like henry hub. In the past, studies like (Hartley, Medlock Iii, Rosthal, & Medlock, 2008a), emphasis on the importance of the substitution between natural gas and petroleum product.

The concept of burner tip parity rule for natural gas is highlighted by (Barron & Brown, 1986) in their study of assessing the market for natural gas. For example, in the case of residual fuel and natural gas, the competitiveness of residual fuel is used through burner tip parity rule. In this rule, prices are linked using three factors. first, the energy content of a barrel of residual fuel. Second, the typical long-run price relationship of residual fuel and WTI crude oil and finally the marginal difference in transportation cost of two commodities. According to (Brown & Yücel, 2008), a barrel (bbl) of residual fuel has the heat content of 6.287 mmBtu. Typically, the price of residual fuel is 85 percent of WTI price. The transportation cost of natural gas is valued within $ 0.1-1.10 per mmBtu. According to Brown and Yucel (2006), from the data of last 15 years, the marginal differential for transportation cost for natural gas is $0.25 per mmBtu more than residual fuel. With these factors, as (Brown & Yücel, 2008) model the equation, from data of weekly Distillate oil price, weekly WTI price and transportation factor of -0.25, we drive equation for Distillate fuel burner tip parity for the natural gas price.

$$P_{HH,t} = -0.25 + 0.1570 P_{WTI,t}$$

Where, $P_{HH,t}$ is the price of natural gas at henry hub at time “t”. $P_{WTI,t}$ is the price of West Texas Intermediate crude oil price at time “t”. With WTI price at $ 45 per barrel, from the burner-tip parity rule, the price of natural gas at henry hub should be $ 5.7 per mmBtu.



Note: The Residual fuel burner tip series is obtained by accounting for the energy content of a barrel of residual fuel, the long-run relationship between prices of residual fuel and WTI crude oil, and the cost differntial in transportation of natural gas and oil. The Henry Hub price series are weekly measured in dollar per MMBtu, and WTI crude oil price series and the Residual fuel price series are weekly measured in dollar per barrel, obtained from Bloomberg Data Terminal at Erasmus University.

Figure 5: Comparison between Predicted price using Burner tip residual fuel rule and Actual Natural gas price at Henry-Hub

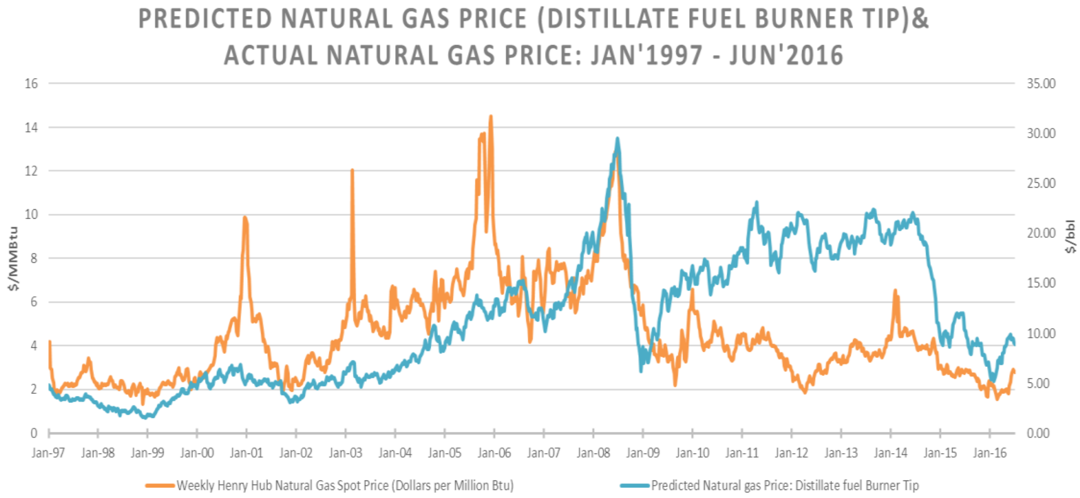
Source: Author using data

In figure:5, the predicted prices of natural gas from the residual burner tip rule is plotted and compared with the actual Henry Hub prices. From the data, it is visible that, till early 2009, the predicted price series using residual burner tip, act as a lower boundary of actual price series. The actual prices at any given time are in most of the cases are above the predicted prices. But this trend changed since the latter half of 2009 when the actual prices of natural gas never attained the value more than the predicted prices by residual fuel burner tip rule. In 2006, the prices of crude oil increased and crossed the value of 50$/bbl. But the natural gas prices also rose with almost same rate, and the two price series move closely. But the same price increase was again observed after crises of 2008, the crude oil prices rose and crossed the 50$/bbl. But this time, the natural gas does not show the same promising results. After 2008, the residual burner tip rule was unable to explain the prices of natural gas and the rule breaks.

As compared to Residual fuel, Distillate fuel is extracted down the line from crude oil and is priced higher than residual fuel. (Brown & Yücel, 2008) examined the data and concluded that distillate fuel is price about 1.24 times more than WTI crude oil price. The heat content of distillate fuel is 5.825 mmBtu per barrel. The marginal differential in transportation is $ 0.80 per mmBtu more for natural gas. With these factors, as (Brown & Yücel, 2008) model the equation, from data of weekly residual oil price, weekly WTI price and transportation factor of -0.80, we drive equation for Distillate fuel burner tip parity for the natural gas price.

$$P_{HH,t} = -0.80 + 0.213 P_{WTI,t}$$

Where, $P_{HH,t}$ is the price of natural gas at henry hub at time "t". $P_{WTI,t}$ is the price of West Texas Intermediate crude oil price at time "t". With WTI price at $ 45 per barrel, from the burner-tip parity rule, the price of natural gas at henry hub should be $ 8.47 per mmBtu.



Note: The Distillate fuel burner tip series is obtained by accounting for the energy content of a barrel of Distillate fuel, the long-run relationship between prices of Distillate fuel and WTI crude oil, and the cost differntial in transportation of natural gas and oil. The Henry Hub price series are weekly measured in dollar per MMBtu, and WTI crude oil price series and the Distillate fuel price series are weekly measured in dollar per barrel, obtained from Bloomberg Data Terminal at Erasmus University.

Figure 6: Comparison between Predicted price using Burner tip Distillate fuel rule and Actual Natural gas price at Henry-Hub

Source: Author using data

Figure:6 illustrated the predicted prices of natural gas from the distillate burner tip rule and compared with the actual Henry Hub prices. The results are similar to the residual fuel burner tip rule. In the case of distillate burner tip methodology, the predicted prices stay much lower to the actual price as compared to residual fuel burner tip method. Also, the distillate burner tip method fails to explain the prices of natural gas after 2008 crises when predicted prices and actual prices move in opposite direction. But as in the case of residual fuel, the two price series again start to converge in after 2013, when the prices of crude oil start to fall. In recent year, the data shows that the distillate burner tip methodology predicts the price much closer than compared to residual fuel. But still the rule break in 2009, and till now the this "rule" does not explain the actual price of natural gas. From the data, we can see that both the burner tip rules are not capable enough to explain the volatility in prices of two commodities. On analysis of burner tip rules, the residual burner tip method is more fitted with data than the distillate burner tip method in capturing the commodity price.

Moreover, it is evident from the figure:5 and figure:6 that when crude oil prices and in turn residual fuel and distillate fuel prices are in higher range, more than 50$/bbl, the burner tip rules are not capable enough to capture the explain natural gas prices. As argued earlier that the crude oil prices crossed the 50$/bbl barrier in 2006, the natural gas increase and followed the same pattern. But the only difference with the rise of prices in 2006 and 2009, were that in 2006, the market had continuous growth since 2000, and the price trajectory was in continuous growth path with few fluctuations. But in 2009, the market starts to recover from the crises of 2008, and the prices of crude oil increased because of the rebuilding of the world economy. But the crude oil prices again fall in 2014 and the gap between predicted prices by burner tip and actual prices reduced.

There is another argument that the natural gas market is more of regional based, and the crude oil market is much more mature and is a global market. The local, regional shocks and variation create volatility in the natural gas market, whereas the global economics crises and shocks have an impact on crude oil prices. Therefore, the natural gas market incorporates the regional variables, making it volatile and only more sophisticated model with local variables can explain the price fluctuation of natural gas. The rules of thumb of burner tip using heat equivalent and transportation cost incorporated in the model do not describe the actual natural gas price series. There is high unexplained volatility and rule of thumb do not take into account of this instability.

Figure: 7 illustrates the comparison between predicted price using various rules and actual Henry Hub natural gas prices. Previously, in the study by (Ramberg & Pasrsons, 2010), the author analyses the error across time and check, out of four rule of thumbs, which one is closest to explain the actual natural gas price. In figure:7 it is clear that residual burner-tip parity rule is closest to actual natural gas price, with an average error of -0.090 (Ramberg & Pasrsons, 2010). The rule of thumb of 10-to-1 is second best series which explain the natural gas price. Also, using historical data predict that the distillate burner tip methodology has the highest difference between predicted price and actual price.

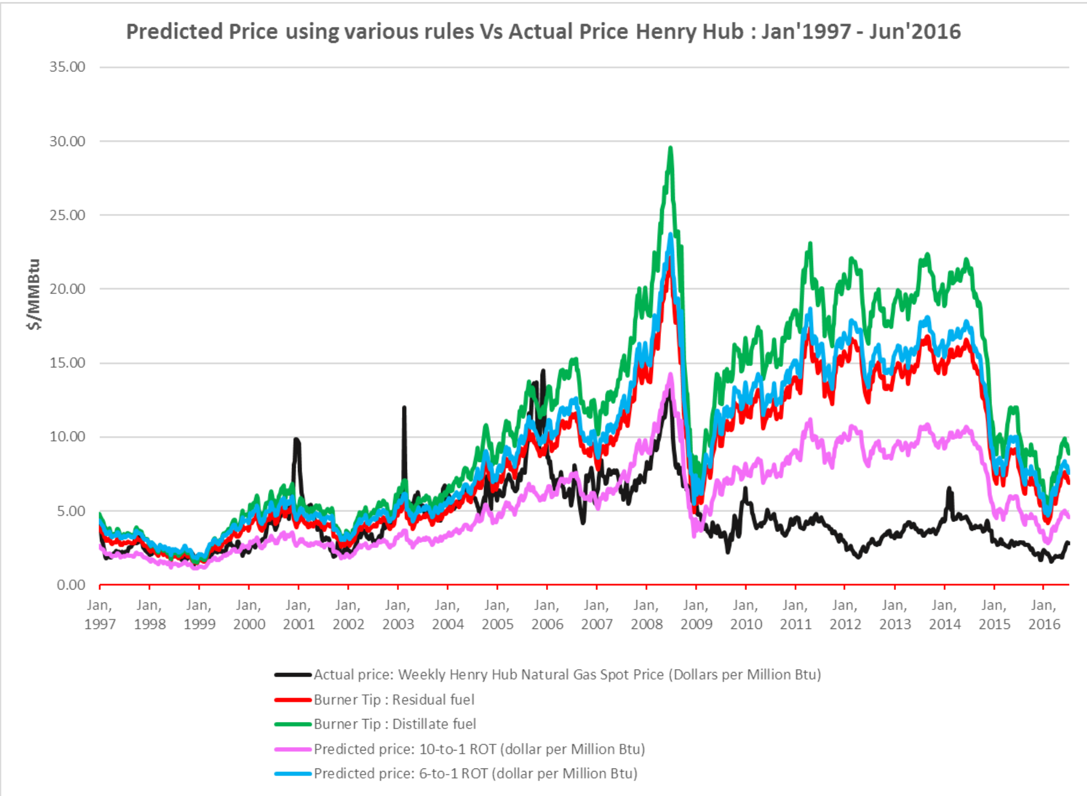


Figure 7: Comparison between predicted price using various rules and actual Henry-Hub spot prices Jan'97-Jun'16

Source: Author using data

3.3 Conclusion

The "rule of thumbs" or "burner tip parity rule" are unable to capture the natural gas price volatility and especially after 2008, there is significant variation between predicted prices and actual prices. In one or another way, all the rules use crude oil prices to predict the natural gas prices. Prior 2008, the natural gas prices were not exactly but to a great extent predicted by rules. In 2008, prices of crude oil fell from more than 140 to less than 40 $/bbl. After this steep drop in prices, these rules were never able to predict the natural gas prices.

Author propose two logical arguments in favour of failure of these rules in predicting the natural gas prices. First, the market size of two commodities is different. Whereas natural gas is a regional commodity and local shocks influence its prices, crude oil is a global commodity which is susceptible to global crises and shocks. So, the global shock has a high impact on oil prices but comparatively less influence on natural gas prices. Second, as these laws are based solely on crude oil prices, they do not capture the effect of variables other than oil. This is an indication that post-2008, natural gas prices volatility cannot merely be explained by volatility in crude oil prices, even after including the transportation factor.

Basis, of these arguments and historical data, it can be concluded that these rules are not capable enough to explain natural gas prices.

4. Meta-Analysis

This section discusses the previous studies and proposed theories in analysing the natural gas price relationship with other commodities. Also, the chapter talks about the various methodologies used in past and the results of the studies.

4.1 Prior Research

Econometricians showed significant interest in finding a concrete relationship between natural gas price and crude oil price since late 90's. (Serletis & Herbert, 1999) tested a trend and relationship between daily Henry Hub and Transco Zone 6 Natural gas prices, as well as of power and fuel prices. Henry Hub and Transco Zone 6 Natural Gas prices (Transco Zone 6 is a significant segment of the Transco pipeline extending from Northern Virginia to New York City, serving the eastern seaboard). The data for power market for electricity price includes the data from Pennsylvania, New Jersey, Maryland (PJM) power market. Important to know that this power market also serves the same geographical area as the Transco Zone 6. Also, the prices of residual fuel oil used as standard reference prices for oil in the Northeast, is employed in the analysis. The author finds that fuel oil price does not show a significant response to change in Henry Hub or Transco Zone 6 Natural gas price, but Transco Zone 6 Natural Gas price show an adjusting tendency of this price series to fluctuation in fuel oil price series at New York Harbour. The study concludes that several factors link the prices of two commodities. One of the reason is the substitution, fuel oil and natural gas is a primary source of energy in industrial boiler and power generation plants.

But in later studies by (Serletis & Rangel-Ruiz, 2004), author fail to find a long-run relationship between natural gas price and crude oil price after analysing the data of 10 years from 1991 to 2001. They focus mainly on the market structure and past development of the natural gas market. The natural gas and crude oil markets evolve and grow significantly since 1980's. The markets were regulated in their early days and then deregulated. (Serletis & Rangel-Ruiz, 2004) argued that due to deregulation of markets the long-run relationship between natural gas and crude oil prices are poised, and the two commodities are decoupled.

After couple of years, the more detailed study is done by (Villar & Joutz, 2006) over the cointegration of natural gas price and WTI crude oil price by analysing the data of 16 years from 1989 till 2005. They found out the long run cointegration relationship between natural gas and WTI crude oil price, with a positive trend. The study implies that the cointegration relationship is evolving with time and not fixed. The study is also important as other exogenous variables and trend term, are used for accommodating the maximum possible volatility of natural gas price. Variables such as Natural gas storage level, the dummy variable for seasonality, dummy variables for the shocks. The important conclusion drawn from the study is that the crude oil price influences the natural gas price, but the natural gas prices have no significant impact on crude oil price.

However, the inclusion of the trend term in the analysis was the drawback of the model. In the same year, (Bachmeier & Griffin, 2006) carried out the research and found the weak linkage between natural gas price and crude oil price. They argued that irrespective of substitution effect of two commodities, the energy market should be seen, in long-run, as a single market of primary energy. At this moment, researchers start to observe that short-term relationship is different from long term, and earlier research failed to account for the drivers of the short-run relationship between prices of two commodities.

Later, in 2008, (Brown & Yücel, 2008) prove that the cointegration relationship between natural gas and crude oil prices does not hold for short time horizon. There are several factors, highlighted in the research which has the significant impact on the fundamental short-run relationship between natural gas and crude oil prices.

In the long run, they found out that there is cointegration between the price of natural gas and WTI crude oil price. In this paper, the author uses Error Correction model (ECM) to counter the unexplained trend in prices by taking in account the WTI crude oil price, weather, seasonality, storage and inventory level and any disturbance in the production process. These variables explain the natural gas price series more efficiently. They found out that weather and inventory levels have the significant impact on the fundamental relationship as these variables tends to divert the natural gas price away from the primary relationship. Also, the natural gas price time series shows a characteristic to adjust as per crude oil prices, but crude oil price series does not show a significant response to natural gas price series.

(Brown & Yücel, 2008), concluded that there exist complicated dynamics in the short run, where other exogenous variables can force the natural gas price to deviate from the fundamental relationship with crude oil. Whereas in the long-term, the relationship is stable.

Where most of the researchers focused on the direct link between natural gas and crude oil prices, (Hartley, Medlock Iii, Rosthal, & Medlock, 2008b) used the different methodology of intermediate commodity pricing. In their research, they proposed a hypothesis that natural gas price is not only a function of crude oil price but also a function of the petroleum products like residual fuel for the primary energy source in the power sector. (Hartley et al., 2008a) argued that natural gas and oil products show a competitive tendency to be an energy source. They used the price of natural gas at Henry Hub, the wholesale price of residual fuel oil and WTI crude oil price to examine a relationship between natural gas price and crude oil price. The weather variables are included to captured the effect of weather on demand and so on price. The Heating degree day (HDD), Cooling degree day (CDD), deviation of HDD and CDD data serve as proxies for weather and seasonality effect. The inventory level data is also used to capture the effect of storage level over demand and hence on price. They found out that the relationship between natural gas price and crude oil price is not direct and residual fuel serves as an intermediate linkage between the two commodities. Additionally, they highlight that crude oil has the significant impact on the natural gas price and residual fuel price and not the vice versa and weather, seasonality and storage level have a significant effect on natural gas price.

4.2 Conclusion

Researchers use different methodologies to analyse the relationship between natural gas prices and crude oil prices. All the previous researches, using variables such as inventory level, heating degree and cooling degree days, and the deviation from the average temperature, conclude a long-term relationship between crude oil and natural gas prices. Also, the variables are responsible for short-run deviation from the fundamental long-term relationship. These variables tend to pull or push the natural gas prices towards or away from the long-run relationship. All the authors confirm that till 2008, there exist a direct or indirect relationship between crude oil and natural gas prices. But since 2008, after the great economic recession, the prices show a random movement and create a suspicion about the breakage of the long-run relationship between the two commodities.

So, there are several questions arise:

Is the relationship between the crude oil and natural gas prices still exists and if the relationship does exist how to define this relationship? What are the variables that can accommodate the volatility in natural gas prices and to what level they can capture the fluctuation in natural gas prices?

5. Factors affecting the natural gas price

As discussed in the earlier chapter, there are variables which are affecting natural gas prices apart from crude oil prices. These variables should be examined and included in the analysis to find a more robust model for the natural gas and crude oil price fundamental long-term relationship. In this chapter, the author discusses the various variables which are affecting the natural gas prices and their significance in evaluating the natural gas prices. Using the data from January 1999 to June 2016, the logged natural gas prices are compared and examined against the various variables. This section proposes a reason to include variables in analysing a natural gas price relationship with crude oil.

5.1 Supply and demand

Economic theory suggests that basic theory of commodity supply and demand link the substitutable commodities. The ability of US industry and power generators to switch fuels makes crude oil and natural gas close substitute. As argued earlier, natural gas and crude oil are primary sources of the energy in power and electricity generation industry; it can be assumed that crude oil and natural gas are linked through supply and demand. Power plants, refineries, factories and other industries switch fuels depending upon the cost of the fuels. This fuel switching phenomenon explains the interdependency of natural gas and crude oil because of competition.
Before 1970, approximately 55% of all natural gas fired power generator in the US were capable of switching fuel to petroleum products. By 1980, this figure reached 71%. These numbers are assumed to increase further due to environmental concern regarding pollution. As natural gas is a clearer form of energy source, the use of natural gas is expected to increase and use of petroleum products is projected to decrease. The change in expected demand for two commodities narrows down the range of opportunity for competition between the fuels in the short run. However, if natural gas prices begin a sustained rise while crude oil prices hold constant, it is possible over the long-run that more fuel switching capability could arise (Costello, Huntingon, & Wilson, 2005).

On the supply side, two commodities, natural gas and crude oil are complementary products, as extraction of crude oil lead to extraction of natural gas as well. The associated natural gas is extracted as a by-product of crude oil. The associated natural gas, a mixture of crude and natural gas, is used in a facility or flared. Though the substantial amount of associated gas flares off, still the 40% of natural gas production is through associated gas well. Therefore, any price increase of crude oil due to growing demand, lead to increase in production of natural gas. Results in a decrease in natural gas prices. Also, there is another type of natural gas, found in the reservoir that only contain natural gas called dry or non-associate gas. So, on one hand, we have associated natural gas, which is complimented relationship, on the contrary, we have dry natural gas which has no petroleum products as a by-product.

In demand side, if the demand of crude oil increases, the producers shift their focus from producing natural gas to producing crude oil. As, resources like labour force and assets like drilling rigs will be focused more towards crude oil, the increase in demand for crude oil will decrease the attention from natural gas and cost of exploring natural gas will rise. This lead to the reduction in natural gas production and exhibits that two commodities, regarding production, are rival in nature.

In general, it can be argued that supply and demand creates a linkage between natural gas and crude oil price in the long term. But from historical data, there were few instances when the prices of two commodities show abnormalities and the movement of prices are in opposite direction as illustrated in the figure:8, where the weekly price of Henry-Hub and WTI weekly crude oil prices are plotted from January 1997 to June 2016.

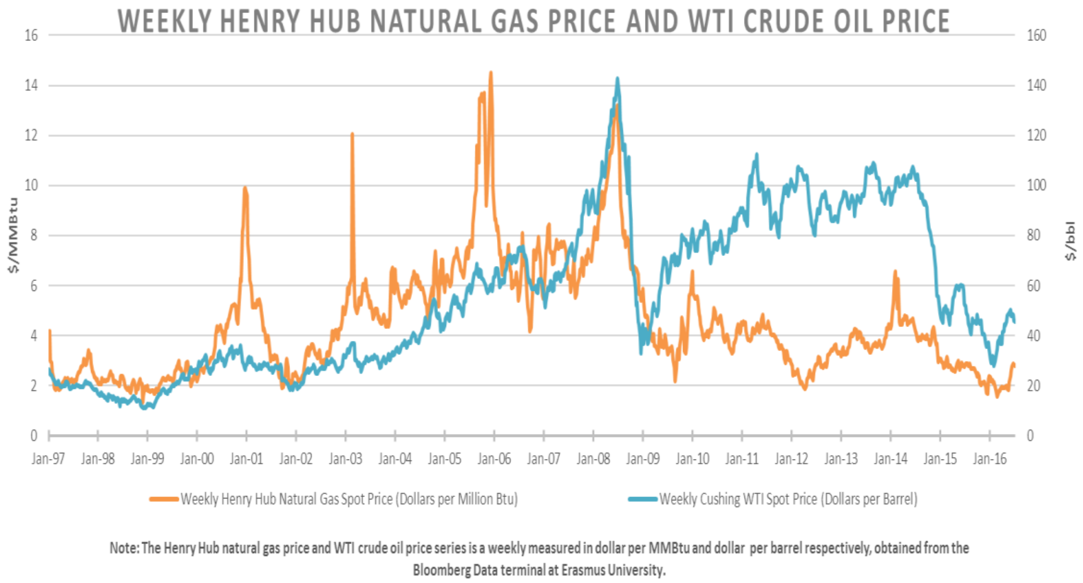


Figure 8: Weekly Henry Hub Natural GAs price and WTI crude oil price

Source: Author using data

Until recently, crude oil and natural gas exploration and recovery were referred to the recovery of crude oil and natural gas through the conventional source and using the traditional techniques. But the recent technological advancements in drilling and recovering technologies along with strong commodity prices levels changed the energy standards and provided a new possible explanation for the divergent trend of natural gas and crude oil price series.
Natural gas is produced in two basic forms, dry and wet. Dry natural gas contains 95% methane (Union Gas, 2011), whereas wet natural gas contains other hydrocarbons such as ethane, propane, butanes and natural gasoline. These hydrocarbons in natural gas are known as Natural gas liquids (NGLs) and in the process of extracting the NGLs, natural gas is a by-product. Various industries use NGLs in the manufacturing of fuel, paints, synthetic rubber, refrigerants, and plastics. These NGLs make the production of wet natural gas more desirable and expensive, leads to increasing the cost of drilling activity. The growing demand of NGLs increases the process of recovering the NGLs, result in increasing the production of natural gas. The increasing supply of natural gas due to the production of NGL could negatively pressurise natural gas price.

(Hartley et al., 2008a), pointed out that the future innovation and technological development will impact the long run relationship between natural gas price and crude oil price to an extent that simple time trend will not be able to explain the relationship. All the relevant studies so far, never account the impact of the recent development.

The development of Shale gas is an example of one these technological advances in the extraction of unconventional natural gas.
In conventional vertical wells, the fuels pass through porous rocks and get accumulated in larger reservoirs. The vertical well is drilled directly into these reservoirs to collect natural gas and crude oil. The efficiency of vertical wells is very in the case of accumulation of crude oil and natural gas in the form of reservoirs because of huge quantity of crude oil and natural gas. Whereas in the case of fine-grained sediment rocks, shale rocks, the crude and natural gas don't pass through them, result in no or small accumulation of crude and natural gas. Therefore, the vertical drilling through shale rocks are inefficient and yield the subtle amount of crude oil and natural gas.

Shale reservoirs required a horizontal drilling. First, a bore is drilled to 5000 meters vertically. When drill enters the shale rocks up to a certain depth, horizontal drilling process is carried out. Through this horizontal drilled hole, a high-pressure mixture of mud, water and chemical injected. This high-pressure mixture creates a crack in the shale rocks and leads to crude oil or natural gas to slip out of the shale rocks. This process of injecting high-pressure mixture is known as hydraulic fracturing. With the help advanced technology, the shale rocks are horizontally drilled and making natural gas and crude oil recovery possible.

5.2 Drilling activities

The supply of natural gas is tremendously boosted by developments in Shale gas production. The strong commodity prices through 2000's drive production of the natural gas and Exploration and Production (E&P) companies focused on shale gas production.

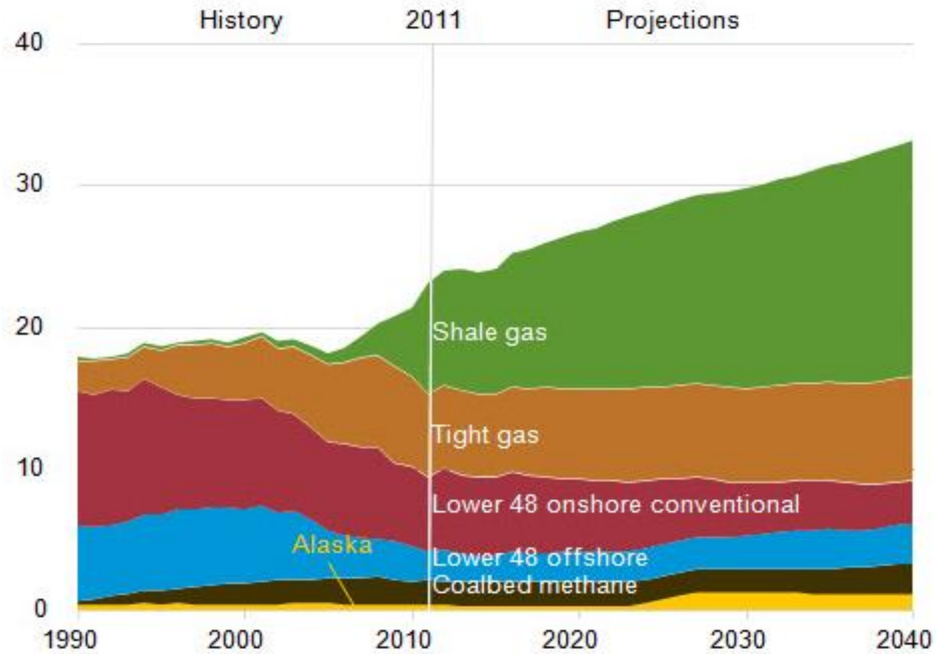


Figure 9: Shale gas market share

Source: (EIA, 2015)

From the figure:9, we can see that the Shale gas market share is increasing continuously since 2004.

The natural gas fairy tale ended in early 2009 when the market dropped down from the price of nearly $12/MMBtu to below $4/MmBtu. In figure:10, the graph is plotted for the monthly operational natural gas rigs and the monthly natural gas price over the period. As observed from the data, the commodity prices drive the exploration and production (E&P) of the commodity. The number of rigs significantly drops as, at the low price, natural gas exploration is not an economically viable option.

Moreover, the natural gas reserves contain only dry natural gas, and the benefits of NGL is not an option in dry natural gas exploration. The prices lead the number of operational rigs over the time. Since 2000, the as the price starts to drop the number of rigs dropped down to below 600. Since, 2001, when natural gas prices again start to increase, the number of rigs starts to grow and apart from sudden momentary shocks in late 2002 and mid of 2005, the number of operational rigs reach significant numbers. In 2008, the market started to crash, and the natural gas prices plunged down. The number of rigs follows the same pattern but with a time lag. Till 2011, the number of rigs follows closely with the price of natural gas, but after that, the prices have no significant impact on the number of operational rigs. This time, it is not the commodity price which drives the number of operational rigs but the development in exploration and production techniques. The development of shale using horizontal drilling techniques has a significant impact on natural gas production. The total natural gas production increased but at the same time the number of rigs decreases. Clearly, the innovation development has significant impact on the price. In the figure, it is clear that with the falling natural gas price, the number of operational rigs all falls.

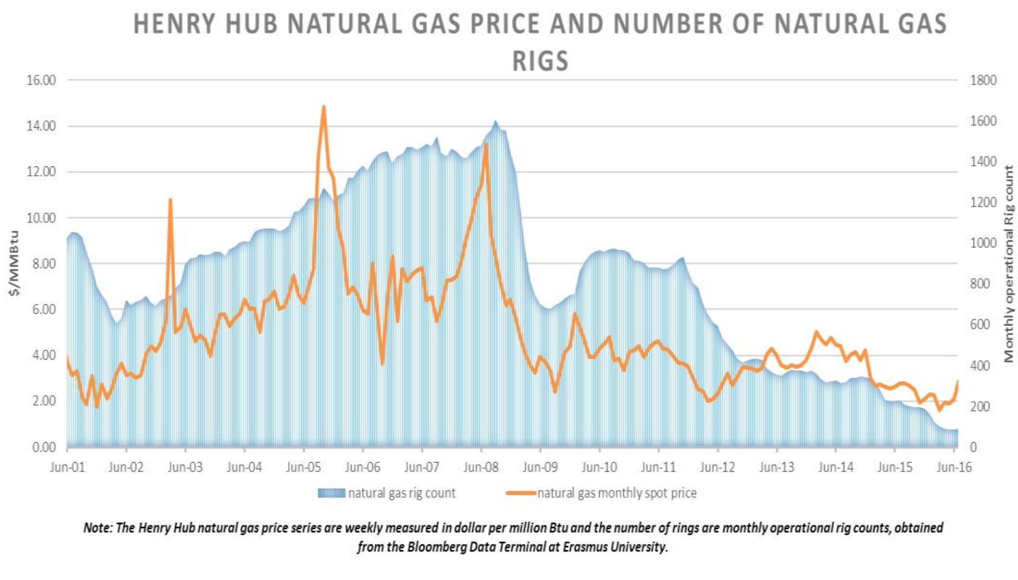


Figure 10: Monthly operational Natural gas rigs and Henry-Hub natural gas spot prices

Source: Author using data

In this aspect, where prices influence the commodity exploration, crude oil shares the same trend with natural gas. In figure:11, the graph is plotted for the monthly operational oil rigs and the monthly WTI crude oil price. Apparently, WTI crude oil price series lead the two series. The number of operational oil rigs increases as the WTI prices increase.

From the data, it is clear that in late 90's and early 2000, the prices of the crude oil show a constant and slightly progressive trends. The prices start to increasing with much faster rate from late 2006 and in one year prices were $140/bbl. But by the end of 2008, prices fall to below $40/bbl. During this period the number of rigs counts drop from 423 to 187. In 2009, the prices started to rebound, and the drillings activities follow the trend. By mid of 2014, the record number of rigs were operational with crude oil price hovering around $100/bbl.

Though the crude oil price was still below the prices of 2008 by $40/bbl, the drilling rates were significantly high as compared to 2008. Few arguments support this phenomenon. First, though the crude oil price is well below the record mark of 2008, the price of $100/bbl was still an economically viable option, and E&P companies invest in oil exploration and production. Moreover, the weak natural gas market, shift the attention of E&P companies away from natural gas and towards crude oil product. But in the next half of the year, crude oil prices fall and so does the number of rigs. The period from 2009 to 2013, saw one of the most significant rises in the number of operational oil rigs. With crude oil prices rise and stay close to the $100/bbl mark, the number of rigs reaches the record number of 1600. But again after the fall in oil price since early 2014, the number of rigs fall tremendously.

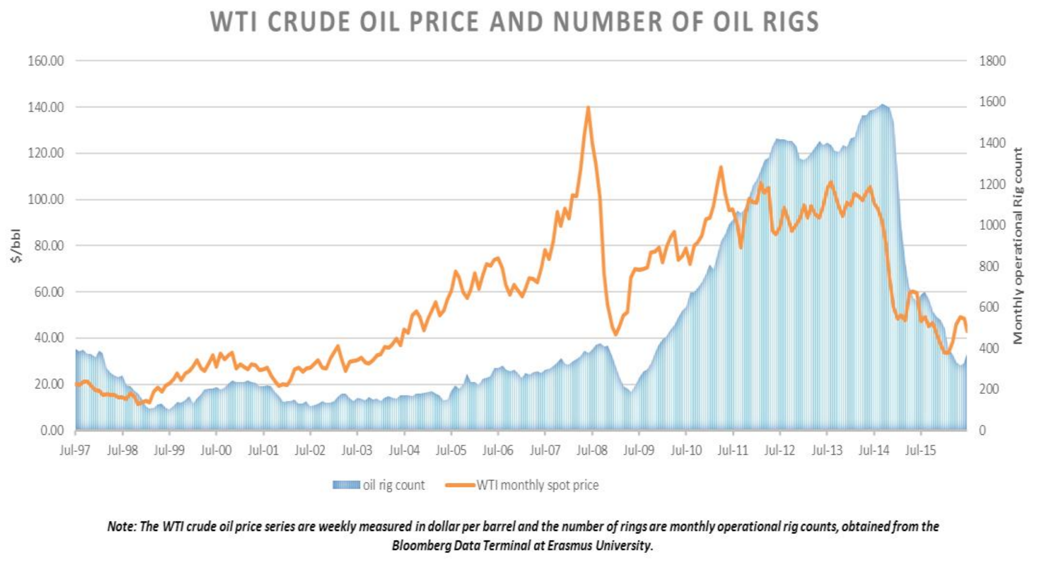


Figure 11:Monthly operational oil rigs and WTI crude oil price

Source: Author using data

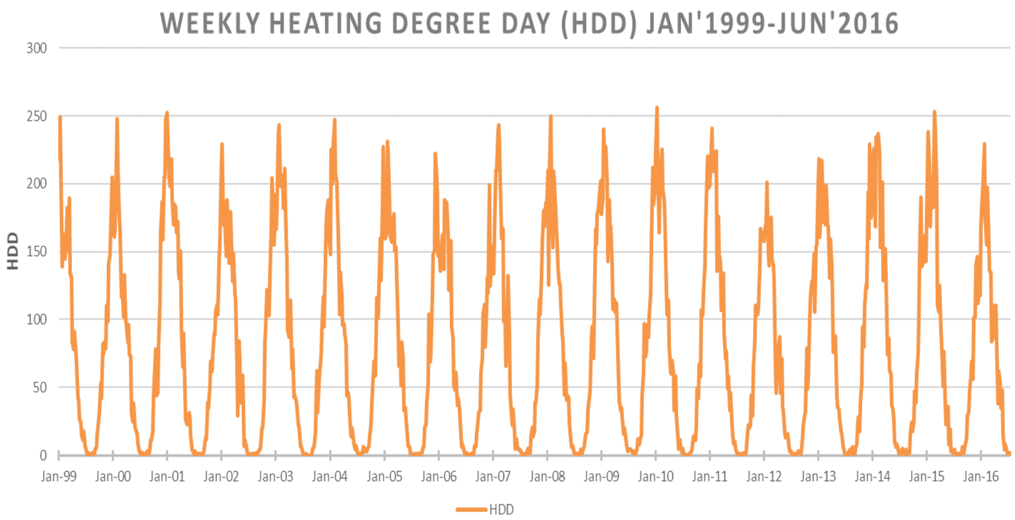
From the data, a theory can be proposed that whenever the price falls, the number of operational rigs count drops. This argument is logical, as the prices fall, the exploration and production (E&P) companies have less incentive to continue exploration and production of crude oil. On the other hand, when the prices are high, E&P companies invest their resources heavily in exploration and production of oil.

The natural gas and the crude oil historical data clearly indicated that the two commodities prices affect the exploration and production activity. Also, if the market of one commodity is weak, the attention of E&P companies shifts to other relevant commodities.

5.3 Weather

The weather patterns significantly affect the natural gas demand. Moreover, the deviation from the expected temperature like heat waves and cold waves in a season drastically fluctuate the demand of natural gas. In residential usage of natural gas, the space heating is the major consumer of energy. The natural gas is used in residential furnaces to provide accommodation heating, water heating, cooking and other miscellaneous uses. As, natural gas is a primary source for the residential heating, sudden deviation of temperature in winters due to cold wave, raise the demand for natural gas and in turn affect the prices of natural gas. So, to analyse the natural gas relationship, we have to keep in mind that weather plays a significant role in demand fluctuation of natural gas.
Similarly, the natural gas is also used in electricity generation power plants. In summers, due to an operation of accommodation air condition, energy demand increases and in turn the natural gas demand increases. With the sudden heat wave in a region, the need for the power increase rapidly and so the demand for natural gas increases.

Apparently, weather plays a significant role in the demand side of natural gas.
To analyse the energy demand due to weather fluctuation, the most common parameters are Heating degree days (HDD) and Cooling Degree days (CDD). Energy sector used these variables to measure the energy demand in a region. The heating requirement in a residential area is directly proportionate with the heating degree days. The heating degree days are defined as the average temperature above the base temperature. The base temperature is taken as 65 degrees Fahrenheit. HDD have used as proxy for the heating requirement on a particular day.

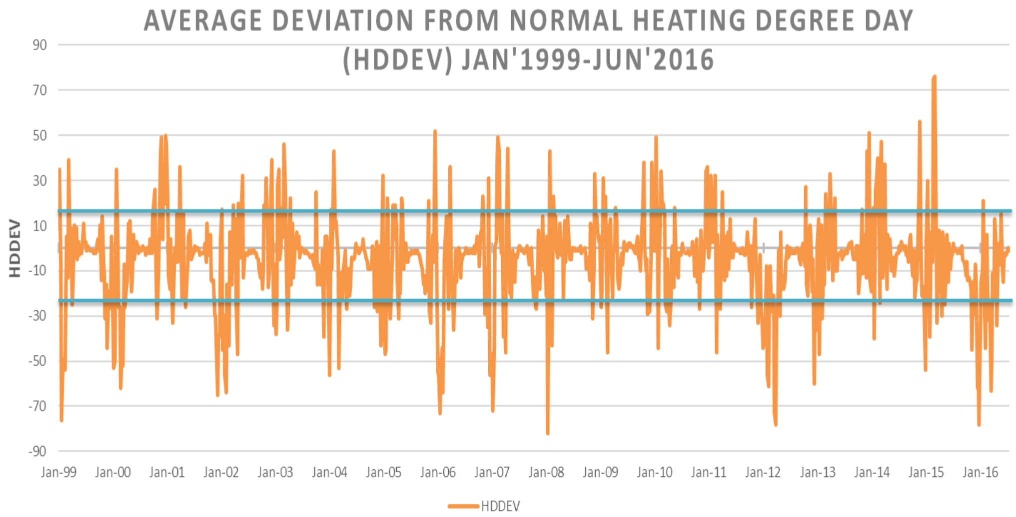


Note: The Heating Degree days is a weekly series collected by the NOAA and weighted by population, HDD defined as the average number of degree below 60 degree Fahrenheit, obtained from the Bloomberg Data terminal at Erasmus University.

Figure 12: Weekly Heating degree day (HDD) January 1999 - June 2016

Source: Author using data

Figure:12 illustrate the Heating degree days (HDD) for the period of January 1999 to June 2016. In winters HDD has the highest value and so does the heating requirement. The HDD data, if follow a general trend provide an expected demand for natural gas. For example, every year, during winter, there is projected increase in natural gas demand and authorities are aware of these growing demand. But, let us assume a particular day in winter is hottest as compared to the same day in last five years, then, there is an increase in power consumption and lead to a sudden increase in natural gas demand. A special day, week or month temperature has deviated from the average and put pressure on natural gas demand. The deviation of Heating degree (HDDEV) days from average increase the demand for natural gas.



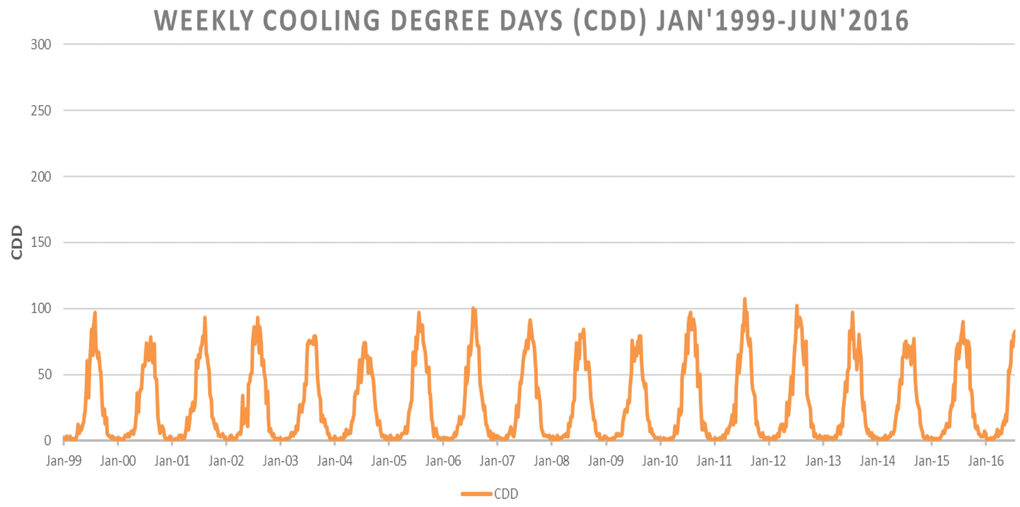
Note: The HDDEV is a weekly series collected by the NOAA and weighted by population, HDDEV difference between average HDD series and normal HDD data series, obtained from the Bloomberg Data terminal at Erasmus University.

Figure 13: Weekly deviation from Normal Heating Degree days (HDDEV) January 1999 - June 2016

Source: Author using data

The figure:13 represent the deviation of HDD from the average temperature. HDDEV is calculated by taking the difference of average HDD data series from normal HDD data series. In the figure one standard deviation above mean and one standard deviation below mean is also plotted.

Similarly, for cooling purpose, cooling degree days (CDD) are used as a benchmark to evaluate the cooling demand of the region. The cooling degree days are defined as the temperature above the base temperature. The vase temperature is 65 degrees Fahrenheit.



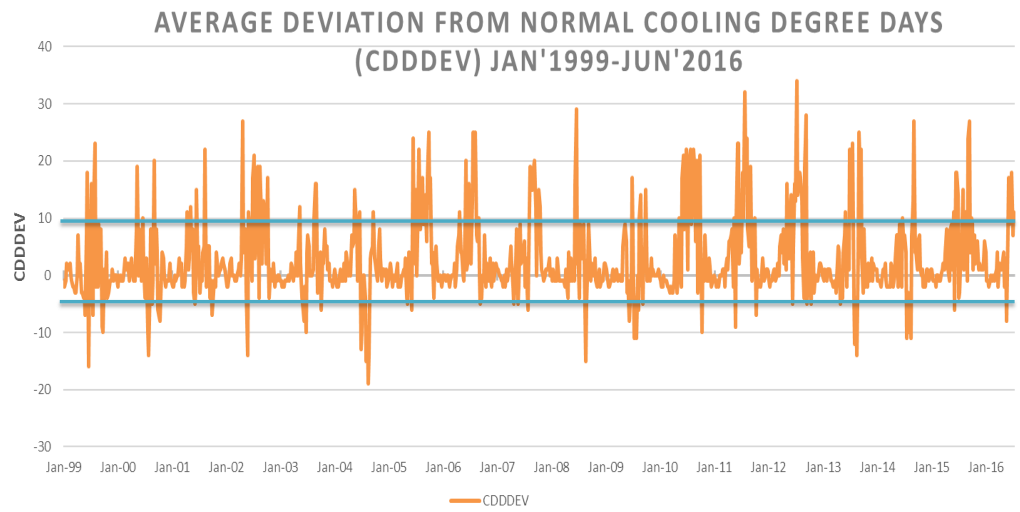
Note: The Cooling Degree days is a weekly series collected by the NOAA and weighted by population, CDD defined as the average number of degree above 60 degree Fahrenheit, obtained from the Bloomberg Data terminal at Erasmus University.

Figure 14: Weekly Cooling degree day (CDD) January 1999 - June 2016

Source: Author using data

The figure: 14 illustrate the CDD over the period of January 1999 to June 2016. CDD value peaked in the summer and had almost zero value in winters.

As discussed for the unexpected cooling day due to cold wave, there are days, week or months which are unexpectedly hotter than the same day, week or month of the previous five years. This unexpected high temperature leads to increase in energy demand and in turn the natural gas demand. To capture the effect of this sudden rise in temperature we draw a graph, where the deviation of temperature from normal CDD is plotted. The figure:15 illustrate the deviation of cooling degree days (CDDDEV) from average CDD. In the figure one standard deviation above mean and one standard deviation below mean is also plotted.



Note: The CDDDEV is a weekly series collected by the NOAA and weighted by population, CDDDEV is difference between average CDD series and normal CDD data series, obtained from the Bloomberg Data terminal at Erasmus University.

Figure 15: Weekly deviation from Normal Cooling Degree days (CDDDEV) January 1999 - June 2016

Source: Author using data

5.4 Inventory Level

The natural gas production witness continuous growth in last 15 years. The high natural gas commodity price, an increase in usage of gas liquid (LGL), and technological advancements in recovering shale gas has pushed the E&P companies to strive for increasing natural gas production. Where natural gas production shows a promising upward trend, natural gas consumption remains relatively stable throughout the period. The figure:16, graphically represents the natural gas production and consumption from 1997 to 2015. As the production is increasing continuously, and consumption is relatively constant, leads to the availability of excess natural gas. This excess natural gas exerts a downward force on natural gas commodity price.

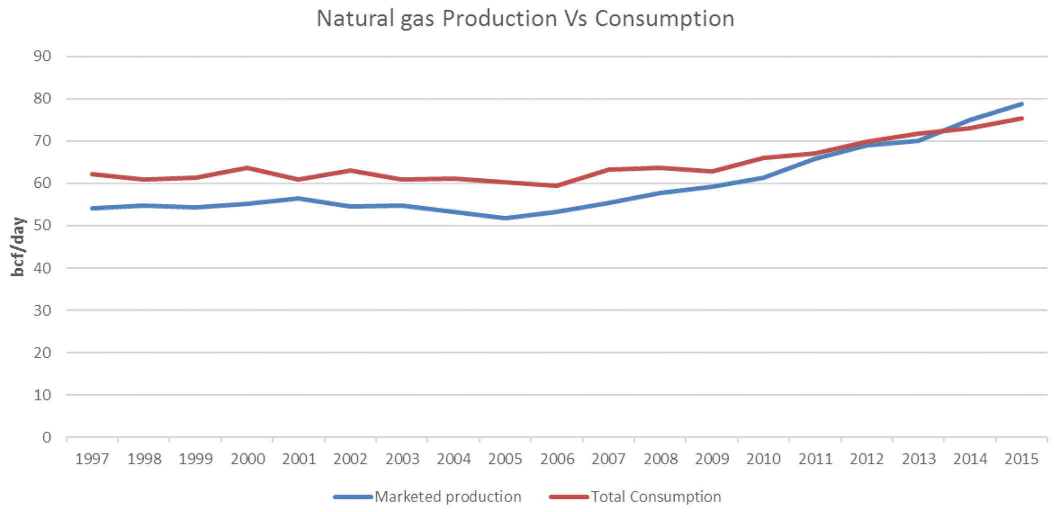


Figure 16:Natural gas consumption and production

Source: Author using data

The excess supply can be stored in various facilities, mainly underground. The deserted rigs and wells are used to store excess natural gas and when demand increase, the natural gas is extracted. The figure:17, shows the weekly natural gas storage levels since 1999 in billion cubic feet (Bcf). The historical data reveals that the storage levels are continuously increasing and in 2016, the gas storage level reached its peak much earlier as compared to last few years. These levels are an indicator of the natural gas supply and demand. Normally, the natural gas consumption increase in winter months in the US, due to heating and decrease in summer months. This imbalance of consumption leads to injection of natural gas in storage facilities in summer months and retrieve back in winter months.

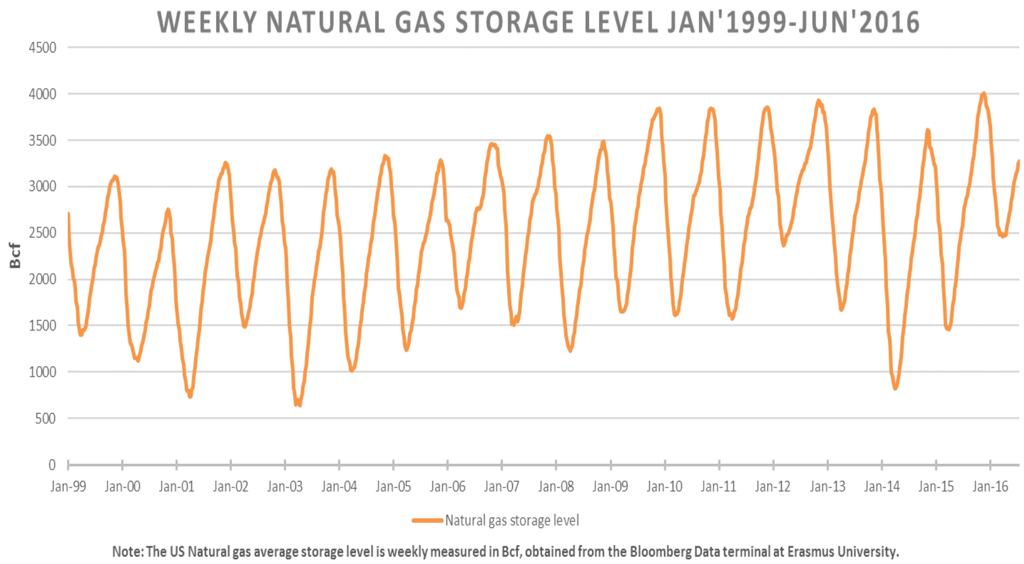


Figure 17: Weekly US natural gas storage level

Source: Author using data

The figure: 18, the graph is a plot to compare the natural gas price at Henry Hub and the inventory level at the same time. The inventory level is measured with the variable STORDIFF. This variable is the difference of weekly storage level and the 5-year running average. All the values of STORDIFF are in billion cubic feet per week (Bcf/week). The average storage level from January'1999 to July'2016 is 2,454 Bcf (2.45 billion mmBtu) on any given week, with the standard deviation of 776 Bcf (798.7 million MMBtu). The average storage differential (STORDIFF) from January'1999 to July'2016 is 134 Bcf (136.9 million MMBtu).

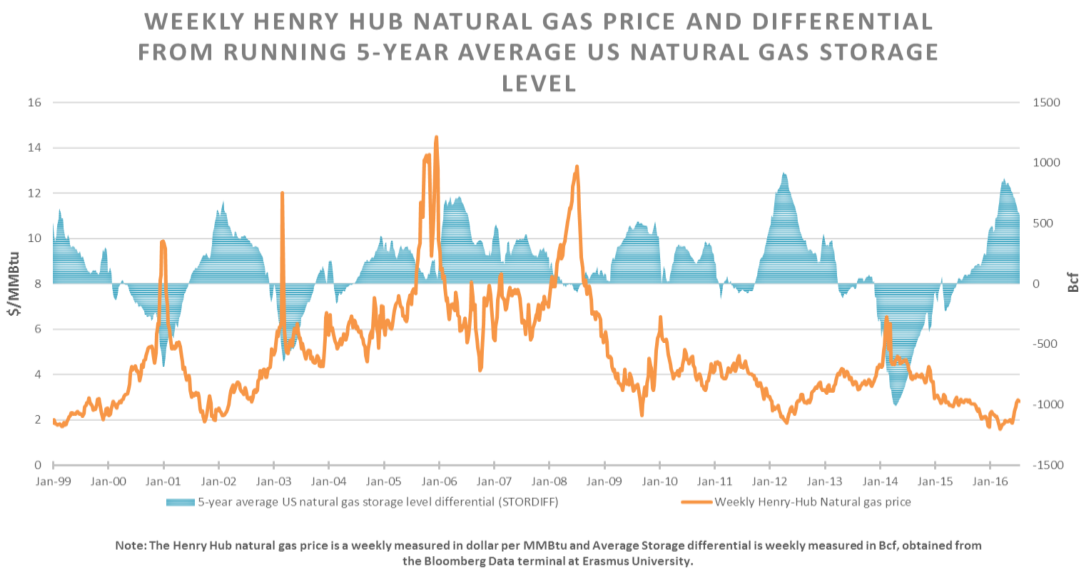


Figure 18: Weekly Henry hub natural gas price and 5-year average US natural gas storage level differential (STORDIFF)

Source: Author using data

5.5 Conclusion

In this research, we include variables such as HDD, CDD, HDDEV, CDDDEV, STORDIFF, and SHUTIN to capture the volatility of natural gas prices. HDD and CDD capture the effect of weather change on the demand for natural gas. As the temperature fall the HDD increases, lead to an increase in the demand for natural gas for heating purpose. Similarly, as temperature increases, CDD increases and causing a rise in the demand for natural gas for electricity generation. The argument is also supported by the historical data as compared in this chapter. To account for deviation in temperature from average; we use HDDEV and CDDDEV. With a cold and hot wave, the temperature fluctuates above and below of average temperature, and this put pressure on the demand for natural gas. From the data, it is evident that the deviation from average has impact on the natural gas prices. From the supply side, we include STORDIFF which is a deviation from average storage level of natural gas. When the natural gas and STORDIFF compared, we found that the prices are inversely proportionate to deviation in storage level. So, STOFDIFF plays an important role in formulating natural gas prices relationship. Also, any disturbance in supply or production also impact the natural gas prices. To measure the impact of disturbance in natural gas production due to a hurricane in Gulf of Mexico, we include SHUTIN variable. It measures the amount of natural gas production reduced due to a storm in the Gulf of Mexico. From historical data, it can visualize that the natural gas prices tend to increase with an increase in SHUTIN quantity.

So, to measure the natural gas price relationship, we believe these variables plays an important role and must be included in analysing the crude oil and natural gas prices relationship.

6. Hypothesis

After theoretically analysing the factors which affect natural gas prices, we can propose a hypothesis to investigate the relationship of natural gas with all the variables. From chapter 3, 4, and 5 we sum-up that along with crude oil prices, there are exogenous variables, which are responsible for the volatility of natural gas prices. We proposed that natural gas supply and demand do not only depend on the rival commodity but also on external factors, it is essential to include these elements in the empirical analysis of natural gas and crude oil price relationship. We propose following two hypotheses:

1. *Natural gas and crude oil price are in long run relationship:*

   As previous research proves that there exists a fundamental relationship between natural gas and crude oil prices, we propose a hypothesis that Henry Hub natural gas and WTI crude oil prices are in a long-run relationship. The long run relationship in times series analysis is also known as cointegration. So, our hypothesis is that Henry-hub natural gas price series and WTI crude oil price series are cointegrated. Also, we assume that as Henry Hub is regional, natural gas price index and WTI is a global benchmark for crude oil prices, crude oil prices influence the natural gas prices and not another way round.

2. *There are other variables that affect natural gas price in short-run apart from crude oil prices:*

   As we discussed earlier, the data shows that the natural gas prices rise and fall at the same time when there is fluctuation in the variables. The variables such as Heating degree days (HDD), Cooling degree days (CDD) which capture the change in natural gas prices due to change in weather shows a symmetrical trend with natural gas prices. Merely visualising the trend of HDD and CDD with Natural gas prices, the two variables showed a positive correlation with Natural gas prices. At the same time, the storage level of natural gas and disturbance in natural gas production also indicate the significant impact on natural gas prices. Visualising the historical data, one can interpret that with an increase in storage level, prices tend to decrease. Regarding the SHUTIN variable, which depicts the disturbance in natural gas production, the prices of natural gas tend to increase with an increase in a disruption in production. All of these variables create a shock and due to these shocks, the natural gas prices fluctuate. But as our first hypothesis predict, there exist a long-run relationship between natural gas prices; these shocks dissipate with time. So, the impact of these exogenous variables is short lived. To sum up our hypothesis, figure 19 illustrate the factors affecting natural gas prices. All the factors need to be included in modelling the natural gas and crude oil price relationship.

Figure:19, illustrate the variables affecting the natural gas prices. Here Henry Hub natural gas spot price is dependent variables and Oil prices, HDD, CDD, HDDEV, CDDDEV, STORDIFF, and SHUTIN are independent variables.

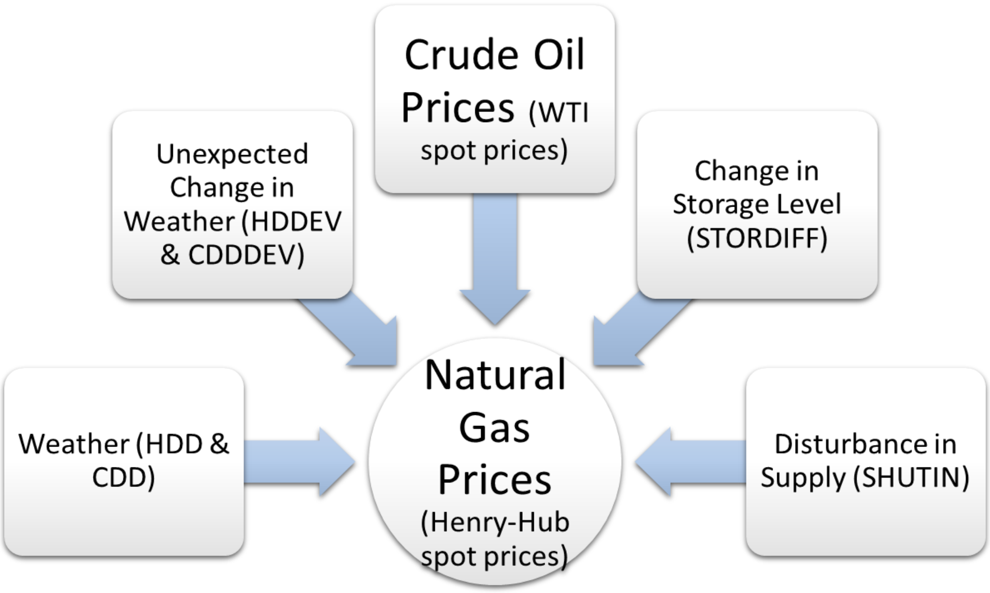


Figure 19: Factors affecting Natural gas prices

Source: Author

To find this long term relationship between natural gas and crude oil prices where natural gas prices and crude oil prices move in the same direction or opposite direction, we use model to find cointegration[1] between natural gas and crude oil prices.

[1] Cointegration is a statistic property of time series which describe a long-run equilibrium or relationship between the nonstationary time series. For example, variable X and Y can float around depending upon their properties, but if these two variables are in a cointegration relationship then, they will always tend to return to long-run equilibrium relationship. Technically, if we have two or more non-stationary time series of order one or higher such that some linear combination of X and Y form a stationary series, then X and Y are said to be cointegrated. Cointegration is a property which examines a long-run relationship between non-stationary time series.

7. Research Methodology and Data

Considering the random behaviour of natural gas time series and crude oil time series in the last couple of years, this chapter aims to introduce the quantitative tools to capture the impact of exogenous variables discussed in Chapter 5, which lead to this random behaviour of prices. These quantitative tools will focus on the cointegration of Henry Hub natural gas price and West Texas Intermediate price, as it will provide results for a clear understanding of the unexplained amount of natural gas price volatility. Also, the quantitative analysis includes exogenous variables to address the unexplained volatility of natural gas price as much as possible.

As this thesis aims to analyse the linkage of natural gas price to WTI crude oil price, it is important to understand the relationship that exists between the natural gas price and various factors that can affect it. The understanding of the fundamental relationship is developed through the analysis of historical statistical data using VECM model, which will identify and characterises the cointegrating relationship(s) between the different time series. Also, the rate at which natural gas push back on long-run equilibrium price is checked by Error correction mechanism.

This chapter aims to introduce the empirical model and data to answer the main research questions.

Summary of methodology:

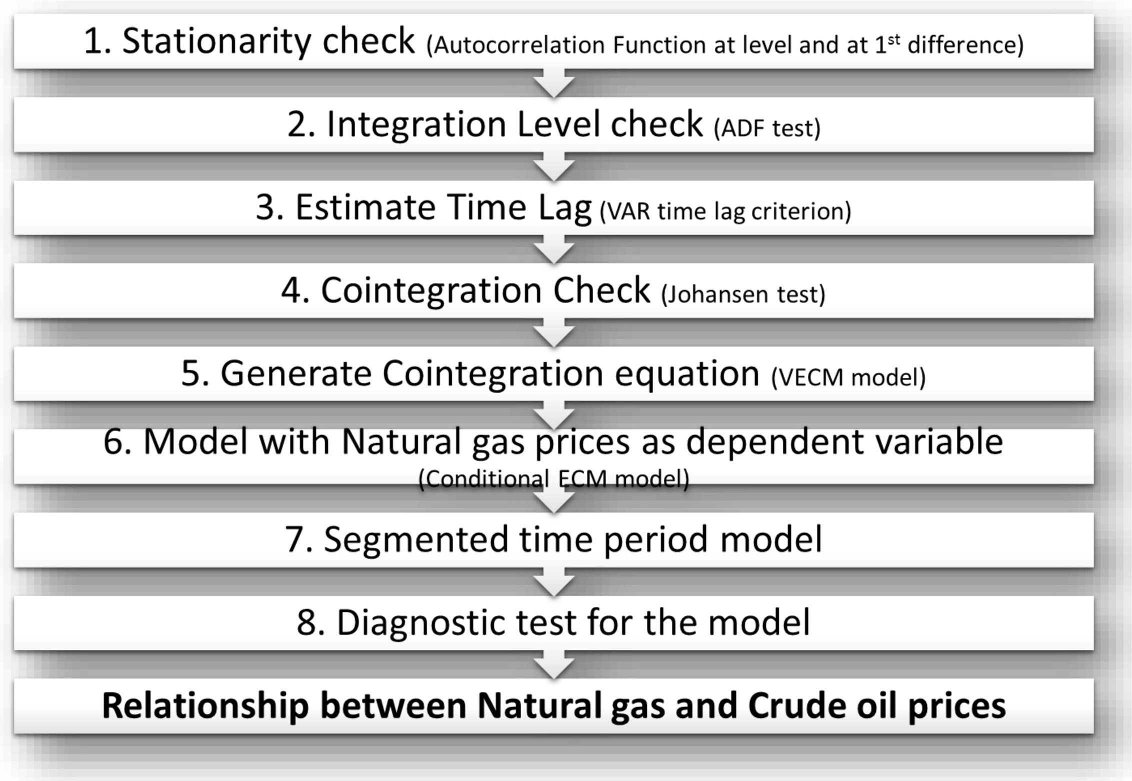


Figure 20: Methodology

Source: Author

Figure 20, illustrate the step by step, the methodology followed in this research. The process of analysing the natural gas prices and crude oil prices relationship is as follows:

1. First, we discuss the basic concept of time series analysis and using our understand of time series; we examine the stationarity of all the dependent and independent variables. We review the logged Henry-Hub natural gas spot prices and WTI crude oil prices. Later we investigate the stationarity of exogenous variables that can affect natural gas price as explained in chapter 5. To check the stationarity of variables, Augmented-dickey Fuller (ADF) test for unit roots and Phillips-Perron unit root test is carried out on the variables.

2. Later, we examine the order of integration of the time-dependent variables (time series). To check the order of integration of variables, Augmented-dickey Fuller (ADF) test for unit roots and Phillips-Perron unit root test is carried out on the variables. We carry out unit root/order of integration test of Logged Natural gas price (ln_HH) and Logged WTI crude oil price (ln_WTI).

3. Once the order of integration of variables is verified, we estimate optimum lag length for the system of variables using VAR lag length criterion test.

4. Later with optimum lag length, Johansen cointegration test is used to determine if there exist a cointegration between the ln_HH and ln_WTI.

5. If the Johansen test indicates cointegration between ln_HH and ln_WTI price series then, Vector error correction model (VECM) is developed to analyse the long-term relationship between Henry Hub natural gas prices and WTI crude Oil prices.

6. Later, we develop Error correction model (ECM) using error correction term from results of Vector error correction model (VECM). The ECM enable us to analyse the impact of WTI crude oil prices and exogenous variables (HDD, HDDEV, CDD, CDDDEV, STORDIFF, and SHUTIN) on Henry Hub natural gas prices.

7. Next, we develop the segmented time-period model to understand the development in the fundamental relationship of natural gas prices and crude oil prices.

8. Finally, we check the fitness of the model using diagnostic tests.

7.1 Stationary and Nonstationary Time series

“Experience with real-world data, however, soon convinces one that both stationarity and Gaussianity are fairy tales invented for the amusement of undergraduates” (Thomson, 1994). In time series analysis, the key aspect is to find out how the observations are distributed over time to distinguish stationary time series from nonstationary time series. The relationship between the observations over time alter the methodological approach to analyse the time series. The concept of analysing if observations are related to each other over time is through analysing the covariance between the elements. Covariance measures the degree of second order variation between the two different elements at different times (Nason, 2006).

Mathematically speaking, the stationary stochastic process is one where for given $t_1, t_2, t_3, \dots, t_l$ the joint statistical distribution of $X_{t1}, X_{t2}, X_{t3} \dots \dots X_{tl}$ is the same as the joint statistical distribution of $X_{t1+T}, X_{t2+T}, X_{t3+T} \dots \dots X_{tl+T}$ for all "l" and "T". A time series is stationary if its mean and all autocovariance are unaffected by change of origin of time (Enders, 2015). It is very important to know the characteristic of time series before selecting the appropriate model. (Box, George, & Jenkins, 1970) famous strategy, aimed to select the appropriate model includes 3 stages. The first stage is identification stage, in which time series is analysed for autocorrelation function and the partial correlation function. The nonstationarity can be addressed by the first difference of time series.

In stationary time series, the shocks by exogenous variables dissipate with time and there are no everlasting effects over the time series (Hamilton, 1994). Whereas in nonstationary time series, shocks by exogenous variables have a permanent effect on the evolution of time series. This can be explained with a mathematical example:

$$y_t = \rho y_{t-1} + e_t$$

Where,
$e_t = random\ variable\ with\ zero\ mean\ and\ standard\ deviation.$

Dependent variable y_t can be solved recursively as a function of y over a period of time and the sum of random variables (e_t) or noise.

$$y_t = \rho^t y_0 + \sum_{i=1}^{t} \rho^{t-i} e_i$$

Depending upon the value of ρ, the series can be stationary or nonstationary. If the value of | ρ | < 1, then over a period of time as *"i"* increases, the effect of shocks to the series dissipates. The series is stationary time series. Also, as the impact of shocks does not persist, with increasing time (t), the variability of autocovariance starts to disappear and converges to constant.

Whereas when, | ρ | = 1, the equation can be reduced to

$$y_t = y_0 + \sum_{i=1}^{t} e_i$$

Here, y_t is the function of the initial value of y (y_0) and sum of all the random shocks on y over the time t ($\sum_{i=1}^{t} e_i$). From the above equation, it can be visualised that the random shocks have a cumulative impact over the variable y. The past random shock, if not removed, have same implication as the random shock at time t. It does not matter when the series experience the shock.
There can be misleading results from the time series analysis if the time series is nonstationary and analysed as stationary. One of the issues with nonstationary time series analysis is the spurious regression. To fit the model, econometrician use value of the coefficient of multiple correlations R_2 as high as possible. With high R_2 value

and extremely low Durbin-Watson statistic, the model has a high degree of fit. The problem with spurious regression is that it shows high R_2 value and low Durbin-Watson statistic value, but still model has no relevance. As explained by Granger & Newbold (1974), the presence of autocorrelation structure in regression equation can lead to the formulation of the model with extremely high R_2 value and extremely low Durbin-Watson statistic value, but still misleading results. This is one of the major issues when dealing with nonstationarity.

In above equation, the stochastic process with | ρ | = 1 is called unit root process. As mentioned earlier, the impact of shock over the process does not dissipate with time. The nonstationary time series with unit root is evolutionary in nature and the value of the variable at any given time “t” depend on upon the cumulative impact of shocks from the beginning of time series. Depending upon the positive or negative cumulative outcome of shocks, the variable can increase or decrease over time. The variance of the unit root process is time dependent and depending upon time, can diverge to infinity.

At any given time "t", a particular observation is derived from the summation of all the preceding random shocks. As, all the random shocks, till time "t", are integrated to give particular observation at time "t", it is called integration. Simply by visual analysis of the equation, we can observe that to remove all the preceding random shocks and account shock at a time "t", we should take a difference of preceding period from the current period. This derived a new regression equation which has no stochastic trend from the previous period, making it stationary time series.

$$\Delta y_t = y_t - y_{t-1} = e_t$$

So, integrated variables can be brought in the form of stationary series by using differencing. But it depends on the order of integration, how many times difference to be taken to make a series stationary. Some series have an order of two and require to take the difference of difference to create a stationary time series. The "k" order integrated variables are denoted by I(k).

When two-time series have a shared stochastic patterns over time, the two-time series can be compared together, and nonstationarity of the series can be removed by generating cointegrated relationship regression equation. Cointegration is similar to integration and the detrending involved removing the shared shocks of stochastic trends. The biggest problem with differencing is that while taking differencing, the data can be lost. There is also a possibility of losing the level of time series. When two-time series are compared and show a characteristic of cointegration, the problem of spurious regression, which was discussed earlier can be solved without differencing. So, the data and level of time series are intact. Cointegration typically refers to a linear combination of nonstationary variables (Enders, 2015). It can be concluded that if a cointegration relationship exists in two-time series, first, the variables shared stochastic trends and second, variables have a long run structural relationship.

7.2 Augmented-dickey fuller (ADF) test

The time series analysis requires to first check if the series is stationary or non-stationary, as explained earlier the chapter. The non-stationary time series if

differenced "d" times before it become stationary time series, then it is said to be of order "d". It means that the series have "d" number of unit roots.
Also, it can be concluded that if a series is checked for unit root and there exist zero unit roots, series is said to be stationary series. So, it is very important to check for the unit root.
The pioneering work on testing of the unit root in time series was done by Dickey and Fuller (Dickey, Fuller, Dickey, & Fuller, 2016).

The aim of the test is to test for the stationarity of series:

$$y_t = \phi y_{t-1} + \mu_t$$

Where, the null hypothesis; H_0: series contains a unit root i.e. $\phi = 1$
alternative hypothesis; H_1: series is stationary, $\phi < 1$

Alternatively, the regression can be written as;

$$\Delta y_t = (\phi - 1)y_{t-1} + \mu_t$$

$$\Delta y_t = \beta y_{t-1} + \mu_t$$

Where, $\beta = \phi - 1 = \phi(1)$. The unit root hypothesis is transformed into:
H_0: $\beta = 0$ which is same as $\phi = 1$, Since $\phi = \beta - 1$ against the alternative hypothesis; H_1: $\beta < 1$.
The dickey fuller test is simply a t- test for the null hypothesis, H_0. Where, the test statistic is:

$$\left(t = \frac{\hat{\beta}}{\widehat{SE(\hat{\beta})}} \right)$$

It should be noted that the test statistics does not follow t-distribution under the null hypothesis, as under null hypothesis the series is nonstationary. So, it is not possible to use the standard t-distribution to calculate the critical value (C.V.). The t-statics have a specific distribution known as dickey-fuller table.
There are three different version of the test, depending upon the property of time series, which is the deterministic time trend and drift (intercept). The versions are:

1. Test for the unit root: Autoregression model with random walk.
   AR equation:
   $$\Delta y_t = \beta y_{t-1} + \mu_t$$
   Hypothesis:
   $$H_0: y_t = y_{t-1} + \mu_t$$
   $$H_1: y_t = \beta y_{t-1} + \mu_t\ ;\ \beta < 1$$

2. Test for the unit root with drift: Autoregression model with random walk and drift.
   AR equation:
   $$\Delta y_t = a_0 + \beta y_{t-1} + \mu_t$$
   Hypothesis:
   $$H_0: y_t = y_{t-1} + \mu_t$$
   $$H_1: y_t = a_0 + \beta y_{t-1} + \mu_t\ ;\ \beta < 1$$

3. Test for the unit root with drift and deterministic time trend:
   AR equation:
$$\Delta y_t = a_0 + a_1 t + \beta y_{t-1} + \mu_t$$
   Hypothesis:
$$H_0 : y_t = y_{t-1} + \mu_t$$
$$H_1 : y_t = a_0 + a_1 t + \beta y_{t-1} + \mu_t \; ; \; \beta < 1$$

The three version of the test gives a different critical value (C.V.) at a same significant level. The null hypothesis of unit root is rejected in favour of the alternative hypothesis of stationary series if the test statistic is more negative than the critical value. Therefore, it is critical to figure out which of the three version of the test is best suited for the particular time series. Depending on the version used to verify the hypothesis, the results can be altered. Using an inappropriate version of the test can lead to discard nonstationarity, where there is one.

Also, the Dickey-Fuller test is only valid if the "μ_t" is only white noise of the regression. If there is autocorrelation[2] exist in the dependent variable, then the structure of autoregression (AR) is more complicated and these complex scenarios are not accounted in the model proposed by the Dickey-fuller. For example, by differencing the trend-stationary series, the non-stationarity can be removed but it also introduces a moving average (MA) structure in the error and it is not accounted in the simple Dickey-Fuller test.

In 1984, Said and Dickey modified the basic dickey fuller test and accommodated the moving average structure introduce in error term due to differencing the series. The author (Said & Dickey, 1984), increases the scope of basic test to include autoregression moving average model (ARMA), and the modified test is known as Augmented Dickey-Fuller test (ADF). The test gives a different critical value (C.V.) at 10%, 5%, and 1% significant level. The null hypothesis of unit root is rejected in favour of the alternative hypothesis of stationary series if the test statistic is more negative than the critical value.
The modified regression is written as:

$$\Delta y_t = \beta y_{t-1} + \sum_{i=1}^{l} \alpha_i \Delta y_{t-i} + \mu_t$$

Where $l = lags\ of\ dependent\ variable$.

But the ADF test has one complexity that is to determine the optimal lag of the dependent variable to counter the problem of autocorrelation. The prior knowledge of the time series is necessary in this case. But when it is not possible to have prior knowledge, various testing strategies can be used to appropriate select the optimal lag of the dependent variable.

There are information criterions (ICs) like Akaike Info Criterion (AIC), Schwarz Info Criterion (SIC), and Hannan-QuinnCriterion (HQC) can be used for selecting the lag

---

[2] Autocorrelation means the correlation of time series with itself at different point of time. It implies that the time series have similarity between its data as a function of time. It shows a pattern in a time series.

length to counter the problem of autocorrelation. These ICs are incorporated in Augmented Dickey-Fuller test environment.

7.3 Johansen cointegration test

The time series can be cointegrated in various ways and depending upon the results of the test; it can be concluded if the time series(s) are cointegrated or not. As checked for the unit root previously, we found there exist unit root in variables, ln_{HH} and ln_{WTI}. So, both the time series are integrated of order one or can be represented as $I(1)$. "The Johansen test can be seen as a multivariate generalisation of the augmented Dickey Fuller test" (Dwyer, 2014).

But, prior running the Johansen test, the number of maximum lag that needs to be included in the test has to be estimated. To determine the optimal number of lagged effects to include in the model, first, we fit the Vector Autoregression (VAR) model using the natural gas price, WTI crude oil price and exogenous variables, discussed earlier.

VAR model with exogenous variable can be empirically written as:

$$P_{HH,t} = a + \sum_{i=1}^{n} b_i P_{WTI,t-i} + \sum_{i=1}^{n} c_i P_{HH,t-i} \sum_{j=1}^{5} d_j X_{j,t} + \varepsilon_t$$

Where,
ε_t denotes a random error term with an expected value tends to zero.
'n' is a number of lags of the previous change in price that effect price at any given time $'t'$.

The log price of Henry hub natural gas ($P_{HH,t}$) is dependent on:

- the previous weeks ranging 1 to 'n' logged price of WTI crude oil ($P_{WTI,t-i}$). The effect of each week is denoted by the coefficient of WTI crude oil price i.e. b_i.
- the previous weeks ranging 1 to 'n' logged price of henry hub natural gas ($P_{HH,t-i}$). The effect of each week is denoted by the coefficient of Henry hub natural gas price i.e. c_i.
- the exogenous variables ($X_{j,t}$), heating degree days (HDD), cooling degree days (CDD), deviations from normal HDD (HDDEV), deviation from normal CDD (CDDEV), and difference of actual to average storage level of natural gas (STORDIFF). The effects of each exogenous variable are accounted by the coefficient d_j.

To estimate the number of lags, we use VAR model. The objective of running VAR model is to fit VAR model with different lag lengths, compare the results of the model and find optimal leg length which explains the data for all the variables used in the model.
The author will fit the VAR model with varying lag lengths up to maximum 12 lags on the Henry Hub natural gas price and WTI crude oil price series and include the exogenous variables in the model. The maximum 12 lag length is selected to capture the effect of previous one season over the weekly natural gas price. VAR includes various lag selection criteria tests like Likelihood Ratio (LR) test, Akaike's Final Prediction Error (FPE), Akaike's Information Criterion (AIC) test, the Schwarz

Bayesian Information Criteria (SBIC) test and the Hannan-Quinn Information Criteria (HQIC) test. Each of the lag selection criteria tests uses different mathematical and statical method to analyse the best suitable lag length for the data.

Once the optimal leg length is estimated using the VAR and leg length criteria test, the natural gas price and WTI crude oil price is analysed to determine if there exist a linear combination of natural gas price and oil price such that the series become stationary in the long run. Implies that there exist a cointegration between Henry Hub natural gas price and crude oil price in the long term. Using the Johansen test for cointegration the natural gas price and crude oil price is checked for cointegration.

The cointegration between two or more variables can be tested using Johansen test for cointegration. Johansen test not only confirms the existence of cointegration but also estimates all cointegration vectors between the variables. If there exist "n" variables then, if cointegration 'n' exist between these variables, there are at most 'n-1' cointegration vectors. The vector autoregression (VAR) in levels with the constant can be written as:

$$X_t = \sum_{1}^{n} A_i X_{t-i} + \varepsilon_t$$

This VAR at level can also be written as

$$\Delta X_t = \delta X_{t-1} + \sum_{i=1}^{n-1} \Pi_i \Delta x_{t-i} + \varepsilon_t$$

Where,
Δ = difference operator
ε_t= Vector residuals.

VECM model has information about the short and long term adjustment in X_t suing the estimated parameters, δ and Π_i.
Π_i can be written in the combination of matrix 'α' and β'. where, β' is a vector of cointegration parameters and 'α' is a vector of error-correction coefficient measuring the speed of convergence to the long-run steady state between the variables.
The coefficients in vector β' are multiplied with the variable to give the linear combination of variables that does not have a unit root, that is $\beta' X_t$. The coefficients in 'α' are multiplied with these cointegrated variables, $\beta' X_t$, to give the response of the variables in the equations when the variables deviate from the long run steady relationship.

So, if the matrix is zero, then there is no cointegration and variables do not have long run steady relationship. To comment on the number of cointegration vectors, the rank of a matrix is checked. If rank is not zero, then there exist a cointegration relationship. In fact, a rank of the matrix gives the value of the number of cointegration vectors. The Johansen test is used to check the rank of the matrix by using maximum Eigenvalue test and track test values. The null hypothesis in these test can be simplified as

H_0: the number of cointegration equation exist is k. where, k is from '0' to 'n' and 'n' is a cointegration equation possible (maximum).

7.4 Vector Error Correction Model (VECM)

Now that we have established a cointegration between Henry Hub natural gas price and WTI crude oil price, we must analyse the impact of the change in the price of one commodity over other. To analyse this, we must build a model this relationship empirically and include exogenous variables that might affect this relationship. The Vector Error Correction Model (VECM) is used to model the fundamental relationship between natural gas and crude oil prices. VECM is a type of Vector Autoregression model (VAR) with Error correction term. Two variables may have common underlying stochastic trends along which they move together on the non-stationary path (Enders, 2015), this is called cointegration of two variables. In the case of natural gas price and crude oil price, the author is interested in finding this cointegration relationship, using data from 1999 till 2016.

In past Econometricians used Vector autoregression model with error correction model to analyse cointegration behaviour of two different time series. In VECM, the equation includes an error-correction term, which measures the previous periods' deviation from the fundamental long-run equilibrium relationship.

In modelling the relationship, we assume that one price time series lead the other, as the effect of one is visible on other after a period. The dependent time series is one which tends to adjust to the change in independent time series. Even though cointegration exists between two different time series, variables tend to deviate from the long-run equilibrium relationship. But, they adjust and return to the fundamental relationship. The error correction term captures the characteristic of the variable to return to the original relationship in the course of time. The error correction term measures the rate or the “speed” of correction of dependent commodity's deviation from long-run equilibrium with the independent commodity in long-run.

In this study, the dependent variable is natural gas price time series, and the independent variable is WTI crude oil price time series. As we have discussed in Chapter 5, there are other exogenous variables - Seasonality, weather, Shut-in due to hurricanes and storage level, which also affect the price of natural gas. To measure the natural gas deviations that are not due to oil prices and also to account for the volatility in the natural gas prices that are not explained by oil prices, the model also fitted with these previously discussed exogenous variables. These variables may push the price of natural gas closer or away from the long-run equilibrium with crude oil price. The difference between the actual natural gas price at Henry Hub and predicted price in long-run equilibrium using VECM model is the deviation of the dependent variable from long-run equilibrium relationship. The deviation of prices from the fundamental relationship is known as error term in VECM model and is denoted by "μ_t". The error correction mechanism pushes the natural gas price closer to the long-run relationship and reducing the gap between the actual price and predicted price from VECM. This correction term corrects the deviation from long-run equilibrium by a certain amount every week.

Once proven that the natural gas prices and crude oil prices are cointegrated using the steps as mentioned in earlier in this chapter, the long-run relationship, including exogenous variables, can be empirically expressed as:

$$P_{HH,t} = \gamma + \beta P_{WTI,t} + \mu_t$$

$$\Delta P_{HH,t} = a + \alpha(\mu_{t-1}) + \sum_{i=1}^{n} b_i \Delta P_{WTI,t-i} + \sum_{i=1}^{n} c_i \Delta P_{HH,t-i} \sum_{j=1}^{6} d_j X_{j,t} + \mathcal{E}_t$$

$$\Delta P_{WTI,t} = a + \alpha(\mu_{t-1}) + \sum_{i=1}^{n} b_i \Delta P_{WTI,t-i} + \sum_{i=1}^{n} c_i \Delta P_{HH,t-i} \sum_{j=1}^{6} d_j X_{j,t} + \mathcal{E}_t$$

Where,

$P_{HH,t} = logged\ henry\ hub\ natural\ gas\ price\ in\ week\ "t"$
$P_{WTI,t} = logged\ West\ Texas\ Intermediate\ crude\ oil\ price\ in\ week\ "t"$
$\gamma = Constant\ (to\ be\ exstimate)$
$\beta = Parameter\ (to\ be\ exstimate)$
$\mu_t = Error\ Term\ in\ week\ "t"$
$\Delta P_{HH,t} = Change\ in\ natural\ gas\ price\ at\ henry\ hub\ from\ previous\ week\ in\ week\ "t"$
$\Delta P_{WTI,t} = Change\ in\ WTI\ price\ from\ previous\ week\ in\ week\ "t"$
$\Delta P_{HH,t-i}$
$= Change\ in\ natural\ gas\ price\ at\ henry\ hub\ from\ previous\ week\ in\ week\ "t-i$
$\Delta P_{WTI,t-i} = Change\ in\ WTI\ price\ from\ previous\ week\ in\ week\ "t-i"$
$\mu_{t-1} = lagged\ set\ of\ equilibirium\ error\ Term, lagged\ by\ 1\ week$
X_j
$= Matrix\ of\ 6\ Exogenous\ variables\ reprentng\ the\ assumed\ variable\ driving\ natural\ gas\ price$
$\mathcal{E}_t = Normal\ Error\ term\ with\ mean\ equals\ to\ zero$
$a, b_i, c_i, d_i = coefficient\ of\ variables\ (to\ be\ estimated)$

Here, the first equation is empirically presentation of the long-run relationship between natural gas and oil prices.
The second equation by VECM have Henry Hub natural gas price as a dependent variable and in the third equation, the WTI crude oil price as a dependent variable. It includes error correction term, the "lagged effect of two price series on the dependent variable, and the effect of six exogenous variables, discussed previously, on the dependent variable.

The VECM model does not only give a result that WTI crude oil prices influence the Henry Hub natural gas prices, but also formulates empirically the results, where in one case oil prices assumed to affect the natural gas prices and in another instance, natural gas prices are assumed to influence the crude oil prices. Therefore, in the model, the dependent variable is not fixed. It generates different regressions wherein each regression the dependent variable is different. So, the regression with WTI crude oil price as a dependent variable and natural gas price as an independent variable is also generated. It allows the effect of exogenous variables on WTI crude oil prices as well since, in this case, the WTI crude oil price is the dependent variable.

Partially, with the help of statistical analysis, we will draw our conclusion whether its natural gas price that is dependent variable or it is oil price that is the dependent variable. Moreover, in this research, we are focused and assume that it is a natural gas price that is dependent variable and oil price is an independent variable.

In this research as we are not going to argue over the "cause and effect" mechanism between natural gas and oil price, VECM model is a best-suited model for this research. The advantage of VECM model is that it depends on the mechanism of a relation between the commodities. The model directly identifies the relationship between the commodities and do not influence by the "cause and effect" mechanism.

7.5 Conditional Error Correction Model (ECM)

As discussed in VECM model analysis, we assume that WTI crude oil prices are at least weakly exogenous and this weak exogeniety of WTI enable us to modify our error correction model. To remove complication that WTI crude oil prices are dependent on Henry-Hub natural gas price and other exogenous variables, we assume that the WTI crude oil prices can be taken as predetermined. As mentioned in (Villar & Joutz, 2006), we can reduce two equation relationship from VECM model to single regression relationship using modified version of error correction model, we called it conditional error correction model (conditional ECM).

The conditional ECM is an autoregressive distributed lag (ARDL) model which measures the relationship between Henry-Hub natural gas price and WTI crude oil price, assuming the WTI prices can take predetermined value and any change in WTI crude oil prices have the immediate effect on Henry-Hub price. Basically, we assume that there are market forces that alter both commodities prices and if the direction of movement of WTI crude oil price is known, one can predict the movement of Henry-Hub natural gas prices. Our conditional ECM model measures:

- the immediate effect of change of WTI crude oil price.
- the effect of the set of six exogenous variables.
- the lagged effect of a change in Henry-Hub price

Moreover, the conditional ECM uses the error term from the long-term relationship generated by VECM model as an exogenous variable. The equation for conditional ECM model is as follows:

$$\Delta P_{HH,t} = a + \alpha\mu_{t-1} + b\Delta P_{WTI,t} + \sum_{i=1}^{n} c_i \Delta P_{HH,t-i} + \sum_{j=1}^{6} d_j X_{j,t} + \varepsilon_t$$

Where a is a constant, α is the coefficient of the error -correction term, μ_{t-1} is error correction term of long-term relationship between Henry-Hub natural gas price and WTI crude oil price generated from VECM model, and b is a coefficient of the predetermined change in WTI crude oil price. The c_i is a series of coefficient for the lagged effect of the change in Henry-Hub natural gas price. Since in VECM model we used the lag length as “9”, we also use the same lag length in conditional ECM model, so the value of n is 9. d_j represent the coefficient of each of the six exogenous variable, previously used in VECM model. $X_{j,t}$ denotes the six exogenous variables, HDD, HDDEV, CDD, CDDDEV, STORDIFF and SHUTIN. ε_t denotes the white noise in the system with expected value of zero

7.6 Segmented Time period

It is a general perception that natural gas prices are linked to crude oil prices, but to understand the phenomena of decoupling of natural gas prices, the entire data series is segmented into two segments. The segmented period model enables to visualise the time when natural gas prices are decoupled in past. We built segmented model for the period of January 1999 - June 2008 and July 2008 - June 2016, and compared with for the period of January 1999-June 2016, to understand the stability of our data and model. In this analyses, the data is used check for cointegration and to build a conditional error correction model (ECM). The series are tested for the unit roots and followed by VAR lag selection criteria test. Later suing the appropriate lag length, we conduct the Johansen test for cointegration. Once the Johansen test confirms the cointegration, we use VECM model to generate the long-term relationship between natural gas and crude oil prices. The results from VECM is analysed and using conditional ECM the change is natural gas price volatility is examined.

7.7 Data

In order to run the time series analysis using VECM and ECM model, it is necessary to collect the relevant data from various sources. This data can be widely gathered through consulting various publicly available databases. A detailed description of the data used in this research can be found in this section. All the data gathered from the Bloomberg Terminal at Erasmus University and analysed using EVIEW software.

All the data collected is measured weekly for to capture the small fluctuation in prices. In the analysis, the weekly estimated Henry-Hub natural gas spot prices and the Weekly estimated WTI crude oil spot prices are endogenous variable. The six exogenous variables are included in this research to capture the unexplained volatility of natural gas prices. The Bloomberg terminal at Erasmus University is used to collect all the data as Bloomberg terminal is a reliable source for the various research in past.

Henry Hub natural gas price series:

The Henry Hub spot prices are used in research to analyse the natural gas price time series. The prices are transformed into natural logarithms to remove the scale effect and to avoid heteroscedasticity of series. As explained by (Villar & Joutz, 2006), the logarithmic transformation has an advantage of constant elasticity which helps in the direct interpretation of results. The logarithmic value has additional benefits that the variables can be used without concerning the units of variables.
The Henry Hub weekly natural gas prices are obtained from Bloomberg Terminal at Erasmus University with the category code of "NGUSHHUB Index."

WTI crude oil price series:

The WTI crude oil spot prices are used in this research to analyse the relationship of oil with natural gas. The WTI prices are also transformed in natural logarithmic for the season explained above. The WTI prices are weekly measured and obtained from the

Bloomberg terminal at Erasmus University with a category code of "USCRWTIC Index."

Heating Degree days:

Heating Degree days (HDD) is used in analyses as a dummy variable to account the change in natural gas prices due to cold temperature in the US. The usage of natural gas is significant in residential heating and its demand increases with increase in HDD value. The HDD is a weekly measure of temperature above the base temperature which is 65 degree Fahrenheit. The HDD data is an explicit representation of the increase in energy utilisation due to cold. The HDD data is accumulated by National Oceanic and Atmospheric Administration (NOAA), US Department of Commerce and obtained from the Bloomberg Terminal at Erasmus University with a category code of "NOAHUSAA Index."

Cooling Degree days:

Cooling degree days (CDD) is used in analyses as a dummy variable to account the change in natural gas prices due to a warm temperature in the US. The usage of natural gas in electricity generation in power plants increases with increase in temperature. With the increase in CDD value, the demand for a natural gas increase as more power is needed by the residential consumer to operate air condition. The CDD is a weekly measure of temperature below the base temperature which is 65 degree Fahrenheit. The CDD data is accumulated by National Oceanic and Atmospheric Administration (NOAA), US Department of Commerce and obtained from the Bloomberg Terminal at Erasmus University with a category code of "NOACNUST Index."

Deviation from average HDD and CDD (HDDEV and CDDDEV):

HDDEV and CDDDEV are used to depict the increase in natural gas demand due to an unexpected change in temperature. The HDDEV and CDDDEV series are deviations from normal HDD and CDD values respectively. The HDDEV value is calculated by taking the difference of average HDD value from normal HDD values. The average HDD and normal HDD value are obtained from the Bloomberg Terminal at Erasmus University with category code "NOAHUSAA Index" and "NOAHUSAN Index" respectively. The average CDD and normal HDD value are obtained from the Bloomberg Terminal at Erasmus University with category code "NOACNUST Index" and "NOACAUST Index" respectively. The unexpectedly high or low temperature pressurise the demand side of natural gas and affect the prices of natural gas.

Storage Level:

The inventory level of natural gas is included in the natural gas analysis as inventories are act as a buffer and can be used to release the pressure on the demand side. The Storage differential with average five-year value (STORDIFF) is employed in the analysis to account for the change in natural gas demand due to change in inventory level. The unexpected change in inventory level from the average implies that the there is an unexpected change in demand for natural gas. This influences the prices and affects the natural gas price relationship. The STORDIFF data is collected from the Bloomberg Terminal at Erasmus University with a category code "DOENUST5 Index."

Disturbance in production:

The disturbance in the Gulf of Mexico due to Hurricane affect the supply side of natural gas and influence natural gas prices. The effect of disturbance in supply side is accounted in this research by SHUTIN variable. SHUTIN data is a weekly reduction in production of natural gas in Billion cubic feet (Bcf) from the Gulf of Mexico and is obtained from Bloomberg Terminal at Erasmus University. The data is collected by U. S. Department of the Interior, Minerals Management Service to measure the reduction in natural gas production.

7.8 Conclusion

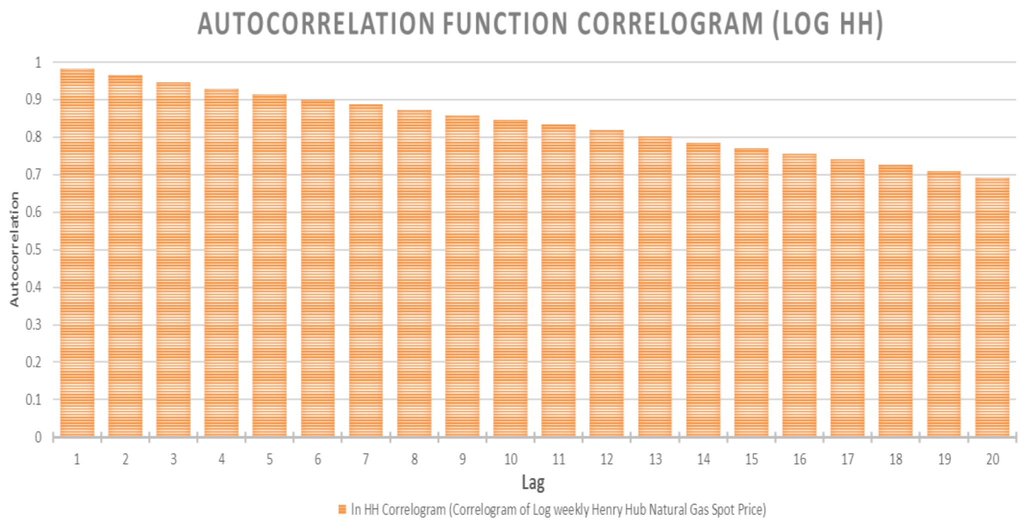
The methodology used in this research consists of several steps to analyse the Natural gas and crude oil price relationship. Few of these steps are prerequisites and requirement of applying a model for examining the cointegration property between two-time series like stationarity, the level of integration, and lag-length. Later, Johansen test for cointegration is used to verify the cointegration between natural gas and crude oil prices. After confirming the cointegration, the cointegration equation is derived using Vector error correction model (VECM). As VECM derive two equations one with natural gas prices as a dependent variable and other with crude oil prices as a dependent variable, we use condition Error correction model (ECM) where natural gas prices are the dependent variable, and the effect of a contemporaneous change in crude oil prices are modelled in it. Later, to examine the relationship over the period, author use segmented conditional error correction model (ECM).

8. Results and Analysis

Following the argumentation above, the VECM model has been employed and used the data mentioned in the previous chapter. The reason to first run the ADF test and Johansen test are that it confirms the cointegration between Henry Hub natural gas price and WTI crude oil price, to select the optimal model for the analysis. The VECM model is applicable only if cointegration exists between the variables. The results from the Dickey-Fuller test, Johansen test and VECM test will be described and analysed in this chapter. A residual diagnostic analysis has also been carried out on the VECM, which is also explained in this chapter.
In addition to the quantitative analysis found in this chapter, a qualitative analysis of results is also included in this section. Also, interpretation of the empirical result is explained in this section.

8.1 Augmented dickey fuller test (ADF)

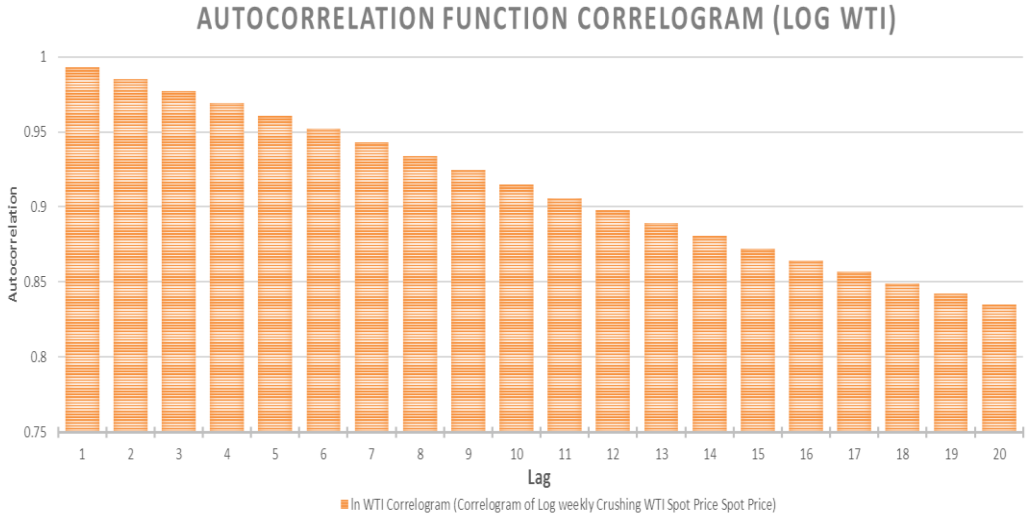
Below, figure:19 and figure:20, represent the results after checking the autocorrelation (footnote) functions (ACF's) in logged Henry Hub natural gas and WTI crude oil price time series. The ACFs through the 20 lags for both series are evaluated and found a trend at a level.



Note: The logged natural gas price is calculated using the Henry Hub price series, measured weekly in dollar per MMBtu, obtained from the Bloomberg Data Terminal at Erasmus University. Autocorrelation function is calculated using "EVIEW" software for maximum Lag length of 20.

Figure 21: Autocorrelation function In HH

Source: Author using data



Note: The logged WTI price is calculated using the WTI crude oil price series, measured weekly in dollar per bbl., obtained from the Bloomberg Data Terminal at Erasmus University. Autocorrelation function is calculated using "EVIEW" software for maximum Lag length of 20.

Figure 22: Autocorrelation function ln WTI

Source: Author using data

The presence of unit root is tested by Augmented Dickey-Fuller (ADF) test. The ADF test checks the stationarity in first difference of variables. The first difference dependent variable can be expressed in a combination of stationary variables and non-stationary variable. If the coefficient of the non-stationary term is zero, then the dependent variable has unit root. Table:1, summarised the result of ADF test and Phillips-Perron Unit root test (footnote).

	Augmented Dickey-Fuller Test, Lag selection using AIC criterion				**Phillips-Perron Unit Root Test, Bandwidth selection using Newwy-West Bandwidth**			
	ln HH		**ln WTI**		**ln HH**		**ln WTI**	
	Lag	t-statistic value for ln HH	Lag	t-statistic value for ln HH	Bandwidth	t-statistic value for ln HH	Bandwidth	t-statistic value for ln HH
	1	2.998137	8	2.740217	14	2.757219	8	2.709539
	Test critical values							
1% significant level	3.44				3.44			
5% significant level	2.86				2.86			
10% significant level	2.57				2.57			

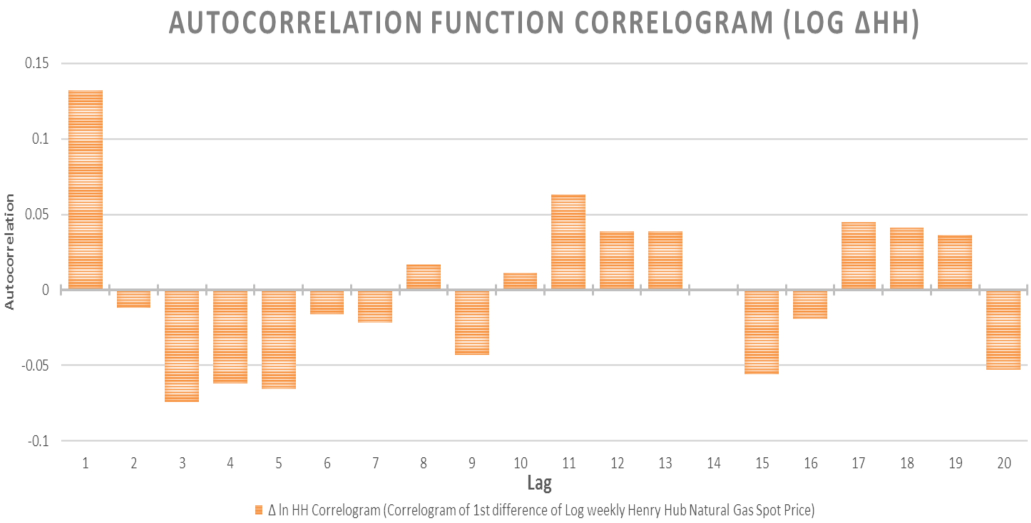
Table 1: ADF and Philips-Perron Unit Root Test

Source: Author using data in "EVIEWS" software

From table:1, it is clear that at a significant level of 5%, the logged Henry-Hub price series (ln HH) support the hypothesis of a unit root. Whereas WTI crude oil price series (ln WTI) support unit root hypothesis at a significant level of 10%. Based on the ADF test and Phillips-Perron unit root test, the Henry Hun natural gas price series and WTI price series appears to be unit root process. It implies that the coefficient of non-stationary term is zero

The results of ADF test and Phillip Perron test indicate that the Henry hub natural gas price and WTI crude oil price series are non-stationary processes.

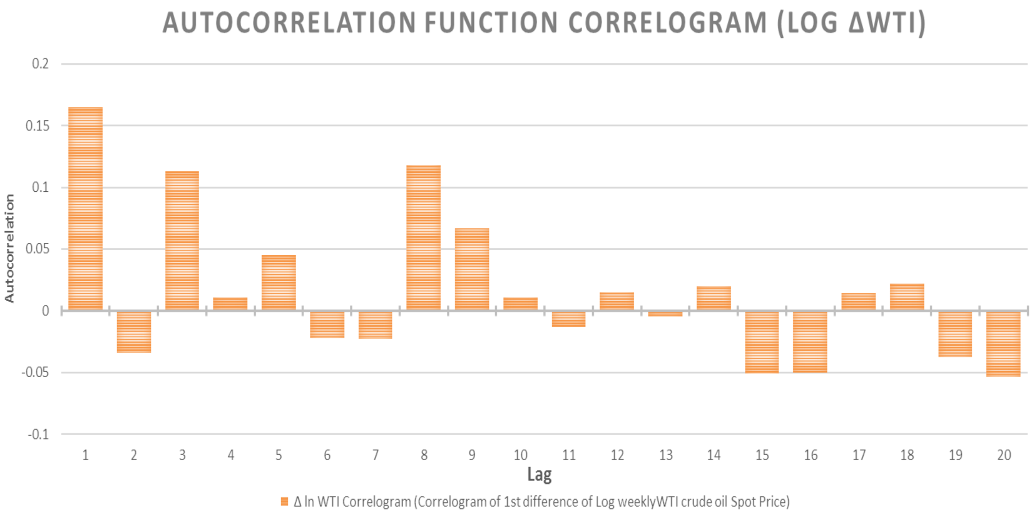
The 1st difference of ln HH and ln WTI are also analysed to confirm the ADF test and Phillip-Perron test result. The figure:21 and figure:22 show the autocorrelation of 1st difference ln HH and ln WTI time series respectively.



Note: The first difference logged natural gas price is calculated by using the Henry Hub price series, measured weekly in dollar per MMBtu, obtained from the Bloomberg Data Terminal at Erasmus University. Autocorrelation function is calculated using "EVIEW" software for maximum Lag length of 20 and unit root in 1st diffference.

Figure 23: Autocorrelation function of 1st difference ln HH

Source: Author using data in "EVIEWS" software



Note: The first difference logged WTI price is calculated by using the WTI crude oil price series, measured weekly in dollar per bbl., obtained from the Bloomberg Data Terminal at Erasmus University. Autocorrelation function is calculated using "EVIEW" software for maximum Lag length of 20 and unit root in 1st difference.

Figure 24: Autocorrelation function of 1st difference ln WTI

Source: Author using data in "EVIEWS" software

Whereas the table:2 indicate the results when the ln HH and ln WTI series are analysed at 1st difference using the ADF and Phillips-Perron unit root tests.

	Augmented Dickey-Fuller Test, Lag selection using AIC				Phillips-Perron Unit Root Test, Bandwidth			
	ln Δ HH		ln Δ WTI		ln Δ HH		ln Δ WTI	
	Lag	t-statistic value for ln HH	Lag	t-statistic value for ln HH	Bandwidth	t-statistic value for ln HH	Bandwidth	t-statistic value for ln HH
	4	15.08841	7	9.089137	19	26.25236	6	25.7151
	Test critical values							
1% significant level	3.44				3.44			
5% significant level	2.86				2.86			
10% significant level	2.57				2.57			

Table 2: ADF and Philips-Perron Unit Root Test in 1st difference

Source: Author using data in "EVIEWS" software

The results are confirmed by analysing the ln HH and ln WTI series at a 1st difference. The autocorrelation as shown in figure:21 and figure:22, clearly indicated that the trend or autocorrelation is not significant at the 1st difference. Also, from the table:2, it is clear that the ADF and Phillips-Perron test indicates no unit root at the 1st difference of series. First differencing the price series results in stationary time series for ln HH and ln WTI time series.

Results indicate that the series are non-stationary and integrated of 1st order, or the series have a unit root. This finding is significant as to analyse the cointegration between the two-time series; both the series must be of same order. Then only we can examine and formulate the cointegration relationship between the two commodities.

8.2 Johansen cointegration test

Prior testing for cointegration using Johansen test, we need to find the optimal lag length for the variables. The lag length is calculated using Vector Autoregression model (VAR). First using VAR model for the data set from January 1999 to June 2016, we examine the lag length of the two-time series. The table:3, results from VAR for calculating the lag length between ln HH and ln WTI series. In the VAR model the exogenous variables (CDD. HDD. CDDDEV, HDDEV, STORDIFF and SHUTIN) are included.

For selecting the lag length, various selection order criteria tests performed over the results from VAR model.
The results from Least Square test (LR), Final Prediction test (FPE), Akaike's Information Criterion test (AIC), the Schwarz Bayesian Information Criteria test (SBIC), and the Hannan-Quinn Information Criteria test (HQIC) are examined for lag selection.
The AIC, HQIC, and the SBIC statistics are using the same methodology to calculate the lag length. Each uses the log likelihood of estimated VAR and a penalty term for the number of the parameter used in estimation.

But as mentioned by (Hamilton, 1994), the results vary due to the level of trade-off between precision and over-parameterization.

From the table:3, it is evident the different criterions result in various lag length. The Least Square (LR) criterion indicate nine lags to be optimal lag, Final prediction error (FPE) and Akaike information criterion (AIC) indicate ten lags, Schwarz information criterion (SC) suggest two lags and Hanna-Quinn information criterion suggest four lags to be optimal.

For this research, we select ten lags, as two tests are supporting the nine lags. Also, we check cointegration for two and five lag length. One important point to remember, Johansen test and VECM model use the leg duration of the endogenous variables at the 1st difference. So, for Johansen test and VECM model, the lag length must be one less than the results of VAR lag order selection test.

VAR Lag Order Selection Criteria
Endogenous variables: LN_HH LN_WTI
Exogenous variables: CDD HDD CDDDEV HDDEV STORDIFF
Date: 08/09/16 Time: 17:47
Sample: 1/01/1999 7/15/2016
Included observations: 904

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-1915.079	NA	0.242506	4.259024	4.312197	4.279332
1	2693.689	9146.162	9.13e-06	-5.928516	-5.854074	-5.900085
2	2707.617	27.57770	8.93e-06	-5.950480	-5.854769*	-5.913925
3	2712.025	8.708327	8.92e-06	-5.951382	-5.834402	-5.906704
4	2725.127	25.82836	8.74e-06	-5.971521	-5.833271	-5.918719*
5	2731.312	12.16350	8.70e-06	-5.976353	-5.816835	-5.915429
6	2735.700	8.611605	8.69e-06	-5.977212	-5.796424	-5.908165
7	2741.271	10.90809	8.66e-06	-5.980688	-5.778631	-5.903517
8	2743.472	4.299591	8.70e-06	-5.976708	-5.753382	-5.891414
9	2750.625	13.94220*	8.64e-06	-5.983684	-5.739089	-5.890267
10	2755.154	8.806832	8.63e-06*	-5.984853*	-5.718989	-5.883313
11	2756.156	1.943785	8.68e-06	-5.978220	-5.691087	-5.868557
12	2760.408	8.231716	8.68e-06	-5.978778	-5.670376	-5.860991

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

Table 3: VAR Lag Order Selection Criteria (1999-2016)

Source: Author using data in "EVIEWS" software

Once we have optimal lag, we run the Johansen test on the dataset, and the results are presented in below table:4. The null hypothesis in Johansen test, there exist no cointegration equation, is rejected by trace test in Johansen test.
The first row of the table tests the null hypothesis, no cointegration exist, against alternative hypothesis. The null hypothesis is rejected at the 1% significant level with trace statistics of 20.88438, indicating that there is a significant cointegration in the system. Hence, the result of Johansen test shows that the Henry-Hub natural gas spot prices and WTI crude oil prices are cointegrated for the period of January 1999-June2016.

Johansen Test for Cointegration: January 1999-June 2016					
Hypothesized No. of CE(s)	Eigenvalue	Trace statistic	10%	5%	1%
0		20.88438	20.75	15.41	20.04
1	0.010078	9.177184	2.69	3.76	6.65
Note: Critical values are found using "EVIEWS" software, which takes the vales from Osterwald-Lenum (1992) table.					

Table 4: Johansen test for cointegration: January 1999 - June 2016

Source: Author using data in "EVIEWS" software

8.3 Vector Error Correction Model (VECM)

After confirming the long run relationship between Henry-Hub natural gas price and WTI crude oil price, we use VECM model to estimate the long-run relationship between two commodities. In VECM model the exogenous variables are also included to account for volatility which is not explained by oil price. Table:5 represents the results of VECM model, assuming Henry Hub prices and, alternatively, WTI crude oil prices as the dependent variable. Using data from January 1999 to June-2016, cointegration regression is estimated using Ordinary Least Square (OLS) estimator. For the single cointegration relationship, with Henry Hub natural gas price as a dependent variable, the regression for long run relationship is:

$$P_{HH,t} = 2.10510 - 0.160421 P_{WTI,t}$$

Also, to include the lagged effect, exogenous variables effect, and the error correction term, the cointegration equations are generated with natural gas price and, alternately WTI price as the dependent variable.

<table>
<tr><th colspan="7">VECM Model including Exogenous Variables (January 1999 - June 2016)</th></tr>
<tr><th colspan="3">Equation</th><th colspan="2">R square</th><th colspan="2">p-value</th></tr>
<tr><td colspan="3">Δln HH (dependent variable)</td><td colspan="2">0.147659</td><td colspan="2">0.0000</td></tr>
<tr><td colspan="3">Δln WTI (dependent variable)</td><td colspan="2">0.103773</td><td colspan="2">0.0000</td></tr>
<tr><th colspan="6">Long-Term Relationship: Single cointegration equation using OLS method.</th><td></td></tr>
<tr><td colspan="3">long term coefficients</td><td colspan="2">values</td><td>p-value</td><td></td></tr>
<tr><td colspan="3">β (independent variable coefficient)</td><td colspan="2">(-) 0.160421</td><td>0.0000</td><td></td></tr>
<tr><td colspan="3">γ (constant)</td><td colspan="2">2.105110</td><td colspan="2"></td></tr>
<tr><th colspan="3">Dependent variable: Henry-Hub Natural gas price</th><td></td><th colspan="3">Dependent variable: WTI crude oil price</th></tr>
<tr><td>Variables</td><td>Values</td><td>p-value</td><td></td><td>Variables</td><td>Values</td><td>p-value</td></tr>
<tr><td>α-coefficient</td><td>(-)0.01853</td><td>0.0008</td><td></td><td>α-coefficient</td><td>(-)0.00022</td><td>0.9472</td></tr>
<tr><td>ΔP HH(t-1)</td><td>0.028442</td><td>0.3994</td><td></td><td>ΔP HH(t-1)</td><td>(-)0.010012</td><td>0.6218</td></tr>
<tr><td>ΔP HH(t-2)</td><td>(-)0.06098</td><td>0.0639</td><td></td><td>ΔP HH(t-2)</td><td>0.005318</td><td>0.7881</td></tr>
<tr><td>ΔP HH(t-3)</td><td>(-)0.08820</td><td>0.0071</td><td></td><td>ΔP HH(t-3)</td><td>(-)0.042821</td><td>0.0298</td></tr>
<tr><td>ΔP HH(t-4)</td><td>(-)0.02186</td><td>0.5033</td><td></td><td>ΔP HH(t-4)</td><td>0.058241</td><td>0.0031</td></tr>
<tr><td>ΔP HH(t-5)</td><td>(-)0.06513</td><td>0.0471</td><td></td><td>ΔP HH(t-5)</td><td>(-)0.012841</td><td>0.5151</td></tr>
<tr><td>ΔP HH(t-6)</td><td>(-)0.01828</td><td>0.5777</td><td></td><td>ΔP HH(t-6)</td><td>(-)0.010537</td><td>0.5939</td></tr>
</table>

ΔP HH(t-7)	(-)0.03276	0.3191	ΔP HH(t-7)	0.019923	0.3140
ΔP HH(t-8)	0.014203	0.6657	ΔP HH(t-8)	(-)0.018362	0.3534
ΔP HH(t-9)	(-)0.06487	0.0475	ΔP HH(t-9)	(-)0.022459	0.2540
ΔP WTI(t-1)	0.129189	0.0222	ΔP WTI(t-1)	0.192121	0.0000
ΔP WTI(t-2)	(-)0.04567	0.4243	ΔP WTI(t-2)	(-)0.104152	0.0025
ΔP WTI(t-3)	0.036611	0.5246	ΔP WTI(t-3)	0.153089	0.0000
ΔP WTI(t-4)	(-)0.06742	0.2451	ΔP WTI(t-4)	(-)0.068219	0.0509
ΔP WTI(t-5)	0.03151	0.5878	ΔP WTI(t-5)	0.064432	0.0658
ΔPWTI(t-6)	0.119548	0.0389	ΔPWTI(t-6)	(-)0.056648	0.1038
ΔP WTI(t-7)	0.088245	0.1245	ΔP WTI(t-7)	(-)0.020874	0.5458
ΔP WTI(t-8)	0.078546	0.1692	ΔP WTI(t-8)	0.10358	0.0026
ΔP WTI(t-9)	0.095814	0.0913	ΔP WTI(t-9)	0.051809	0.1291
HDD(t)	4.62E-05	0.4203	HDD(t)	3.22E-05	0.3509
HDDEV(t)	0.001505	0.0000	HDDEV(t)	1.85E-05	0.8243
CDD(t)	(-)0.00039	0.0160	CDD(t)	(-)5.49E-06	0.9553
CDDDEV(t)	0.002797	0.0000	CDDDEV(t)	0.000539	0.0349
STORDIFF(t)	4.95E-06	0.5093	STORDIFF(t)	6.40E-06	0.1504
SHUTIN(t)	3.25E-06	0.2933	SHUTIN(t)	(-)5.03E-06	0.0069
Joint Significance: Henry-Hub natural gas price			**Joint Significance: WTI crude oil price**		
Variable	*Chi2 Stat*	*p-value*	*Variable*	*Chi2 Stat*	*p-value*
lagged ΔP HH	21.00197	0.0126	lagged ΔP HH	16.18384	0.0631
lagged ΔP WTI	22.3463	0.0078	lagged ΔP WTI	68.48336	0.0000
Exogenous variables	95.13984	0.0000	Exogenous variables	14.95595	0.0206

Table 5: VECM results January 1999-June 2016

Source: Author using data in "EVIEWS" software

Table:5, illustrate the results of VECM model for the period of January 1999-June 2016. There are two equations, one with Henry Hub natural gas prices as a dependent variable and other with WTI crude oil prices as a dependent variable.
The equation with natural gas price as the dependent variable, the R^2 value is 0.147659. Whereas for the equation with WTI crude oil price as the dependent variable, the R^2 value is 0.103773. The R^2 values imply that in the case of Henry-Hub prices, approximately 14.76% of the volatility of Henry-Hub prices can be explained through the volatility of WTI crude oil and exogenous variables, included in the equation. Whereas in the case of WTI crude oil price, it is 10.37% volatility of WTI crude oil price that can be explained through volatility in Henry-Hub natural gas price and six exogenous variables, included in the equation. So, the model supports our hypothesis and is better at explaining the Henry Hub natural gas price volatility than WTI crude oil price volatility.

Now we will analyse the change in Henry Hub prices, as illustrated in the table:5, the cointegration coefficient (alfa-coefficient) is -0.018538 with p-value 0.0008. The value of alfa-coefficient is very significant, as it defines the reversion characteristic of the equation. If the coefficient is negative and have significant p-value, it can be said that there exists a long term relationship, and the dependent variable has a tendency to return to this long run relationship. If all the condition and variables remain same, based on the cointegration coefficient, any fluctuation in Henry Hub natural gas price from the long-run relationship would be nullified (corrected) by half of initial error value[3] in about 37 weeks.

The significance of the effect of exogenous variables on the Henry Hub natural gas price varies widely. Variable such as HDDEV and CDDDEV which are accounted for the seasonality, showed a p-value of 0.0000. Also, the variable CDD which is accounted for the weather, shown a p-value 0.0160, which is significant at 5% level. Implies that in the short run these exogenous variables have an impact on natural gas price volatility. Where the HDDEV and CDDDEV have a positive coefficient, the CDD have a negative coefficient. It means, the HDDEV and CDDDEV, push natural gas price away from the long-run relationship and CDD push natural gas price towards the long-run relationship. However, the combined effect of the set of six exogenous variables returns a high chi^2 statistic value of 95.139 at a p-value of 0.0000.

Also, in the short run, the model includes the effect of price change to natural gas at a different lag. From the table:5, we can see that for dependent variable Henry-Hub natural gas price, the lagged price change of natural gas from a three weeks prior, from five weeks prior, and from previous nine weeks was significant within 5% p-value and the second lag was significant at 10% level. But, when all the lagged natural gas prices are jointly analysed, the effect of all nine lagged price changes have a p-value of 0.0126, which is well at the 5% significant level.
When the lagged effect of WTI crude oil price change on natural gas price change is analysed, the lagged price change of WTI crude oil from one week prior, and from six weeks prior were significant within 5% p-value and the ninth lag was significant at 10% level. But when the combined effect of all nine lag of WTI crude oil price change is analysed, the model returns the p-value of 0.0078, which is below 1% level.

8.4 Conditional Error Correction Model (Conditional ECM)

The results from Conditional ECM are illustrated and analysed in this section. Table 6 illustrate the results from the conditional ECM model. The R^2 statistic has increased slightly with 0.150016, meaning that over 15% of the volatility in Henry-Hub natural gas price can be explained through the volatility in exogenous variables, the change in WTI price, and the lagged effect of nine lagged weekly price change in Henry-Hub natural gas price. The chi^2 statistic is also increased slightly from 143.43 for VECM to 147.60 for conditional ECM.

The most important point to be remembered in this conditional ECM model is that this model is not determining the long-run relationship and still the long-run predicted cointegration relationship is estimated by VECM. The cointegration error correction term from VECM model is transferred in conditional ECM with one lag difference. So, the cointegration equation is not estimated by ECM. A conditional ECM model works

[3] The half-life of error term implies the duration in which the dependent variable correct itself by 50% of the deviation from the fundamental relationship. It can be calculated by using following equations:

$$(1-|\alpha|)^n = (0 \cdot 5)$$

Where, $\alpha = error\ correction\ term, n = duration\ of\ half - life\ (unit\ same\ a\ unit\ of\ data)$

on the assumption that the effect of the price difference in WTI crude oil price immediately transferred to Natural gas price and the WTI crude oil price take predetermined values.

Estimation from model suggests that a 10% increase in contemporaneous change in WTI crude oil price lead to 2.68% change in Henry-Hub prices for the period of January 1999 - June 2016. The finding is significant; the natural gas prices tend to change by more than 26% for very 100% contemporaneous change in WTI crude oil prices. As ECM model is not generating the long term relationship, the long-term relationship formulated by VECM still holds. For exogenous variables, the results are similar to VECM model. The variables such as HDDEV and CDDDEV have the significant impact on the change in natural gas prices with p-value 0.0000. The coefficient of each variable depicts the multiplier in short-run equilibrium.

Conditional ECM Model including Exogenous Variables (January 1999-June 2016)

Equation	**R square**	**p-value**
D_ln HH (dependent variable)	0.150016	0.0000

Long-Term Relationship: Single cointegration equation using OLS method.

long term coefficients	values	p-value
β (independent variable coefficient)	-0.160421	0.0000
γ (constant)	2.105110	

Dependent variable: Henry-Hub Natural gas price

Variables	*coefficient*	*p-value*
Constant "a"	0.008656	0.3112
ΔP HH(t-1)	0.047781	0.1459
ΔP HH(t-2)	(-)0.06038	0.0598
ΔP HH(t-3)	(-)0.07568	0.0181
ΔP HH(t-4)	(-)0.04502	0.1601
ΔP HH(t-5)	(-)0.05156	0.1057
ΔP HH(t-6)	0.00059	0.9853
ΔP HH(t-7)	(-)0.02864	0.3701
ΔP HH(t-8)	0.02981	0.3498
ΔP HH(t-9)	(-)0.049807	0.1164
ΔP WTI(t)	0.268581	0.0000
HDD(t)	(-)3.52E-06	0.9493
HDDEV(t)	0.001062	0.0000
CDD(t)	(-)0.00044	0.0064
CDDDEV(t)	0.002585	0.0000
STORDIFF(t)	6.88E-06	0.3475
SHUTIN(t)	3.23E-06	0.2867
Error Correction term (t-1)	(-)1.73E-02	0.0017

Joint Significance: Henry-Hub natural gas price

Variable	*Chi2 Stat*	*p-value*
lagged ΔP HH	*21.43162*	0.01
Exogenous variables	87.66673	0.0000
lagged + Exogenous	147.5958	0.0000

Table 6: Conditional ECM results January 1999-June 2016

Source: Author using data in "EVIEWS" software

For the lag effect of previous week price change, the third-week prior price change has an impact on present week price change at 5% significant level. Rest of the past week's price changes have no impact on current week price changes. The walt-test is used in the analysis of coefficient and to find significant joint level.

8.5 Segmented Time Model and Economical Impact of Variables

We have developed an empirical model to analyse the long-run and short-run relationship between Henry Hub natural gas prices and other variables. As compared to previous studies, the results are similar but modified. We are now interested in segmenting our period of data, January 1999-June 2016, into two parts, January 1999-June 2008 and July 2008-June 2016.

We test the unit root test to confirm the stationarity of time series in both the time frame. The results indicate that the Henry Hub and WTI crude oil price series are integrated of 1st order for, January 1999-June 2008 and July 2008-June 2016.
Later same procedure followed as discussed in methodology, the test assessing the lag length using VAR lag selection criterion followed by the test for cointegration using Johansen-test. During the both the segmented period, Henry Hub natural gas and WTI crude oil prices are observed to be cointegrated. We develop the cointegration equation and error correction model in VECM framework. The VECM model returned long-term relationship, cointegration equation, along with short-term exogenous and lagged effect of times series on each other. The short-run relationship terms capture the volatility in Henry-Hub natural gas prices and the impact of idiosyncratic shocks and disturbance present in the natural gas market due to weather, seasonality, storage levels, and production disruptions.

The results are similar to the ones we found for the in the model:1[4]. The Henry Hub prices are observed to adjust itself to a long-term relationship, whereas WTI crude oil prices are not susceptible to change in Henry Hub natural gas prices. Clearly, the WTI is a global benchmark, and there are global factors which can explain the change in WTI crude oil prices, whereas the Henry-Hub is more of a regional index for natural gas.
Using casualty run on VECM model predict that WTI prices are weekly exogenous to the system. Therefore, the long-run equilibrium is disturbed due to random shocks in the crude oil market. The Henry-Hub is a regional commodity and still there is not significant export to an international market, whereas the crude oil is an international traded commodity and is more of an international benchmark price. So, the regional shocks and Henry-hub natural gas volatility do not affect the crude oil prices, and the results of the model support this theory.

So, we develop a conditional error correction model (ECM) to focus on the Henry Hub price change, as discussed in the section:7.4. A Conditional ECM model is generated using the error term from VECM model. The results illustrate the change in natural gas prices due to the lag effect of previous weeks' natural gas price change, the contemporaneous change in WTI crude oil price change, and the exogenous variables. The results from all the models are compiled and illustrated in Table:7. The

[4] Model:1 is a VECM model for the period of January 1999-June 2016.

comparison is drawn between three models, model:2[5], model:4[6] and model:6[7]. The table is formulated for the Henry Hub price change from 4 $/MMBtu. A negative sign denotes the reduction in Henry Hub prices for an increase in a variable. Table:7, clearly shows the economic impact of the change in variables to Henry Hub prices.

Variables	**Model:2 (conditional ECM: January 1999-June 2016)**		**Model:4 (conditional ECM: January 1999-June 2008)**		**Model:2 (conditional ECM: July 2008-June 2016)**	
	Change in HH from $4/MMBtu per Standard Deviation increase in variable in $	**Percentage change in HH price**	**Change in HH from $4/MMBtu per Standard Deviation increase in variable in $**	**Percentage change in HH price**	**Change in HH from $4/MMBtu per Standard Deviation increase in variable**	**Percentage change in HH price**
HDD	*-0.0011*	*-0.03%*	*0.0383*	*0.96%*	*-0.0059*	*-0.15%*
HDDEV	*0.0856*	*2.14%*	*0.0932*	*2.33%*	*0.0816*	*2.04%*
CDD	*-0.0508*	*-1.27%*	*-0.0365*	*-0.91%*	*-0.0431*	*-1.08%*
CDDDEV	*0.0722*	*1.81%*	*0.0839*	*2.10%*	*0.0501*	*1.25%*
STORDIFF	*0.0093*	*0.23%*	*-0.0549*	*-1.37%*	*0.0098*	*0.24%*
SHUTIN	*0.0107*	*0.27%*	*0.0448*	*1.12%*	*0.0109*	*0.27%*
Δ PHH (-1)	*0.0141*	*0.35%*	*0.0176*	*0.44%*	*0.011*	*0.28%*
Δ PHH (-2)	*-0.0178*	*-0.45%*	*-0.0157*	*-0.39%*	*-0.0117*	*-0.29%*
Δ PHH (-3)	*-0.0223*	*-0.56%*	*-0.0090*	*-0.23%*	*-0.0385*	*-0.96%*
Δ PHH (-4)	*-0.0133*	*-0.33%*	*-0.0235*	*-0.59%*	-	-
Δ PHH (-5)	*-0.0152*	*-0.38%*	-	-	-	-
Δ PHH (-6)	*0.0002*	*0.00%*	-	-	-	-
Δ PHH (-7)	*-0.0088*	*-0.22%*	-	-	-	-
Δ PHH (-8)	*0.0088*	*0.22%*		-	-	-
Δ PHH (-9)	*-0.0147*	*-0.37%*	-	-	-	-
Δ PWTI	*0.0465*	*1.16%*	*0.0549*	*1.37%*	*0.0308*	*0.77%*

Table 7: Economic significance of models

Source: Author using data in "EVIEWS" software

[5] Model:2 is a Conditional Error correction model (ECM) for the period of Jan'1999-Jun' 2016.
[6] Model:4 is a Conditional Error correction model (ECM) for the period of Jan'1999- Jun'2008.
[7] Model:6 is a Conditional Error correction model (ECM) for the period of Jul'2008-Jun'2016.

Estimation from model suggests that a 10% increase in contemporaneous change in WTI crude oil price lead to 2.68% change in Henry-Hub prices for the period of January 1999 - June 2016. Whereas when segmented period analysis carried out, for the period of January 1999-June 2008, the contemporaneous change in WTI by 10% account for more than 3.39% change in natural gas prices. For the period of July 2008 - June 2016, the contemporaneous change of 10% in WTI prices lead to 1.66% change in natural gas prices.

For Model:2 (January 1999-June 2016), when the prior week change also known as lagged effect of natural gas price is analysed, except three-week previous price change, no other have any impact on Henry-Hub price change. The three-week prior price change is also not sufficiently capture the volatility in natural gas prices. Whereas for model:4 (January 1999-June 2008), the earlier week's change in henry-hub prices found to be insignificant. But in the model:6 (July 2008-June 2016), three-week prior change in natural gas price have a significant impact on the Henry Hub natural gas price.

Similar to (Brown & Yücel, 2008), the seasonality effect of hot and cold degree days (HDDEV and CDDDEV) have a significant impact on short-run dynamics of henry-hub natural gas prices. For model:2, the unit increase in standard deviation of HDDEV and CDDDEV lead to 2.14% and 1.80% increase in Natural gas price respectively. For the segmented period, in the model:4, the HDDEV and CDDDEV result in 2.33% and 2.10% increment in natural gas prices respectively, and in the model:5, HDDEV and CDDDEV lead to 2.04% and 1.25% change in prices. In all the three conditional ECM models (model:2, 4 and 6), the impact of HDDEV and CDDDEV is significant, and these results are expected. Logically, with cold waves and heat waves in the US, the demand for natural gas for heating and electricity increases.

The number of cooling degree days was found to be significant in the model:2, with 1.27% decrease in natural gas price for a unit standard deviation increase in CDD. But in Model:4, the value of CDD is not significant and in the model:6, the value is only significant at 10% level. CDD is a proxy for a seasonal shift in demand for natural gas due to summer. The results are logical, as natural gas is used for the household heating purpose, and the increase in CDD implies a decrease in the domestic heating lead to a reduction in demand for natural gas.

The disturbance in natural gas production in the Gulf of Mexico, in the model:2, shows no significant impact on natural gas prices. But when segmented periods are analysed, in the model:4, the dummy variable used to capture the effect of shut-in the production of natural gas at the Gulf of Mexico, SHUTIN, shows a significant impact on the natural gas price change. In model:4, for unit standard deviation increase in SHUTIN, the prices increase by 1.12%. Whereas for model:6, there is no significant impact of SHUTIN.

This result is, to some extent, expected. The onshore shale gas reserves are producing way more than the offshore reserves in the Gulf of Mexico. The offshore production dropped down drastically from more than 8Bcf in 2007 to less than 3 Bcf in 2015. At the same time, the shale gas production increases rapidly from less than 5 Bcf in 2007 to more than 40 Bcf in 2015. That is an increase of more than 500%. Therefore, the share of offshore reserves in total production falls and less likely the Hurricanes have an impact on natural gas prices.

In the case of long-run equilibrium, the speed at which the natural gas prices pulled back towards this equilibrium is measured by error correction term coefficient. In model:2, this coefficient is estimated as -0.017287. Implies that in a case of shocks, the prices correct any deviation from a long-run relationship with a rate of 1.728% per week. But for model:4, this rate was 10.4% and for model:6 it is 5.5%. Apparently wherein model:4, the half-life of shock is 6.3 weeks, in the model:6, it increases to 12.2 weeks.

The period covered in the model:4, January 1999-June 2008, shows a strong long-run relationship and prices are anchored in their long run relationship. But during July 2008-June 2016, as shown in the model:6, the prices are less susceptible to change. The relationship becomes weaker as compared to previous years.

8.6 Conclusion:

The results indicate that there exists a relationship between natural gas and crude oil prices. The segmented model compared the two periods, January 1999-June 2008 and July 2008-June 2016. The result of this model indicates that the relationship is evolving and changing continuously.

9. Conclusion

After analysing the results of the research, it can be concluded that there exist a long-run relationship between natural gas and crude oil prices. The results imply that the two commodities share a relationship which is changing continuously and at the same time becoming weak. There could be several reasons for this change; one could argue that the shale gas revolution, which changes the production process of natural gas could be the cause for this change. Also, with an increase in output of natural gas, the export facilities are developing, shaping the Henry Hub pricing index one of the global benchmarks for natural gas pricing.
But when shale gas technology came into the picture, the world was facing global economic recession. So, it is hard to segregate the impact of global recession and shale gas and point out the cause of the disparity. But at the same time, data shows that recession such as that of 2001, shows no significant impact on the relationship between prices.

It will be interesting to watch this relationship between natural gas and crude oil prices in coming few years with the US becoming a net exporter of natural gas, and Henry Hub index plays a significant role in pricing natural gas globally. At this moment the relationship does exist but becoming weak as compared to last decade.

9.1 Key Findings

In the case of a long-term relationship between natural gas and crude oil prices, the conditional ECM model has proven that there exist a cointegration between the two commodities. Almost 15% of the volatility in the natural gas prices are captured by conditional variables and crude oil prices. The conditional ECM model allows us to analyse the mechanics of fluctuation in natural gas prices in long-term and short-term time frame and examine the natural gas prices volatility in response to disturbance in supply, unexpected change in weather with heat and cold waves, and deviation in average storage level of natural gas.

The question, whether the price series of two commodities decoupled, we examined the entire period in two parts, January 1999-June 2008 and July 2008-June 2016. For the period of January 1999-June 2008, the model predicts, up to 21.89% volatility can be explained by conditional variables and crude oil prices. But we found a drop in this percentage for the period of July 2008-June 2016. The model predicts, up to 16.25% volatility capture by independent variables and WTI crude oil prices.

Regarding the impact of crude oil prices on natural gas prices, for the entire period of January 1999-June 2016, model predict that up to 26.85% change in natural gas prices for the contemporaneous change in WTI crude oil prices, if rest all variables remains same. But for the period of January 1999-June 2008, the contemporaneous change in WTI crude oil prices are accounted for, up to 33.96%. After the crises of 2008, the contemporaneous change in WTI crude oil prices impact on natural gas prices is dropped to 16.68%. Apparently, the change in WTI prices is losing their grip on natural gas prices.

Regarding the decoupling of natural gas prices and crude oil prices, it is apparent that there are instances when the natural gas prices move away from the long run relationship. But for the period of January 1999 to June 2008, the decoupling of natural gas prices was temporary and short-term. The prices return to long run relationship with a half-life of error as 6.29 weeks. But at the same time, the relationship between the two commodities has gradually evolved over the period. There are clear signs of change in the relationship. For the period of July 2008-June 2016, the prices decouple with less susceptible to return to the original long-run relationship. The rate of correcting the deviation and return to the fundamental relationship has increased, and half-life of error rose to 12.37 weeks. The clearly indicates that apart from crude oil prices and conditional variables, there is a significant impact of other factors which are not accounted in the studies. These new factors are significantly dragging the natural gas prices away from a long-term relationship with crude oil prices.

In last, the research concludes that despite few bumps in long-run relationship between natural gas and crude oil prices, there persist a long-run relationship, but the relationship is changing and weakening. However, prior 2010, the relationship changes, found a new equilibrium and become stable, with the new equilibrium. But after 2010, natural gas prices are drifting from the anchored price relationship with oil. Could be the factors like shale gas evolution and Henry Hub breaking a barrier as a regional price index and accepted globally, are changing the long-lived theory of oil and gas relationship.

9.2 Limitation of the Research

The results of this paper have few limitations, which must be kept in mind to fully accept and interpret the results. Firstly, we choose lag of time series using VAR lag selection criteria and prefer AIC criterion. All the criteria for selection lag length are equally acceptable and accommodate all the theoretical aspect of lag selection. But for avoiding complication, in this research, we select lag by AIC criterion. The results can be different with another approach but using the results from all the test in Johansen test confirm the cointegration.

Additionally, the exogenous variables choose in this research are solely based on previous studies and supported by the theoretical backup. There could be other exogenous variables that can also explain the volatility of natural gas prices more significantly. Because there could be many variables, we restrict our variables to six exogenous variables and perform our data analysis based on these six exogenous variables.

Also, using the data from different sources could lead to different results. The data utilised in this research is verified from multiple sources but still found some disputed data from various sources.

9.3 Suggestion for Future Research

As this research is focused on analysing the natural gas and crude oil relationship, there are multiple directions in which one can extend this research. First, this research only uses six exogenous variables and analyse the relationship between two commodities based on these variables. As we found out that for the period July 2008-June 2016, the natural gas prices are reluctant to return to a long-term relationship, there is scope for including other dummy variables to capture the natural gas price volatility. The recent development of US as an exporter of natural gas and increased shale gas production, lead Henry-Hub as an increasingly globally acceptable pricing index.

This paper, limit its focus only on natural gas and does not include the impact of Liquefied natural gas (LNG) market. LNG is liquefied form of natural gas, used in long distance transportation of natural gas. With rising US natural gas production and recent development in Liquification capacity in the US, any disturbance in LNG market significantly impacts the natural gas market. The future research could include global variables as well in analysing natural gas prices along with dynamics involve in LNG market.

In the last, with serious environmental concern and global hue and cry to reduce pollution, natural gas has a bright future. For last several decades, globally, Oil and Coal are leading sources, but countries are investing heavily in natural gas infrastructure. The price relationship between natural gas and crude oil has given the market power to anticipate the future in the energy sector. But with increasing natural gas acceptance as a global commodity and developing a global natural gas pricing index, market needs to find a new linkage with the natural gas price to anticipate future. As such, the way energy market is developed so far, it is on the verge of a major transaction.

Bibliography

Adelman, M., & Watkins, G. (1996). The value of United States oil and gas reserves (MIT-CEEPR (Series) No. 96-004WP). Retrieved from http://hdl.handle.net/1721.1/50224

Arabia, S. (2005). EXAMINING THE LONG-RUN RELATION AMONG SPOT PRICES OF.

Bachmeier, L. J., & Griffin, J. M. (2006). Testing for Market Integration Crude Oil , Coal , and Natural Gas. Energy, 55–71. Retrieved from http://www.jstor.org/stable/23297019

Barron, M. J., & Brown, S. P. A. (1986). Assessing the Market for Natural Gas. Texas A & M Business Forum (Vol. Fall).

Bashiri Behmiri, N., & Pires Manso, J. R. (2013). Crude Oil Price Forecasting Techniques: A Comprehensive Review of Literature. SSRN Electronic Journal, 30–48. http://doi.org/10.2139/ssrn.2275428

Bloomberg. (2016a). Has Natural Gas Hit Bottom? Retrieved April 10, 2016, from http://www.bloomberg.com/news/videos/2016-03-31/has-natural-gas-hit-bottom

Bloomberg. (2016b). Next Stop for U.S. Natural Gas Is 20-Year Low Amid Warm Weather. Retrieved June 20, 2016, from http://www.bloomberg.com/news/articles/2015-12-15/natural-gas-futures-extend-slide-to-14-year-low-on-mild-december

Box, George, E. ., & Jenkins, G. (1970). Time series analysis: Forecasting and control.

Brown, S. P. a., & Yücel, M. K. (2008). What drives natural gas prices? Energy Journal, 29, 45. Retrieved from http://mail1.dallasfed.org/research/papers/2007/wp0703.pdf

Costello, K., Huntingon, H. G., & Wilson, J. F. (2005). After the Natural Gas Bubble: An Economic Evaluation of the Recent U.S. National Petroleum Council Stud. The Energy Journal, 26(2), 89–110.

Dickey, D. A., Fuller, W. A., Dickey, B. Y. D. A., & Fuller, W. A. (2016). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. Econometric Society Stable, 49(4), 1057–1072. Retrieved from http://www.jstor.org/stable/1912517

Dwyer, G. P. (2014). The Johansen Tests for Cointegration, (April), 1–7.

Enders, W. (2015). Applied Econometric Time Series. CEUR Workshop Proceedings (Vol. 1542). http://doi.org/10.1017/CBO9781107415324.004

Engle, R. F., & Granger, C. W. J. (1987). Co-Integration and Error Correction : Representation , Estimation , and Testing. Econometrica, 55(2), 251–276. http://doi.org/10.2307/1913236

Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. Journal of Econometrics, 16(1), 121–130. http://doi.org/10.1016/0304-4076(81)90079-8

Hamilton, J. (1994). Time Series Analysis. pup.princeton.edu. Retrieved from https://pup.princeton.edu/titles/5386.html

Hartley, P. R., Medlock Iii, K. B., Rosthal, J. E., & Medlock, K. B. (2008a). The Relationship of Natural Gas to Oil Prices. Source: The Energy Journal, 29(3), 47–65.

Hartley, P. R., Medlock Iii, K. B., Rosthal, J. E., & Medlock, K. B. (2008b). The Relationship of Natural Gas to Oil Prices. Source: The Energy Journal, 29(3), 47–65. Retrieved from http://www.jstor.org/stable/41323169\nhttp://www.jstor.org/page/

Hendry, D. F., & Juselius, K. (2000). Explaining Cointegration Analysis: Part 1. The Energy Journal, 21(1), 1–42. Retrieved from http://www.jstor.org/stable/41322853

Hendry, D. F., & Juselius, K. (2001). Explaining Cointegration Analysis: Part II. The Energy Journal, 22(1), 1–34. http://doi.org/10.5547/ISSN0195-6574-EJ-Vol22-No1-4

Krichene, N. (2002). World crude oil and natural gas: A demand and supply model. Energy Economics, 24(6), 557–576. http://doi.org/10.1016/S0140-9883(02)00061-0

Nason, G. P. (2006). Stationary and Non-stationary Time Series. Statistics in Volcanology, (1994), 129–143. Retrieved from http://www.cas.usf.edu/~cconnor/geolsoc/html/chapter11.pdf

Neumann, A. (2009). Linking natural gas markets - Is LNG doing its job? Energy Journal, 30(SPECIAL ISSUE 1), 187–200. http://doi.org/10.5547/ISSN0195-6574-EJ-Vol30-NoSI-12

Panagiotidis, T., & Rutledge, E. (2007). Oil and gas markets in the UK: Evidence from a cointegrating approach. Energy Economics, 29(2), 329–347. http://doi.org/10.1016/j.eneco.2006.10.013

Pankratz, A. (1983). Forecasting with Univariate Box - Jenkins Models: Concepts and Cases. http://doi.org/10.1002/9780470316566

Phillips, P. C. B. (1986). Understanding spurious regressions in econometrics. Journal of Econometrics, 33(3), 311–340. http://doi.org/10.1016/0304-4076(86)90001-1

Rakesh Upadhayay. (2016). Can The Natural Gas Rally Continue? Retrieved July 10, 2016, from http://oilprice.com/Energy/Energy-General/Can-The-Natural-Gas-Rally-Continue.html

Ramberg, D. J., & Pasrsons, J. E. (2010). The Weak Tie Between Natural Gas and Oil Prices.

Runion, M. L. (2002). the History of, (November), 2000–2002. http://doi.org/Book Review

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika. http://doi.org/10.1093/biomet/71.3.599

Serletis, A., & Herbert, J. (1999). The message in North American energy prices. Energy Economics, 21(5), 471–483. http://doi.org/10.1016/S0140-9883(99)00015-8

Serletis, A., & Rangel-Ruiz, R. (2004). Testing for common features in North American energy markets. Energy Economics, 26(3), 401–414. http://doi.org/10.1016/j.eneco.2004.04.007

Smith, J. L. (2004). Petroleum Property Valuation. Encyclopedia of Energy, 811–822. http://doi.org/http://dx.doi.org/10.1016/B0-12-176480-X/00136-4

Thomson, D. (1994). Jackknifing multiple-window spectra. Acoustics, Speech, and Signal Processing, 1994. ICASSP-94., 1994 IEEE International Conference on, vi, VI/73 – VI/76. http://doi.org/10.1109/ICASSP.1994.389899

U.S. Energy Information Administration. (2015). Electric Power Monthly: with data for December 2014. EIA. Retrieved from http://www.eia.gov/electricity/monthly/epm_table_grapher.cfm?t=epmt_1_01

Union Gas. (2011). Chemical composition of natural gas. Retrieved July 10, 2016, from https://www.uniongas.com/about-us/about-natural-gas/Chemical-Composition-of-Natural-Gas

Villar, J. A., & Joutz, F. L. (2006). The relationship between crude oil and natural gas prices. Information Administration, Office of Oil and Gas, 37. Retrieved from http://aceer.uprm.edu/aceer/pdfs/CrudeOil_NaturalGas.pdf

Appendix 1: Results from "EVIEWs" (January 1999-June 2016)

Vector Autoregression Estimates

Vector Autoregression Estimates
Date: 08/11/16 Time: 15:15
Sample (adjusted): 1/15/1999 7/15/2016
Included observations: 914 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	LN_HH	LN_WTI
LN_HH(-1)	1.027267	-0.004474
	(0.03366)	(0.02036)
	[30.5211]	[-0.21968]
LN_HH(-2)	-0.046036	0.011276
	(0.03354)	(0.02029)
	[-1.37259]	[0.55566]
LN_WTI(-1)	0.113412	1.150413
	(0.05479)	(0.03315)
	[2.07010]	[34.7060]
LN_WTI(-2)	-0.106622	-0.153866
	(0.05468)	(0.03308)
	[-1.94997]	[-4.65097]
CDD	-0.000282	2.62E-05
	(0.00015)	(9.1E-05)
	[-1.87236]	[0.28735]
HDD	4.60E-05	2.92E-05
	(5.1E-05)	(3.1E-05)
	[0.90675]	[0.95090]
CDDDEV	0.002385	0.000470
	(0.00042)	(0.00025)
	[5.66361]	[1.84407]
HDDEV	0.000923	-9.25E-07
	(0.00013)	(8.0E-05)
	[7.02578]	[-0.01164]
STORDIFF	1.04E-05	1.05E-05
	(7.4E-06)	(4.5E-06)
	[1.39091]	[2.32236]
SHUTIN	2.23E-06	-6.25E-06
	(3.1E-06)	(1.9E-06)
	[0.72308]	[-3.35495]
R-squared	0.976130	0.994045
Adj. R-squared	0.975893	0.993986
Sum sq. resids	4.442605	1.626300
S.E. equation	0.070103	0.042415
F-statistic	4107.559	16767.23
Log likelihood	1137.342	1596.596
Akaike AIC	-2.466831	-3.471764
Schwarz SC	-2.414119	-3.419052
Mean dependent	1.459623	3.966777
S.D. dependent	0.451501	0.546927

Determinant resid covariance (dof adj.)	8.62E-06
Determinant resid covariance	8.43E-06
Log likelihood	2745.427
Akaike information criterion	-5.963735
Schwarz criterion	-5.858312

Table 8: VAR model January 1999-June 2016

Source: Author using data in "EVIEWS" software

VAR Lag Order Selection Criteria
Endogenous variables: LN_HH LN_WTI
Exogenous variables: CDD HDD CDDDEV HDDEV STORDIFF SHUTIN
Date: 08/11/16 Time: 15:15
Sample: 1/01/1999 7/15/2016
Included observations: 904

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-1833.632	NA	0.203417	4.083258	4.147065	4.107627
1	2701.587	8990.170	9.01e-06	-5.941565	-5.856488*	-5.909072
2	2714.297	25.13881	8.84e-06	-5.960834	-5.854489	-5.920218
3	2719.110	9.496971	8.82e-06	-5.962632	-5.835017	-5.913892
4	2731.308	24.01922	8.66e-06	-5.980770	-5.831886	-5.923907*
5	2737.642	12.44278	8.62e-06	-5.985933	-5.815779	-5.920947
6	2741.743	8.038773	8.62e-06	-5.986156	-5.794734	-5.913047
7	2747.987	12.21288	8.57e-06	-5.991122	-5.778431	-5.909890
8	2750.252	4.420042	8.61e-06	-5.987284	-5.753323	-5.897928
9	2756.972	13.08161*	8.55e-06	-5.993300	-5.738070	-5.895821
10	2761.420	8.641379	8.55e-06*	-5.994292*	-5.717793	-5.888690
11	2762.409	1.916355	8.60e-06	-5.987630	-5.689862	-5.873905
12	2766.820	8.528575	8.60e-06	-5.988539	-5.669502	-5.866690

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

Table 9: VAR Lag selection criteria January 1999 - June 2016

Source: Author using data in “EVIEWS” software

Vector Error Correction Estimates

Vector Error Correction Estimates
Date: 09/04/16 Time: 19:20
Sample (adjusted): 3/12/1999 7/15/2016
Included observations: 906 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1	
LN_HH(-1)	1.000000	
LN_WTI(-1)	0.160421	
	(0.23436)	
	[0.68449]	
C	-2.105110	
Error Correction:	**D(LN_HH)**	**D(LN_WTI)**
CointEq1	-0.018538	-0.000220
	(0.00553)	(0.00333)
	[-3.35418]	[-0.06623]
D(LN_HH(-1))	0.028420	-0.010012
	(0.03371)	(0.02029)
	[0.84306]	[-0.49351]
D(LN_HH(-2))	-0.060983	0.005318
	(0.03287)	(0.01978)
	[-1.85546]	[0.26889]
D(LN_HH(-3))	-0.088201	-0.042821
	(0.03271)	(0.01968)
	[-2.69685]	[-2.17566]
D(LN_HH(-4))	-0.021868	0.058241
	(0.03266)	(0.01965)
	[-0.66962]	[2.96338]
D(LN_HH(-5))	-0.065131	-0.012841
	(0.03276)	(0.01972)
	[-1.98783]	[-0.65121]
D(LN_HH(-6))	-0.018288	-0.010537
	(0.03283)	(0.01976)
	[-0.55703]	[-0.53332]
D(LN_HH(-7))	-0.032761	0.019923
	(0.03286)	(0.01978)
	[-0.99695]	[1.00743]
D(LN_HH(-8))	0.014203	-0.018362
	(0.03286)	(0.01977)
	[0.43225]	[-0.92855]
D(LN_HH(-9))	-0.064877	-0.022459
	(0.03269)	(0.01968)
	[-1.98438]	[-1.14149]
D(LN_WTI(-1))	0.129189	0.192121
	(0.05639)	(0.03394)
	[2.29101]	[5.66140]
D(LN_WTI(-2))	-0.045675	-0.104152
	(0.05715)	(0.03439)
	[-0.79928]	[-3.02853]
D(LN_WTI(-3))	0.036611	0.153089
	(0.05751)	(0.03461)
	[0.63656]	[4.42305]
D(LN_WTI(-4))	-0.067425	-0.068219
	(0.05797)	(0.03489)
	[-1.16303]	[-1.95535]
D(LN_WTI(-5))	0.031510	0.064432
	(0.05812)	(0.03498)
	[0.54219]	[1.84222]
D(LN_WTI(-6))	0.119548	-0.056648
	(0.05781)	(0.03479)
	[2.06789]	[-1.62824]
D(LN_WTI(-7))	0.088245	-0.020874
	(0.05740)	(0.03454)
	[1.53742]	[-0.60431]
D(LN_WTI(-8))	0.078546	0.103560
	(0.05708)	(0.03435)
	[1.37603]	[3.01526]
D(LN_WTI(-9))	0.095814	0.051809
	(0.05668)	(0.03411)
	[1.69057]	[1.51901]
C	0.002697	-0.002851
	(0.00877)	(0.00528)
	[0.30761]	[-0.54039]
HDD	4.62E-05	3.22E-05
	(5.7E-05)	(3.4E-05)
	[0.80631]	[0.93333]
HDDEV	0.001050	1.85E-05
	(0.00014)	(8.3E-05)
	[7.58793]	[0.22205]
CDD	-0.000393	-5.49E-06
	(0.00016)	(9.8E-05)
	[-2.41454]	[-0.05607]
CDDDEV	0.002797	0.000539
	(0.00042)	(0.00026)
	[6.59928]	[2.11305]
STORDIFF	4.95E-06	6.49E-06
	(7.5E-06)	(4.5E-06)
	[0.66013]	[1.43953]
SHUTIN	3.25E-06	-5.03E-06
	(3.1E-06)	(1.9E-06)
	[1.05149]	[-2.70740]
R-squared	0.147659	0.103773
Adj. R-squared	0.123445	0.078312
Sum sq. resids	4.211771	1.525353
S.E. equation	0.069182	0.041634
F-statistic	6.098038	4.075762
Log likelihood	1147.575	1607.668

Table 10: VECM model January 1999-June 2016

Source: Author using data in "EVIEWS" software

Dependent Variable: D(LN_HH)
Method: Least Squares (Gauss-Newton / Marquardt steps)
Date: 08/13/16 Time: 19:44
Sample (adjusted): 3/12/1999 7/15/2016
Included observations: 906 after adjustments
D(LN_HH) = C(1)*(LN_HH(-1) + 0.160420864685*LN_WTI(-1) - 2.10510960617) + C(2)*D(LN_HH(-1)) + C(3)*D(LN_HH(-2)) + C(4) *D(LN_HH(-3)) + C(5)*D(LN_HH(-4)) + C(6)*D(LN_HH(-5)) + C(7) *D(LN_HH(-6)) + C(8)*D(LN_HH(-7)) + C(9)*D(LN_HH(-8)) + C(10) *D(LN_HH(-9)) + C(11)*D(LN_WTI(-1)) + C(12)*D(LN_WTI(-2)) + C(13)*D(LN_WTI(-3)) + C(14)*D(LN_WTI(-4)) + C(15)*D(LN_WTI(-5)) + C(16)*D(LN_WTI(-6)) + C(17)*D(LN_WTI(-7)) + C(18) *D(LN_WTI(-8)) + C(19)*D(LN_WTI(-9)) + C(20) + C(21)*HDD + C(22)*HDDEV + C(23)*CDD + C(24)*CDDDEV + C(25) *STORDIFF + C(26)*SHUTIN

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.018538	0.005527	-3.354181	0.0008
C(2)	0.028420	0.033711	0.843059	0.3994
C(3)	-0.060983	0.032867	-1.855462	0.0639
C(4)	-0.088201	0.032705	-2.696852	0.0071
C(5)	-0.021868	0.032658	-0.669615	0.5033
C(6)	-0.065131	0.032765	-1.987835	0.0471
C(7)	-0.018288	0.032831	-0.557027	0.5777
C(8)	-0.032761	0.032861	-0.996951	0.3191
C(9)	0.014203	0.032859	0.432249	0.6657
C(10)	-0.064877	0.032694	-1.984381	0.0475
C(11)	0.129189	0.056390	2.291009	0.0222
C(12)	-0.045675	0.057145	-0.799284	0.4243
C(13)	0.036611	0.057513	0.636559	0.5246
C(14)	-0.067425	0.057973	-1.163035	0.2451
C(15)	0.031510	0.058117	0.542186	0.5878
C(16)	0.119548	0.057812	2.067895	0.0389
C(17)	0.088245	0.057398	1.537423	0.1245
C(18)	0.078546	0.057082	1.376026	0.1692
C(19)	0.095814	0.056676	1.690571	0.0913
C(20)	0.002697	0.008767	0.307607	0.7585
C(21)	4.62E-05	5.73E-05	0.806308	0.4203
C(22)	0.001050	0.000138	7.587934	0.0000
C(23)	-0.000393	0.000163	-2.414537	0.0160
C(24)	0.002797	0.000424	6.599275	0.0000
C(25)	4.95E-06	7.49E-06	0.660132	0.5093
C(26)	3.25E-06	3.09E-06	1.051491	0.2933

R-squared	0.147659	Mean dependent var	0.000561
Adjusted R-squared	0.123445	S.D. dependent var	0.073893
S.E. of regression	0.069182	Akaike info criterion	-2.475884
Sum squared resid	4.211771	Schwarz criterion	-2.337876
Log likelihood	1147.575	Hannan-Quinn criter.	-2.423181
F-statistic	6.098038	Durbin-Watson stat	1.914275
Prob(F-statistic)	0.000000		

Table 11: VECM model equation January 1999-June 2016

Source: Author using data in "EVIEWS" software

Dependent Variable: D(LN_HH)
Method: ARDL
Date: 08/16/16 Time: 19:31
Sample (adjusted): 3/12/1999 7/15/2016
Included observations: 906 after adjustments
Dependent lags: 9 (Fixed)
Dynamic regressors (0 lag, fixed): D(LN_WTI)
Fixed regressors: HDD HDDEV CDD CDDDEV STORDIFF SHUTIN ECT(-1) C

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
D(LN_HH(-1))	0.047781	0.032833	1.455265	0.1459
D(LN_HH(-2))	-0.060385	0.032044	-1.884416	0.0598
D(LN_HH(-3))	-0.075682	0.031950	-2.368758	0.0181
D(LN_HH(-4))	-0.045023	0.032021	-1.406017	0.1601
D(LN_HH(-5))	-0.051566	0.031842	-1.619415	0.1057
D(LN_HH(-6))	0.000590	0.031929	0.018463	0.9853
D(LN_HH(-7))	-0.028641	0.031939	-0.896722	0.3701
D(LN_HH(-8))	0.029810	0.031869	0.935400	0.3498
D(LN_HH(-9))	-0.049807	0.031695	-1.571443	0.1164
D(LN_WTI)	0.268581	0.053637	5.007371	0.0000
HDD	-3.52E-06	5.53E-05	-0.063580	0.9493
HDDEV	0.001062	0.000137	7.769730	0.0000
CDD	-0.000440	0.000161	-2.733093	0.0064
CDDDEV	0.002585	0.000421	6.144384	0.0000
STORDIFF	6.88E-06	7.32E-06	0.939870	0.3475
SHUTIN	3.23E-06	3.03E-06	1.065928	0.2867
ECT(-1)	-0.017287	0.005478	-3.155378	0.0017
C	0.008656	0.008543	1.013291	0.3112

R-squared	0.150016	Mean dependent var	0.000561
Adjusted R-squared	0.133743	S.D. dependent var	0.073893
S.E. of regression	0.068774	Akaike info criterion	-2.496312
Sum squared resid	4.200127	Schwarz criterion	-2.400769
Log likelihood	1148.830	Hannan-Quinn criter.	-2.459826
F-statistic	9.219127	Durbin-Watson stat	1.925655
Prob(F-statistic)	0.000000		

*Note: p-values and any subsequent tests do not account for model selection.

Table 12: Conditional ECM model January 1999-June 2016

Source: Author using data in "EVIEWS" software

Appendix 2: Results from "EVIEWs" (January 1999-June 2008)

Vector Autoregression Estimates

Vector Autoregression Estimates
Date: 08/22/16 Time: 00:44
Sample (adjusted): 1/15/1999 6/27/2008
Included observations: 494 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	LN_HH	LN_WTI
LN_HH(-1)	0.972977	0.017323
	(0.04573)	(0.02464)
	[21.2765]	[0.70293]
LN_HH(-2)	-0.043732	-0.013755
	(0.04454)	(0.02400)
	[-0.98182]	[-0.57302]
LN_WTI(-1)	0.093503	1.145097
	(0.08464)	(0.04561)
	[1.10474]	[25.1050]
LN_WTI(-2)	-0.058699	-0.148631
	(0.08452)	(0.04555)
	[-0.69448]	[-3.26303]
HDD	3.63E-05	6.56E-05
	(7.3E-05)	(3.9E-05)
	[0.49860]	[1.67307]
HDDEV	0.001225	1.73E-05
	(0.00019)	(0.00010)
	[6.59943]	[0.17335]
CDD	-0.000516	0.000149
	(0.00022)	(0.00012)
	[-2.35935]	[1.26142]
CDDDEV	0.003599	0.000531
	(0.00062)	(0.00033)
	[5.79325]	[1.58657]
STORDIFF	-2.13E-05	9.17E-06
	(1.3E-05)	(7.2E-06)
	[-1.59210]	[1.27227]
SHUTIN	8.80E-06	-1.98E-06
	(4.1E-06)	(2.2E-06)
	[2.15246]	[-0.90004]
R-squared	0.977123	0.994008
Adj. R-squared	0.976698	0.993896
Sum sq. resids	2.646233	0.768523
S.E. equation	0.073942	0.039848
F-statistic	2296.973	8920.530
Log likelihood	590.7058	896.1018
Akaike AIC	-2.351035	-3.587457
Schwarz SC	-2.265964	-3.502385
Mean dependent	1.606295	3.667323
S.D. dependent	0.484387	0.510040
Determinant resid covariance (dof adj.)		8.44E-06
Determinant resid covariance		8.10E-06
Log likelihood		1493.902
Akaike information criterion		-5.967216
Schwarz criterion		-5.797073

Table 13: VAR model January 1999-June 2008

Source: Author using data in "EVIEWS" software

VAR Lag Order Selection Criteria
Endogenous variables: LN_HH LN_WTI
Exogenous variables: HDD HDDEV CDD CDDDEV STORDIFF SHUTIN
Date: 08/22/16 Time: 00:44
Sample: 1/01/1999 6/27/2008
Included observations: 484

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-723.4783	NA	0.071605	3.039166	3.142854	3.079910
1	1457.633	4290.119	8.87e-06	-5.957160	-5.818909*	-5.902836*
2	1463.883	12.24235	8.79e-06	-5.966459	-5.793645	-5.898553
3	1469.562	11.07612	8.73e-06	-5.973396	-5.766020	-5.891910
4	1473.031	6.737294	8.75e-06	-5.971202	-5.729263	-5.876134
5	1480.837	15.09708	8.61e-06*	-5.986932*	-5.710430	-5.878283
6	1481.558	1.387000	8.73e-06	-5.973379	-5.662315	-5.851149
7	1487.009	10.45286	8.68e-06	-5.979378	-5.633751	-5.843567
8	1489.725	5.185003	8.72e-06	-5.974072	-5.593883	-5.824680
9	1490.983	2.390784	8.82e-06	-5.962740	-5.547988	-5.799767
10	1493.665	5.074940	8.87e-06	-5.957292	-5.507977	-5.780738
11	1495.591	3.630124	8.95e-06	-5.948724	-5.464847	-5.758589
12	1501.459	11.00904*	8.88e-06	-5.956444	-5.438004	-5.752728

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

Table 14: VAR Lag selection criteria January 1999-June 2008

Source: Author using data in "EVIEWS" software

Johansen Cointegration Test

Date: 08/18/16 Time: 20:33
Sample (adjusted): 2/05/1999 6/27/2008
Included observations: 491 after adjustments
Trend assumption: Linear deterministic trend
Series: LN_HH LN_WTI
Exogenous series: HDD HDDEV CDD CDDDEV STORDIFF SHUTIN
Warning: Critical values assume no exogenous series
Lags interval (in first differences): 1 to 4

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.071742	37.86107	15.41	20.04
At most 1	0.002661	1.308204	3.76	6.65

Trace test indicates 1 cointegrating equation(s) at both 5% and 1% levels
*(**) denotes rejection of the hypothesis at the 5%(1%) level

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.071742	36.55286	14.07	18.63
At most 1	0.002661	1.308204	3.76	6.65

Max-eigenvalue test indicates 1 cointegrating equation(s) at both 5% and 1% le
*(**) denotes rejection of the hypothesis at the 5%(1%) level

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):

LN_HH	LN_WTI
-5.376557	4.059369
-0.667688	2.615783

Unrestricted Adjustment Coefficients (alpha):

D(LN_HH)	0.019121	-0.001088
D(LN_WTI)	-0.001248	-0.001973

1 Cointegrating Equation(s): Log likelihood 1505.915

Normalized cointegrating coefficients (standard error in parentheses)

LN_HH	LN_WTI
1.000000	-0.755013
	(0.06433)

Adjustment coefficients (standard error in parentheses)

D(LN_HH)	-0.102805
	(0.01775)
D(LN_WTI)	0.006711
	(0.00950)

Table 15: Johansen test for cointegration January 1999-June 2008

Source: Author using data in "EVIEWS" software

Vector Error Correction Estimates

Vector Error Correction Estimates Date: 08/18/16 Time: 20:35 Sample (adjusted): 2/05/1999 6/27/2008 Included observations: 491 after adjustments Standard errors in ( ) & t-statistics in [ ]		
Cointegrating Eq:	CointEq1	
LN_HH(-1)	1.000000	
LN_WTI(-1)	-0.755013 (0.06433) [-11.7370]	
C	1.161989	
Error Correction:	D(LN_HH)	D(LN_WTI)
CointEq1	-0.102805 (0.01775) [-5.79288]	0.006711 (0.00950) [0.70620]
D(LN_HH(-1))	0.059027 (0.04457) [1.32447]	0.012280 (0.02386) [0.51460]
D(LN_HH(-2))	-0.043316 (0.04371) [-0.99087]	0.021560 (0.02341) [0.92109]
D(LN_HH(-3))	-0.044437 (0.04330) [-1.02634]	-0.046039 (0.02318) [-1.98584]
D(LN_HH(-4))	-0.044873 (0.04319) [-1.03891]	0.066969 (0.02313) [2.89558]
D(LN_WTI(-1))	0.062905 (0.08604) [0.73112]	0.194823 (0.04607) [4.22881]
D(LN_WTI(-2))	-0.043880 (0.08692) [-0.50480]	-0.150054 (0.04654) [-3.22386]
D(LN_WTI(-3))	0.052225 (0.08650) [0.60376]	0.100307 (0.04632) [2.16568]
D(LN_WTI(-4))	-0.098262 (0.08557) [-1.14826]	-0.105813 (0.04582) [-2.30922]
C	0.001554 (0.01224) [0.12697]	-0.005154 (0.00655) [-0.78646]
HDD	0.000147 (8.0E-05) [1.83293]	5.36E-05 (4.3E-05) [1.24574]
HDDEV	0.001154 (0.00019)	3.16E-05 (0.00010)

Vector Error Correction Estimates

	[5.94786]	[0.30375]
CDD	-0.000282 (0.00023) [-1.20825]	0.000122 (0.00013) [0.97811]
CDDDEV	0.003407 (0.00063) [5.42681]	0.000599 (0.00034) [1.78160]
STORDIFF	-4.34E-05 (1.5E-05) [-2.96626]	1.21E-05 (7.8E-06) [1.54380]
SHUTIN	1.11E-05 (4.1E-06) [2.69339]	-2.97E-06 (2.2E-06) [-1.34817]
R-squared	0.193900	0.090861
Adj. R-squared	0.168444	0.062152
Sum sq. resids	2.541019	0.728556
S.E. equation	0.073140	0.039164
F-statistic	7.617107	3.164835
Log likelihood	595.5835	902.2757
Akaike AIC	-2.360829	-3.610084
Schwarz SC	-2.224081	-3.473337
Mean dependent	0.004055	0.004883
S.D. dependent	0.080207	0.040441
Determinant resid covariance (dof adj.)		7.94E-06
Determinant resid covariance		7.43E-06
Log likelihood		1505.915
Akaike information criterion		-5.995579
Schwarz criterion		-5.704991

Table 16: VECM Model January 1999-June 2008

Source: Author using data in "EVIEWS" software

Dependent Variable: D(LN_HH)
Method: ARDL
Date: 08/18/16 Time: 20:49
Sample (adjusted): 2/05/1999 6/27/2008
Included observations: 491 after adjustments
Dependent lags: 4 (Fixed)
Dynamic regressors (0 lag, fixed): D(LN_WTI) ECT(-1)
Fixed regressors: HDD HDDEV CDD CDDDEV STORDIFF SHUTIN C

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
D(LN_HH(-1))	0.054768	0.042976	1.274387	0.2031
D(LN_HH(-2))	-0.048898	0.042015	-1.163832	0.2451
D(LN_HH(-3))	-0.028042	0.041728	-0.672017	0.5019
D(LN_HH(-4))	-0.073230	0.041698	-1.756209	0.0797
D(LN_WTI)	0.339555	0.081672	4.157566	0.0000
ECT(-1)	-0.104280	0.017229	-6.052685	0.0000
HDD	0.000128	7.88E-05	1.620778	0.1057
HDDEV	0.001133	0.000188	6.012258	0.0000
CDD	-0.000326	0.000228	-1.432288	0.1527
CDDDEV	0.003185	0.000612	5.202695	0.0000
STORDIFF	-4.75E-05	1.44E-05	-3.306796	0.0010
SHUTIN	1.20E-05	4.04E-06	2.974906	0.0031
C	0.003315	0.012010	0.276050	0.7826

R-squared	0.218924	Mean dependent var	0.004055
Adjusted R-squared	0.199315	S.D. dependent var	0.080207
S.E. of regression	0.071770	Akaike info criterion	-2.404585
Sum squared resid	2.462136	Schwarz criterion	-2.293477
Log likelihood	603.3255	Hannan-Quinn criter.	-2.360952
F-statistic	11.16469	Durbin-Watson stat	1.883659
Prob(F-statistic)	0.000000		

*Note: p-values and any subsequent tests do not account for model selection.

Table 17: Conditional ECM January 1999-June 2008

Source: Author using data in “EVIEWS” software

Appendix 3: Results from "EVIEWs" (July 2008-June 2016)

Vector Autoregression Estimates

Vector Autoregression Estimates
Date: 09/04/16 Time: 20:04
Sample (adjusted): 7/18/2008 7/15/2016
Included observations: 418 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	LN_HH	LN_WTI
LN_HH(-1)	0.976388	-0.063492
	(0.05040)	(0.03684)
	[19.3733]	[-1.72362]
LN_HH(-2)	-0.036394	0.063431
	(0.04888)	(0.03573)
	[-0.74451]	[1.77535]
LN_WTI(-1)	0.110449	1.127343
	(0.06689)	(0.04889)
	[1.65128]	[23.0601]
LN_WTI(-2)	-0.090695	-0.126488
	(0.06713)	(0.04906)
	[-1.35111]	[-2.57812]
HDD	-3.49E-05	-3.12E-05
	(7.1E-05)	(5.2E-05)
	[-0.49346]	[-0.60230]
HDDEV	0.000974	7.54E-05
	(0.00018)	(0.00013)
	[5.32087]	[0.56344]
CDD	-0.000383	-0.000168
	(0.00021)	(0.00016)
	[-1.80411]	[-1.08203]
CDDDEV	0.001665	0.000562
	(0.00055)	(0.00040)
	[3.00657]	[1.38739]
STORDIFF	4.76E-07	8.73E-06
	(9.2E-06)	(6.7E-06)
	[0.05164]	[1.29576]
SHUTIN	4.78E-06	-1.42E-05
	(5.0E-06)	(3.7E-06)
	[0.95381]	[-3.88403]
R-squared	0.965281	0.982821
Adj. R-squared	0.964515	0.982443
Sum sq. resids	1.532357	0.818599
S.E. equation	0.061284	0.044793
F-statistic	1260.400	2593.621
Log likelihood	579.0966	710.1329
Akaike AIC	-2.722950	-3.349918
Schwarz SC	-2.626408	-3.253375
Mean dependent	1.281092	4.315972
S.D. dependent	0.325335	0.338045
Determinant resid covariance (dof adj.)		7.42E-06
Determinant resid covariance		7.07E-06
Log likelihood		1292.334
Akaike information criterion		-6.087724
Schwarz criterion		-5.894638

Table 18: VAR model July 2008-June2016

Source: Author using data in "EVIEWS" software

VAR Lag Order Selection Criteria
Endogenous variables: LN_HH LN_WTI
Exogenous variables: HDD HDDEV CDD CDDDEV STORDIFF SHUTIN
Date: 08/22/16 Time: 01:00
Sample: 6/30/2008 7/11/2016
Included observations: 408

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-581.1379	NA	0.062774	2.907539	3.025517	2.954223
1	1252.740	3595.839	7.98e-06	-6.062450	-5.905146*	-6.000205*
2	1258.445	11.13023	7.92e-06	-6.070808	-5.874177	-5.993001
3	1261.462	5.857019	7.95e-06	-6.065991	-5.830034	-5.972622
4	1272.297	20.92552	7.69e-06*	-6.099493*	-5.824210	-5.990563
5	1274.996	5.187224	7.74e-06	-6.093118	-5.778509	-5.968626
6	1277.944	5.635676	7.78e-06	-6.087961	-5.734025	-5.947907
7	1282.324	8.330648	7.77e-06	-6.089824	-5.696562	-5.934209
8	1284.165	3.483551	7.85e-06	-6.079240	-5.646653	-5.908064
9	1289.954	10.89706*	7.78e-06	-6.088010	-5.616097	-5.901273
10	1291.185	2.304097	7.89e-06	-6.074434	-5.563194	-5.872135
11	1292.157	1.810909	8.01e-06	-6.059592	-5.509026	-5.841731
12	1296.454	7.963131	8.00e-06	-6.061051	-5.471158	-5.827628

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

Table 19: Lag selection criteria July 2008-June 2016

Source: Author using data in "EVIEWS" software

Johansen Cointegration Test

Date: 08/18/16 Time: 20:59
Sample (adjusted): 8/01/2008 7/15/2016
Included observations: 416 after adjustments
Trend assumption: Linear deterministic trend
Series: LN_HH LN_WTI
Exogenous series: HDD HDDEV CDD CDDDEV STORDIFF SHUTIN
Warning: Critical values assume no exogenous series
Lags interval (in first differences): 1 to 3

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.051936	22.83256	15.41	20.04
At most 1	0.001551	0.645731	3.76	6.65

Trace test indicates 1 cointegrating equation(s) at both 5% and 1% levels
*(**) denotes rejection of the hypothesis at the 5%(1%) level

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.051936	22.18683	14.07	18.63
At most 1	0.001551	0.645731	3.76	6.65

Max-eigenvalue test indicates 1 cointegrating equation(s) at both 5% and 1% le
*(**) denotes rejection of the hypothesis at the 5%(1%) level

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):

LN_HH	LN_WTI
4.287677	-2.388054
1.707384	2.550846

Unrestricted Adjustment Coefficients (alpha):

D(LN_HH)	-0.013193	-0.000726
D(LN_WTI)	0.001952	-0.001684

1 Cointegrating Equation(s): Log likelihood 1301.298

Normalized cointegrating coefficients (standard error in parentheses)

LN_HH	LN_WTI
1.000000	-0.556958
	(0.16169)

Adjustment coefficients (standard error in parentheses)

D(LN_HH)	-0.056569
	(0.01268)
D(LN_WTI)	0.008369
	(0.00932)

Table 20: Johansen test for cointegration July 2008-June 2016

Source: Author using data in "EVIEWS" software

Vector Error Correction Estimates

Vector Error Correction Estimates
Date: 08/22/16 Time: 01:09
Sample (adjusted): 7/28/2008 7/11/2016
Included observations: 416 after adjustments
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1	
LN_HH(-1)	1.000000	
LN_WTI(-1)	-0.556958 (0.16169) [-3.44466]	
C	1.124747	
Error Correction:	**D(LN_HH)**	**D(LN_WTI)**
CointEq1	-0.056569 (0.01268) [-4.46059]	0.008369 (0.00932) [0.89811]
D(LN_HH(-1))	0.031054 (0.04855) [0.63964]	-0.061525 (0.03567) [-1.72469]
D(LN_HH(-2))	-0.025585 (0.04727) [-0.54125]	-0.020277 (0.03473) [-0.58380]
D(LN_HH(-3))	-0.150761 (0.04683) [-3.21905]	-0.032906 (0.03441) [-0.95620]
D(LN_WTI(-1))	0.088586 (0.06694) [1.32339]	0.139505 (0.04919) [2.83628]
D(LN_WTI(-2))	-0.142406 (0.06761) [-2.10633]	-0.070935 (0.04968) [-1.42792]
D(LN_WTI(-3))	-0.066987 (0.06718) [-0.99707]	0.155919 (0.04937) [3.15847]
C	0.008453 (0.01138) [0.74252]	0.005317 (0.00836) [0.63560]
HDD	-4.10E-05 (7.5E-05) [-0.54956]	-4.21E-05 (5.5E-05) [-0.76828]
HDDEV	0.001088 (0.00019) [5.87922]	0.000108 (0.00014) [0.79428]
CDD	-0.000420 (0.00021) [-1.95918]	-0.000187 (0.00016) [-1.18635]
CDDDEV	0.001763 (0.00055)	0.000632 (0.00040)

Table 21: VECM Model July 2008-June 2016

Source: Author using data in "EVIEWS" software

Dependent Variable: D(LN_HH)
Method: ARDL
Date: 08/18/16 Time: 21:04
Sample (adjusted): 8/01/2008 7/15/2016
Included observations: 416 after adjustments
Dependent lags: 3 (Fixed)
Dynamic regressors (0 lag, fixed): D(LN_WTI) ECT(-1)
Fixed regressors: HDD HDDEV CDD CDDDEV STORDIFF SHUTIN C

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
D(LN_HH(-1))	0.042324	0.048160	0.878817	0.3800
D(LN_HH(-2))	-0.044756	0.046461	-0.963318	0.3360
D(LN_HH(-3))	-0.147411	0.046203	-3.190511	0.0015
D(LN_WTI)	0.166811	0.066353	2.513975	0.0123
ECT(-1)	-0.054494	0.012514	-4.354664	0.0000
HDD	-1.94E-05	7.41E-05	-0.261069	0.7942
HDDEV	0.001037	0.000184	5.636973	0.0000
CDD	-0.000361	0.000214	-1.686148	0.0925
CDDDEV	0.001688	0.000547	3.085699	0.0022
STORDIFF	6.34E-06	8.64E-06	0.734090	0.4633
SHUTIN	4.00E-06	4.96E-06	0.807172	0.4200
C	0.005666	0.011336	0.499813	0.6175

R-squared	0.162518	Mean dependent var	-0.003035
Adjusted R-squared	0.139715	S.D. dependent var	0.064962
S.E. of regression	0.060254	Akaike info criterion	-2.752088
Sum squared resid	1.466719	Schwarz criterion	-2.635818
Log likelihood	584.4343	Hannan-Quinn criter.	-2.706115
F-statistic	7.127132	Durbin-Watson stat	1.961589
Prob(F-statistic)	0.000000		

*Note: p-values and any subsequent tests do not account for model selection.

Table 22: Conditional ECM model July 2008-June 2016

Source: Author using data in "EVIEWS" software