The Relationship between Crude Oil and Natural Gas Spot Prices and its Stability Over Time

by

David J. Ramberg

B.A. with College Honors, Economics (1995)

University of Washington

Submitted to the Engineering Systems Division
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Technology and Policy

at the

Massachusetts Institute of Technology

June 2010

Signature of Author ____

Technology and Policy Program
Engineering Systems Division
May 7, 2010

Certified by ____

John E. Parsons
Senior Lecturer in Applied Economics, Sloan School of Management
Executive Director, Center for Energy and Environmental Policy Research (CEEPR)
Thesis Supervisor

Accepted by ____

Dava J. Newman
Professor of Aeronautics & Astronautics and Engineering Systems
Director, Technology and Policy Program

The Relationship between Crude Oil and Natural Gas Spot Prices and its Stability Over Time

by

David J. Ramberg

Submitted to the Engineering Systems Division
On May 7, 2010 in Partial Fulfillment of the
Requirements for the Degree of Master of Science in
Technology and Policy

ABSTRACT

The historical basis for a link between crude oil and natural gas prices was examined to determine whether one has existed in the past and exists in the present. Physical bases for a price relationship are examined. An econometric modeling exercise seeks to establish whether a stable price relationship exists and to define it through the use of a vector error correction model. The model identifies strong evidence of cointegration between the crude oil and natural gas spot price series in the United States. It conditions the predicted natural gas price volatility through exogenous variables related to weather and supply. Once identified, the relationship is clarified more efficiently through the implementation of a conditional error correction model. The model is then utilized to simulate the effects of weather shocks, seasonality, supply deviations and hurricane activity on the cointegrating relationship between crude oil and natural gas.

Finally, an analysis is conducted to test whether the relationship shifts over time to new equilibria. The results of the series of exercises suggest that crude oil and natural gas prices have moved together historically and statistical analysis supports the assumption that the two price series continued to be cointegrated through the end of 2008. The analysis presents evidence that the relationship shifts over time to new equilibria, and the data suggest that these new eqilibria are likewise stable.

Thesis Supervisor: John E. Parsons
Title: Senior Lecturer in Applied Economics, Sloan School of Management and
Executive Director, Center for Energy and Environmental Policy Research (CEEPR)

Table of Contents

1. Introduction

This thesis examines the conventional wisdom that there is a price relationship between crude oil and natural gas in the United States. Section 2 delves into the overlapping history of each commodity. It explores the actual physical and economic basis upon which the idea of an oil-gas linkage emerged. How well this history and integrated market structure supports the idea of a linkage is likewise examined, to the extent possible, in light of available data.

Section 3 reviews the approaches of previous researchers on the subject and compares both their methodologies and results. The thrust is to potentially identify fruitful avenues for exploring the oil-gas price relationship. The other goal is to determine the consensus among other researchers as to whether a price relationship between the two commodities indeed exists.

The first part of Section 4 focuses on developing a methodology to statistically test whether a linkage between the two commodities exists, and how to describe and quantify it. This econometric analysis utilizes both a Vector Error Correction Model and a vector autoregression called a Conditional Error Correction Model. Each is described in greater detail below. The data used to conduct the analysis are described, and the methodology is covered in detail. The findings on whether the two price series move together (are cointegrated) are presented in this section.

The rest of Section 4 examines the implications of the model results. It translates how the long-run relationship between the two commodities responds to events that provoke price movements in natural gas but not crude oil. These include seasonality, weather-related temperature shocks, gas storage surpluses or shortages, and hurricane activity.

The end of Section 4 explores whether the relationship is stable or can shift over time. If and when the long-run relationship is broken, do the price series drift without any relationship, or do the price movements settle into a new equilibrium? Is the strength of these new equilibria greater or weaker than the previous relationship?

The concluding section summarizes the findings and provides guidance for further investigation, as well as critiques the weaknesses of the approach described in this thesis.

2. Are oil and natural gas prices related?

Is there a relationship between crude oil and natural gas prices? Do the prices of the two commodities track each other in predictable ways? Why would anyone assume that they do? Why would the identification of a relationship between the two commodities matter to anyone? What are the dangers of ignoring a possible relationship? Does the relationship between oil and natural gas prices still exist, or has it ended? These questions are all addressed in the following paragraphs.

2.1. Why is defining the relationship important?

One might rightfully question why any effort should be expended on defining and identifying a relationship between crude oil and natural gas. After all, the two commodities largely have different uses and operate in separate markets. A brief discussion of why it might matter is explored here from two perspectives: that of industry participants and that of policymakers.

2.1.1. Industry participants

There are numerous reasons why industry participants might find an understanding of the crude oil-natural gas price relationship useful. The players with the most at stake in the issue are international energy majors, independent power producers and utilities, and energy marketers and traders.

International energy majors

International energy majors operate both in oil and natural gas exploration and marketing. Project lifetimes are measured in decades and investment levels are measured in hundreds of millions or even billions of dollars. An understanding of how price movements in one commodity are expressed in the price of the other could prove useful in predicting price behavior over the longer run, facilitate project planning and profit maximization, and identify potential hedging strategies. This thesis simply explores another potential instrument for quantifying the relationship and providing some predictive ability for project managers and developers. It could potentially prevent large investment losses or the failure to take advantage of imminently favorable market conditions for one commodity or the other.

Since natural gas is a feedstock to petroleum refining operations, the models explored here could provide refiners with some predictive ability as to input costs for their operations. However, industry players might argue that they know better than anybody what the effect of their demand on natural gas for refining has on the overall price of natural gas.

Power producers

A single power plant can cost between tens of millions to over one billion dollars, and the operating expenses are dominated by fuel costs. A prudent choice of fuels is essential before ground is even broken on the building site if one is to minimize the risk of losses and maximize potential profit streams. The question of fuels has steadily shifted over the years from the choice between petroleum-based fuel oils and natural gas to one between coal and natural gas. However, knowing how natural gas prices are likely to be affected by movements in the much larger and more liquid crude oil industry could be instrumental in helping project developers make a decision on which type of plant to build.

Other energy-dependent industries

Much like power producers, large factories are often faced with a choice of fuels for their boilers. This usually distils to a decision between fuel oil-, coal-, or natural gas-fired boilers. Knowledge of the relationship between crude oil and natural gas prices enables informed decision-making on the part of industrialists. The equipment itself is high cost, and is long-lived. Many industrial processes are also high-temperature, fuel-intensive operations. The fuel decision thus has a high impact on overall operating costs for decades.

2.1.2. Policymakers

The modeling approach to be explored here may prove of most use to policymakers. Those in industry have a working knowledge of how their sector functions and have at hand a cache of operating data to assist in decision-making about commodity prices and fuel choice. Policymakers are generally tasked with decision-making that affects broad swaths of the economy, yet are at a disadvantage when it comes to industry-specific knowledge and expertise. The approach used in this thesis attempts to compensate for that shortcoming by requiring only data on commodity prices, weather events and temperature readings to characterize the relationship between crude oil and natural gas prices.

The modeling approach detailed below could also help policymakers tasked with climate change mitigation to quantify the effect of their decisions on the commodity markets of crude oil and natural gas. A model that is sufficiently flexible to allow policymakers to predict how policies affect the oil-gas relationship would allow for a more realistic analysis of the likely impacts of such policies.

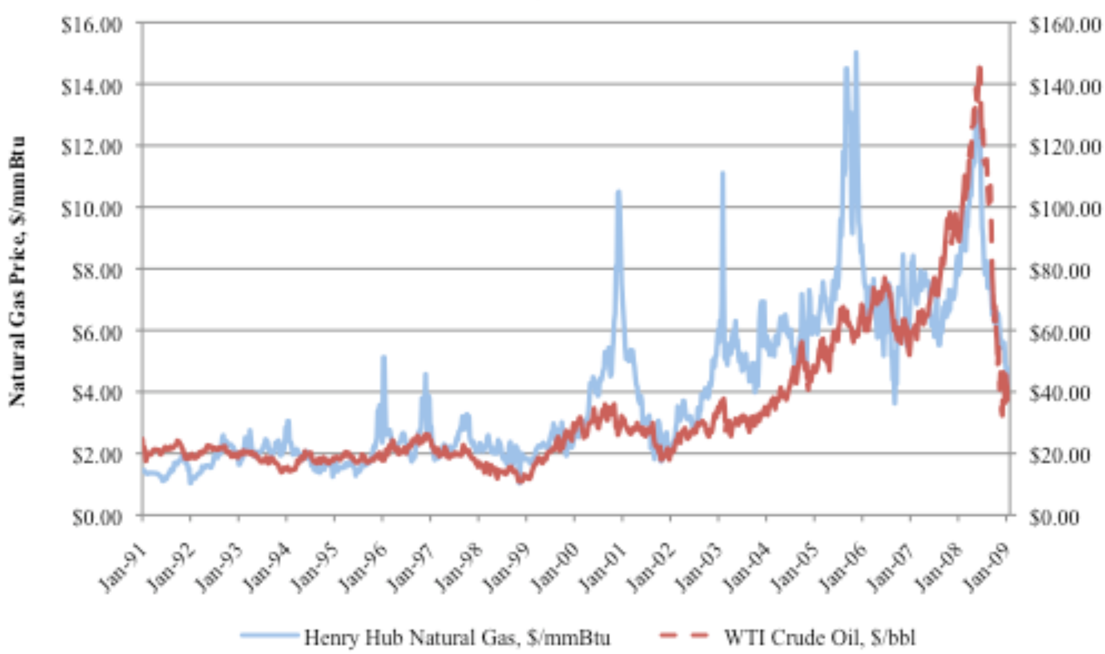
2.2. Can the oil-gas price relationship shift?

Another goal of this thesis is to examine whether the crude oil-natural gas pricing relationship (if it exists) can shift to alternate equilibria. Presumably, changes in policy or technology, either for consumers or producers, could affect the relationship. If these changes are drastic enough, the link could conceivably be severed altogether. Some in industry claim that this is already the case (Blas and Hoyos, 2009). The popular notion of the day is that natural gas prices are now more closely linked with coal prices. This thesis briefly explores whether the oil-gas relationship has shifted in the past, and makes some statistical interpretations surrounding such shifts. It takes a much briefer examination of the coal-gas question.

The thesis also explores possible catalysts for such shifts in the commodity price equilibria. It asks whether they can be observed as they happen or merely interpreted after the fact as new data become available.

Petroleum and natural gas producers, marketers, and traders operate on the assumption that crude oil prices and natural gas prices exhibit a more or less stable historical relationship. This linkage spans both supply and demand-side connections between the two commodities. Before delving too deeply into the possible roots of the pricing relationship, it is useful to examine the actual price series of the two commodities. Figure 1 presents weekly data on natural gas and crude oil

Figure 1 - Henry Hub Natural Gas and WTI Crude Oil Prices, 1991-2009



prices in the U.S. from 1991-2009. The gas price data are the weekly spot price averages at Henry Hub in Louisiana as reported by Bloomberg on its data terminal.

The oil price data are also from the Bloomberg terminal, and are the spot price averages of West Texas Intermediate (WTI) crude. WTI crude is a high-quality, high-volume light sweet crude oil stream. The pricing hub is Cushing, Oklahoma, which is very close to Henry Hub. Their proximity eliminates the need for a transportation differential to be considered. Figure 1 plots the Henry Hub natural gas price in dollars per million Btu ($/mmBtu) on the left-hand y-axis, and the WTI crude oil price in dollars per barrel ($/Bbl) on the right-hand y-axis. In order to better align the price series for comparison, the oil price has been scaled so that the oil price axis is ten times the gas price axis, arbitrarily representing a 10-to-1 price relationship. The gas price is represented by the light blue line, while the oil price is represented with a dark red plot.

What is immediately apparent is that both of these price series seem to share the same general trend. Another simple observation is that the natural gas price series is much more volatile than the crude oil price series. In fact, when the price volatility is annualized, natural gas spot price volatility is nearly twice as high as crude oil price volatility. For natural gas, the figure is 69%, while the annual volatility for crude oil is just 37%. Furthermore, while gas prices seem to stray quite often from the general trend of the crude oil price series, the prices tend to return to the oil price relationship rather rapidly. During the 2008 price run-up in crude oil, natural gas prices look as if they are particularly correlated with the oil price movements. What causes the volatility in natural gas and why gas prices nonetheless seem to return to some sort of relationship with crude oil prices is the motivation for this study.

2.3. Why might crude oil and natural gas prices be related?

On the supply side, natural gas is often found mixed with oil in oil wells. In the early days of the oil industry, before natural gas could be transported, it was simply flared. Once pipelines were capable of transporting natural gas in the 1930s, competition could commence (Davis and Killian, 2008). At that point, exploration began for natural gas independently of whether it was associated with oil reservoirs. Natural gas competed with fuel oils for heating, and with either fuel oils or crude oil itself as a boiler fuel for factory operations. At the turn of the century, "town gas" was either coal gas or locally produced natural gas (Yergin, 1991). When the Federal Power Commission (FPC) imposed price controls on natural gas in the mid-1950s, gas supplies dropped as producers declined to search for it. The linkage between the fuels was not likely to hold very steadily, except in states where gas was actually produced, since intrastate production did not fall under federal control and was free of the federal price caps (Davis and Killian, 2008).

The above scenario illustrates a potential supply-side link: if both markets are free from either price or quantity controls, or if the controls are applied evenly, then the fuels should exhibit some price relationship at the Btu level[1] (although prices of by-products like sulfur could still be independent). This is because the equipment for drilling and the cost of exploration for the fuels are similar, and producers are prompted to explore for either oil or natural gas deposits by price signals. Those with the ability to explore for either gas or oil will pursue the commodity with the greatest expected return on the investment (Finon, 2008). When the returns are not equal, more

[1] A Btu (British thermal unit) is a measure of energy content. It is approximately the amount of energy required to heat one pound of water by one degree Fahrenheit.

resources will be spent on exploration of the commodity with the greatest return until returns on both are once again comparable.

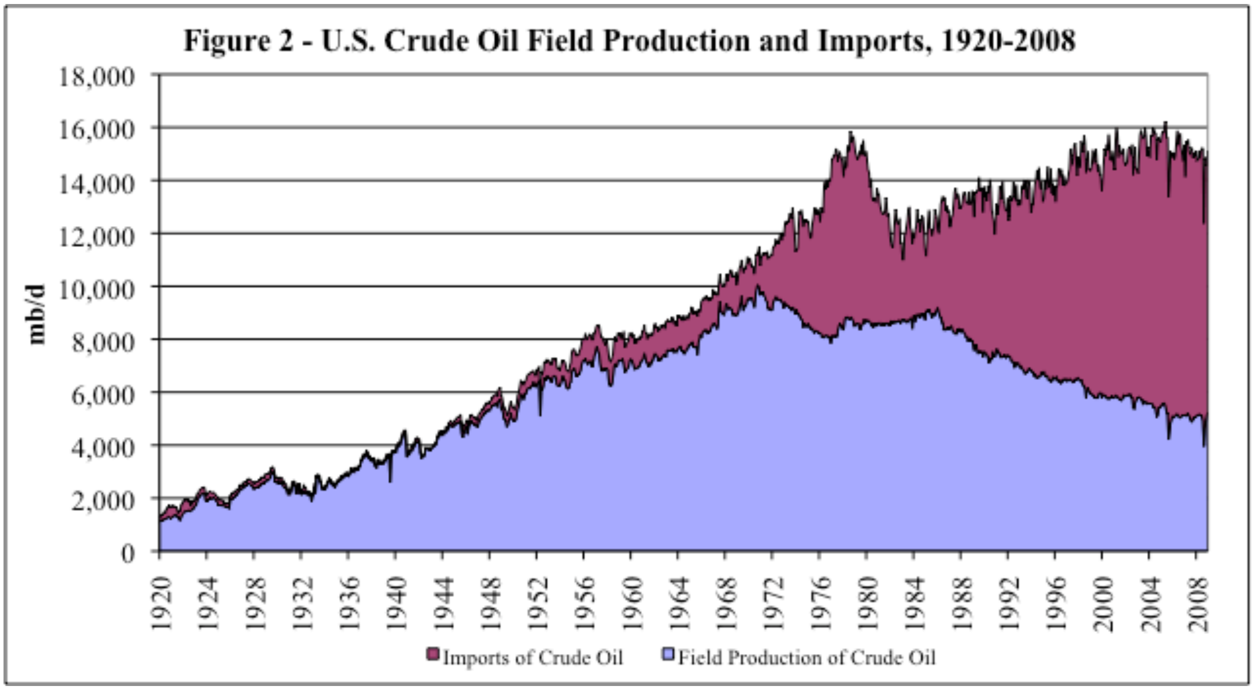
One must also take into account expectations of future regulatory action. If the risk of punitive regulatory action is higher for one commodity than the other, the riskier commodity will be less attractive. However, considering the time lags involved in oil or gas exploration projects, sudden announcements regarding policy changes are likely to affect only long-term planning.

2.3.1. Structural similarities between crude oil and natural gas markets

An understanding of the respective historical developments surrounding each commodity might provide useful insights as to why the two fuel prices exhibit a linkage in the present. Much of the development of the petroleum product and natural gas marketplace occurred in tandem and there is significant overlap. Nonetheless, the respective history of each commodity will be reviewed separately.

Oil history

The oil market in the United States developed in the late 1800s in Pennsylvania. In a short time, the simplest distillate of crude oil, kerosene, was a competitor to whale oil for lighting. Once oil was discovered in Texas and California, however, more and more uses were found for it, and it proceeded to power factories and electrical generating units, motor vehicles and locomotives (Yergin, 1991). Over time, refining evolved from a simple distillation process into a complex system of engineering and chemistry. Much of American industry has relied on petroleum in one form or another for most of the 20th century. However, as new technologies and pollution regulations have been adopted, oil has retreated into the transportation sector, where it remains the dominant fuel globally. In the U.S., reliance on imports has grown dramatically since the 1970s, when domestic production peaked and began to decline. Figure 2 tracks U.S. crude oil



production and imports in thousand barrels per day since 1920. It was compiled from data provided by the U.S. Department of Energy's (DOE's) Energy Information Administration (EIA). Figure 3 shows the path of product demand and the growth of each in thousand barrels per day, as far back as 1936 in the case of residual fuel oil (U.S. DOE website (e)).

Crude oil is now a global commodity, shipped via pipeline, tanker truck, rail lines, and seagoing oil tankers. Local markets have become global markets, and global oil prices are set primarily according to two major “marker” crudes: Brent North Sea crude, from Great Britain, and West Texas Intermediate (WTI). Other crudes are priced according to these high-volume crude streams through a comparison of their quality as measured by their API gravity.[2] Once the province of actual oil industry participants in search of cross-hedging opportunities with other producers, the crude oil futures markets are now the forums for a diverse set of traders. Brent is traded in London on the International Petroleum Exchange (IPE). WTI is part of a basket of crudes in the light sweet crude oil contract on the New York Mercantile Exchange (NYMEX). Thus, the high-quality crude stream from West Texas serves as a major global indicator of the value of crude oil worldwide.

Figure 3 - U.S. Crude Oil and Petroleum Product Demand, 1936-2009

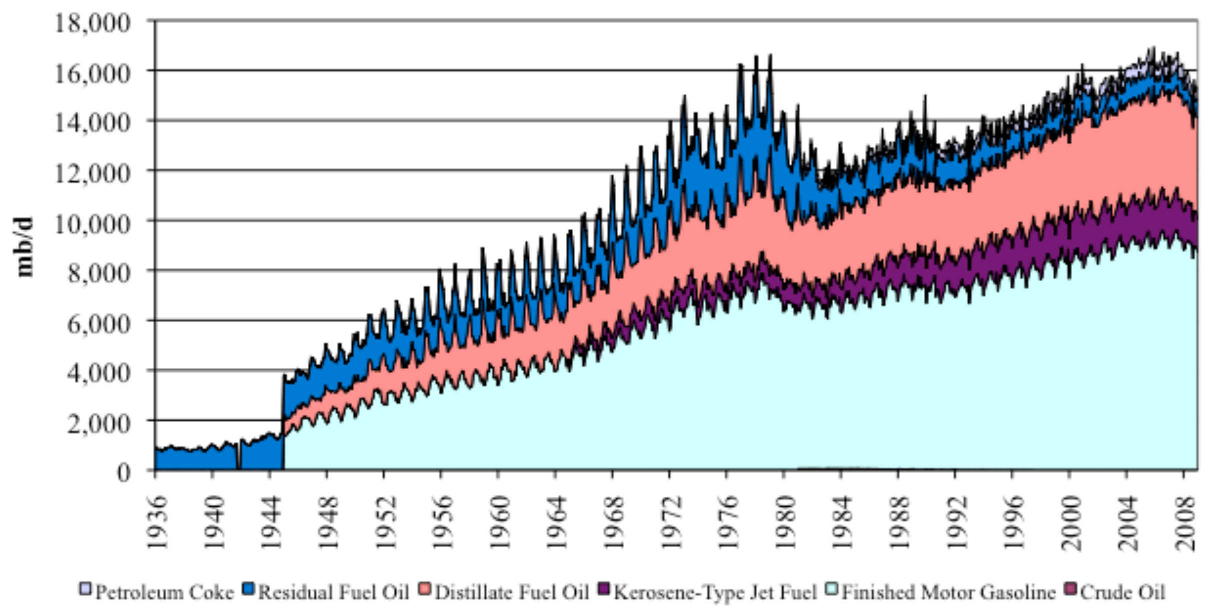


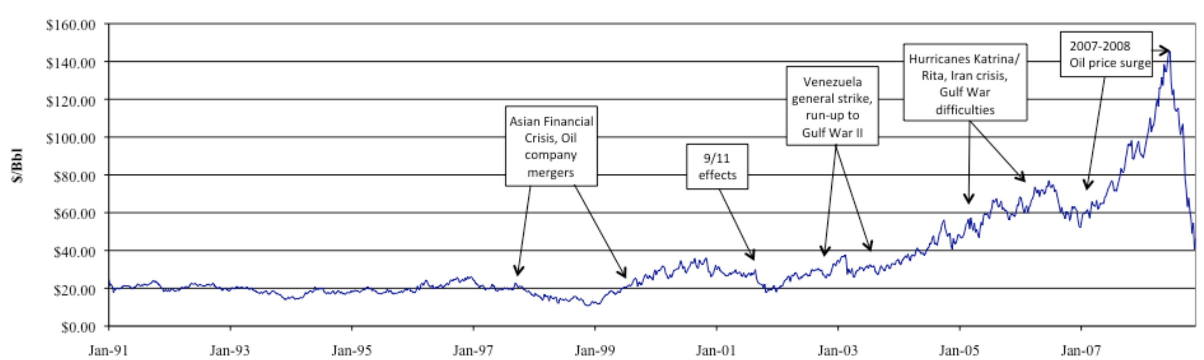
Figure 4 illustrates the weekly spot prices for WTI crude oil since the last week of January of 1991. The trend is fairly mundane until 1998, when the effect of the Asian financial crisis

[2] API (American Petroleum Institute) gravity is a scale developed to compare the densities of petroleum liquids. Higher numbers represent less dense crude oil streams, which are more valuable.

reflected the decline in demand for oil, and the price fell to $10-11/Bbl. From there, the WTI price began a steady rebound as global economic expansion resumed, led by the U.S. economy's boom in tech and Internet stocks. The Yugoslav war heated up, and news that Iraq was flouting the restrictions on its oil exports emerged. By 2000 and 2001, many oil companies completed mergers that had begun when oil was trading at less than a third of its price just two years before.

The subsequent stock market crash and the wave of bankruptcies in Internet companies corresponded with a visible dent in the price rise. The oil market stabilized until the recession at the end of 2001 that coincidentally followed the 9/11 attacks. Oil prices again dipped to $20/Bbl.

Figure 4 - WTI Cushing Spot Weekly Price and Historical Events, $/bbl



A subsequent global economic expansion ran through the summer of 2006, with strong growth in Asia and the U.S., especially for construction. These forces exerted upward pressure on oil prices. The period was punctuated with distinct spikes during Venezuela's general strikes, when Chavez took over the foreign-held oil exploration and development contracts, and during the run-up to the second Iraq war. Additional spikes occurred with the onslaught of Hurricanes Katrina and Rita, which shut in much of Gulf of Mexico oil production for months, and under concerns that Iran might become involved in the wider conflicts in the Persian Gulf.

Just when oil prices neared $80/Bbl, they retrenched to the upper $50/Bbl range. They were again driven upward on increased demand from India and China for both transportation and construction through the summer of 2008. Oil peaked at over $145/Bbl in the beginning of July. At that point, confidence in the global financial system collapsed, along with financing for the ambitious development plans worldwide – especially in developing economies that were taking advantage of low-cost credit. When the sources of funds dried up, so did the strength in the oil industry, and prices crashed to below $40/Bbl.

The next section will discuss the history of natural gas prices. It will be seen that while the drivers of oil and gas price fluctuations had little in common, the two series nonetheless seemed to follow a similar underlying trend.

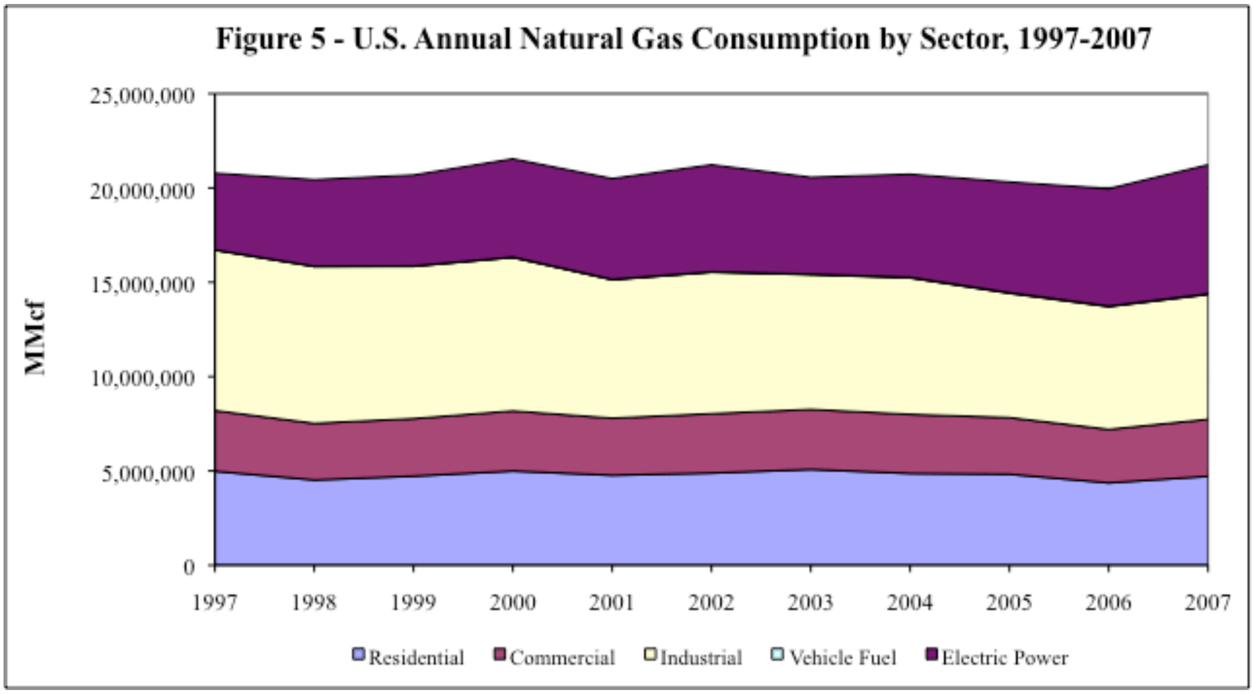
Gas history

Natural gas was originally a by-product of the crude oil exploration process. While it was a potentially useful fuel, the fact that it was gaseous made transport nearly impossible due to technological limitations. Pipelines of the time, though suited for liquid crude oils, were not capable of containing natural gas without significant leakage and product loss. The early gas finds were thus generally flared off of oil drilling rigs if no immediately local use could be found for them. However, town gas, derived from coal, preceded the use of natural gas and helped to establish its usefulness in many areas, including street lighting and heating. By the 1930s, the issue of pipeline leakage had been resolved, and natural gas pipelines began to be built to service regions far afield from where the gas was originally discovered and produced (Davis and Killian, 2008).

Over time, more and more uses for gas were discovered – it first replaced town gas for heating, cooking, and street lighting. Later it was used as a fuel in industrial processes and electricity generation. Now, in limited form, natural gas has even penetrated the transport sector. Price controls from the 1950s through the late 1970s in the U.S. prevented its widespread adoption in industry due to scarcity issues arising from an inability of producers to recover the costs of exploration and production (Breyer and MacAvoy, 1973). Once price controls were lifted, the market again expanded as homeowners increasingly chose natural gas for heating and electric utilities found it a lower-cost option than nuclear, solar, or wind energy. It is also a lower-pollution option than coal or oil for power generation.

Over the last ten years, the quantity of natural gas consumed in the U.S. has been stable. Figure 5, based on data from the EIA website, illustrates how the share of natural gas devoted to the power generation sector has grown. At the same time, industrial consumption declined by a similar amount (U.S. DOE website (c)).

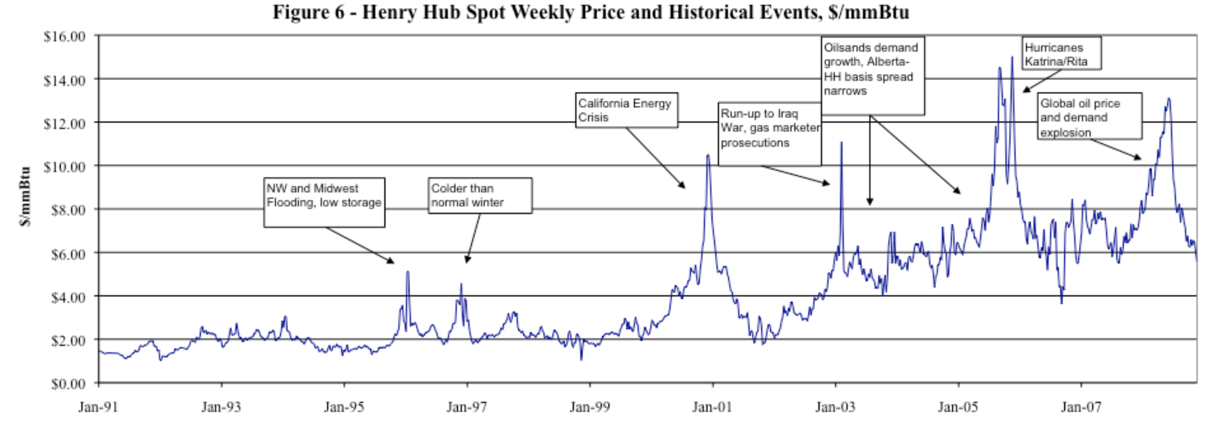
The period of heaviest usage for natural gas is during the winter, when it is used for heating. In contrast, crude oil is mostly used for plastics, chemicals and transportation fuel. Thus, the uses of natural gas are more seasonal than those for oil products. This seasonality is imparted to the price of natural gas over the course of the year despite the fact that whenever it is winter in one part of the world, it is summer somewhere else.



Part of the reason for this is that the natural gas market is regional and segmented due to the difficulty of overseas transportation, so uses of gas tend to be relatively close to the point of production. Crude oil, on the other hand, is a global market, with a largely fungible supply and a fairly common price level worldwide. Since the oil price encompasses global demand, crude oil prices do not exhibit seasonal patterns. This is one reason that Henry Hub natural gas prices are more volatile than WTI crude oil prices. Figure 6 illustrates the weekly spot price movements for Henry Hub natural gas from the last week of January 1991 through December 2008. There are a few spikes before 1998 coinciding with forecasts for very cold weather, flooding in the Northwest and Midwest, and storage levels significantly below the 5-year average. Interestingly enough, the combination of these factors seemed sufficient to provoke a steep rise in prices, while the any one of these conditions on their own were not always enough to trigger a market response.

The first significant price spike occurred between early 2000 and continued until after the summer of 2001. This coincided with the California Energy Crisis, where a combination (often called a "perfect storm") of factors brought natural gas prices to historic highs – above $10/mmBtu. California was at the peak of a long business cycle, since it was the center of the

Figure 6 - Henry Hub Spot Weekly Price and Historical Events, $/mmBtu



tech and dot-com industries, so demand for electricity was exceptionally high even without weather effects. Furthermore, the winter was especially cold, while the summer was significantly warmer than average.

Normally, California electricity demand was greater than its capacity to meet it, and summer supplies were augmented by imports from the Northwest. However, a drought in 1999-2000 kept reservoir levels well below normal, severely curtailing the ability of the Northwest hydropower system to meet California's electricity needs. Due to environmental regulations, most of the generating capacity added to the California grid in the late 1990s was gas-fired, and since electricity consumption was so high, most of these marginal units were operating. In addition, it was later learned that Enron and other natural gas marketers were actively gaming the market, exacerbating its volatility. Losses were so large for the economy of the West that it almost certainly contributed to the recession that ensued (information on the California Energy

Crisis provided by O'Neil, ed., *The Energy Market Report* (1999-2001), Borenstein (2002), and Joskow (2001, 2002)).

Price fluctuations through 2001 were all related to weather and gas-specific issues. The spikes in 2002 and 2003 more likely reflected the implicit linkage between the oil and natural gas markets. In the run-up to the second Iraq War, oil prices rose steadily to reach nearly $40/Bbl, with expectations that supply shortages would cause prices for crude oil to rise even further worldwide. Gas prices also rose.

One explanation for why a rise in crude oil prices could prompt an increase in natural gas prices is the use of natural gas in petroleum refining and exploration. The oil sands industry is centered in Alberta, Canada. According to company press releases, production of synthetic crude oil is profitable when oil prices break about $40/Bbl (depending on the cost of natural gas, which is a feedstock in the process). The bulk of U.S. natural gas imports, as always, come from Canada. Prices at the AECO and Empress hubs on the Canadian border historically traded at a discount to domestic natural gas. As soon as high oil prices induced investment in the oil sands fields, the price advantage disappeared. Figures 7 and 8, showing the price discounts for both AECO and Empress Canadian hubs compared to Henry Hub natural gas prices, confirm this observation (Bloomberg terminal). Gas that had been selling at discounts of about $2/mmBtu in the mid-to-late 1990s suddenly traded within a dollar of Henry Hub prices. It appeared that the era of cheap gas from Canada had indeed ended. The perception of scarcity, along with strong competition from the oil sands industry, is a likely culprit for the sustained rise in natural gas prices until 2005. That summer, the quick succession of Hurricanes Katrina and Rita caused natural gas prices at Henry Hub to soar to over $15/mmBtu, far above any levels yet recorded. Gas production in the Gulf of Mexico was, by all accounts, a shambles. Until production returned to near pre-hurricane levels, Canadian gas again traded at a discount to Henry Hub natural gas, but only because Gulf-sourced gas had become much more expensive than before.

Figure 7 - Price Discount - AECO-Henry Hub Natural Gas

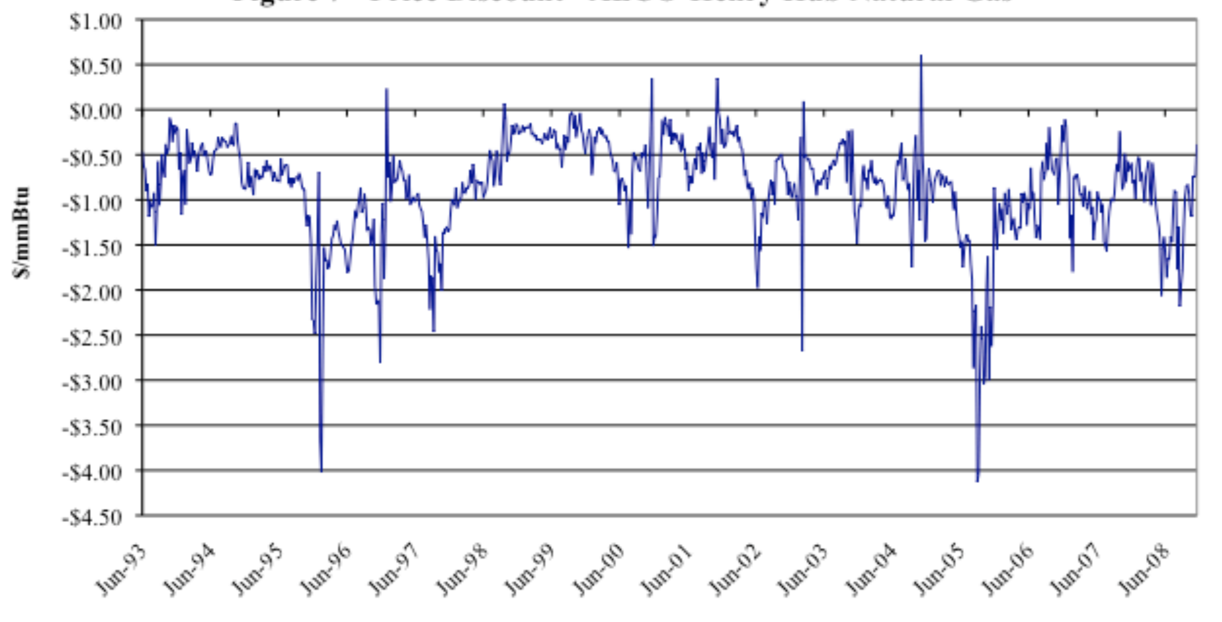
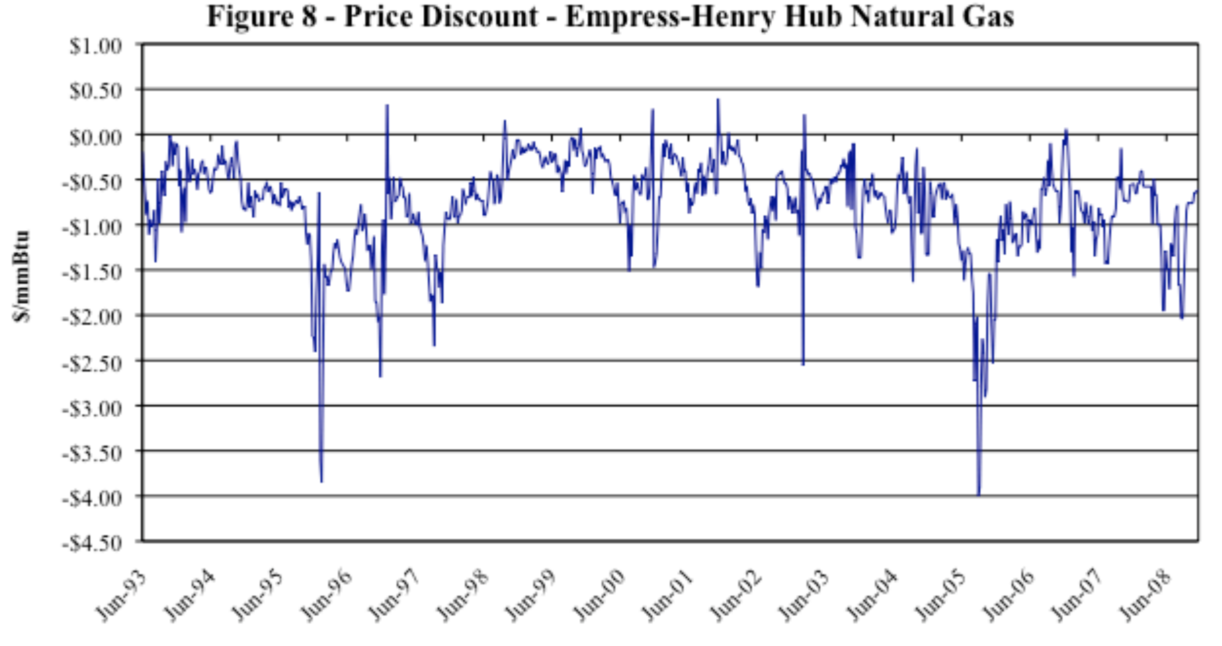


Figure 8 - Price Discount - Empress-Henry Hub Natural Gas



The trend of rising gas prices between 2007 and 2008 mirrored the explosion in global crude oil prices. It is possible that increased activity in both conventional crude oil refining and in production of synthetic crudes put pressure on gas supplies in competition with uses for heating and power generation. The National Petroleum Council's September 2003 report predicted that over the next 20 years, the amount of natural gas demand that would be met by North American producers would fall from about 96% to only about 75%, with the remainder representing imports of LNG if demand does not adjust accordingly (NPC, 2003).

Insights from history

The historical development of the two commodities provides clues to where the two industries might transmit pricing information in such a way that the oil and gas price series move together. The introduction suggested one possibility, that of direct competition for drilling resources at the wellhead. But the oil and gas histories have also illuminated a few other areas of overlap: portions of the manufacturing and generating sectors use machinery with the ability to switch between oil products and natural gas as fuels. Another potential factor linking the two price series is the use of natural gas as a feedstock for petroleum refining and oil sands operations. Since U.S. imports of natural gas draw off a major portion of Canadian natural gas production and Canadian imports constitute a large fraction of total gas consumed in the U.S., Canadian gas demand increases can arguably provoke price increases in U.S. natural gas. These will be discussed and examined in more detail below, along with an additional vector through which prices could be transmitted internationally: the LNG market.

The history of the oil and gas commodities also illuminates another aspect of their pricing relationship. The events that occurred while oil prices spiked or fell were generally global in nature and a reflection of broad economic fundamentals. There was little to suggest direct causality between specific events and oil price movements. The natural gas price series, on the contrary, appeared to react to specific regional or even local events that would not be expected to have a global impact. This is both a reflection of the segmented nature of the natural gas market (in contrast to the global oil market) and a possible explanation for the higher volatility of the gas price series.

2.3.2 Fuel substitutability

Given the history of the petroleum and natural gas industries, identifying areas where they compete or complement each other is straightforward. One avenue through which a price relationship between crude oil and natural gas might be sustained is direct competition. This was most common from the 1930s to the 1950s along the natural gas transmission system, from the 1950s through the 1980s in states that had indigenous gas production, and then nationwide from the 1980s through nearly 2000 (Huntington, 2007). During this period, many power generation units and much of the machinery in industry had the ability to shift at low cost between natural gas and petroleum product fuels. An increasing burden on polluters starting in the 1970s and the boom in natural gas power generation in the 1990s has narrowed the range of opportunities for direct competition between the fuels in the short run. However, if gas prices embark on a sustained rise while oil prices hold steady, it is conceivable over the long-run that more fuel-switching capability could arise (Costello, Huntington, and Wilson, 2004). However, recent evidence suggests that gas will be relatively cheaper than oil going forward.

Before 1970, nearly 56% of natural gas-fired power generators had the capability to switch to petroleum products as fuels. From 1980 to 1984, nearly 71% of new gas-fired power plants had the capability to switch to petroleum product fuels – mainly residual or distillate fuel oils (including diesel). By the 2000-2004 period, only 16% of new natural gas power plants could switch to petroleum product fuels. As of 2007, 31.5% of all gas-fired power plants reported an operational capability, including all necessary equipment and storage, to switch to either distillate or residual fuel oil as an alternative fuel. However, of these plants, 67.9% reported either environmental regulatory restrictions or some other factor that would limit fuel switching (DOE, 2009b, pp. 6).

The EIA's Manufacturing Energy Consumption Survey noted that about 17.7% of all U.S. manufacturing processes using natural gas could switch to petroleum products in 2002 (DOE, 2002, and U.S. Bureau of the Census, 2002). That same year, the manufacturing sector accounted for 24.5% of total U.S. natural gas consumption. These two figures imply that the fuel-switching capability in the manufacturing sector amounted to just over 4.3% of total U.S. natural gas consumption (DOE, 2009). The power generation and manufacturing sectors together account for about 50% of total U.S. natural gas consumption. The proportions of each sector with the ability to switch from natural gas to petroleum product fuels imply that 13-14% of total natural gas consumption in the U.S. can be substituted with petroleum products within a single season (barring environmental regulatory constraints). Taking emission constraints into account would shrink that figure to about 7% of total U.S. gas consumption. Table 1 details the potentially displaced volumes of natural gas due to gas-to-petroleum fuel switching in the power generation and manufacturing sectors (and vice versa). It also expresses these volumes as a percentage of total natural gas usage in the U.S. Note that a complete shift from petroleum products to natural gas fuels could result in an *increase* in natural gas usage of just over 2%.

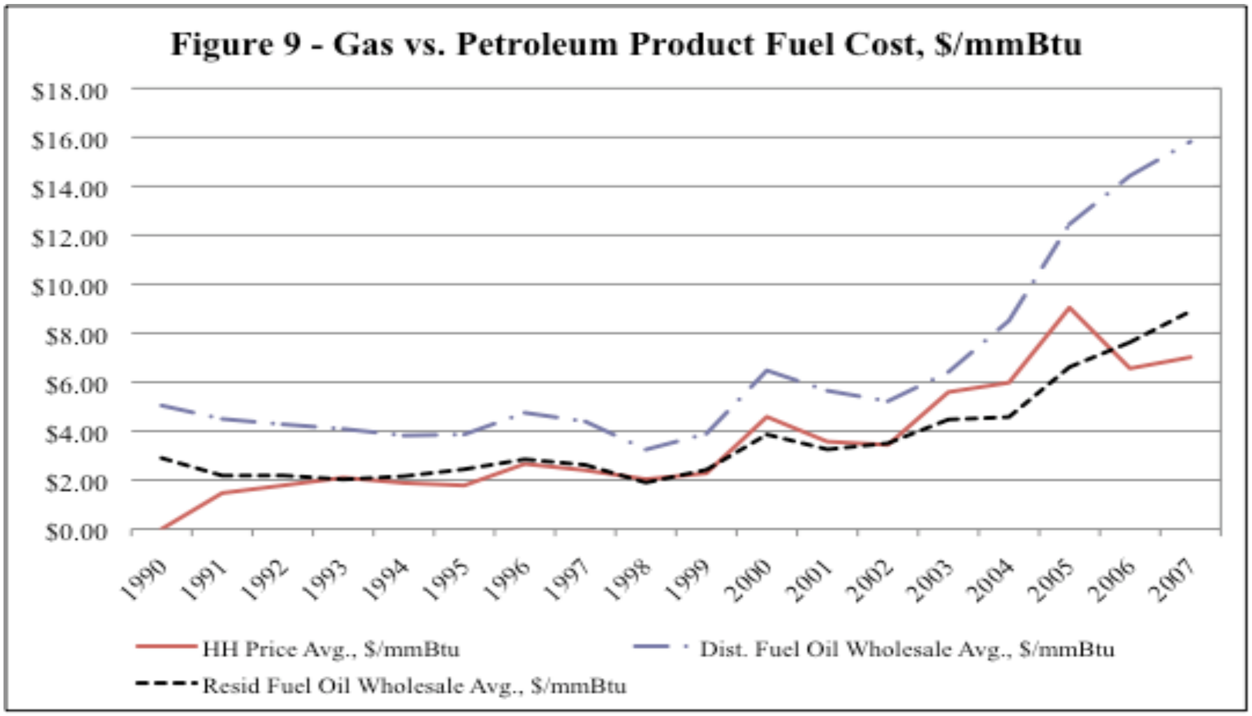
Table 1 - Gas/Petroleum Product Fuel Switching Effects due to Physical Substitutability								
Sector	**Data Year**	**Subsector**	**Amount of Gas Affected by Switchability (no environmental/other constraints) (bcf)**	**Amount of Gas Affected by Switchability incl. environmental/ other constraints) (bcf)**	**Effect on Natural gas Usage (+ or -)**	**Total U.S. Gas Consumption - relevant year (bcf)**	**Percentage of U.S. Natural Gas Usage Affected (w/o constraints)**	**Percentage of U.S. Natural Gas Usage Affected (w/ constraints)**
Manufacturing	2002	Gas-to-Petroleum	998	998	-	23,007	-4.34%	-4.34%
Manufacturing	2002	Petroleum-to-gas	87	87	+	23,007	0.38%	0.38%
Power Generation	2007	Gas-to-Petroleum	2,157	693	-	23,047	-9.36%	-3.01%
Power Generation	2007	Petroleum-to-gas	383	334	+	23,047	1.66%	1.45%
Total, All Switch to Petroleum Products	02/07 hybrid		-3,155	-1,691	-		-13.70%	-7.34%
Total, All Switch to Natural Gas	02/07 hybrid		470	421	+		2.04%	1.83%

Table 1 reports the portion of natural gas consumption in play through physical switching at the equipment and machinery level. The physical substitutability described above is simply a catalogue of the volumes of gas or petroleum products that *could* be switched. The analysis does not take into account the price points at which gas would be substituted with residual or distillate fuel oils or vice versa. Nor does it account for the cost of switching itself. As a result, the volumes listed as substitutable would only be fully utilized when natural gas is either sufficiently expensive or sufficiently cheap in comparison to the petroleum products with which it competes to cover this additional cost.

In 2003 the National Petroleum Council provided a rough guide to the points at which fuel switching will occur. However, they made an error to be discussed below. At \$3.75/mmBtu, its report states, industrial boilers unconstrained by environmental restrictions tend to switch to residual fuel oil if they have the capability; unconstrained residual fuel oil-enabled power generators will switch from natural gas before the price reaches \$4. At \$4/mmBtu, methanol and ammonia plants begin shutting down. At \$4.75/mmBtu, power generators switch to distillate fuel oil. At $5/mmBtu, ethylene plants begin shutting down. At $5.50/mmBtu, processes begin to be adjusted to shift away from natural gas and toward distillate fuel oils. By $6/mmBtu, even boilers subject to environmental penalties will choose to switch from natural gas to residual fuel oil (NPC (2003), p. 23-24). Using these figures one would expect to see shifting away from natural gas and into petroleum product fuels near the high end of the estimates in Table 1 if natural gas reaches $6/mmBtu.

Two points are important to note, however. First, the NPC study was conducted in 2003, when prices for crude oil, residual fuel oil, and distillate fuel oils were hovering in the $30/Bbl range. The NPC did not take into account the possibility of a price tradeoff between the fuels, but rather held petroleum product prices constant while allowing the natural gas price to vary. If there is a stable price relationship between the two types of fuels, it is unlikely that natural gas prices would move dramatically higher for an extended period without some corresponding movement in the oil-based commodities. It is possible that the natural gas price points at which fuel switching would occur when crude oil is trading above $30/Bbl are much higher than those depicted in the NPC report. An improvement to the analysis would use price points *relative* to petroleum product prices, rather than absolute levels.

Second, Table 1 is static. It does not account for growth in gas-fired or petroleum product-fired capacity. The proportion of gas usage that could be switched for petroleum products would shrink over time if gas usage grows faster than petroleum product usage in power generation and industrial processes. The following is an empirical test of both the NPC price assumptions and the level of substitution predicted at the breakpoint prices detailed in the 2003 report. Figure 9 plots the actual historical course of fuel costs between natural gas, distillate fuel oil, and residual fuel oil in dollars per million Btu.



As Figure 9 illustrates, during the six months prior to the publication of the NPC report, average monthly Henry Hub natural gas prices ranged from $4.64 to $5.99/mmBtu. According to the report, there should have been switching from natural gas to both residual and distillate fuel oils in power production.[3] However, even though natural gas traded well above even the $6/mmBtu breakpoint from 2004 through 2007, there was no corresponding rush to petroleum product-fired generation. The share of power generation from natural gas fuel increased fairly steadily from about 12% to nearly 22% from 1990 through 2007, as illustrated in Figure 10 (U.S. DOE (2009)). In 2003, however, the prices of both sets of fuels were very close, and gas prices were

[3] The NPC report also predicted shutdowns in methanol plants, ammonia plants, and ethane-based ethylene plants, but due to industry-specific data constraints, this thesis examines only substitution in the power generation sector.

rising faster than the prices of fuel oils. Thus the concern that natural gas could be abandoned in favor of fuel oils was not unfounded.

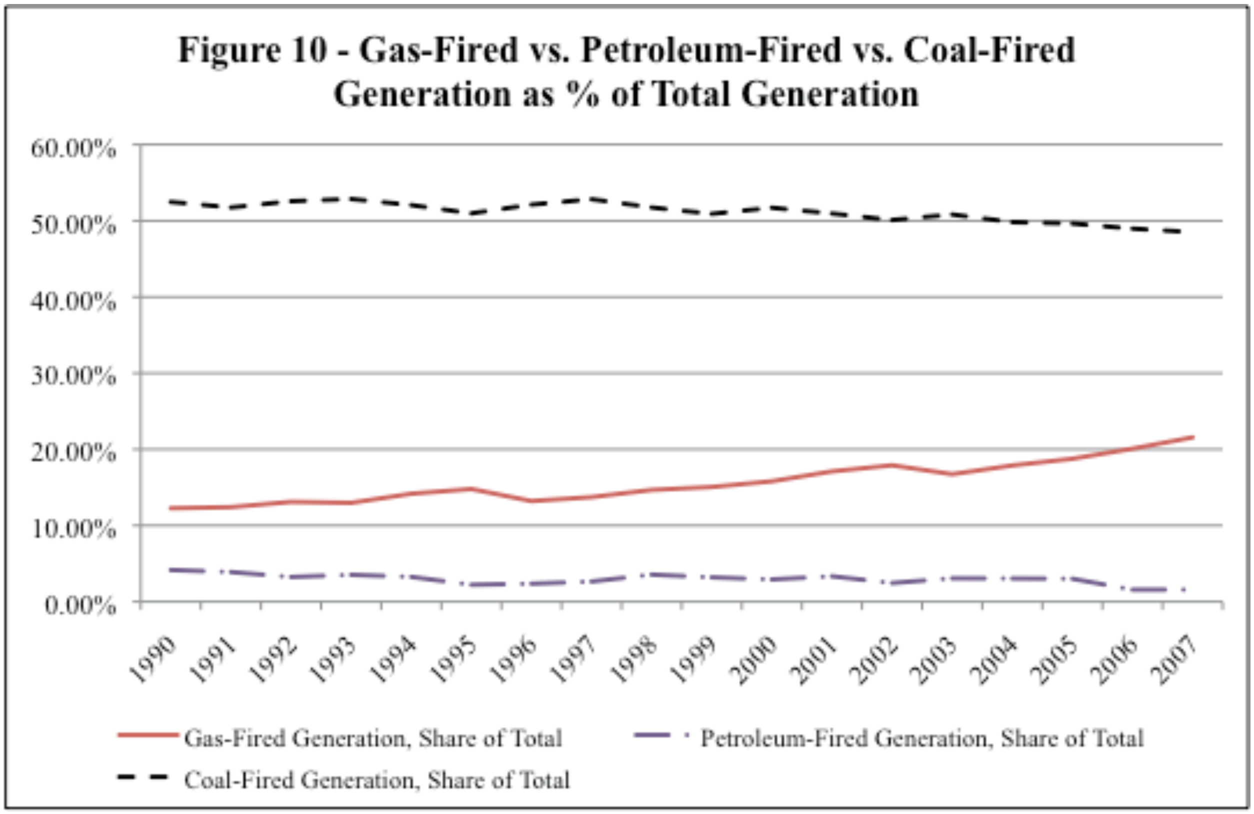


Figure 10 provides the share of total generation for natural gas-fired, petroleum product-fired, and coal-fired generation. Despite rising gas prices, both coal- and petroleum-fired generation have declined as a proportion of total power generation. Fuel switching within individual plants is but a portion of these figures. Another aspect of interfuel competition is through the dispatch order, in which the power plant with the lowest cost is called upon to provide power before plants with higher cost.

2.3.3. Dispatch-order competition in the power sector

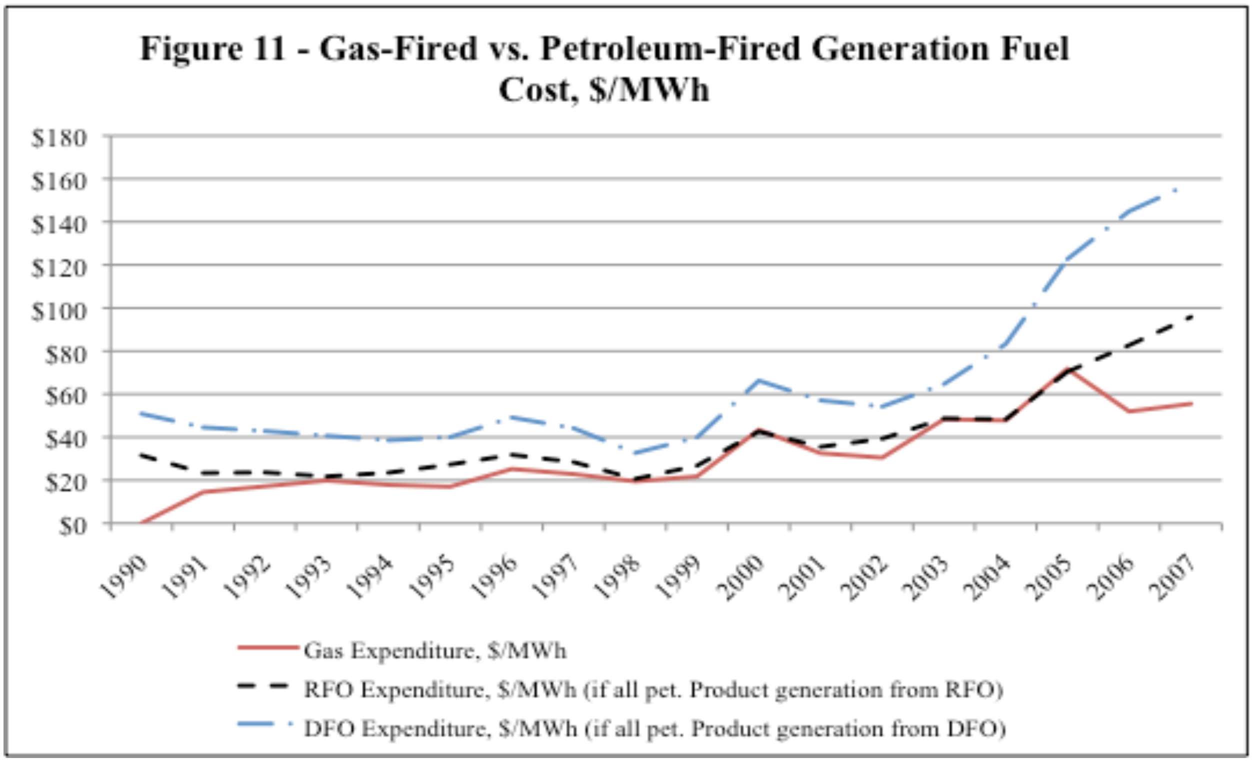
In the 21st century, the energy-generating dispatch order is the more common form of competition between gas and other fuels in the power sector. Owners of both gas-fired and coal-or oil product-fired plants will only dispatch the plants that are profitable according to fuel and operating costs on any given day. In energy markets with merit-order dispatch,[4] the competition is not even between plants within the portfolio of an individual power company. At present, the generating portfolio aspect of interfuel competition is more commonly between gas and coal than between gas and oil, but some opportunities for gas-oil competition remain. Gas-fired power generation appears to be well entrenched, however. In both 1996 and 2003, the years in which natural gas-fired power production retreated as a share of total power production, petroleum-product-fired and coal-fired generation advanced only imperceptibly, as shown in Figure 10. At the same time, 1997 and 2003 were both years in which the price of natural gas exceeded the price of residual fuel oil on a $/mmBtu basis. Instead of a large-scale flight to fuel oils, there was only slight substitution away from natural gas and into petroleum products (and coal) as fuels for power generation.

The above analysis suggests that setting a static breakpoint at which gas falls out of favor and generators shift to petroleum products for fuels is not a useful method to determine which commodity is more attractive for fueling power generation. The cost of providing power to the grid is not simply the per-unit cost of the fuel itself. The average variable cost per megawatt-hour (MWh) of electricity production takes into account fuel efficiency and operations and maintenance costs. The EIA's 2007 Electric Power Annual provides all of the necessary data.

---

[4] In markets with merit-order dispatch, all potential generators submit their bids for the price and amount of power they are willing to provide to the grid on a given day. The market operator stacks bids from the lowest to the highest price, and bids are accepted in that order until the forecasted load is covered. The bid price of the final accepted unit becomes the $/megawatt-hour price that all generators accepted for dispatch receive. In this low-cost auction setting, whether or not to generate is based solely on the relative costs of the units and unit owners' preferences for profit margins.

Annual consumption of fuel by fuel type (in barrels for petroleum products and in mmBtu for natural gas) is multiplied by the average fuel price per unit to determine total fuel expenditures for power generation for each type of fuel. This figure is then divided by the total annual number of MWh produced to arrive at a price per MWh per fuel. This method accounts for the relative fuel efficiencies of the plants using the various fuels as well as the cost of the fuel itself. A plot of the series from 1990 through 2007 is illustrated in Figure 11.



When this method is used to compare generation costs, natural gas consistently tracks below the cost of both residual fuel oil (RFO) and distillate fuel oil (DFO) wholesale prices for the entire period. At a few points, RFO reaches price parity with natural gas costs per MWh – specifically in 2000 and from 2003 to 2005 – but these are for short periods and in these rough calculations, neither environmental restrictions on emissions nor the small portion of production that might actually be charged at retail prices are taken into account.

This metric reveals that natural gas consistently remained competitive with petroleum-based fuel oils in power generation from 1990 through 2007 regardless of its $/mmBtu price. This explains why so little substitution away from natural gas was recorded. Throughout the 1990s natural gas-fired power plants made great strides in efficiency through a shift to combined-cycle power generation.[5] The result is that even when the price per unit of energy content rises for natural gas in comparison to the fuel oils, it has still maintained, *at least*, parity with variable RFO-fired generation costs. This characterization would be strengthened further if environmental restrictions and a mix of retail and wholesale prices were taken into account. Pyrdol and Barron examine fuel switching during the 2000-2001 natural gas price spike in their 2003 working paper for the Energy Modeling Forum. They found that although gas prices averaged over $10/mmBtu nationally, only about one-third of the units that were capable of doing so abandoned natural gas in favor of petroleum product fuels.

The analysis above illustrates how little the level of gas-fired generation has been affected in the past by competition with coal and petroleum product-fired generators. But how much gas *could* be displaced by petroleum products through dispatch-order competition? The EIA's 2008 Annual Electric Generator Report recorded net summer capacities of gas-fired generation of 397,432 MW and petroleum-fired generation of 57,445 MW (U.S. DOE, 2010). A comparison of these capacities reveals that even a full dispatch of petroleum-fired generation at the expense of gas-fired generation only amounts to 14% of the installed capacity of gas-fired generation. At that level, the displacement due to gas-to-oil switching would affect nearly 4% of total annual

[5] A combined-cycle gas plant has both a gas turbine and a steam turbine. The waste heat from the burning of natural gas for the gas turbine heats the steam to power the steam turbine.

U.S. natural gas consumption. Combining this figure with the data from Table 1 suggests that between physical and dispatch-order substitution, nearly 18% of total natural gas consumption in the U.S. could be displaced by a shift to petroleum products. Figure 10 shows that when substitution from gas to oil does occur, however, it is actually on a much smaller scale.

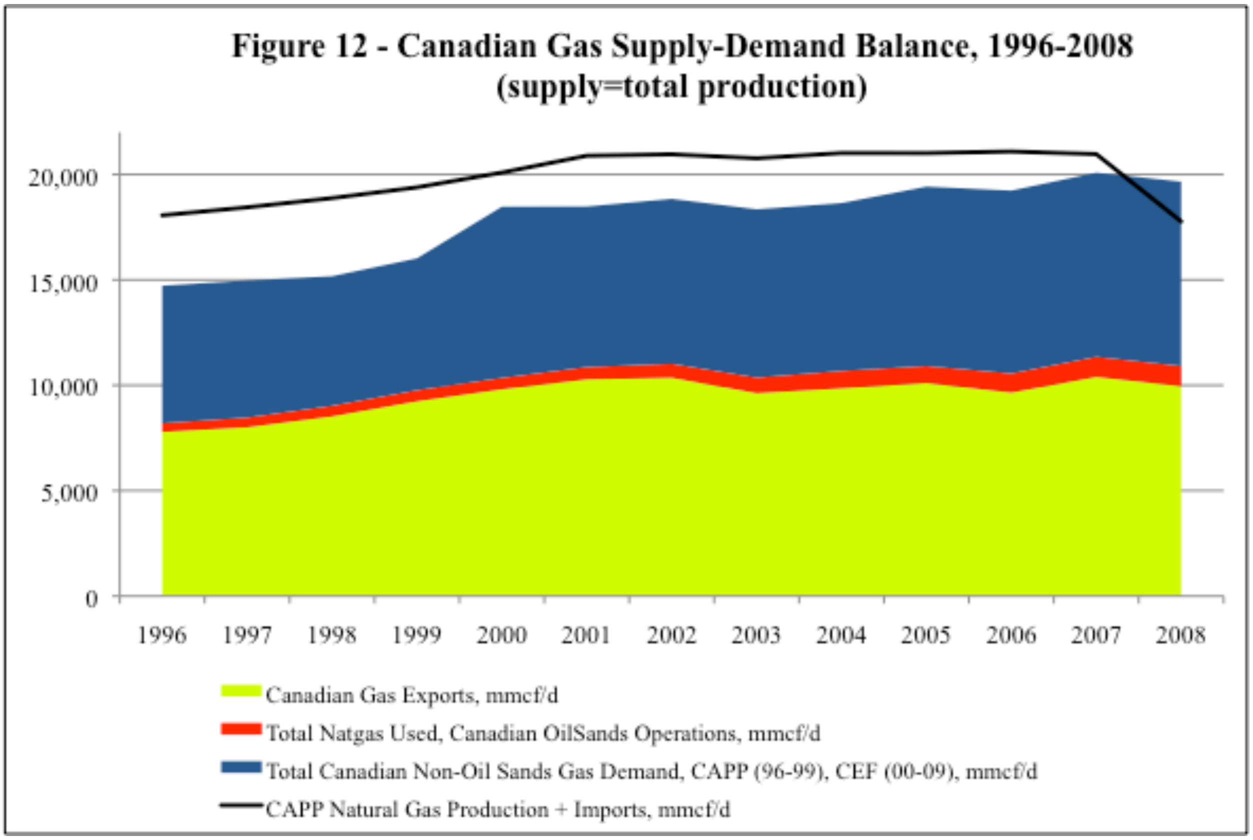
The benefits of gas-fired generation go beyond a simple energy content price-parity comparison between alternate fuels. Unlike coal units, gas-fired units can ramp up quickly to meet demand. They emit about half the carbon dioxide and less than half the total pollution of coal units, and have an emissions advantage over petroleum-fired units as well. Furthermore, the efficiency gains in gas-fired units over the years have far outstripped the gains made in coal- or petroleum product-fired generation. The likely trajectory of these factors that are not directly related to fuel costs suggest that dispatch order substitution in power generation will be a shrinking factor through which oil products and natural gas compete. The findings of this section give weight to those who claim that the oil-gas price relationship may be weakening.

2.3.4. Gas as a Refining Feedstock

There are other links between natural gas and oil commodities that represent a portion of the natural gas demand that can be switched for petroleum products. As mentioned earlier, natural gas is a feedstock in refinery operations (Huntington, 2007). While some natural gas can be replaced with synthetic gas residues of the refinery processes, gas remains a preferred feedstock due to its superior quality and Btu content. As long as refiners need natural gas as a feedstock, an increase in the value of refined products should increase the demand (and thus price) for both crude oil and natural gas. This is perhaps the most persistent link between the commodities, especially in light of the development of the newest technologies for unconventional oil production. However, the actual volume of natural gas used for petroleum refining is not a very large portion of total gas demand. While refiner consumption of gas is a justification for assuming that a persistent link exists, less than 1% of natural gas consumption in the U.S. was devoted to hydrogen production in refineries (U.S. DOE, website(i)).

The oil sands projects in Alberta use a large volume of natural gas both for extraction of the tar sands and in the refining process for synthetic crude. The oil sands industry is thus more natural gas intensive than the conventional crude oil industry. In the extraction process, natural gas is used to heat steam, which is injected into wells to melt the tarry residues mixed with the sand and soil. The residues drip down into a deeper, horizontal well core where the liquefied tar is then extracted. In order to refine the tars into even a low-quality crude oil, a large amount of natural gas is required yet again. In the twelve-year period from 1996 to 2008, natural gas demand for oil sands activities rose from 406 million cubic feet of gas per day to nearly 952 million cubic feet per day (Canadian Association of Petroleum Producers (2009), National Energy Board (2007)). This represents about 1.5% of total U.S. gas consumption in 2008.

With Canadian gas demand growing and its production relatively stagnant, supply pressures have given rise to theories that gas usage in the oil sands industry may push Canadian gas export prices above their historical discounts and into parity with U.S.-sourced gas. Figure 12 illustrates the growing demand pressure on Canadian-supplied natural gas. Adding the share of gas consumption devoted to the refining and oil sands industries to the running total of the share of gas consumption that could potentially be displaced through interfuel competition means that about 20.4% of total U.S. natural gas usage can somehow interact with the oil industry, either through gas usage or through competition between the fuels.



2.3.5. International Linkages in Crude Oil and Natural Gas Prices

A major reason that this study focuses on spot oil and gas prices in the United States is that it is perhaps the only place in the world where any correlation between the two spot-price series can arguably be due to underlying market conditions and not due to a convention in contracting. Many European long-term natural gas contracts as well as most global long-term contracts for

LNG deliveries are priced using crude oils as a benchmark. In Asia, it is the Japanese Customs Clearing price (JCC)[6] that serves as the main price component for LNG deliveries. However, most contracts also have both floor and ceiling prices to control volatility and to mitigate producer and consumer risk (Jensen (2003), p. 12, 33). If the decision had been made to analyze the possibility of a linkage between crude oil and natural gas prices anywhere but in the U.S., where each market is deregulated, independent, and operating with very liquid hubs for spot trading, finding a relationship would have been certain. Because the relationship would include the influence of a link by contract design, however, it would have been impossible to separate the effects of the two distinct components of linkage. Nonetheless, even in the U.S. there is a basis through the LNG trade by which a portion of the gas market could have its prices tied to crude oil and/or petroleum products. (Jensen (2003), p. 3).

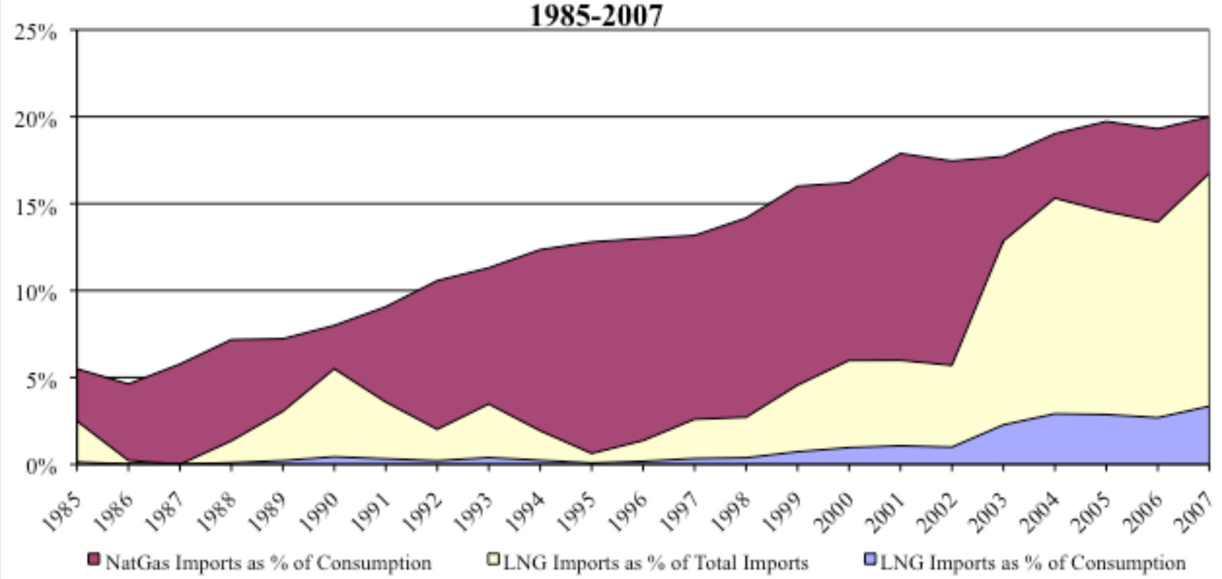
A probable reason for using oil as the benchmark for gas prices in long-term contracting is precisely because oil prices are less volatile. A secondary reason could be because both oil and gas serve some of the same uses, so movements in oil prices should at least partially reflect the same underlying forces that move the price of natural gas. A third reason might simply be because there is no local gas production capacity of any significance in Japan or Europe to which to benchmark prices.

The U.S. accounts for a major portion of global LNG spot cargoes (Jensen, 2003). These prices are set in the world market for LNG. If the world LNG spot price is influenced by the oil-indexed long-term LNG contracts, then there is a possibility that the LNG spot price at least

[6] Colloquially known as the "Japanese Crude Cocktail."

partially reflects movements in the international crude oil price. The amount of influence in the U.S. is likely to be rather small. Figure 13 provides a brief illustration of why one would not expect an oil price pass-through from the LNG component of U.S. gas consumption to affect the spot price relationship very much.

Figure 13 - Natural Gas Imports and LNG Imports as Percentage of Consumption and LNG Imports as Percentage of Total Imports, 1985-2007



The graph shows three key fractions regarding natural gas imports. First, it shows that total imports of natural gas (including Canadian pipeline imports) as a percentage of total consumption have risen from about 5% in 1985 to about 20% in 2007. It also shows that LNG imports have grown to about 17% of total imports. The growth in LNG imports mirrors the decline in pipeline imports from Canada. Together they account for over 99% of total U.S. imports every year except for 2007, when they account for nearly 99%. This translates to the figure illustrated by the bottom plot: LNG imports have grown to nearly 3.5% of total U.S. gas consumption. Total U.S. gas consumption in 2007 was 23 trillion cubic feet (tcf), of which LNG made up just 770 billion cubic feet (bcf) (U.S. DOE website (c)). If gas prices are a weighted average of the prices of the various sources of natural gas, then only about 3.5% of the average natural gas price in the U.S. is a result of LNG import prices. There are offshore and coastal LNG terminals in the Gulf of Mexico and on the Louisiana coast. These are all connected to the regional pipeline system. Therefore, it is not inconceivable that the prices of these LNG imports could be reflected in the Henry Hub price.

3. Literature

Section 2 provided a justification for why one might expect there to be a relationship between crude oil and natural gas prices. There is a baseline at which demand for petroleum is directly related to demand for natural gas through the use of natural gas as a petroleum refining feedstock. There is an aspect of upstream competition at the wellhead for resources, in which one could expect decisions to drill for one commodity to affect the supply of the other commodity. This would thus belatedly affect the price of the unexplored commodity so that it moves toward the price of the commodity for which exploration took place. There is also an element of competition between the two fuels, with the true overlap lying between the two extremes cited above. In short, the measurable oil-gas linkages could affect between 7% and 25% of total gas usage.[7]

Assuming that a price relationship exists, how would one identify it and how should one quantify it? What kinds of tools should be used? Prior literature addresses these questions, though the approaches to analysis and the conclusions drawn are not uniform.

[7] The price effect from wellhead competition was not measured in Section 2.

Serletis and Herbert (1999) examined spot natural gas prices at Henry Hub and another hub in the Northeast between Northern Virginia and New York City called Transco 6. These prices were compared to distillate fuel oil prices in the mid-Atlantic region (New York Harbor deliveries). They used a vector autoregression (VAR) model with an error correction mechanism. This type of model assumes two things: that prior observations have an effect on observations that follow them, and that deviations from a long-run relationship diminish with time. This type of model will be discussed in greater detail in Section 4.

Serletis and Herbert concluded that any of the two series compared as a pair were cointegrated. That is, Henry Hub was cointegrated with Transco 6, Transco 6 was cointegrated with NY Harbor fuel oil, and Henry Hub was likewise cointegrated with NY Harbor fuel oil. Cointegration is a term that applies to time series variables such as prices. Usually, time series variables exhibit a time-related trend. They are not stationary, i.e., the mean of the change of each series is not zero. When two or more series that are not stationary can be made stationary through a linear combination of the two series, they are considered cointegrated. This implies that price movements in one of the series are strongly correlated with the other series along the same trend. Serletis and Herbert concluded that the coordinated price movements reflect the effect of arbitrage in distributing market pressures among the fuels. Specifically, they estimated that Henry Hub price changes reflect about 91.5% of the changes observed in Transco 6 prices, with Transco 6 both causing Henry Hub prices *and* vice versa. When Henry Hub prices were compared to fuel oil prices, the authors found that generally a 1% change in fuel oil prices would effect a 1.8% change in prices at Henry Hub. However, the linkage between fuel oil prices and natural gas prices was only weakly supported in terms of statistical significance. In other words, they found Henry Hub prices to be somewhat dependent on fuel oil prices, with changes in fuel oil prices being magnified and transferred to Henry Hub (Serletis and Herbert, 1999, p. 478). For the error correction term, only the relationship between Transco 6 and Henry Hub turned out to be statistically significant; they concluded that 16.1% of the price difference between Henry Hub and Transco 6 would be eliminated *each day* (ibid., p. 481). However, the notion that natural gas prices at two distinct hubs would tend to move together in markets physically linked by pipelines is hardly a revelation.

A later study by Serletis and Rangel-Ruiz (2004) examined WTI and Henry Hub prices. They developed a bounds testing approach based on the work of Pesaran et. al. as well as a shared cycles test developed by Engle and Kozicki (1993) and Vahid and Engle (1993). They used these methods to test cointegration between oil and gas using a variation of the vector error correction model[8] that they called an autoregressive distributed lag. In the course of their investigation, they found evidence that there was a long-run relationship between the WTI crude oil and Henry Hub natural gas markets.

They concluded that the lagged prices of each commodity seem to have a predictive ability on near-term future prices – i.e., there was some kind of momentum in price trends, so that past movements tended to influence future ones. In other words, past natural gas prices had an effect on later natural gas prices and past crude oil prices had an effect on later crude oil prices. However, when they attempted to determine whether this commodity-specific momentum was

[8] The vector error correction model is explored in detail in Section 4.

part of a common trend *between* the two commodities, they could not conclude that crude oil and natural gas shared a common cycle.

The authors then examined whether the prices might in fact be linked to a common cycle but with different time lags, (a phenomenon they termed *codependent cycles*). This model also failed to prove a common relationship between the price series. Serletis and Rangel-Ruiz concluded that the oil and natural gas markets in the United States had become "decoupled." They hypothesized that the deregulation of oil and gas markets in the U.S. had removed whatever factor might have been causing the two price series to share a common relationship in the past.

More recently, Villar and Joutz (EIA, 2006) contended that Henry Hub natural gas prices follow WTI crude oil prices because they are substitutes in consumption and both rivals and complements in the production sector. This is the same as the justification that was developed in Section 2 of this thesis. They also postulated that oil prices lead gas prices because oil has a global market, while natural gas has a regional one (Villar and Joutz, p. 2). Villar and Joutz use a VAR model and an error correction mechanism to estimate a long-run relationship between the two commodities as well as price feedbacks between the series.

In addition, they include in their estimate the effects of some of the fundamental variables for natural gas prices. They used data on working inventories of gas and their levels relative to the 5-year average, heating-degree days, cooling-degree days, and heating- and cooling-degree day variations from normal. For event-related supply or demand shifts and seasonal impulses, the authors used dummy variables. As a final refinement of their model, they used the deviations from the long-run predicted oil-gas price relationship to fit a VAR model that omits the insignificant variables in their first regression. They called this the conditional error-correction model (Conditional ECM).

Villar and Joutz concluded, in contrast to Serletis and Rangel-Ruiz, that the two commodities are cointegrated and that their long-run relationship is stable after accounting for a constant time trend. Their model postulated a dissipating shock transfer, where a 20% shock to WTI prices makes an immediate 5% impact on the Henry Hub natural gas price, which dissipates to 2% in two months. A permanent shock to WTI prices was expected to result in a 16% increase in the Henry Hub price one year later if all else were held equal. However, their model also assumed a 0.52% independent and positive trend in natural gas prices. This meant that, if not for the error correction mechanism pulling gas back in line with oil prices, natural gas prices would independently increase without end. Villar and Joutz cited this as a principal weakness in their model, and the reason that the EIA would not be able to implement the model in their *Annual Energy Outlook* and *Short-Term Energy Outlook* reports. While their study concluded that a link exists, the mechanism by which the oil price is passed on to the gas price was not explored.

Bachmeier and Griffin (2006) use both error-correction and vector error-correction models to examine possible cointegration between global crude, U.S. coal, and U.S. natural gas markets. Their findings strengthen the theory that the oil markets are globally integrated into a single liquid pool, but they find little evidence that gas is integrated with either coal or oil. While their "simple tests for statistical significance on the long run cointegrating coefficients implies all three fuels are cointegrated, …this is misleading." (Bachmeier and Griffin (2006), p. 68). The strongest relationship they could find between WTI oil and natural gas prices was that a price shock in WTI was only passed on to gas contemporaneously at a rate of 2.2%, and that on a BTU-parity level, a $1 per mmBtu price change in natural gas was reflective of only a $0.60 per mmBtu price change in WTI. They concluded that the linkage was far from a truly cointegrated marketplace between oil and natural gas. That is, Bachmeier and Griffin dispute the notion that there is a primary energy market in which all relevant fuels compete. However, their *a priori* hypothesis – that all fuels compete in a global energy marketplace – was much stronger than that of the other authors.

Brown and Yücel (2008) accept the probability that there is a long-term relationship between oil and gas prices, led by oil. They attempt to describe the deviations from the oil price in the gas price through changes in the fundamentals of the natural gas market. The exogenous variables whose effects are examined by Brown and Yücel are what the traders themselves use to judge supply and demand factors in their transactions. Their model tested for a cointegrating relationship between Henry Hub natural gas and WTI crude oil prices. They sought to determine if the uncorrelated movements in the Henry Hub natural gas price could be explained through changes in the variables of supply and demand that pertain to natural gas alone: heating and cooling degree days, deviations from normal heating and cooling degree days, the storage differential from the five-year average, and shut-in production. This "fundamentals" approach has much appeal, since it attempts to explain natural gas price volatility using the same variables that traders use as criteria when making transactions.

Much of this thesis draws upon their modeling approach. They use a variation of the VAR called a vector error-correction model (VECM) that can simultaneously account for the oil and gas price series, the lagged effects of the commodity prices on later observations, and the fundamental variables. The VECM also includes an error correction mechanism. Brown and Yücel concluded that a cointegrating relationship indeed exists between WTI crude oil and Henry Hub natural gas.

Most recently, Hartley, Medlock and Rosthal (2008) developed a VECM that examined cointegration between monthly Henry Hub natural gas and residual fuel oil prices with strong results. Here WTI acts on the price of Henry Hub natural gas only through its effect on residual fuel oil. They account for the trend that Villar and Joutz identified in their paper as the incremental improvements in the heat rates of gas-fired power generators.[9] Hartley's group used this variable as part of the estimation of the competition between residual fuel oil and natural gas on a portfolio scale. They found that the change in Henry Hub monthly natural gas prices reflects almost 56% of the contemporaneous change in residual fuel oil prices, and that the error-correction mechanism makes up for the difference between the actual and the long-run equilibrium price at a rate of 23% per month. The authors used deviations from normal heating- and cooling-degree days in their equation as fundamentals, but not the levels themselves. The variables showed a very small but statistically significant effect on the price of natural gas, which presented itself noticeably only when the figures were extreme. This is a significant departure from the Brown-Yücel model, and this decision will be discussed in greater detail later on.

Hartley, Medlock and Rosthal had another stage included in their model – the residual fuel oil price was determined by the WTI price. While this is not surprising, they argue that it is through

[9] A heat rate is a measure of a natural gas plant's efficiency. It is measured as the amount of Btus that must be burned in order to generate a kilowatt-hour (kWh) of electricity.

the mechanism of the effect that residual fuel oil has on natural gas prices that crude oil prices appear to determine the natural gas price.

4. Finding the Oil-Gas Price Relationship

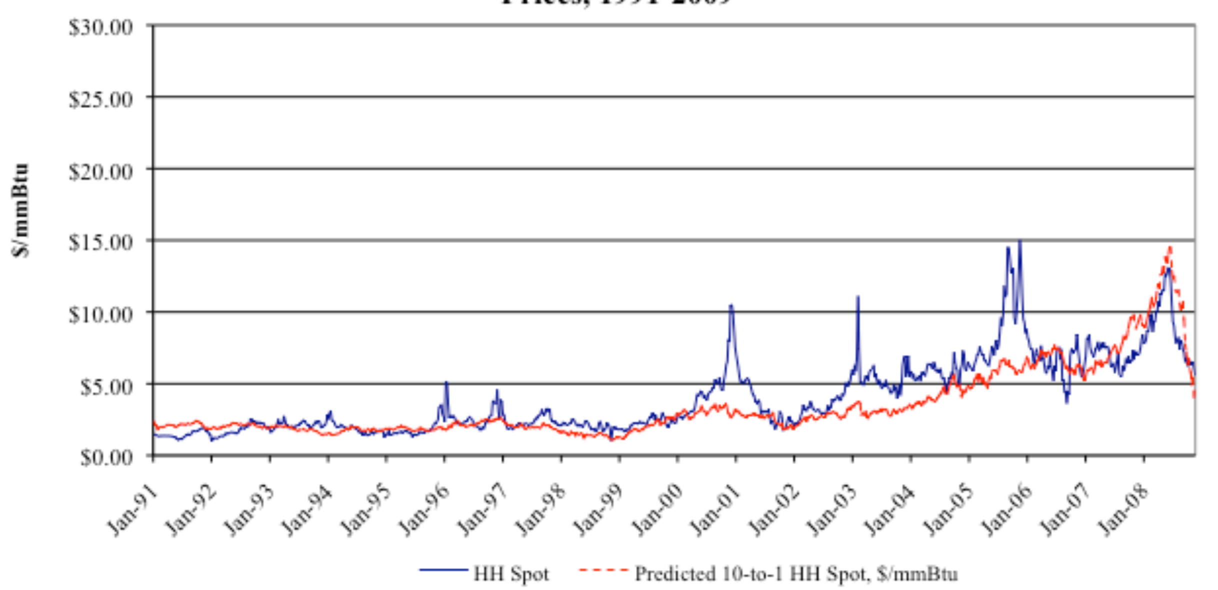
The Literature section provided a review of the methodologies and approaches pursued by those involved in researching the crude oil-natural gas pricing relationship. All used econometric methods as the principal methodology. Why such a decision was made will be explored later in this section. Not all of the authors are in agreement. Two of the papers listed decided that a linkage did not exist, while the rest claimed that oil and gas prices were linked. This study will take a fresh look. First, however, the historical methods for determining the pricing relationship between crude oil and natural gas will be reviewed.

4.1. Rules of Thumb

Over the years, traders and refiners have developed simple methods for estimating the long-term price relationship between crude oil and natural gas. Brown and Yücel (2007, 2008) detailed the most common. This discussion of these rules of thumb is largely drawn from theirs. For each, the predicted price according to the rules of thumb are compared against the actual natural gas price at Henry Hub in $/mmBtu from January 25, 1991 through February 20, 2009. These dates represent the earliest date for which natural gas prices became available on the Bloomberg data terminal and continue until after the end of the oil price run-up of 2008.

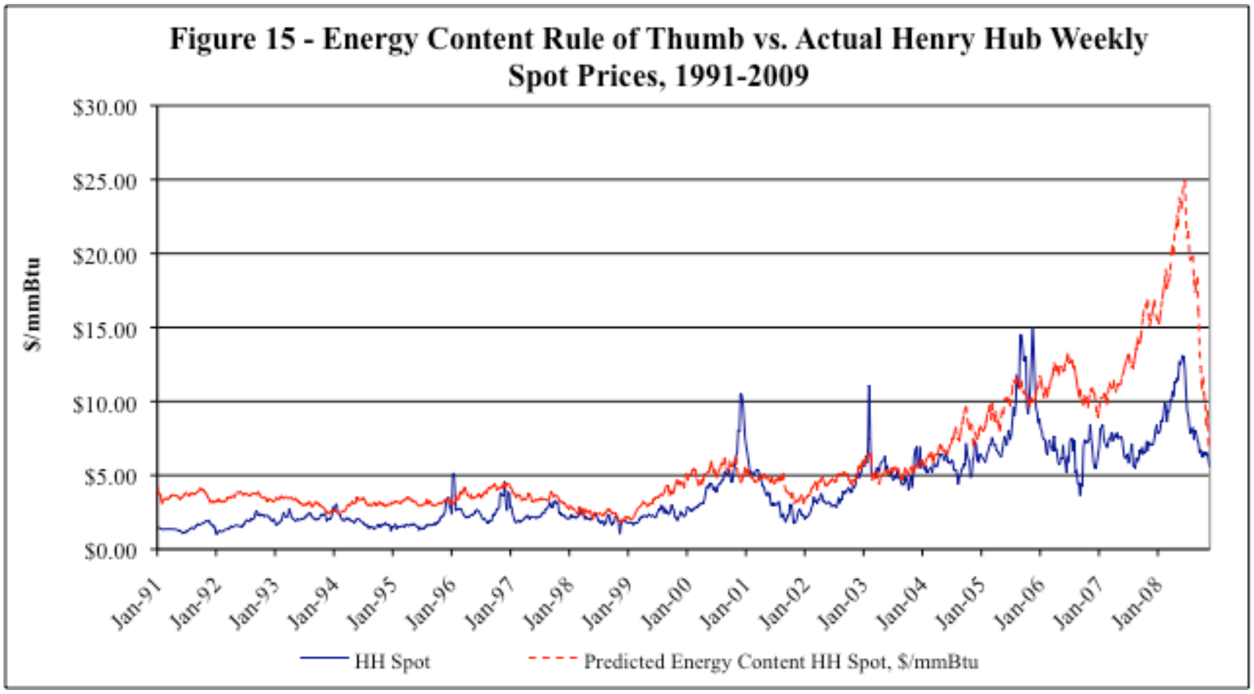
The first rule of thumb to examine is the 10-to-1 rule, which assumes that the natural gas price per mmBtu is simply one-tenth of the crude oil price per barrel. Hartley, Medlock and Rosthal identified the 10-to-1 rule as likely to be the oldest, and one of the best simple relationships between oil and natural gas prices for the 1990 through 2000 period. Historically, they identify the relationship between crude oil and natural gas prices as ranging from 4-to-1 to 12-to-1 (Hartley, Medlock and Rosthal, 2008). Figure 14 shows how the 10-to-1 rule has performed against the actual natural gas price at Henry Hub using WTI crude oil prices since 1991. The solid blue line is the actual Henry Hub price for natural gas. The broken red line is the less-volatile prediction derived from dividing the WTI oil price by 10. The 10-to-1 rule is, at best, an imprecise estimator of actual prices in the mid-range.

Figure 14 - 10-to-1 Rule of Thumb vs. Actual Henry Hub Weekly Spot Prices, 1991-2009



Another simple rule of thumb is what is sometimes called the Energy Content rule. While it has the ostensible advantage of comparing the Btu content of a barrel of oil (5.825 mmBtu) and an mmBtu of natural gas, it too fails to accurately predict the actual gas price when given the WTI crude oil price under most circumstances. It appears to fit only within a certain price range, as is evident from an examination of Figure 15. Again, the solid blue plot represents actual Henry

Hub natural gas spot prices, while the dashed red line represents the gas price as predicted by the Energy Content rule. In short, it is apparent that natural gas is priced at a discount to oil on an energy content basis.

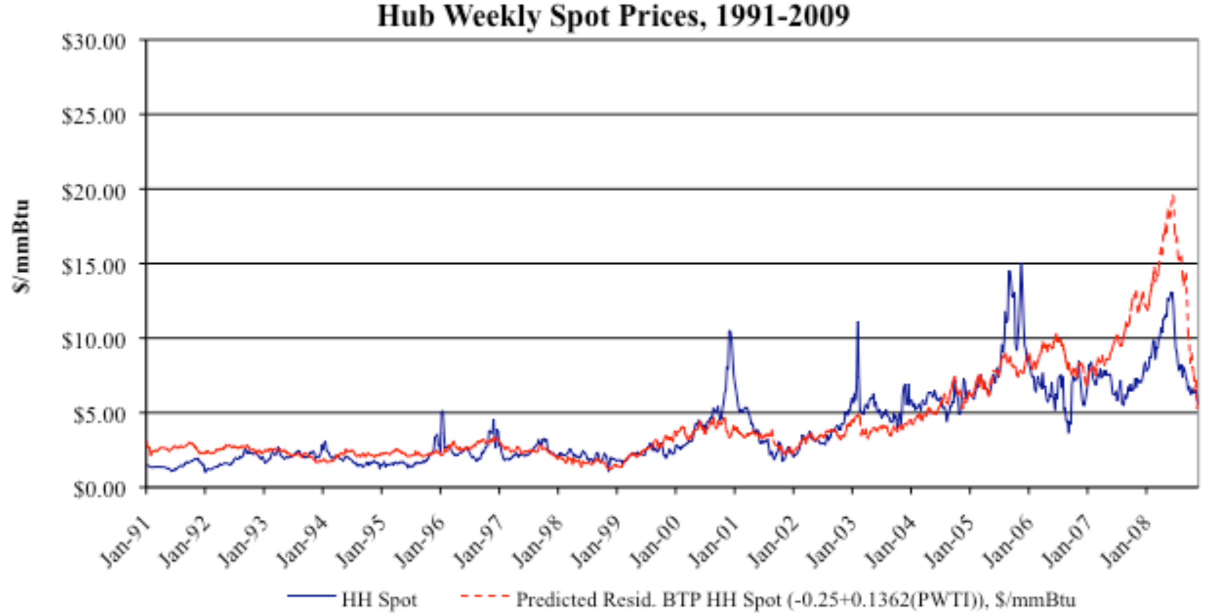


In addition to these two simple rules, there evolved more complex attempts to estimate the long-term price relationship between oil and gas. Brown and Yücel covered both of these in their paper, and they are repeated here for the benefit of comparison with the econometric relationship. The burner-tip parity rules are economic-minded efforts to define a price relationship between oil and natural gas that reflect actual market conditions. Each first equalizes the price per mmBtu of the nearest competitor fuel for natural gas at the market of greatest competition. That relationship is then used as a predictor of the natural gas price from the price of the relevant petroleum product. The first of these is called the residual fuel burner-tip parity rule. It involves two steps: first, since residual fuel is a byproduct of petroleum refining, a price adjustment must be made. Brown and Yücel cited a Barron and Brown (1986) study that used a historical average residual fuel oil price of about 85% of the crude oil price. A quick examination shows that the average residual fuel oil price at the retail level is 85.6% from May 1983 through December 2008, with a standard deviation of 11.1% (U.S. DOE, May 2009b). For Btu parity, this price must be divided by the Btu content of a barrel of residual fuel oil (approximately 6.827 mmBtu/bbl). The price for a million Btu of residual fuel oil is thus 0.1362 times the price of a barrel of WTI crude oil (P_{WTI}). Brown and Yücel also account for a transportation differential in favor of Henry Hub natural gas of about $0.25 per mmBtu (Brown and Yücel, 2007). The final residual fuel burner-tip parity rule of thumb is thus expressed:

$$P_{HH,t} = -0.25 + 0.1362 \text{ x } P_{WTI,t}$$

Here, $P_{HH,t}$ is the Henry Hub natural gas price at time *t*, and $P_{WTI,t}$ is the price of WTI crude at time *t*. A WTI price of $40 per barrel under this rule of thumb would translate to a Henry Hub natural gas price of 5.20$/mmBtu. Figure 16 plots the price graph of this modified version of the Energy Content rule of thumb over the period for which Henry Hub prices were available.

Figure 16 - Residual Burner-Tip Parity Rule of Thumb vs. Actual Henry Hub Weekly Spot Prices, 1991-2009



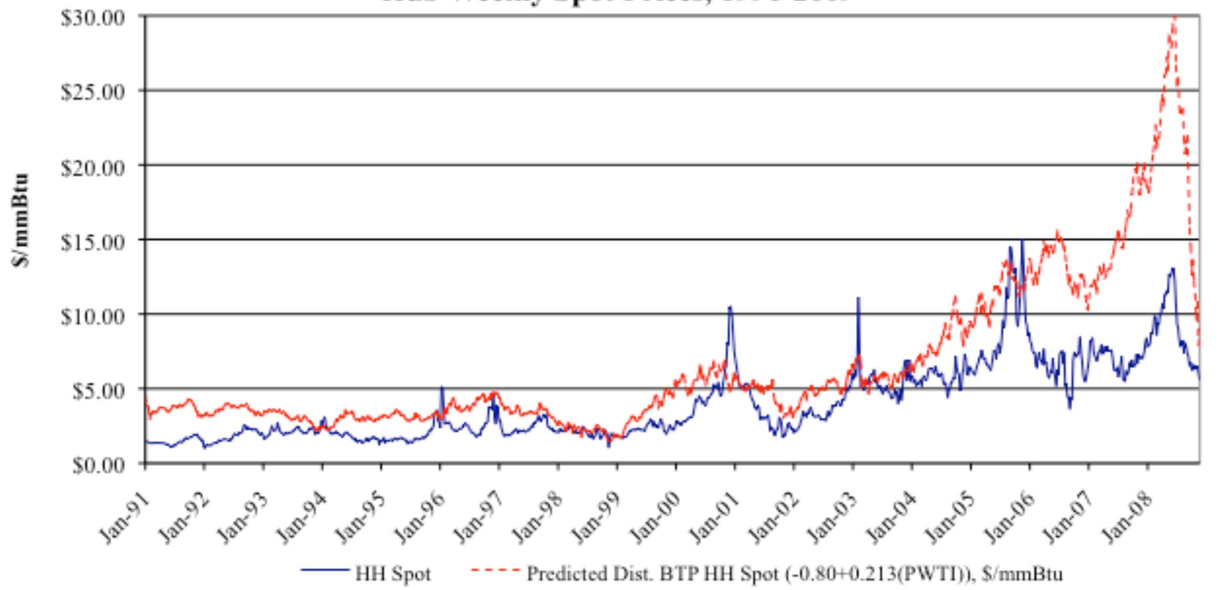
The refinement, represented by the broken red line, nonetheless tends to overestimate the price of Henry Hub natural gas when oil prices are high. It is, however, less distorting than the price predicted by the Energy Content rule. Part of the problem here is that, between 1983 and 2008, the price of residual fuel oil as a percentage of WTI crude oil prices at the retail level ranged from 55.1% to 164% per barrel of WTI crude oil (U.S. DOE, May 2009b).

Another potential competitor for natural gas, to a lesser extent than residual fuel, is distillate fuel oil. This is a higher-quality product than residual fuel oil, and it is priced accordingly. Within certain price ranges, especially where delivery through a natural gas pipeline system is limited, distillate fuel oil may serve as an imperfect substitute for natural gas. Thus, a potential relationship between the two fuels exists, especially when the price of natural gas is high compared to the oil price. Over the same 1983 to 2008 period, a barrel of distillate fuel oil averaged 123.9% of the price of a barrel of WTI crude oil, with a standard deviation of 12.1% (U.S. DOE, May 2009b). With a Btu content of 5.825 mmBtu per barrel of distillate fuel oil, this implies a relationship with crude oil of 0.213 x P_{WTI}. Brown and Yücel's calculated transportation differential is about $0.80 per mmBtu, also in favor of the pipeline-delivered natural gas. Taking these figures into account, the distillate fuel oil burner-tip parity rule of thumb is expressed:

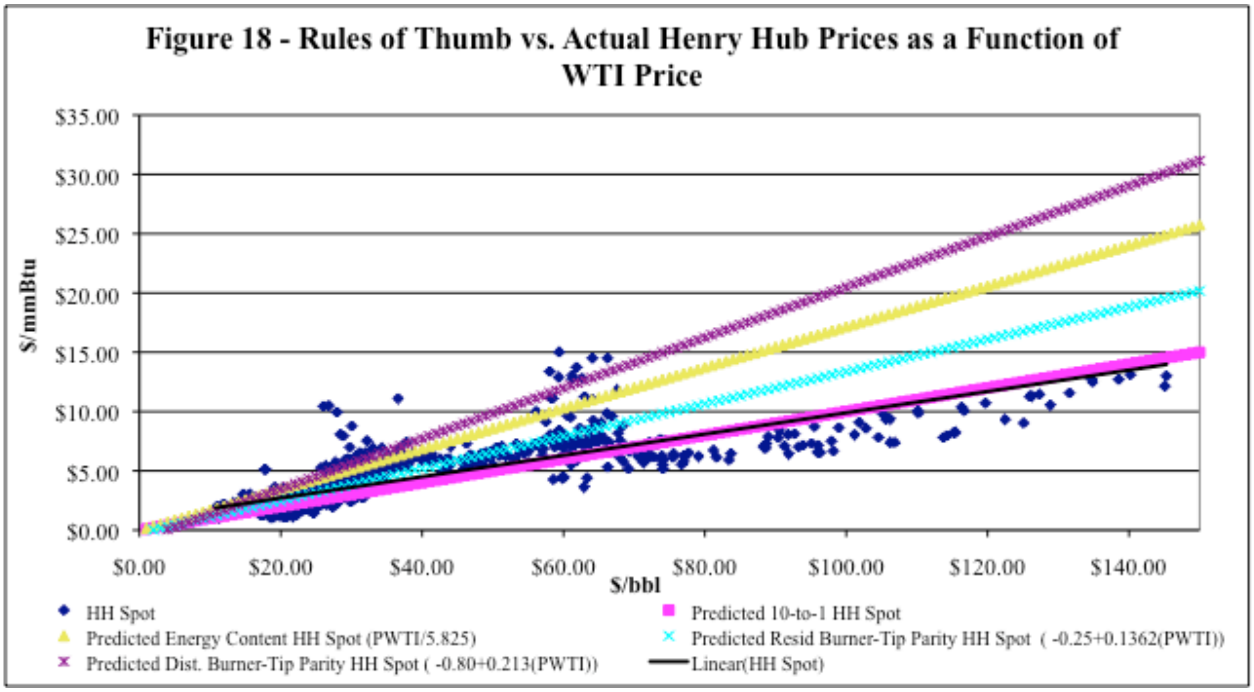
$$P_{HH,t} = -0.80 + 0.213 \times P_{WTI,t}$$

where $P_{HH,t}$ and $P_{WTI,t}$ represent the same prices as in the residual burner-tip parity equation above. A $40 barrel of WTI under the distillate burner-tip parity rule would predict a Henry Hub natural gas price of 7.72$/mmBtu, which is traditionally a fairly high number for natural gas prices when oil prices are as low as $40 per barrel. Figure 17 plots the price graph of the distillate burner-tip parity rule over the time period for which natural gas prices were available for purposes of comparison with the actual Henry Hub price. The Distillate Burner-Tip Parity (BTP) rule is the least accurate of the rules of thumb, with routine overestimation of the natural gas price. Like the Residual BTP rule, the problem is that the actual price of distillate fuel oil ranged from 91.6% to 206.8% of the WTI crude oil price per barrel (U.S. DOE, May 2009b).

Figure 17 - Distillate Burner-Tip Parity Rule of Thumb vs. Actual Henry Hub Weekly Spot Prices, 1991-2009



There is perhaps a better way to illustrate how the rules of thumb perform as predictors of the Henry Hub natural gas spot price. A graph can be constructed in which the X-axis represents WTI crude oil prices in dollars per barrel and the Y-axis represents Henry Hub natural gas prices in $/mmBtu. The rules of thumb, being linear relationships, all appear as straight lines on the graphic. To judge their accuracy, the actual Henry Hub natural gas prices are presented in a scatterplot. The rules of thumb are plotted against actual Henry Hub natural gas prices in just such a fashion in Figure 18. A cursory look should make clear that none of the rules of thumb accurately account for the wide range of natural gas prices over the entire range of WTI prices. The highest line represents the distillate BTP rule. The next highest line is the Energy Content rule. Below the Energy Content rule lies the Residual BTP rule in light blue. The purplish plot near the bottom represents the 10-to-1 rule.



Certain of the rules seem to fit certain price ranges of WTI better than others. The 10-to-1 rule is the only one that seems very accurate at the high-WTI price levels. In the $18-to-$40/bbl range, it appears that almost any of the four rules of thumb might work. In fact, a simple linear regression trend line almost perfectly matches the 10-to-1 rule of thumb trend line, with an equation of $P_{HH} = 0.9075 + 0.09 \times P_{WTI}$. It is represented by the black line. What is immediately apparent is that the actual Henry Hub prices are much more volatile than the ability of any of the rules of thumb to capture accurately.

There are many possible reasons that some of the rules of thumb work better than others at various price ranges. Many of them point to the earlier discussion of the origins of the historical relationship between the two commodities. For the burner-tip parity and energy content rules, the Henry Hub natural gas price must be high relative to the WTI crude oil price – higher than historically usual.

The validity of the energy content and burner-tip parity rules of thumb could also be considered from a theoretical viewpoint. According to this theory, the "competing" petroleum product must also be able to serve as an *actual* substitute for natural gas. This means that the two products must both occupy a supply and demand region where substitution is feasible. Not only must there be opportunities for actual physical substitution, but the fuel oil price itself must also lie in a specific price band with the crude oil price in order for the rules to accurately predict the natural gas price. This is particularly true of residual fuel oil, since it can be upgraded to higher-value petroleum products if the cost of doing so is lower than the price differential between residual fuel oil and the higher-valued products (Van Vactor, 1982). When this is in fact the case, the value of residual fuel oil will exceed the range in which it is actually competitive with natural gas, and neither the assumption that the residual fuel oil price is 85.6% of the crude oil price per mmBtu nor the rule of thumb based on this assumption will hold. As a result, the theory predicts that in this case neither the energy content nor the residual fuel oil burner-tip parity rules can be expected to make accurate predictions about the Henry Hub natural gas price.

The above drawbacks highlight the problem with all of the rules of thumb examined so far. They are all based on a simple adjustment to the crude oil price. None of them can accurately capture the idiosyncrasies of the natural gas market at any given point in time. Attempts to condition when the rules should apply and when they should not, as in the last paragraph, become very complicated. The core of the problem is that none of the rules are capable of accounting for the volatility in natural gas prices that is not already present in crude oil prices. Considering that natural gas prices are roughly twice as volatile as crude oil prices, this leaves a considerable portion of the determinants of the natural gas price unaccounted for. This is why a more complex model is necessary.

4.2. The search for a methodology

The performance of the rules of thumb as predictors of the natural gas price makes obvious the need for a model that is able to account for the volatility in natural gas prices independently of price movements in crude oil. This necessitates a more complex approach. The papers addressing this issue have all used some form of time-series econometric model, but could other methods be appropriate?

The oil and gas sectors of the economy, and the demand sectors with which they are integrated, are remarkably complex and interrelated. This brings to mind the possibility of using a dynamic optimization model. Dynamic optimization would allow for a detailed breakdown of the determinants of supply and demand in the oil and gas industries. The resulting model could take into account the factors that would be likely to shift the price relationship between the fuels, and thus provide some predictive ability.

The oil and gas industries are likewise characterized by a high level of integration in which decisions at any point in the process can affect the outcomes at other points, which in turn affect yet more portions of the process chain in a cascading fashion. System dynamics is often utilized to analyze such complex systems. A further benefit of system dynamics is that the resulting relationships need not be linear. Its promise for useful predictive ability is high – if the model can be correctly specified and identified.

It is this last point that makes both dynamic optimization and system dynamics unfeasible for the current problem in the current timeframe. First of all, the system in which the crude oil and natural gas industries operate is incredibly complex. The two commodities and their respective industries affect and are affected by an enormous number of factors spread across virtually every energy-intensive industry and well into commercial and residential demand sectors. Complex models along the lines of dynamic optimization or system dynamics would require not only painstaking construction but also copious amounts of data. Considering the complexity of the model inputs, would the results be simple and illuminating enough to be useful?

The real barrier to the creation of these models, however, is data. In order to construct either the dynamic optimization or the system dynamics models, detailed data about producers, consumers, and everyone in between would be required. Much of this data would need to reach down to the level of individual firms or sectors. In competitive industries companies are careful about sharing proprietary data. Gathering, organizing and compiling this data into a consistent and usable form is thus an impractical undertaking. While those avenues show promise for the potential insights they could provide, the data issue is the most likely reason that no other researchers have taken dynamic optimization or system dynamics approaches to the oil-gas price relationship.

4.3. The Solution: Econometric Modeling

That brings the focus back to econometrics. The benefits of econometrics to such a problem are numerous. Above all, econometric models make efficient use of data to derive relationships between data sets. While the systems within which the oil and gas industries operate are extremely complex, the outcomes of the entire process, prices, are readily available. Any additional data that could either help characterize the price relationship or serve as proxies for data that would do so can be easily integrated into the model. There is also a long history in econometrics of treatment of potential problems in which observations in the data set actually depend on outcomes in prior time periods. Controlling for such correlations allows for more precise estimation. Even data that do not directly relate to the price series, but which are correlated with data that do, can be used as regressors in econometric methods.

4.4. Finding an appropriate model

The simple regression by ordinary least squares (OLS) on the crude oil and natural gas spot price series returned a relationship that was not very different from the 10-to-1 rule of thumb. The reason is that the only predictor of the natural gas price in such a regression is the crude oil price. However, it has already been noted that natural gas volatility is much greater than that of crude oil. What is needed is a method for capturing the idiosyncrasies between the two commodities. Nobody expects the oil price to be the only determinant of the natural gas price. A reliable model will need to better account for the actual relationship between the two commodities yet adjust to factors that influence the gas price but have no effect on the oil price.

Brown and Yücel, Villar and Joutz, and Hartley et. al. likewise noted the shortcomings of these simple rule of thumb models. What is needed is a two-stage approach. A long-term price prediction can be estimated based on the historical relationship between WTI crude oil prices and Henry Hub natural gas prices in the first stage. This would resemble the OLS simple regression that was run earlier. For the second stage, variables that correspond to the fundamentals of the natural gas market can serve as exogenous variables in the model. This will adjust the natural gas price either closer or further away from the long-run relationship in response to the pressures of the gas market. The idiosyncratic volatility of natural gas prices can be captured in this fashion.

There is another aspect of the crude oil-natural gas pricing relationship that also needs to be addressed. Returning to Figure 1, it is obvious that after some shock causes the natural gas prices to diverge from the long-run relationship, they tend to be drawn back to the old relationship with the oil price. An error-correction mechanism can thus measure this behavior of natural gas prices with regard to the predicted long-run relationship.

A final piece of the puzzle is left: a mechanism for determining the extent to which past price movements affect prices in the present. This would reveal whether price movements have momentum, or whether they are truly independent of all past price activity.

Brown and Yücel, Villar and Joutz, and Harley et. al. used the Vector Error Correction Model (VECM) to address these issues. Through the VECM, the long-run relationship between oil and gas prices can be established; the rate at which prices return to the relationship after shocks can be measured, and idiosyncratic shocks in the gas price can be accounted for, as can the momentum in price movements. Engle and Granger were the principal developers of the VECM. They created it to account for a complex scenario such as the one described above (Engle and Granger, 1987). A variation of this model will be used to identify and quantify the spot price relationship between crude oil and natural gas.

The benefit of the VECM is that it makes sense from a theoretical perspective, and does not even require that one accept that the WTI price is "causing" the Henry Hub natural gas price. It is possible that the relationship between the two commodities is due to some shared factor, which could be a mechanism of any of the physical linkages discussed in the first section of this paper, or perhaps even some unexplored mechanism. The model will function in the same descriptive fashion regardless. Its purpose is to identify relationships between data sets. Whether oil prices are the cause of the gas price movements, or whether the oil price simply reflects underlying economic realities sooner than gas prices due to its greater liquidity and globalized structure, is irrelevant to the mechanics of identifying whether such a relationship exists using econometric methodology.

4.4.1. Data selection

Before delving into which data to utilize, the nature of the volatility in natural gas needs to be addressed. Specifically, why are natural gas prices more volatile than oil prices, and is the cycle of volatility related to the oil market?

Oil and gas prices are on different cycles

Returning, once again, to Figure 1, one notices that the oil price shows a long-run trend with random deviations from the trend. One might also notice that the gas price, though apparently on a similar trend, exhibits a regular peak and trough over the course of each year. These peaks and troughs are consistently more extreme than the price movements in crude oil. This reflects the fact that the bulk of gas demand occurs in the winter, when it is needed for home and commercial heating.

What is described in the previous paragraph is *seasonality*. Demand for natural gas is seasonal. Therefore, when demand is high, the price responds by rising. Part of the reason for this phenomenon is that natural gas markets are local and segmented. Gas that is stored or produced on the other side of the world has little ability to flow to the areas of high demand in time to meet that demand. This is due to a lack of pipeline and LNG tanker (and terminal) capacity. This is not the case for oil, which can move across the globe to the markets with the highest prices, thus adjusting to demand differentials across borders. Oil prices are not seasonal.

The model thus requires some ability to account for the seasonality in natural gas prices. Once again, the literature provides possible solutions. Villar and Joutz used monthly dummy variables to control for the seasonal fluctuations in gas prices (Villar and Joutz, 2006). This would resolve the seasonality issue by allowing coefficients on monthly variables to account for the monthly variability in natural gas prices. However, the model being considered will use weekly data to allow for a higher-resolution image of the oil-gas relationship, and a 51-dummy variable set corresponding to each week in the year (minus one)[10] would be rather unwieldy. It would also remove a considerable number of degrees of freedom from the model, which are useful in establishing whether the results are significant or not.

A more appealing approach was utilized by Brown and Yücel in their 2008 paper (Brown and Yücel, 2008). They made a convincing case for their choice of variables to capture seasonality. The most relevant fundamental determinant of the price of natural gas, as alluded to in the discussion of seasonality, is the temperature. Consequently, Brown and Yücel chose heating- and cooling-degree days (HDD and CDD) as the exogenous variables to account for seasonality. They serve as proxies for heating and cooling demand, and will be discussed in detail below.

Gas prices react to weather-related shocks

Another aspect of the oil and gas price series can be gleaned from Figures 4 and 6. The graphs depict oil and gas price series separately, and include historical events that occurred near the price movements. It is immediately apparent from a look at the 1995-1997 period that natural gas prices reacted both to a flooding event precipitated by an earlier-than-usual thaw, and again to a cold snap. A glance at the oil price series over the comparable period shows no reaction whatsoever. Clearly, natural gas prices are responsive to unexpected heat waves and cold snaps. This characteristic of natural gas prices is another important aspect that the model should capture. Again, Brown and Yücel (and Hartley, et. al.) provide a useful dataset: the deviations from normal HDD and CDD. These signal when the weather is different than what would be

[10] One week would not need a dummy variable, since the model will account for the price behavior of the first week by the *absence* of a dummy variable.

expected at a given time of year. Deviations from normal HDD and CDD will be described in detail below.

Another prominent difference between the oil and gas price series is apparent in 2005, when Hurricanes Rita and Katrina struck the Gulf Coast in quick succession. The effect on Henry Hub natural gas prices was extreme – the hurricanes caused the single greatest price spike in natural gas history. The comparable effect on oil prices was much more muted. What is needed is a method for capturing hurricane events. One possibility is the dummy variable. However, the problem with a dummy variable is that while the hurricane event itself may only last a single week, the damage caused by the hurricane is much longer-lived, as is the effect of the event on prices. A superior variable would thus take this longer-term effect into account. The literature provides the answer in the form of the data series from the Minerals Management Service (MMS) and its reports on shut-in natural gas production in the Gulf of Mexico (Brown and Yücel, 2008, Hartley, et. al., 2008). This data series is also treated in greater detail below. The benefit of using shut-in production data is that it translates a short-lived weather event into a longer-lived supply disruption that should provide an improved accounting of the effects of hurricanes on the natural gas spot price.

Accounting for other natural gas supply imbalances

Supply imbalances in natural gas can have longer-lived price effects than what one would expect to find due to short-run shocks. To account for this, Villar and Joutz, Brown and Yücel, and Hartley's group all make use of natural gas storage data from the Energy Information Administration (EIA). Brown and Yücel use deviations from normal storage levels as their metric. This approach is appealing for its ability to identify periods when gas storage levels are contrary to what was expected by the market at the time. This dataset will also be discussed further below.

4.4.2. Data availability and reliability

The advantage of the data described above is twofold. First, all are widely available for download from the Internet from government agencies. The HDD and CDD data, and the HDD and CDD norms, are all available on the website of the National Oceanic and Atmospheric Administration (NOAA) of the National Weather Service (NWS). The shut-in natural gas production dataset can be constructed from the Gulf region website of the MMS in their press releases. The natural gas storage levels are available for download on the DOE's EIA website in the storage section of their natural gas page.

These constitute a set of variables that can be considered independent of the price series that are to be estimated. There is no basis to assume that weather patterns or storms are determined by either oil or natural gas prices. Thus the weather-related data series fit the definition of exogenous variables. There is a possible argument that gas storage levels are indeed affected by gas prices, though none of the authors in the literature considered this an issue. There is no challenge to the inclusion of deviation from average storage levels as exogenous variables in this treatment, either.

Another point in favor of the exogenous variable datasets is that the sources for each are official U.S. government agency archives. Owing to the fact that the government publishes its methodology for data collection and routinely reviews data to correct errors, these datasets can be considered reliable measurements of actual weather, production, and storage conditions.

4.5. The Data

With the data needed for the VECM procedure identified, a description of the source, theoretical effect, and behavior of each dataset is in order. This section will examine each set separately and in detail.

4.5.1. WTI and Henry Hub prices

As discussed in Section 2.3.5, U.S. spot prices were chosen for econometric modeling to identify a potential price relationship between the two commodities because of the deregulated and independent nature of the oil and natural gas markets in the U.S. There is no contracting convention linking the two commodities explicitly, and the use of spot prices serves to filter out any other long-term contract pricing. In short, choosing the U.S. spot markets allows for a fair assessment of whether a linkage exists. Since a linkage is not explicitly created through contracting conventions, any linkage discovered would be due to shared market fundamentals or because of actual opportunities for competition or complementarity between the two fuels.

For the West Texas Intermediate (WTI) oil price series, weekly volume-weighted average spot prices for WTI-Cushing from a market sample by Bloomberg were collected. Cushing, Oklahoma is a principal trading hub for the WTI stream of crude oil. WTI is a benchmark crude in the U.S. and is the most heavily-weighted portion of the basket of crudes traded in the New York Mercantile Exchange (NYMEX) light sweet crude contract.

The gas price series was likewise downloaded from MIT's Bloomberg data terminal; weekly volume-weighted average Henry Hub spot prices were used (Bloomberg). Like WTI for crude oil, Henry Hub is the benchmark hub for pricing natural gas in the U.S. It, too, has a futures contract traded on NYMEX. The trading hub is based at the locus of natural gas gathering lines from wells both offshore and onshore in the Gulf region in Louisiana. Gas traded there is sourced from Louisiana, the Gulf of Mexico, Texas, and Oklahoma. Many long-distance pipelines originate there to deliver gas up the East Coast and into the Midwest.

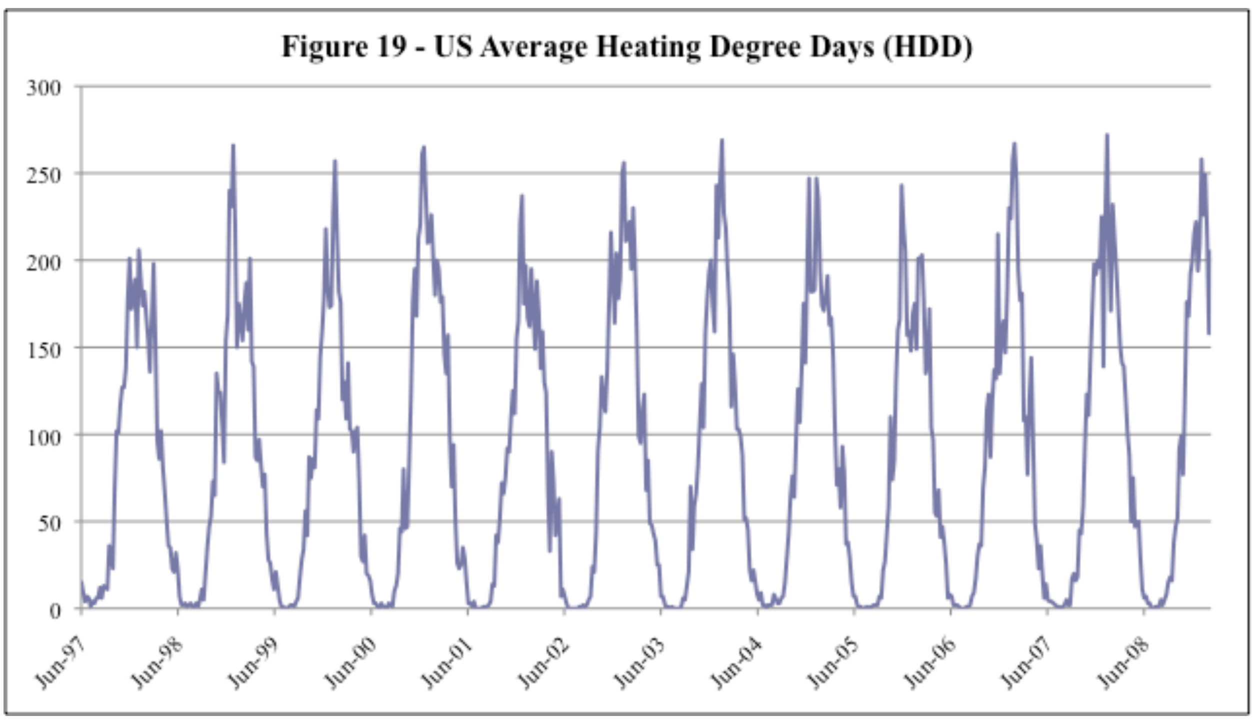
The behavior of each price series has already been exhaustively examined, both in Figure 1 and in Figures 4 and 6. The weekly price series for natural gas and crude oil run from January 25, 1991 through February 20, 2009. For the econometric analysis, at every stage the *natural logarithm* of the prices was used. This meant changes in each variable were in terms of magnitude, so that the difference in scaling would be less pronounced. In other words, it makes it easier to see that a $1 price shift is a greater change for the gas price than for the oil price. Where appropriate, the logged prices have been converted back into actual dollars so that the economic implications of the model can be better understood. Unless otherwise noted, all mentions of price in the remainder of this thesis should be considered the natural log of the price.

4.5.2. The exogenous variables and their behavior

Due to data limitations, the range of dates for which the HDD, CDD, and deviations from normal HDD and CDD could be collected was June 13, 1997 through February 20, 2009. This also forced a shortening of the window for the econometric model to the same dates. Six and a half years of the oil-gas pricing relationship thus had to be abandoned in order to construct a model that could account for natural gas volatility through the conditioning of the exogenous variables. Though they have been mentioned above, a detailed account of their source and behavior is provided below.

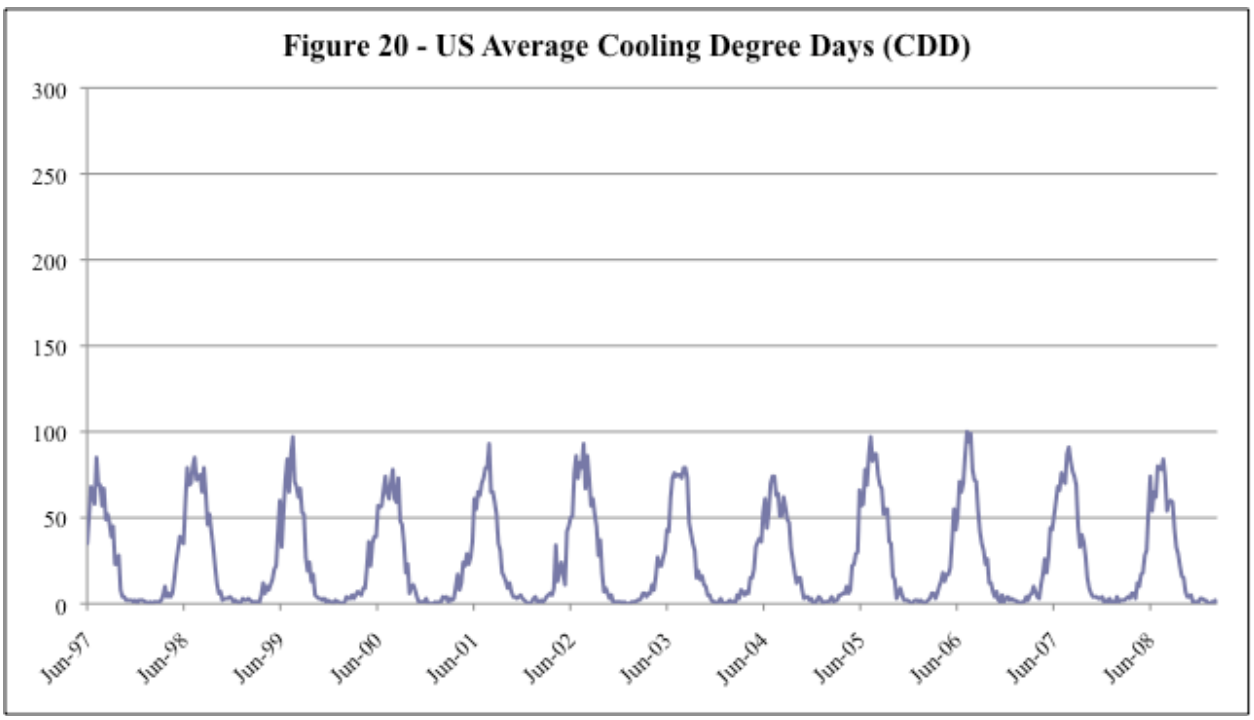
HDD and CDD data

Data on heating-degree days (HDD) and cooling-degree days (CDD) was gathered from the Climate Prediction Center of the National Weather Service under the National Oceanic and Atmospheric Administration (NOAA). Both are calculated as a weighted average of weather station temperature data. The average number of degrees below 65°F is the HDD figure. The larger the HDD figure, the greater the demand for heat. The CDD figure is calculated as the average number of degrees above 65°F, representing the demand for air conditioning. In the series collected for the modeling exercise, CDD is weighted by population and HDD is weighted by both regional gas usage and population. This project uses the gas-weighted HDD series.



Weekly figures represent a weekly accumulation of each day's HDD and CDD.

Figures 19 and 20 depict the weekly HDD and CDD data over the 1997-2009 period. The seasonal pattern is obvious. HDD levels tend to peak about three times higher than CDD levels. They also move in opposing directions. The peak for HDD is in the winter, with minimum values over the summer, while the converse is true of CDD figures. Based on the 1997-2009 data series, HDD ranged from an average high of 205 in January to an average low of 1 in July. The overall average HDD figure was 87.8, with a standard deviation of 79.8. Over the year, CDD typically ranged from a low of about 1 in January to a high of about 75 in July. CDD figures were much less volatile than HDD figures. The average CDD was 25.4, with a standard deviation of 28.2.

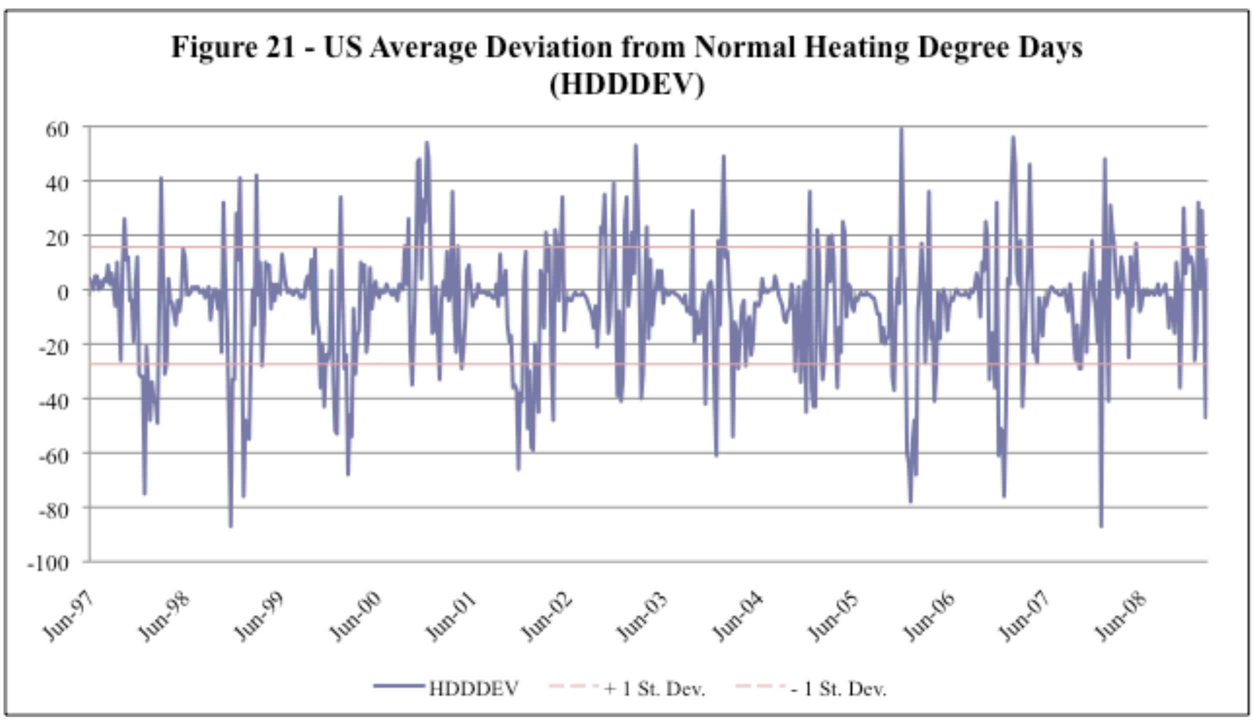


The point of including HDD and CDD data is to provide a basis to model the seasonal variation in natural gas prices over the year. Since HDD figures peak in the winter, and CDD figures peak in the summer, this correlation with the seasonal pattern in natural gas prices should provide a reasonable weather-related set of statistics on which the price seasonality can be estimated.

Deviations from normal HDD and CDD

Deviations from normal HDD and CDD levels were included in order to characterize unexpected temperature deviations from the seasonal norms. These have been labeled HDDDEV and CDDDEV. They were calculated by subtracting the 10-year average HDD and CDD figures from the actual observations in each given week. The HDDDEV and CDDDEV figures were calculated by the NOAA and presented in their weekly reports as deviations from normal HDD and CDD.

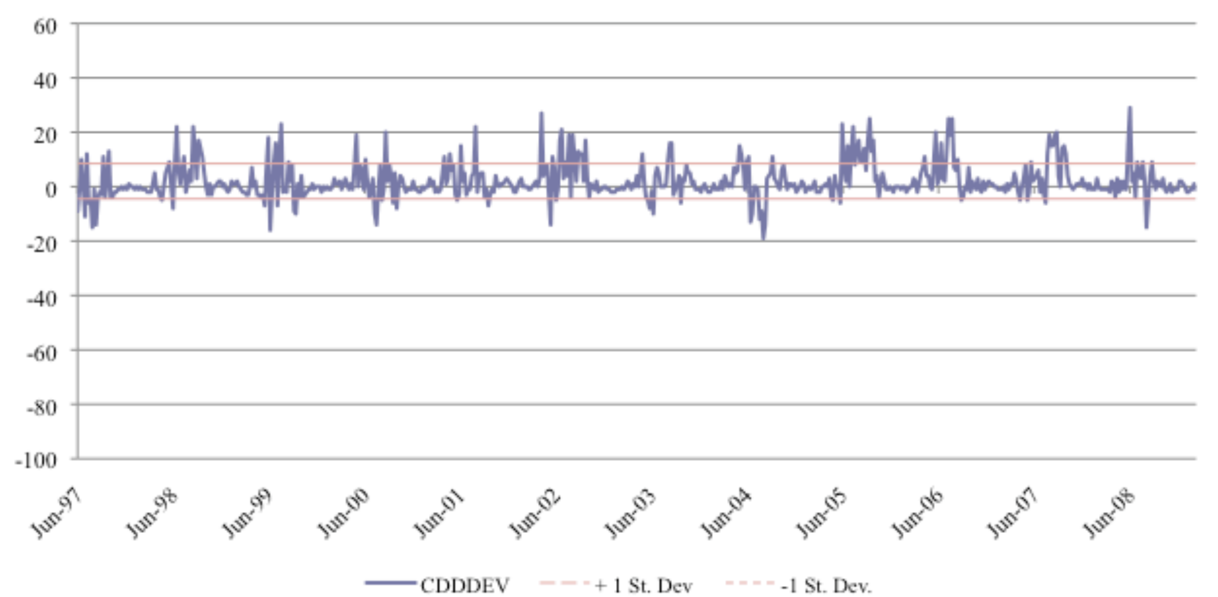
Figures 21 and 22 depict the deviations from the normal, or expected, HDD and CDD figures over the course of the year, and they include plots of one standard deviation both above and



below the mean for each. These variables are included in an attempt to model the price spikes that accompany a temperature-related shock to the gas market. The coefficient will attempt to model how much each degree of unexpected difference from the expected temperature affects the natural gas price at Henry Hub.

HDDDEV shocks tended to last about 2-3 weeks on average, reflecting the typical profile of a cold snap. It is interesting to note that over the 1997-2009 period, deviations from normal HDD were, on average, negative, at -5.9, with a standard deviation of 21.5. HDD deviations tended to be minimized over the summer, and maximized in the winter. However, over the year, negative HDD shocks (denoting warmer-than-expected temperatures) outweighed positive ones (reflecting cold snaps) so that the average HDDDEV figure in January, when the average was of greatest magnitude, was -19.

Figure 22 - US Average Deviation from Normal Cooling Degree Days (CDDDEV)



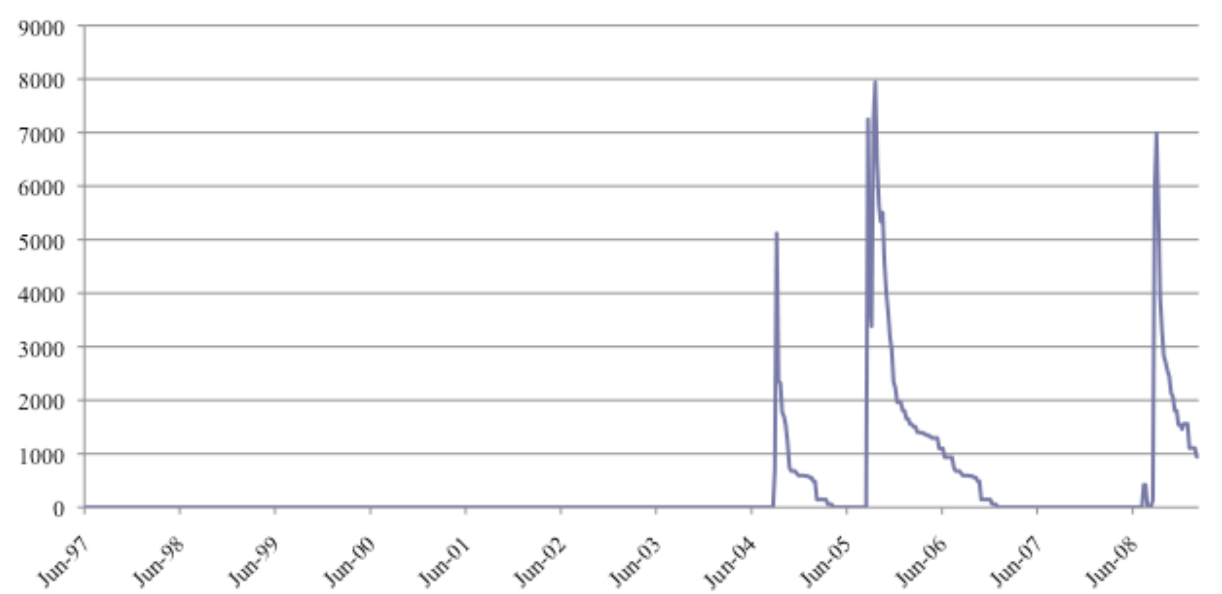
In contrast, CDDDEV measured shocks such as heat waves, with positive values denoting when temperatures exceeded expectations and negative values denoting when temperatures were below norms. They tended to last 1-2 weeks before returning to normal. From 1997-2009, the average

CDDDEV figure was 2, with a standard deviation of 6.5. Shocks tended to be highest in September, with an average value of 5, and lowest in the winter, where they tended to be zero.

Shut-in natural gas production in the Gulf of Mexico

Figure 23 plots the amount of natural gas production capacity that was curtailed in the Gulf of Mexico in million cubic feet (mmcf). The MMS posts weekly shut-in production statistics on its Gulf of Mexico webpage under the Press Releases/Reports heading whenever hurricane activity prompts oil and gas producers to halt production at their offshore platforms.

Figure 23 - Shut-in Natural Gas Production in the Gulf of Mexico (SHUTIN), mmcf



Gulf of Mexico natural gas production is roughly 10% of total U.S. gas production over the course of the year (according to the EIA). Over the 1997-2009 period, Gulf of Mexico natural gas production ranged from about 7 to 10 Bcf per day (MMS website). Hurricanes Katrina and Rita managed to shut in over 80% of Gulf natural gas production (MMS website). Hurricane disruptions to Gulf gas production have a characteristic pattern: a large spike in the week preceding a hurricane impact, followed by a gradual decline in shut-in production. This reflects the initial evacuation before hurricane contact, the immediate return to production of undamaged wells after the hurricane has passed, and the gradual return to production of rigs that suffered varying degrees of damage.

The shortest such disruption in the 1997-2009 dataset was 34 weeks in response to Hurricane Ivan in 2004. The other two hurricane events in the dataset were combinations of two hurricane impacts in each case – Hurricanes Katrina and Rita in 2005 and Hurricanes Gustav and Ike in 2008. There is thus no reliable method to calculate an "average" production impact for a single hurricane from the dataset – there is only one observation. For the model, the shut-in production statistics have been labeled SHUTIN.

Natural gas storage differential

Natural gas is stored at various locations throughout the United States, ranging from depleted oil and/or gas fields to LNG storage facilities. The EIA collects and sums the data on storage levels on a weekly basis and reports it on its website. The variable labeled STORDIFF represents the difference between a weekly gas storage level and the 5-year running average in billion cubic feet (Bcf). The average storage level from 1997-2009 was 2,256 Bcf on any given week, with a standard deviation of 716 Bcf. The average storage differential from 1997-2009 was 133 Bcf, and the average amount of time that storage remained out of sync with the normal storage level was about 39 weeks. The median duration of a storage differential was 17 weeks. Figure 24 depicts the differences from the 5-year average storage levels. Economic theory would predict that when there are greater-than-average storage levels, prices should be lower due to a glut of supplies. The converse would be expected when storage levels linger below their average.[11]

Figure 24 - Differential from Normal US Natural Gas Storage Levels (STORAGE DIFF), Bcf

4.6. The Vector Error Correction Model

Once the data were collected and organized, the Vector Error-Correction Model (VECM) was utilized to map out the relationship between Henry Hub natural gas prices and WTI crude oil prices. Stata 8 Intercooled was utilized for its versatility and ease of use. The VECM is a three-equation model. A first stage establishes a long-run relationship between the crude oil price and the natural gas price. The second stage models the change in price, both from the perspective of the natural gas price as the dependent variable and with the crude oil price as the dependent

[11] One possible issue with this reasoning is that, over time, more storage capacity for natural gas is added, so the "normal" level is on a long-term increasing trend. This could pose potential problems when modeling the oil-gas relationship using these variables as exogenous modifiers of the natural gas price.

variable. An examination of the statistics will reveal which of these formulations is most credible.

The second stage equations include a number of terms: a constant, an error-correction coefficient, coefficients representing the effects of lagged price changes of both WTI and Henry Hub prices in natural logs, and coefficients on the values of the six exogenous variables described above (HDD, CDD, HDDDEV, CDDDEV, SHUTIN, and STORDIFF). The error-correction mechanism measures the rate at which the dependent commodity variable "corrects" toward the long-term price relationship each week as the gap between actual prices and the forecasted equilibrium prices narrows. Theory and practice both predict that in this case, Henry Hub prices should be at least weakly dependent on WTI prices. The inclusion of the six exogenous variables representing the fundamentals of the natural gas price serve to model the volatility in Henry Hub prices that the rules of thumb or simple OLS regression failed to take into account.

The mathematical representation of the VECM is as follows:

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

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

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

The top equation depicts the long-term relationship between Henry Hub and WTI prices in the first stage. This is called the cointegrating relationship. $P_{HH,t}$ is the logged Henry Hub natural gas price in week t, $P_{WTI,t}$ is the logged West Texas Intermediate crude oil price in time t, γ is a constant to be estimated, and β is a parameter to be estimated. μ_t is an error term in week t. The equation designates the "target" toward which Henry Hub prices move over time.

The second equation incorporates the error correction mechanism as well as the lagged effects of the two price series on Henry Hub logged prices, followed by the effects of the exogenous (seasonal) variables on the Henry Hub price. μ_{t-1} is the lagged set of equilibrium errors in the estimated cointegrating equation. These are the μ-series from the top equation, but lagged one week. X_j is the matrix of exogenous variables representing the fundamental drivers of the Henry Hub natural gas price. ε_t is the normal error term. Finally, a, b, c, and d_j are coefficient parameters of each of the variables – they will likewise be estimated by the model regression.

The third equation is identical to the second one in structure, except that this time WTI crude oil prices are the dependent variable. Examination of the resulting probability statistics will enable a decision to be made about which of the two models is a better characterization of the actual relationship.

Testing for unit roots

Before the VECM could be estimated, a number of statistical tests had to be conducted to lay the groundwork for the estimation algorithm. The first step was to determine if a unit root existed in any of the data series. If a series has a unit root, it is autocorrelated. Autocorrelation implies that, after an OLS regression of a time series, the error terms from one time period to the next are related to one another. This means that subsequent observations are affected by previous ones, so the series will exhibit a trend rather than a sequence of random movements. Such a series is considered *non-stationary*. For the VECM to provide useful results, time series data must be stationary – especially the price series for both WTI crude oil and Henry Hub natural gas. A number of statistical methods have been developed to test for non-stationarity (i.e., unit roots). Two of them were utilized for this study.

The WTI crude oil prices, the Henry Hub natural gas prices, and the HDD, CDD, HDDDEV, CDDDEV, SHUTIN and STORDIFF series were tested with the Augmented Dickey-Fuller (ADF) test and with the Phillips-Perron test.[12] The results are presented in Tables 2 and 3.

Table 2 - Augmented Dickey-Fuller Tests

Dates of analysis (weekly): 6/13/97 - 2/20/09

Variable	Levels	Significance %	1st Differences	Significance %
lnhh	-2.441	13.1%	-24.984	0%
lnwti	-1.477	54.5%	-27.185	0%
hdd	-3.786	0%	NA	
hdddev	-15.358	0%	NA	
cdd	-3.611	0.6%	NA	
cdddev	-16.675	0%	NA	
stordiff	-2.716	7.1%	NA	
shutin	-6.148	0%	NA	

Both the logged WTI crude oil and the Henry Hub natural gas prices failed to reject the null hypothesis that there is a unit root and that the data are non-stationary. The ADF coefficient for the natural log of the Henry Hub price series was -2.441, with a p-value of 13.1%. The ADF coefficient for the natural log of the WTI crude oil price was -1.477, which corresponds to a p-value of 54.5%. Identical conclusions can be drawn from the Phillips-Perron test results. For the logged Henry Hub natural gas prices, the Z(rho) statistic is -9.103 and the Z(t) statistic is -2.262.

[12] On Stata 8 Intercooled, the Augmented Dickey-Fuller test is called the Interpolated Dickey-Fuller test. The command is "dfuller". For the Phillips-Perron test, the command is "pperron."

These correspond to a p-value of 18.4%. In the case of logged WTI crude oil prices, the Z(rho) statistic is -3.204 and the Z(t) statistic is -1.399. These correspond to a p-value of 58.3%. The p-value is the probability that the prices observed are the prices one would expect to observe if the null hypothesis were true. Normally, to reject a null hypothesis would require a p-value below the 5% level. The ADF and Phillips-Perron test statistics thus indicate that both the Henry Hub natural gas prices and the WTI crude oil prices exhibit unit root processes.

Table 3 - Phillips-Perron Tests

Dates of analysis (weekly): 6/13/97 - 2/20/09

Variable	Levels Z(rho)	Levels Z(t)	Significance Z(t) %	1st Differences Z(rho)	1st Differences Z(t)	Significance Z(t) %
lnhh	-9.103	-2.262	18.4%	-558.254	-25.164	0%
lnwti	-3.204	-1.399	58.3%	-664.208	-27.207	0%
hdd	-33.882	-4.111	0%	NA	NA	
hdddev	-331.427	-15.223	0%	NA	NA	
cdd	-40.339	-4.499	0%	NA	NA	
cdddev	-430.649	-17.27	0%	NA	NA	
stordiff	-15.161	-2.802	5.8%	NA	NA	
shutin	-64.148	-5.842	0%	NA	NA	

Often, time series data that has a unit root can be made stationary by differencing. That is, instead of modeling with weekly natural gas and crude oil price levels, one can examine the price *changes* each week. The Augmented Dickey-Fuller and Phillips-Perron tests confirmed that the differenced WTI crude oil and Henry Hub natural gas prices were stationary at the 1% level. For the Henry Hub natural gas prices, the first differences yielded an ADF test statistic of -24.984 and a Phillips-Perron Z(rho) statistic of -558.254. The Phillips-Perron Z(t) statistic was -25.164. The equivalent statistics for the differenced WTI crude oil price series were -27.185, -664.208, and -27.207 respectively. Since both the ADF and Phillips-Perron tests showed that the first-differenced time series were stationary, the two are said to be autocorrelated of order one – they each have one unit root. The results indicate that one should be able to use a regression model on the first differences of the two price series and avoid distortion due to the effect that previous price levels have on subsequent observations.

All but one of the six exogenous variables rejected the null hypothesis that a unit root existed when tested in levels. The exception was the storage differential, STORDIFF. The ADF test statistic was -2.72, with a p-value of 7.1%, while the Z(rho) statistic from the Phillips-Perron test was -15.16 and the Z(t) statistic was -2.80. The latter two corresponded to a p-value of 5.8%. As gas usage has increased over the years, more capacity for gas storage has come online. This could account for a positive trend in the storage differential. A trend line added to the apparently random fluctuations of the storage differential indeed showed a slightly rising slope over the period studied. Since the storage differential from the 5-year average is already a differenced figure, and since the p-values were so close to the preferred 5% threshold, the series was still included in the estimation exercise.

Determining lag length

One of the features of the VECM is the fact that there are coefficients for the *lagged* price changes in both Henry Hub natural gas and WTI crude oil. In fact, the price data were differenced in order to filter out the effects of past prices on the present prices. This implies that lagged prices do help determine price levels in the present. But how many lags should be included?

In order to determine the appropriate number of lagged effects to include in the VECM a vector autoregression (VAR) model must first be fit using the prices and exogenous variable series in levels and then a series of selection order criteria tests must be conducted. The VAR model (with the exogenous variables included) is as follows:

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

Here, the log price of Henry Hub natural gas is determined by the previous 1 to *n* weeks' prices of WTI crude oil in logged dollars per barrel ($P_{WTI,t-i}$), with each week's effect denoted by the corresponding coefficient b_i; by the previous 1 to *n* weeks' prices of Henry Hub natural gas in logged dollars per mmBtu ($P_{HH,t-i}$), with each week's effect denoted by the corresponding coefficient c_i; and by the contemporaneous set of six exogenous variables (HDD, CDD, HDDDEV, CDDDEV, SHUTIN, and STORDIFF). The effect of each of the exogenous variables is denoted by the coefficient d_j. ε_t corresponds to a random error term with an expected value of zero.

The point of running the VAR selection order criteria test is to determine the number of lags, *n*, of previous price effects at which the modeled gas price exhibits the smallest deviation from the actual Henry Hub natural gas prices. The VAR selection order criteria test was run with a maximum of 12 lags on the entire series of logged Henry Hub natural gas and WTI crude oil prices and included the exogenous variables in the VAR.[13] Since the data are of weekly frequency, using up to 12 lags in the model allows for the historical oil and gas prices as far back as 12 weeks to influence the contemporaneous price of natural gas. The convention allows for the effects of approximately one season's duration to feed into the determination of the actual week's natural gas price. The selection order criteria tests are the Likelihood Ratio (LR) test,

[13] The command for the selection order criteria test is "varsoc" in Stata 8 Intercooled.

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 test represents a slightly different mathematical method of statistically determining the model whose predicted results match most closely with the observed data. Table 4 details the results of the VAR Selection Order Criteria tests. The Likelihood Ratio Test, the Final Prediction Error, and the Akaike Information Criteria tests all showed the closest fit at ten lags. The Hannan-Quinn Information Criteria test selected two lags, while the Schwartz Bayesian Information Criteria selected just one lag. Ten lags were chosen for use in the model. A ten-lag VAR is equivalent to a nine-lag VECM.

Table 4 - Selection Order Criteria Tests (fitting a Vector Autoregression)
Dates of analysis (weekly): 6/13/97 - 2/20/09

No. of Lags	Log Likelihood	Likelihood Ratio	p-value	Final Prediction Error (FPE)	Akaike's Information Criterion (AIC)	Hannan-Quinn Information Criterion (HQIC)	Schwatrz Bayesian Information Criterion (SBIC)
0	-379.52			0.013	1.31611	1.35616	1.41897
1	1435.07	3629.200	0.000	3.00E-05	-4.73935	-4.68786	-4.60710 *
2	1442.72	15.317	0.004	3.00E-05	-4.75159	-4.68866 *	-4.58995
3	1448.04	10.633	0.031	2.90E-05	-4.75599	-4.68162	-4.56497
4	1452.90	9.722	0.045	2.90E-05	-4.75887	-4.67306	-4.53846
5	1453.96	2.111	0.715	3.00E-05	-4.74902	-4.65177	-4.49922
6	1457.71	7.505	0.111	3.00E-05	-4.74820	-4.63950	-4.46900
7	1460.33	5.232	0.264	3.00E-05	-4.74357	-4.62342	-4.43499
8	1466.19	11.732	0.019	3.00E-05	-4.74981	-4.61822	-4.41184
9	1470.53	8.683	0.070	3.00E-05	-4.75095	-4.60792	-4.38359
10	1477.22	13.364 *	0.010	2.90E-05 *	-4.75992 *	-4.60545	-4.36317
11	1480.10	5.771	0.217	2.90E-05	-4.75619	-4.59028	-4.33006
12	1481.42	2.647	0.619	3.00E-05	-4.74724	-4.56989	-4.29172

Testing for cointegration

The VECM operates under the assumption that the two price series in the long-run relationship equation are cointegrated. Cointegration is a condition in which two time series datasets are non-stationary, but a linear combination of the two is stationary. The Johansen test can identify the presence of cointegration.

Cointegration between the oil and gas price series, and the number of cointegrating relationships, were checked for using the Johansen test.[14] The results of the Johansen test using the ten lags suggested by the selection order criteria test are detailed in Table 5. Running the Johansen test using nine lags provided similar results and allowed identical conclusions to be drawn.

Table 5 - Johansen Tests for Cointegration

Logged Prices - including exogenous variables, 6/13/97 to 2/20/09

Max. Rank (h_0=p)	Log Likelihood	Eigenvalue	Trace Statistic	Max Statistic	SBIC	HQIC	AIC
0	1467.331		31.621	29.811	-4.358	-4.582	-4.724
1	1482.236	0.048	1.809**	1.809	-4.376***	-4.613***	-4.764
2	1483.141	0.003			-4.368	-4.610	-4.764
* = significant at the 5% level, ** = significant at the 1% level, *** = "best fit" according to various criteria							

The Johansen test indicated a rank of one (at a significance of 1%) based on the trace statistics, and both the SBIC and HQIC likewise implied a rank of one. The results provide strong evidence of a single (rank=1) cointegrating relationship between Henry Hub natural gas and WTI crude oil prices if exogenous variables acting on the Henry Hub price are included. This falls into line with the predictions based on theory.

Estimating the VECM

Finally, the VECM algorithm was run. The VECM model treats the Henry Hub price series and, alternately, the WTI price series as the dependent variable. Note that the VECM does not simply assume that Henry Hub prices are determined by WTI prices. It mechanically returns results separately as if each price series were dependent on the other. As such, the model also performs the regressions as if crude oil prices were the dependent variable and natural gas prices were the

[14] The command for the Johansen test is "vecrank" in Stata 8 Intercooled.

independent variable. That allows the effects of the exogenous variables to be measured on WTI crude oil. It also allows for an assessment of the possibility that it is actually oil prices that are dependent on natural gas prices. An analysis of the statistics that measure the degree to which the model's predicted results match the actual dataset will help to determine whether WTI crude oil prices influence Henry Hub natural gas prices or vice versa. This discussion focuses on the hypothesized relationship, in which gas prices are dependent on oil prices.

4.7. Results of the VECM

The statistics resulting from the VECM regression of Henry Hub natural gas prices on the WTI crude oil price, the six exogenous variables, and nine lagged price changes in each of the two commodities is detailed in Table 6, along with a simultaneous regression of the WTI prices on Henry Hub and the same set of variables. Each of the components is discussed in greater detail below.

4.7.1. The cointegrating equation

For the single cointegrating relationship, the β coefficient is 0.7342, and the γ coefficient is -1.0493. The β coefficient is highly statistically significant, with a p-value of 0.000. The equation for the long-term relationship is as follows:

$$log(P_{HH,t}) = -1.0493 + 0.7342 log(P_{WTI,t})$$

Over the 1997 through 2009 period in question, the weekly WTI crude oil price ranged from \$10.83 per barrel to \$145.29 per barrel. The predicted Henry Hub price based on the cointegrating relationship above would thus range from \$2.01 per mmBtu at the low WTI price to $13.55 per mmBtu at the high end. In actuality, Henry Hub prices ranged from a low of $1.03 to a high of $15.02 per mmBtu.

Table 6 - VECM Model including Exogenous Variables (6/13/97-2/20/09)

Equation	Parms	RMSE	R-sq	chi2	P>chi2
D_lnhh	26	0.0916	0.1411	94.2729	0.0000
D-lnwti	26	0.0576	0.1143	74.0499	0.0000

Long-Term Variables:	Values	P-Values	
β	0.7342	0.0000	**
γ	-1.0493		

Henry Hub Effects (D_lnhh):

Short-Term Variables:	Values	P-Values	
constant (*a*)	-0.0031	0.832	
cointegrating term (t-1) (α)	-0.0828	0.000	**
$\Delta P_{HH(t-1)}$	-0.0559	0.195	
$\Delta P_{HH(t-2)}$	-0.0710	0.096	+
$\Delta P_{HH(t-3)}$	-0.0931	0.026	*
$\Delta P_{HH(t-4)}$	-0.0590	0.156	
$\Delta P_{HH(t-5)}$	-0.0740	0.074	+
$\Delta P_{HH(t-6)}$	0.0267	0.518	
$\Delta P_{HH(t-7)}$	-0.0631	0.124	
$\Delta P_{HH(t-8)}$	0.0687	0.093	+
$\Delta P_{HH(t-9)}$	-0.1012	0.014	*
$\Delta P_{WTI(t-1)}$	0.0376	0.574	
$\Delta P_{WTI(t-2)}$	0.0370	0.583	
$\Delta P_{WTI(t-3)}$	0.0562	0.407	
$\Delta P_{WTI(t-4)}$	-0.0185	0.783	
$\Delta P_{WTI(t-5)}$	-0.0101	0.882	
$\Delta P_{WTI(t-6)}$	0.0876	0.197	
$\Delta P_{WTI(t-7)}$	0.0983	0.149	
$\Delta P_{WTI(t-8)}$	0.0519	0.460	
$\Delta P_{WTI(t-9)}$	0.1077	0.121	
$HDD_{(t)}$	1.14E-04	0.191	
$HDDDEV_{(t)}$	1.02E-03	0.000	**
$CDD_{(t)}$	-4.23E-04	0.116	
$CDDDEV_{(t)}$	3.43E-03	0.000	**
STORAGE $DIFF_{(t)}$	-2.97E-05	0.076	+
SHUT $IN_{(t)}$	4.49E-06	0.263	

+ = 0.1, * = 0.05, ** = 0.01 significance levels

Joint Significance:

Variable	Chi^2 Stat	P-Value	
Lagged HH	24.13	0.0041	**
Lagged WTI	7.04	0.6333	
Lagged HH & WTI	30.28	0.0348	*
Exogenous Vars.	51.15	0.0000	**
Exog + HH Lag	71.71	0.0000	**
Exog + WTI Lag	54.61	0.0000	**
Lagged + Exogs	75.86	0.0000	**

WTI Effects (D_lnwti):

Short-Term Variables:	Values	P-Values	
constant (*a*)	-0.0081	0.380	
cointegrating term (t-1) (α)	0.0318	0.011	*
$\Delta P_{HH(t-1)}$	0.0733	0.007	**
$\Delta P_{HH(t-2)}$	0.0164	0.540	
$\Delta P_{HH(t-3)}$	-0.0191	0.466	
$\Delta P_{HH(t-4)}$	-0.0001	0.997	
$\Delta P_{HH(t-5)}$	0.0101	0.698	
$\Delta P_{HH(t-6)}$	0.0308	0.235	
$\Delta P_{HH(t-7)}$	0.0163	0.526	
$\Delta P_{HH(t-8)}$	0.0006	0.982	
$\Delta P_{HH(t-9)}$	-0.0283	0.273	
$\Delta P_{WTI(t-1)}$	-0.1314	0.002	**
$\Delta P_{WTI(t-2)}$	-0.1024	0.016	*
$\Delta P_{WTI(t-3)}$	0.0891	0.036	*
$\Delta P_{WTI(t-4)}$	-0.0195	0.645	
$\Delta P_{WTI(t-5)}$	0.0357	0.401	
$\Delta P_{WTI(t-6)}$	-0.0625	0.143	
$\Delta P_{WTI(t-7)}$	-0.0818	0.056	+
$\Delta P_{WTI(t-8)}$	0.1035	0.019	*
$\Delta P_{WTI(t-9)}$	0.1126	0.010	**
$HDD_{(t)}$	4.45E-05	0.419	
$HDDDEV_{(t)}$	-1.66E-04	0.195	
$CDD_{(t)}$	1.97E-04	0.244	
$CDDDEV_{(t)}$	3.70E-04	0.418	
STORAGE $DIFF_{(t)}$	2.44E-05	0.021	*
SHUT $IN_{(t)}$	-7.23E-06	0.004	**

+ = 0.1, * = 0.05, ** = 0.01 significance levels

Joint Significance:

Variable	Chi^2 Stat	P-Value	
Lagged HH	11.28	0.2567	
Lagged WTI	44.28	0.0000	**
Lagged HH & WTI	50.72	0.0001	**
Exogenous Vars.	17.61	0.0073	**
Exog + HH Lag	30.37	0.0107	*
Exog + WTI Lag	62.34	0.0000	**
Lagged + Exogs	70.24	0.0000	**

The R^2 of the equation for *changes* in the Henry Hub price (the second equation in the VECM) is 0.1411, while the R^2 statistic of the equation for changes in the WTI price (the third equation in the VECM) is 0.1143. This implies that in the case of Henry Hub prices, about 14.1% of the volatility in Henry Hub prices can be described through the volatility of the WTI price and the values of the six exogenous variables. In the case of the WTI price-change equation, about 11.4% of the volatility in WTI prices can be explained by volatility in the Henry Hub price and the values of the six exogenous variables. In line with the original hypothesis, the model is better at explaining Henry Hub natural gas price movements than it is at explaining WTI crude oil price movements.

4.7.2. The error-correction term

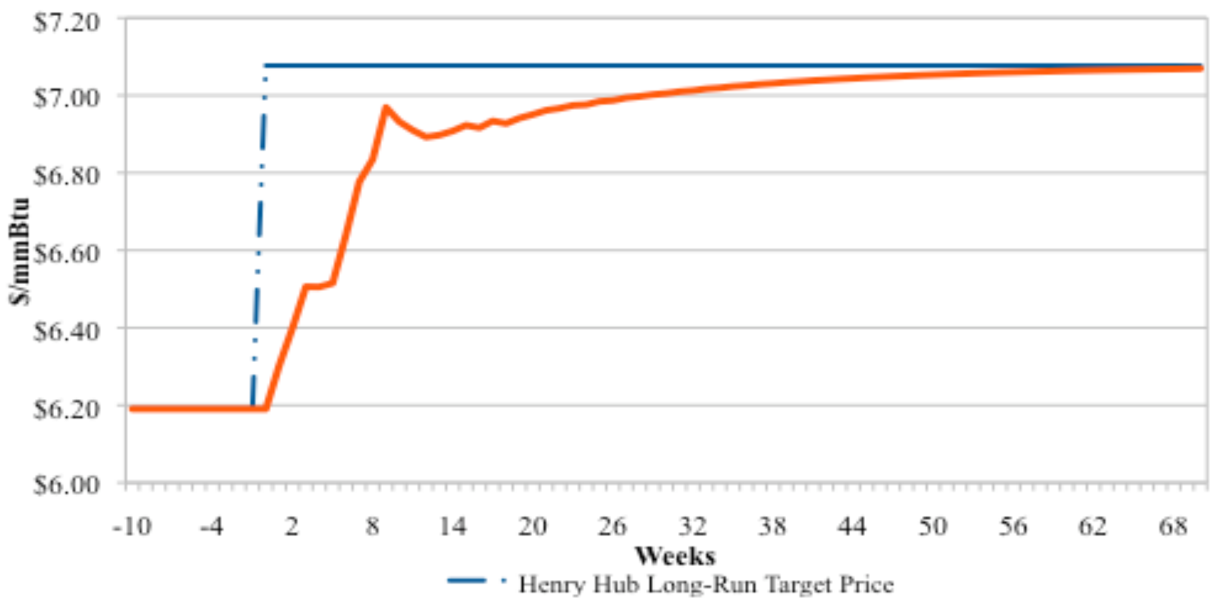
Another major element of the VECM is the error-correction term. The effect on the change in Henry Hub logged prices in the price-change equation due to the error-correction term is denoted by α. The modeled α-coefficient for reversion to the long-run predicted price relationship is -0.0828, with a p-value of 0.0000. Based on this single coefficient, the implication is that if all else held equal, any spike in Henry Hub prices from the long-run relationship would be "corrected" (be diminished) by half of the original error value in about 8.4 weeks. This figure represents the *half-life* of the error correction term.[15] However, the model also includes lagged price effects for Henry Hub and WTI prices, each of which affect the Henry Hub price, and thus the rate at which the Henry Hub price can correct back to the predicted long-run relationship.

[15] The half-life is calculated using the equation: $|\ln(2)/\alpha|$, where α is the error-correction coefficient.

When lagged effects are included, one must consider specific scenarios in which the Henry Hub price can diverge from the long-run equilibrium.

Consider, for example, a case in which the WTI crude oil price rises from $50 to $60/Bbl and then holds steady at the $60/Bbl price. This represents a 20% rise in price for crude oil. In this scenario, all of the other exogenous variables that can perturb the Henry Hub gas price are held constant. The effect of the changes in the Henry Hub price due to the price changes of the previous nine weeks in Henry Hub and WTI prices accelerate the error correction process.

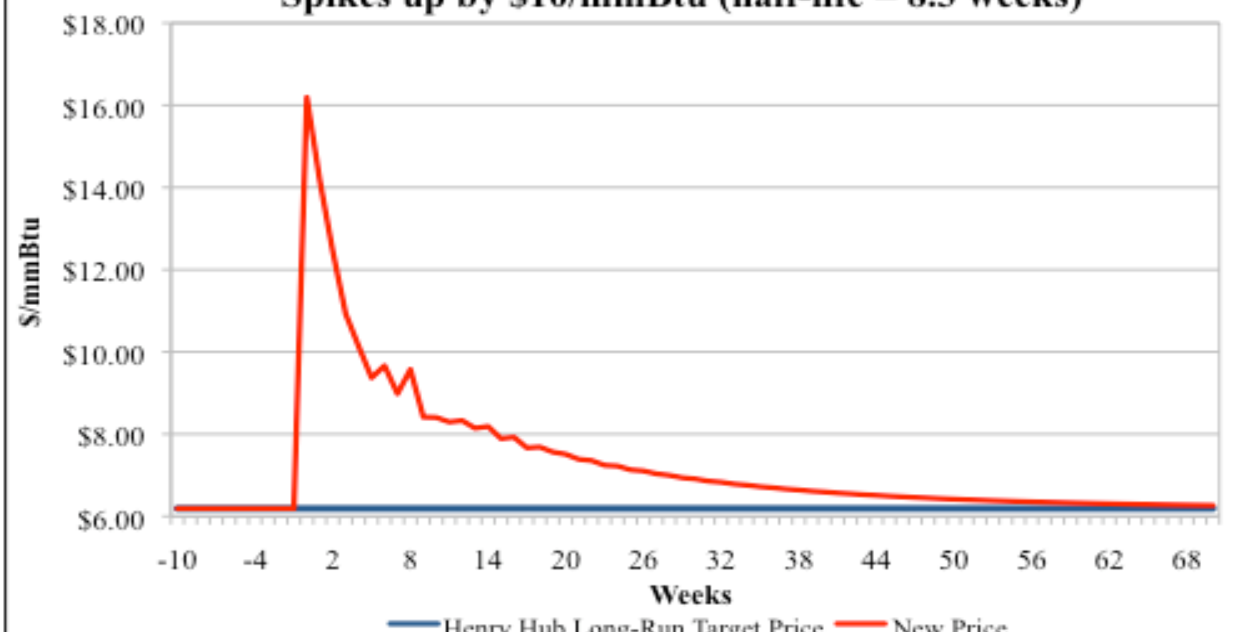
Figure 25 - Behavior of Henry Hub Price when WTI Crude Oil Rises Permanently from $50 to $60/barrel (half-life = 5.7 weeks)



However, since the signs for each lag are not identical, the half-life calculation can vary depending on the time window being examined. In this case, the half-life is 5.7 weeks if one counts 13 weeks (¼ of a year) from the first week after the price shift occurred. Figure 25 illustrates this process. It also shows how the half-life can vary depending upon which segment of the curve is the focus. The curve is not as smooth as predicted by a standard half-life calculation because the effects of the lagged price changes on price movements sometimes act against the movement of the error-correction term. The standard half-life calculation depends on only a single variable (the error-correction term), while including the nine lags in both Henry Hub and WTI prices make for a half-life calculation based on 19 distinct coefficients. The Henry Hub price path is the jagged red curve in the figure, while the broken blue line represents the new equilibrium price for natural gas based on the cointegrating equation.

Likewise, spikes in the Henry Hub price when the WTI price remains constant have a half-life of 8.4 weeks based solely on the α-coefficient. But when the Henry Hub price spikes by $10/mmBtu from its long-run equilibrium and WTI crude oil is constant at $50 per barrel, the added effects of the lagged changes in the Henry Hub price series implies a half-life of 8.3 weeks. Here again, one is counting from the first week after the spike and extending 13 weeks,

Figure 26 - Behavior of Henry Hub Price After Gas Price Spikes up by $10/mmBtu (half-life = 8.3 weeks)



or ¼ of a year. This is illustrated in Figure 26. The horizontal blue line reflects the fact that the WTI price has not changed. Thus, according to the cointegrating equation, the long-run natural gas price should not change at all. The spiking red line represents the Henry Hub natural gas price in this scenario. In short, in each of these cases, the return to the long-run equilibrium is accelerated by the inclusion of the lagged price change effects of both WTI and Henry Hub commodity prices on Henry Hub prices.

4.7.3 The statistical significance of the exogenous and lagged price change variables

The significance of the effects of the exogenous variables representing the fundamentals of the natural gas price varies widely. Only the variables accounting for unseasonal temperature events – HDDDEV and CDDDEV – show p-values below the 1% threshold. Of the rest of the variables, only the STORDIFF coefficient had a p-value within 10%, which is not considered individually significant. However, the joint statistical significance of the set of six exogenous variables was high, with a Chi^2 statistic of 51.15. That corresponds to a joint p-value of 0.0000. In the end all six exogenous variables were included in the Henry Hub price change equation.

The effects of the lagged price changes of Henry Hub and WTI on the contemporary price change at Henry Hub were also retained in the model. Individually, only the Henry Hub price change from three weeks prior and from nine weeks prior were statistically significant within the 5% p-value range, and the second, fifth, and eighth lags were significant at the 10% level. However, jointly the nine lagged price changes have a p-value of 0.0041, which is well within the 1% significance range.

None of the lagged WTI price changes were statistically significant, either alone or jointly. The joint significance test of the WTI lagged price changes returned an insignificant p-value of 0.6333, which implies that the probability that the combined effects of the nine lagged WTI prices are actually zero are around 63%. However, when included with the lagged Henry Hub price changes and the six exogenous variables (the entire set of all variables in the regression) the Chi^2 statistic is 75.86, corresponding to a statistically robust p-value of 0.0000. Thus, all of the coefficients in the modeling exercise were included for Henry Hub price changes.

4.7.4 The economic significance of the exogenous and lagged price change variables

Table 6 on page 84 presents the coefficients of each of the variables on the change in logged Henry Hub natural gas prices. The decision to use natural logs was justified above. But determining the economic significance of these variables is difficult when the results are presented in natural logs. The VECM provides coefficients on the effects of HDD, CDD, HDDDEV, CDDDEV, STORDIFF and SHUTIN as well as nine weeks of lagged changes in the logged prices of Henry Hub and WTI to determine how the logged price at Henry Hub changes, but what does that mean in terms of \$/mmBtu? Table 7 provides an example of how a one-unit increase in the relevant variable will change the price of Henry Hub natural gas from a \$7/mmBtu starting point. A detail of note is that the effect of a one-unit increase in any of the six exogenous variables is much, much smaller than the effect of a one-unit increase in either lagged Henry Hub or lagged WTI crude oil prices when prices are in natural logs.

To correct for this mismatch in units, the effects of a *one-standard deviation* increase in each of the variables is detailed in the last column of Table 7. The general magnitude of effect of a one-standard deviation increase on the price of Henry Hub natural gas is similar across all but two of the variables, ranging from about 3 to 6 cents/mmBtu. The two big outliers are HDDDEV and CDDDEV, which can each shift the price of Henry Hub natural gas up by 16 cents/mmBtu when they increase by a single standard deviation and the Henry Hub price starts at $7/mmBtu.

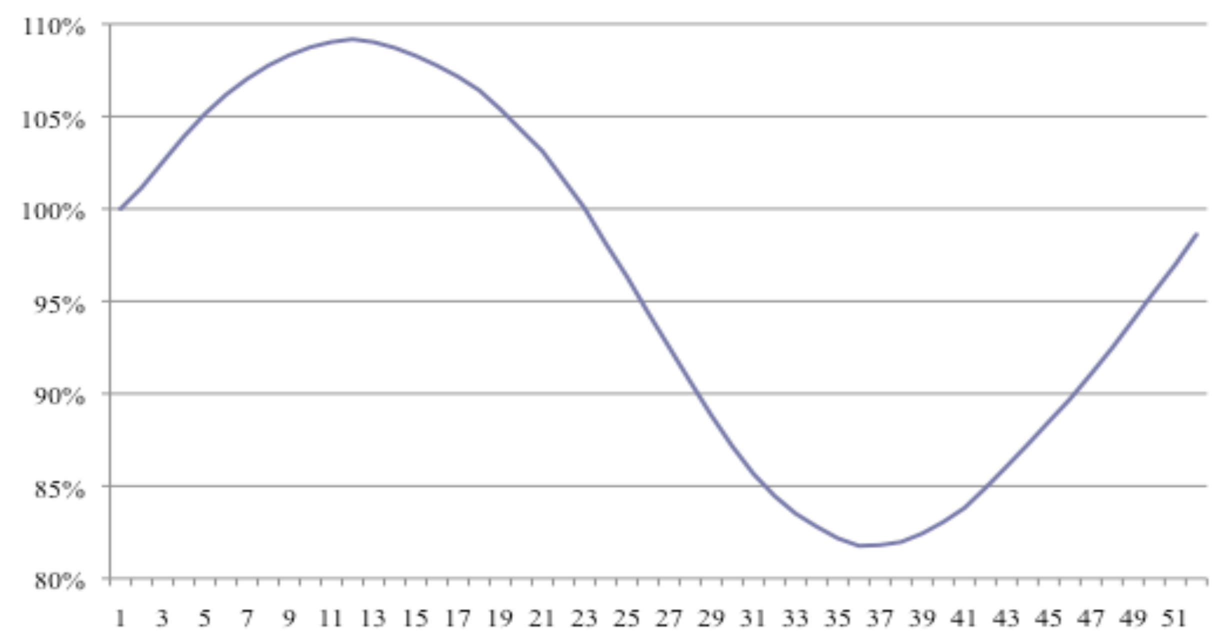
Table 7 - VECM Effects of Variables on Price of Henry Hub in $/mmBtu

Variable	Change in HH from $7/mmBtu per One-Unit Increase in Variable	p-value of exogenous coefficient	% probability that variable's coefficient is really zero	Maximum Value in Data Set	Minimum Value in Data Set	Standard Deviation of Values in Data Set	Change in HH from $7/mmBtu per Standard Deviation Increase in Variable
HDD	$0.000800	0.191	19.10%	272	0	79.83	$0.06
HDDDEV	$0.007169	0.000	0.00%	59	-87	21.52	$0.16
CDD	-$0.002962	0.116	11.60%	100	0	28.23	-$0.08
CDDDEV	$0.024027	0.000	0.00%	29	-19	6.48	$0.16
STORDIFF	-$0.000208	0.076	7.60%	1251	-838	278.61	-$0.06
SHUTIN	$0.000031	0.263	26.30%	7941	0	1036.88	$0.03
$\Delta P_{HH,t-1}$	-$0.380349	0.195	19.50%	0.480	-0.569	0.096	-$0.04
$\Delta P_{HH,t-2}$	-$0.479979	0.096	9.60%	0.480	-0.569	0.096	-$0.05
$\Delta P_{HH,t-3}$	-$0.622234	0.026	2.60%	0.480	-0.569	0.096	-$0.06
$\Delta P_{HH,t-4}$	-$0.401293	0.156	15.60%	0.480	-0.569	0.096	-$0.04
$\Delta P_{HH,t-5}$	-$0.499145	0.074	7.40%	0.480	-0.569	0.096	-$0.05
$\Delta P_{HH,t-6}$	$0.189468	0.518	51.80%	0.480	-0.569	0.096	$0.02
$\Delta P_{HH,t-7}$	-$0.428020	0.124	12.40%	0.480	-0.569	0.096	-$0.04
$\Delta P_{HH,t-8}$	$0.497794	0.093	9.30%	0.480	-0.569	0.096	$0.05
$\Delta P_{HH,t-9}$	-$0.673426	0.014	1.40%	0.480	-0.569	0.096	-$0.07
$\Delta P_{WTI,t-1}$	$0.268482	0.574	57.40%	0.359	-0.312	0.060	$0.02
$\Delta P_{WTI,t-2}$	$0.263806	0.583	58.30%	0.359	-0.312	0.060	$0.02
$\Delta P_{WTI,t-3}$	$0.404783	0.407	40.70%	0.359	-0.312	0.060	$0.02
$\Delta P_{WTI,t-4}$	-$0.128501	0.783	78.30%	0.359	-0.312	0.060	-$0.01
$\Delta P_{WTI,t-5}$	-$0.070292	0.882	88.20%	0.359	-0.312	0.060	-$0.00
$\Delta P_{WTI,t-6}$	$0.641104	0.197	19.70%	0.359	-0.312	0.060	$0.04
$\Delta P_{WTI,t-7}$	$0.723169	0.149	14.90%	0.359	-0.312	0.060	$0.04
$\Delta P_{WTI,t-8}$	$0.372864	0.460	46.00%	0.359	-0.312	0.060	$0.02
$\Delta P_{WTI,t-9}$	$0.796316	0.121	12.10%	0.359	-0.312	0.060	$0.05

4.7.5 The implicit seasonality in HDD and CDD data

Despite their weak economic effect and dubious individual statistical significance, HDD and CDD exhibit a predictable seasonal pattern which, when aggregate effects are taken into account, provides a basis for measuring the seasonal price shifts in Henry Hub natural gas prices. To illustrate how these variables account for gas price seasonality, the HDD and CDD values were averaged over the dataset by week. What results is a single year of weekly observations, in which each observation is the average value in that week for HDD or CDD over the 1997-2009 period. The next step required the assumption of a stable WTI price, and HDDDEV, CDDDEV, STORDIFF, and SHUTIN values of zero. The VECM coefficients for HDD and CDD were then multiplied by the set of average HDD and CDD coefficients over the course of each year. The simulation was extended out for 90 years in order to settle the perturbations caused by the lagged price changes in Henry Hub prices. The resulting price series in predicted logged Henry Hub prices was examined. To normalize, the logged price for week 1 was subtracted from each of the 52 weeks in the annual dataset. The result was raised as a power of the natural exponent. This provided normalized coefficients, in which week 1 takes on a value of 1 (100%), and each subsequent week takes a value corresponding to that week's relationship to week one. The process is equivalent to first taking the natural exponent of the logged price series and then dividing each week's value by the value of week 1.

Figure 27 - Weekly Seasonality from HDD+CDD coefficient Effects on change in lnHH



The results are plotted in Figure 27. The seasonality implied by the HDD and CDD values when logged Henry Hub natural gas spot prices are regressed on them returns a roughly sinusoidal curve peaking at about 9% above January values in March and falling to a low in September that is roughly 82% of January's price.

4.7.6. The VECM with WTI as the dependent variable

Thus far the focus has been on the second equation in the VECM: the effects of the lagged Henry Hub and WTI price changes and the six exogenous variables on the change in Henry Hub prices. However, the model also provides an examination of the effects of those same variables on the change in prices for WTI crude oil in the third equation. The working hypothesis has been that the WTI crude oil price series is exogenous to the system. The VECM conveniently allows this hypothesis to be tested statistically.

The error-correction coefficient (α) in the WTI case has a p-value of 0.011, which is statistically significant at the 5% level. However, the coefficient is a *positive* 0.0318. This would imply that when the WTI price deviates from the price predicted by the cointegrating relationship, the model would "correct" the error by making it larger. This is clearly a nonsensical result that leads one to doubt the validity of the model in determining the price movements of WTI crude oil.

The hypothesis likewise supposed that the effect of the exogenous variables on WTI prices would be zero, but that was not universally the case. The storage differential was significant within the 5% range with a p-value of 0.021, while the shut-in production statistic was highly statistically significant, with a p-value of 0.004. However, economic theory predicts that when storage levels go up, prices should go down due to supply gluts, and when shut-in production figures rise, prices should follow due to the implied supply shortage. The sign in both cases here is counterintuitive. The VECM for changes in the WTI price predicts that the price would rise as storage rises, and that prices would fall as more production is taken offline in the Gulf. While there is no reason that crude oil storage levels should be related to natural gas storage levels, shut-in oil production and shut-in gas production should be strongly correlated in the Gulf of Mexico since both are responses to the same hurricanes. Thus, these results lead to the conclusion that the predicted relationships for the only two exogenous variables with any statistical significance are spurious.

Of the lagged Henry Hub price change effects on the change in WTI prices, only the one-week lagged Henry Hub price change was statistically significant, with a p-value of 0.007. However, jointly the nine lagged Henry Hub price changes only had a p-value of 0.2567 when tested for joint significance. This is not robust enough for inclusion in the model. In fact, only the lagged WTI price changes had any significant effect on the contemporaneous price of WTI crude oil. That result, however, is to be expected.

The coefficients in the VECM model when applied to changes in the WTI crude oil price were counterintuitive and spurious. This suggests that the model does not hold any ability to estimate the price movements in WTI crude oil based on the values of the exogenous and lagged natural gas price variables. In other words, one can conclude from the data in the VECM that the variables being regressed have no predictive ability on the price of WTI crude oil. WTI crude oil prices in this model can thus be considered at least weakly exogenous to the system. That conclusion will allow for consideration of a modified version of the error correction model called the Conditional Error Correction Model (Conditional ECM).

4.8. Improving the VECM – the Conditional ECM

The insights provided by examining the VECM allow for an improved model to be created using the strongest elements of the VECM and a simpler estimation system. This improvement is called the Conditional ECM. A similar model was used by Villar and Joutz in their 2006 paper.

The Conditional ECM is based on the inference from the VECM that the change in the logged WTI price imparts an immediate effect on the logged price of Henry Hub natural gas. This is really an assumption that there are market factors that affect both commodities contemporaneously such that knowing the WTI price change can allow one to infer what the change in price of Henry Hub gas ought to be. Furthermore, the Conditional ECM uses the cointegrating errors from the VECM as an exogenous variable. The cointegrating errors are the actual Henry Hub logged prices minus the long-run logged Henry Hub prices predicted by the VECM for each week. Finally, the six exogenous variables utilized in the original VECM are retained due to their high level of joint significance on the change in Henry Hub prices.

The functional form for the Conditional ECM is a vector autoregression (VAR). The error-correction term is provided by taking the long-run predicted cointegrating relationship estimated in the VECM as a given. Thus, the cointegrating equation is not directly estimated in the Conditional ECM. The estimated equation follows:

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

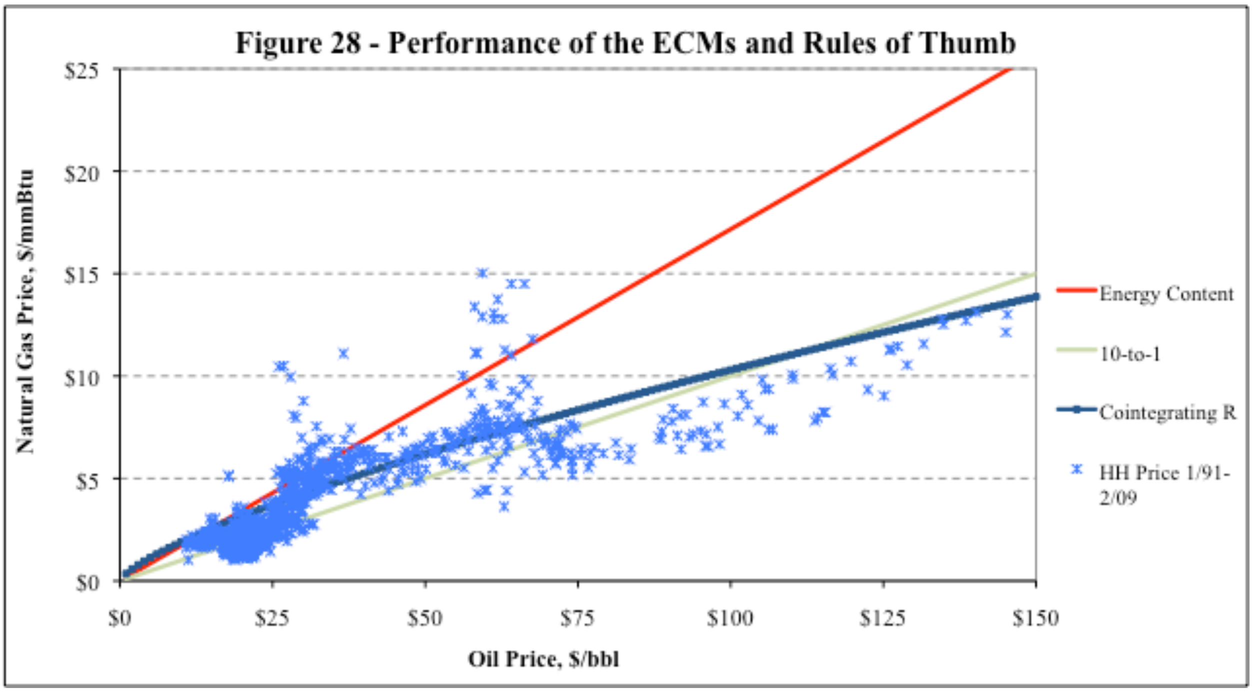
In the Conditional ECM, a is a constant, α is the error-correction coefficient on the lagged errors in the cointegrating relationship from the original VECM (μ_{t-1}), and b is the coefficient on the *contemporaneous* change in the logged WTI crude oil price. The c_i series of coefficients represent the effects of the lagged changes in the logged Henry Hub price. Since 9 lags were used in the VECM, all were included in the Conditional ECM, so $n = 9$. d_j represent the coefficients for each of the exogenous variables used in the VECM (here denoted as X_j to represent HDD, CDD, HDDDEV, CDDDEV, STORDIFF, and SHUTIN). ε_t is, as before, a standard white noise error term with an expected value of zero.

4.8.1. The cointegrating equation remains the same

As mentioned above, the cointegrating relationship between Henry Hub natural gas and WTI crude oil as estimated in the VECM was taken as a given for the Conditional ECM. For ease of reference, the relationship is reproduced here:

$$log(P_{HH,t}) = -1.0493 + 0.7342log(P_{WTI,t})$$

Figure 28 depicts the cointegrating relationship between oil and gas prices as a comparison to Figure 18. To ease interpretation, only the Energy Content and 10-to-1 rules of thumb remain on the graph, as does the scatterplot of actual Henry Hub natural gas prices on the gas price-oil price plane. The cointegrating relationship is the curve that bows up into the thick of the price points in the $25-$60/Bbl oil price range and then curves down as the oil price rises, falling below even the 10-to-1 plot when oil prices reach about $120/Bbl.



The superiority of a regression in logs is obvious. Instead of a straight line, the VECM and Conditional ECM's long-run relationship can curve to best fit the noisy price data over the 1997-2009 period. It avoids a wild divergence from the actual prices, as exhibited in the plot of the Energy Content rule, which predicts prices for natural gas over $25/mmBtu when oil prices near $150 per barrel.

4.8.2. The results of the Conditional ECM

Table 8 reports the results of the Conditional ECM. The R^2 statistic has increased slightly to 0.1606, meaning that over 16% of the volatility in the Henry Hub natural gas price can be described through the fluctuations of the exogenous variables, the change in the WTI price, and the effects of the nine lagged weekly price changes in Henry Hub natural gas. The Chi^2 statistic has increased to 114.8 from its previous value in the VECM of 94.3.

Table 8 - Conditional ECM including Exogenous Variables (6/13/97-2/20/09)

Equation	Parms	RMSE	R-sq	chi2	P>chi2
dlnhh	18	0.0899	0.1606	114.8139	0.0000

Long-Term Variables:	Values	P-Values	
β	0.7342	0.0000	**
γ	-1.0493		

Henry Hub Effects (dlnhh)

Short-Term Variables:	Values	P-Values	
constant (a)	0.0015	0.9160	
cointegrating term (t-1) (α)	-0.0952	0.0000	**
$\Delta P_{HH(t-1)}$	-0.0610	0.1350	
$\Delta P_{HH(t-2)}$	-0.0566	0.1550	
$\Delta P_{HH(t-3)}$	-0.0723	0.0640	+
$\Delta P_{HH(t-4)}$	-0.0536	0.1690	
$\Delta P_{HH(t-5)}$	-0.0747	0.0540	+
$\Delta P_{HH(t-6)}$	0.0334	0.3890	
$\Delta P_{HH(t-7)}$	-0.0419	0.2750	
$\Delta P_{HH(t-8)}$	0.0864	0.0240	*
$\Delta P_{HH(t-9)}$	-0.0816	0.0340	*
ΔP_{WTI}	0.2872	0.0000	**
$HDD_{(t)}$	7.660E-05	0.3560	
$HDDDEV_{(t)}$	1.051E-03	0.0000	**
$CDD_{(t)}$	-4.704E-04	0.0690	+
$CDDDEV_{(t)}$	3.263E-03	0.0000	**
$STORAGE\ DIFF_{(t)}$	-3.370E-05	0.0370	*
$SHUT\ IN_{(t)}$	5.120E-06	0.1820	

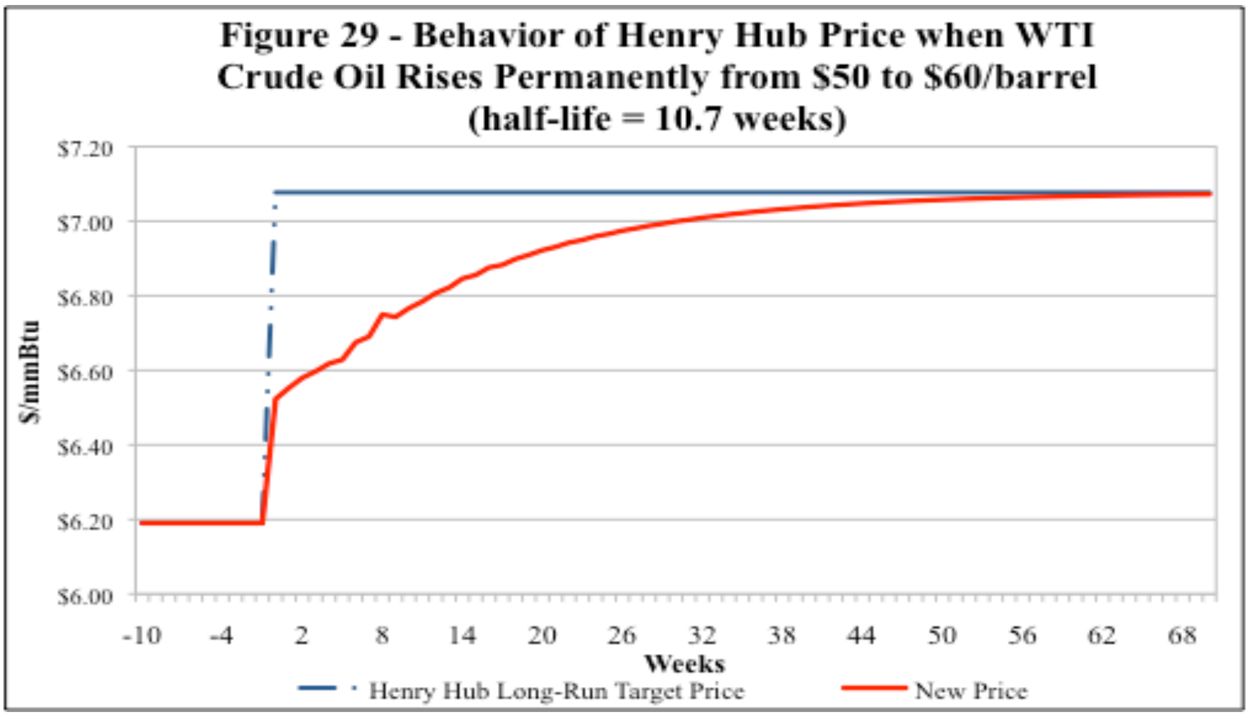
Joint Significance:

Variable	Chi^2 Stat	P-Value	
Lagged HH	24.58	0.0035	**
Exogenous Vars.	54.39	0.0000	**
Exog + WTI	72.44	0.0000	**
Lagged + Exogs	76.49	0.0000	**
Lag + Exog + WTI	95.16	0.0000	**

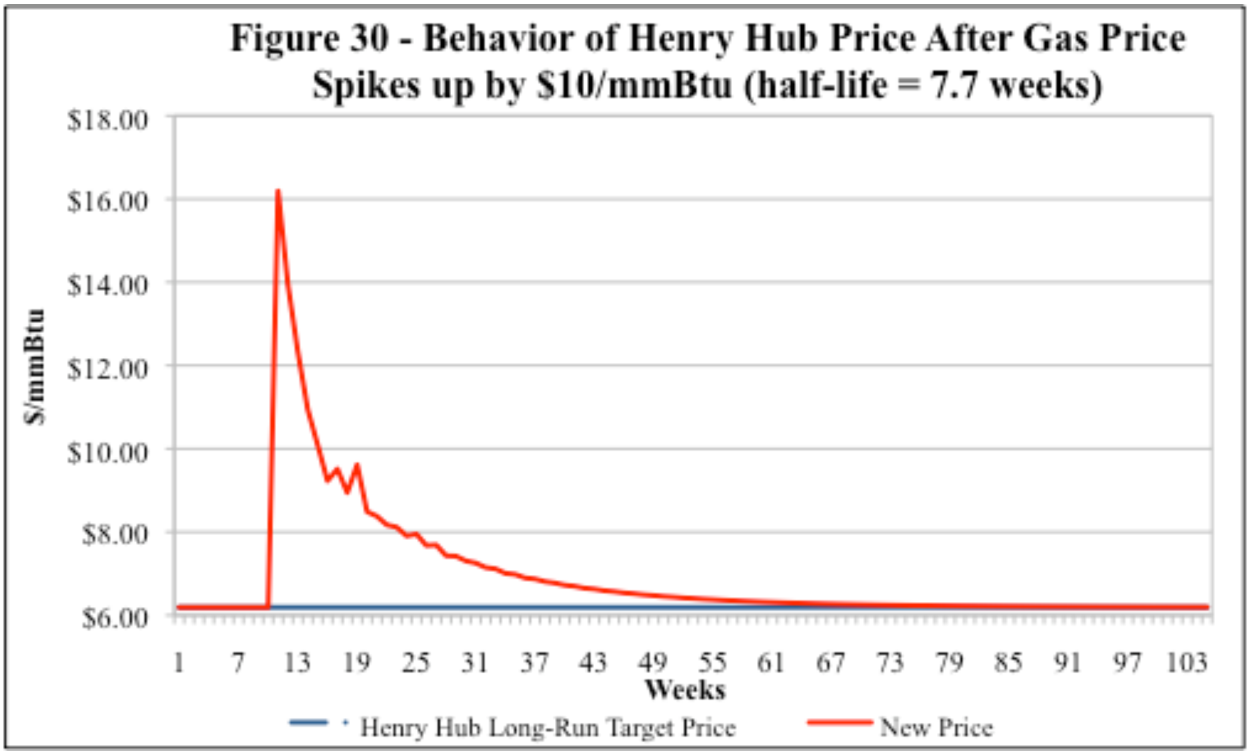
+ = 0.1, * = 0.05, ** = 0.01 significance levels

4.8.3. The new error-correction term

The α coefficient for the error correction has intensified to -0.0952 and has remained statistically significant, with a p-value of 0.0000. The figure reflects a half-life for the error correction mechanism to eliminate differences from the long-run predicted relationship of 7.3 weeks if the effects of the lagged Henry Hub price changes are ignored. Figures 29 and 30 illustrate the behavior of the Henry Hub price under the two price shock scenarios discussed for the VECM.



In the first scenario, the WTI crude oil price rises from $50 to $60 per barrel and all other factors are held constant. Under these circumstances, the half-life measured from 13 weeks after the start of the disturbance is 10.7 weeks due to the perturbations of the change in oil price and the lagged changes in the gas price. In the second scenario, the Henry Hub price spikes by $10/mmBtu, and Figure 30 traces its return to long-run equilibrium. Here, the half-life measured 13 weeks after the shock is 7.7 weeks.



4.8.4. The statistical significance of the variables

The statistical significance of the exogenous variables has increased for each of the variables that were significant in the VECM, with the STORDIFF p-value of 0.037 bringing it into the 5% significance range and the CDD p-value improving to within the 10% significance range at 0.069. Furthermore, the signs for each coefficient are what one would expect from a change in any of the variables: increases in HDD, HDDDEV, CDDDEV, and SHUTIN provoke increases in the Henry Hub price, while increases in CDD and STORDIFF provoke decreases in the Henry Hub price. The price effect of a change in the WTI price is both strong, at 0.2872, and statistically significant at 0.0000.

Furthermore, the Chi^2 statistics for each of the joint variable significance tests has improved considerably in every case. In the Conditional ECM, the joint lagged Henry Hub price changes, the joint exogenous variables, and the combinations of lagged, WTI, and exogenous variables in joint significance tests all return p-values implying statistical significance at the 1% level or better.

4.8.5. The economic significance of the variables

Table 9 illustrates the economic effects of the exogenous variables in the Conditional ECM on the change in price at Henry Hub in \$/mmBtu. Essentially, the effects of HDD, CDD, HDDDEV, CDDDEV, STORDIFF and SHUTIN, as well as the effects of the lagged price changes at Henry Hub, are only slightly different than in the VECM.

Table 9 - Cond. ECM Effects of Variables on Price of Henry Hub in \$/mmBtu

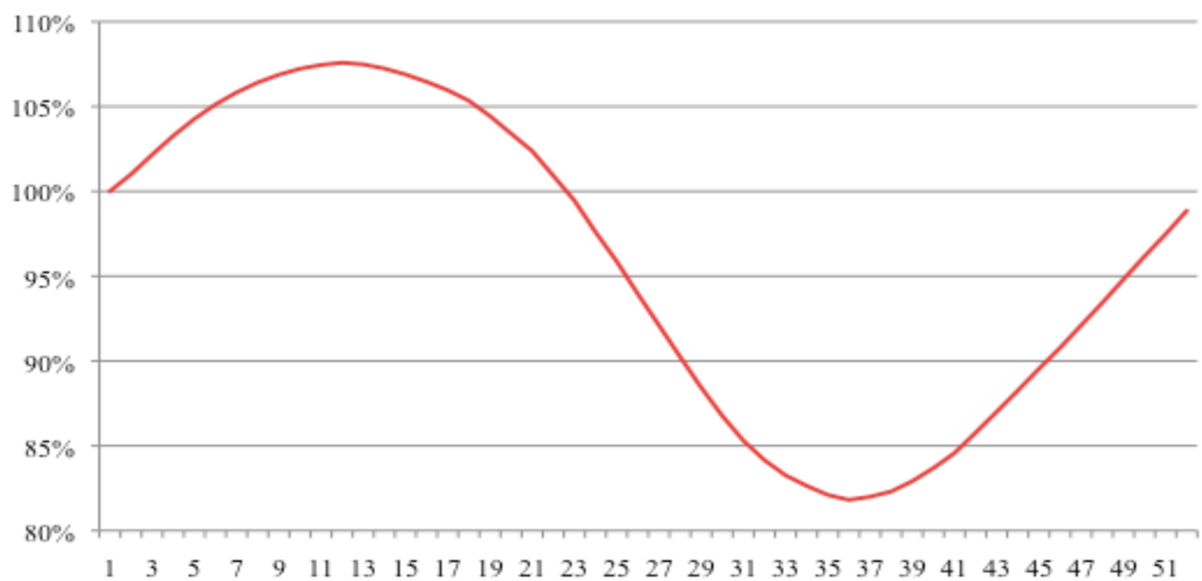
Variable	Change in HH from \$7/mmBtu per One-Unit Increase in Variable	p-value of Exogenous Coefficient	% Probability that Variable's Coefficient is Actually Zero	Maximum Value in Data Set	Minimum Value in Data Set	Standard Deviation of Values in Data Set	Change in HH from \$7/mmBtu per Standard Deviation Increase in Variable
HDD	\$0.000536	0.356	35.60%	272	0	79.83	\$0.04
HDDDEV	\$0.007358	0.000	0.00%	59	-87	21.52	\$0.16
CDD	-\$0.003292	0.069	6.90%	100	0	28.23	-\$0.09
CDDDEV	\$0.022881	0.000	0.00%	29	-19	6.48	\$0.15
STORDIFF	-\$0.000236	0.037	3.70%	1251	-838	278.61	-\$0.07
SHUTIN	\$0.000036	0.182	18.20%	7941	0	1036.88	\$0.04
$\Delta P_{HH,t-1}$	-\$0.414223	0.135	13.50%	0.480	-0.569	0.096	-\$0.04
$\Delta P_{HH,t-2}$	-\$0.384941	0.155	15.50%	0.480	-0.569	0.096	-\$0.04
$\Delta P_{HH,t-3}$	-\$0.488012	0.064	6.40%	0.480	-0.569	0.096	-\$0.05
$\Delta P_{HH,t-4}$	-\$0.365563	0.169	16.90%	0.480	-0.569	0.096	-\$0.04
$\Delta P_{HH,t-5}$	-\$0.503645	0.054	5.40%	0.480	-0.569	0.096	-\$0.05
$\Delta P_{HH,t-6}$	\$0.237745	0.389	38.90%	0.480	-0.569	0.096	\$0.02
$\Delta P_{HH,t-7}$	-\$0.287392	0.275	27.50%	0.480	-0.569	0.096	-\$0.03
$\Delta P_{HH,t-8}$	\$0.631679	0.024	2.40%	0.480	-0.569	0.096	\$0.06
$\Delta P_{HH,t-9}$	-\$0.548612	0.034	3.40%	0.480	-0.569	0.096	-\$0.05
$\Delta P_{WTI,t}$	\$2.328795	0.000	0.00%	0.359	-0.312	0.060	\$0.12

However, the coefficient on the contemporaneous change in WTI price is much stronger than that of lagged WTI price changes as presented in the VECM. Now, a one-standard deviation increase in WTI prices when Henry Hub prices are at \$7/mmBtu would provoke a 12-cent/mmBtu increase in the Henry Hub price. Clearly, the change in WTI prices is now economically on par with HDDDEV and CDDDEV in terms of its effect on the price at Henry Hub.

4.8.6. HDD and CDD implied seasonality in the Conditional ECM

To reinforce the notion that the Conditional ECM has not changed the coefficients of the exogenous variables or the effects of the lagged changes in price at Henry Hub to a significant degree, Figure 31 is presented, which re-creates the seasonality implied by the model through the HDD and CDD values.

Figure 31 - Weekly Seasonality from HDD+CDD coefficient Effects on change in lnHH in Cond. ECM



The methodology was identical to that for the VECM seasonality, negating the need to repeat it here. The resulting single-peak and single-trough seasonality curve is remarkably similar to that of Figure 27. The peak still occurs in March and the trough still occurs in September. However, the effect at the extremes is somewhat dampened, so that the price at the peak is only about 7.5%

higher than the price in week 1, while the price at the trough is still about 82% of the week 1 price.

4.9. Comparing the VECM and the Conditional ECM to the models in the literature

In order to compare how these models measure up to other models of price changes, Table 10 is provided (partially sourced from IEA World Energy Outlook, 2009). The table compares four models in two scenarios. The four models are the Villar-Joutz model used in their 2006 paper (Villar and Joutz, 2006), the Brown-Yücel model from their 2008 paper (Brown and Yücel, 2008), and the VECM and Conditional ECM as described in this thesis. The two scenarios are: one in which the WTI price spikes up by 20% and holds steady thereafter (identical to that explored in Figures 25 and 29), and one in which the WTI price spikes for one period up by 20% and then returns to its original value. The table shows how the price of Henry Hub natural gas should change, percentage-wise, from its original value given the price movements in WTI. The discrepancies in the predicted price changes in Henry Hub are rather small, especially considering the mismatch in the time periods being studied and the fact that Villar and Joutz use monthly data (and monthly seasonal dummy variables) while Brown and Yücel use weekly data (and HDD and CDD data as proxies for seasonal trends), which is the same approach used here.

Table 10 - Impacts of Oil Price Changes on Gas Prices

Effect of a Permanent change in the price of crude oil

Researcher	Period (months)	Period (weeks)	Change in price of WTI (%)	Change in gas price at Henry Hub (%)
Villar-Joutz (1989-2005 monthly)	0	NA	20.0	5.0
	1	NA	0.0	7.8
	2	NA	0.0	9.8
	12	NA	0.0	16.0
Brown-Yücel VECM (6/13/97-6/8/07 weekly)	0	0	20.0	0.0
	0	1	0.0	3.4
	1	4	0.0	4.5
	2	8	0.0	8.5
	12	52	0.0	15.8
Ramberg-Parsons VECM (6/13/97-2/20/09 weekly)	0	0	20.0	0.0
	0	1	0.0	1.8
	1	4	0.0	5.0
	2	8	0.0	10.0
	12	52	0.0	13.2
Ramberg-Parsons Conditional ECM (6/13/97-2/20/09 weekly)	0	0	20.0	5.4
	0	1	0.0	5.7
	1	4	0.0	6.5
	2	8	0.0	8.4
	12	52	0.0	13.2

Effect of a Transitory change in the price of crude oil

Researcher	Period (months)	Period (weeks)	Change in price of WTI (%)	Change in gas price at Henry Hub (%)
Villar-Joutz (1989-2005 monthly)	0	NA	20.0	5.0
	1	NA	-16.7	2.8
	2	NA	0.0	2.1
	12	NA	0.0	0.6
Brown-Yücel VECM (6/13/97-6/8/07 weekly)	0	0	20.0	0.0
	0	1	-16.7	3.4
	1	4	0.0	-1.2
	2	8	0.0	1.2
	12	52	0.0	0.2
Ramberg-Parsons VECM (6/13/97-2/20/09 weekly)	0	0	20.0	0.0
	0	1	-16.7	1.8
	1	4	0.0	0.0
	2	8	0.0	0.9
	12	52	0.0	0.1
Ramberg-Parsons Conditional ECM (6/13/97-2/20/09 weekly)	0	0	20.0	5.4
	0	1	-16.7	0.6
	1	4	0.0	0.5
	2	8	0.0	1.1
	12	52	0.0	0.3

4.10. Using the Conditional ECM: modeling the effects of exogenous variables

So far, the modeling exercise has only served to provide a guide to how certain factors affect the gas price and its behavior regarding its long-run relationship with the oil price. But embedded in the model's results are useful tools for traders, marketers, and policymakers both within and outside of industry.

Through the Conditional ECM one can examine the predicted effects of the four exogenous variables relating to weather and supply shocks: HDDDEV, CDDDEV, STORDIFF and SHUTIN. How the model predicts the behavior of Henry Hub prices under typical shocks experienced over the 1997-2009 period are detailed below. As with the examination of the error-correction mechanism, when analyzing the effect of a single variable in the shock, the effect of the shock is measured by examining the difference in the behavior of Henry Hub prices under their normal seasonal pattern (as defined by HDD and CDD) and their behavior when affected by the shock. All other variables are held constant. In order to smooth out the lagged price effects so that the seasonal pattern becomes consistent, in each case the shock was implemented in the 90th year of the simulation.

4.10.1. HDDDEV and CDDDEV – temperature shocks

To model the HDDDEV shocks, the effect of a typical cold snap in early March on Henry Hub prices was simulated. Over the 1997-2009 period, most cold snaps lasted two weeks. Furthermore, the first week was usually more severe than the second. Figure 32 depicts the effect in which the first week of the cold snap exhibits a HDDDEV of +20 and the second week has a value of +15. These are of typical magnitude over the 12-year period under study. The interesting point here is that while the temperature effect lasts only two weeks, the half-life of the price disturbance is about 7 weeks, with measurable differences from the price series without the shock lasting over 38 weeks. At its peak, the effect on the change in Henry Hub price for a single week was 23 cents/mmBtu. The effect of a negative HDDDEV shock was identical but of the opposite sign, and was not displayed here.

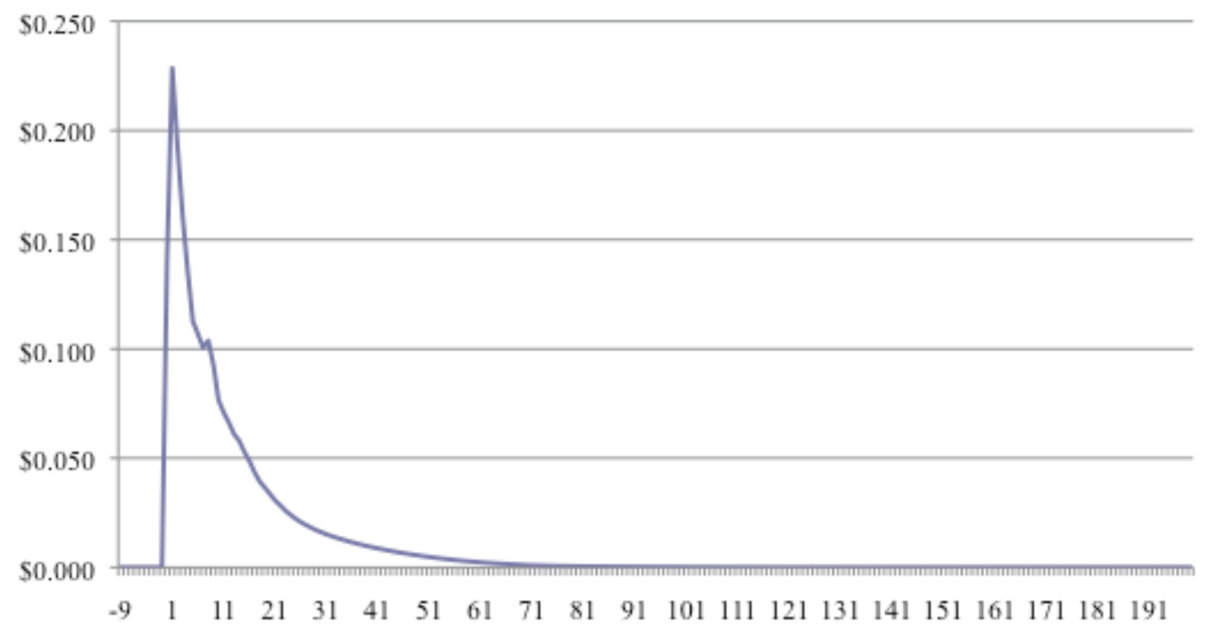
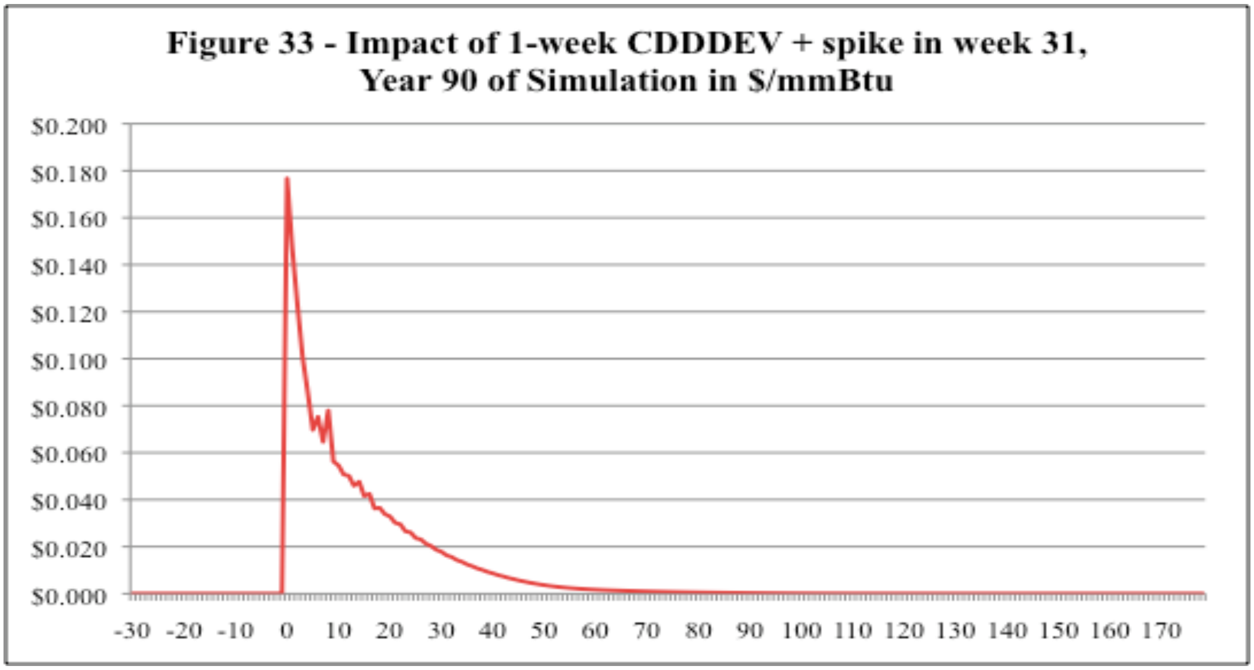


Figure 32 - Impact of 2 week HDDDEV + spike in week 10, Year 90 of Simulation in $/mmBtu

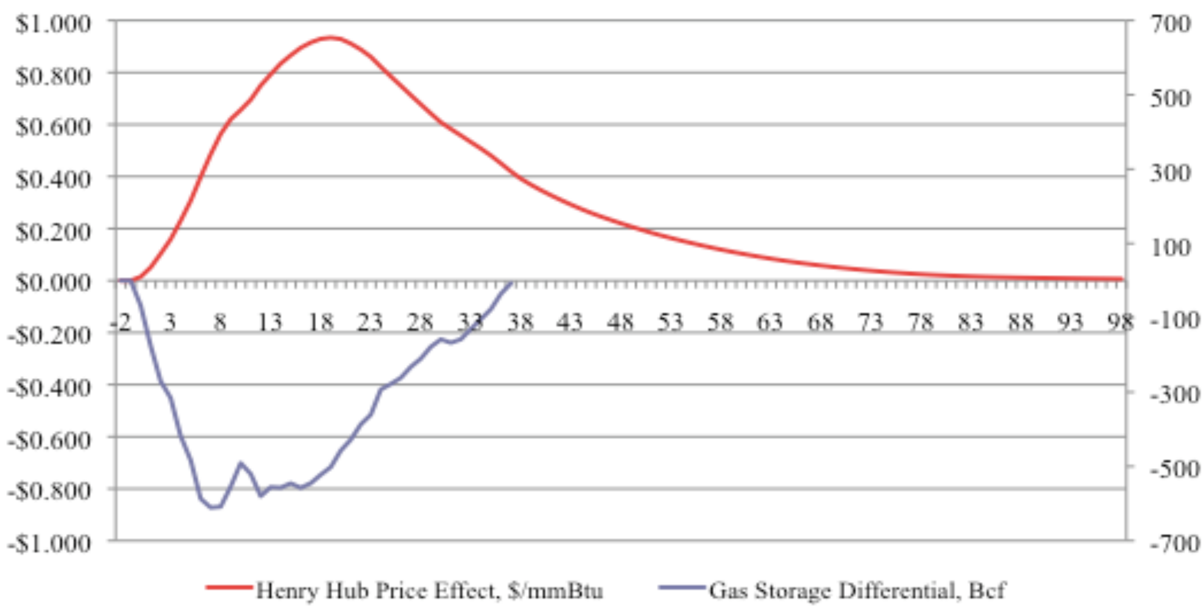
The effect of a typical CDDDEV shock was also modeled. The exercise simulated a heat wave occurring in late July on the price of Henry Hub natural gas. The timing of the shock was the summer, and the magnitude and duration of the shock represented typical CDDDEV shocks (a +10 degree deviation lasting only a single week). However, the half-life was unchanged at 7 weeks and the measurable effect continued for 38 weeks. At its peak, the effect on the change in Henry Hub gas prices was 18 cents/mmBtu. Figure 33 depicts the effect of a typical heat wave on Henry Hub prices. A negative CDDDEV shock that was also modeled but not produced in a graphic provided a mirror image price plot.



4.10.2. STORDIFF – abnormally large or small storage levels

Storage differentials tended to occur for longer periods than cold snaps or heat waves. Figure 34 illustrates the modeled effect of an actual 38-week supply shortfall that occurred in January of 2003 on Henry Hub natural gas prices. At its extreme, the supply shortfall was 610 Bcf below normal storage levels. This is approximately 25% of average storage levels, so the shortfall was not insignificant. The average shortfall over the 38-week episode was 350 Bcf. Figure 34 plots both the modeled storage differential (on the right-hand y-axis) and its price effect on Henry Hub natural gas (on the left y-axis). In this case, the half-life was 22 weeks and the maximum single-week price effect on Henry Hub natural gas was an increase of 93 cents/mmBtu. Even 91 weeks after the shortfall began, there was still a persistent 1-cent/mmBtu effect on the Henry Hub price.

Figure 34 - Impact of 38 week supply shortfall in Year 90 of Simulation in $/mmBtu

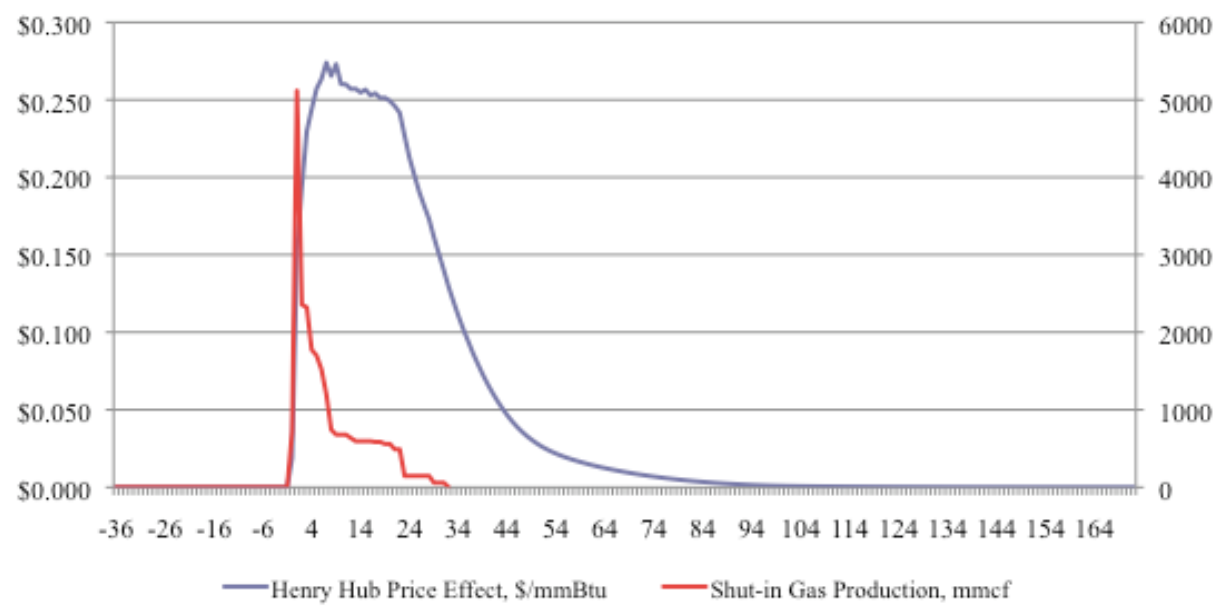


4.10.3. SHUTIN – the effect of hurricanes

The last simulation modeled the effects of a typical hurricane on Henry Hub natural gas prices. In the dataset, there is only a single observation – that of Hurricane Ivan in 2004. It was the only hurricane event in which a hurricane impacted Gulf gas production and was not followed by a second hurricane before the supply disruption could be resolved. At its peak, Hurricane Ivan shut-in over 5 Bcf of Gulf natural gas production, representing over half of the gas production capacity in the Gulf at the time. The average shut-in production volume over the 33-week supply disruption was about 812 million cubic feet.

Figure 35 illustrates the impact of Hurricane Ivan on Henry Hub natural gas prices and on Gulf of Mexico natural gas production. The right-hand y-axis measures the amount of natural gas that was shut in, while the left-hand y-axis measures the weekly price effect on Henry Hub natural gas. The biggest single-week impact on Henry Hub prices was about 27 cents/mmbtu. The half-life of the disturbance due to the hurricane was 34 weeks. The figure, like Figure 34, shows that the effect of the disturbance continues to affect prices long after the physical curtailment has been resolved. In this case, the effect of the hurricane on Henry Hub natural gas prices was still 1 cent/mmBtu 68 weeks after impact.

Figure 35 - Impact of Hurricane Ivan in Year 90 of Simulation in $/mmBtu



4.10.4. Implications for practitioners and policymakers

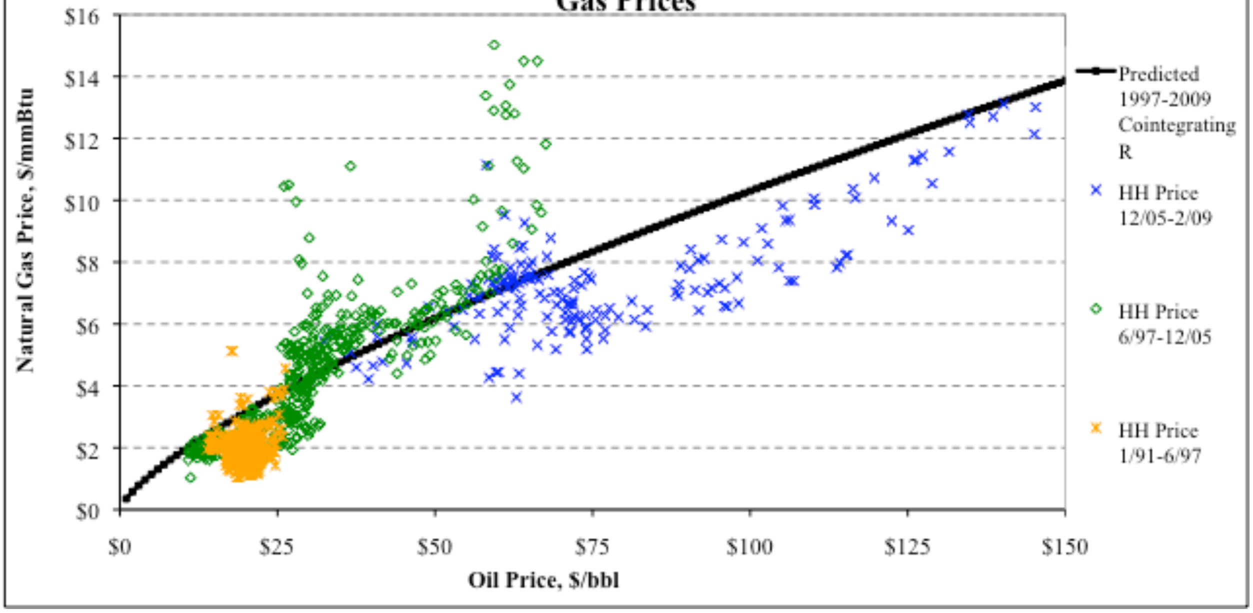
The simulations of shocks to the natural gas price described above show how the model can be used to predict the longer-run effects of the shocks as they unfold or even before they occur. This alone should provide utilities or other load-serving entities (LSEs) an idea of the price reaction they might expect from such occurrences. It would also provide the government with a benchmark for determining the damage caused by the shock by analyzing the difference in price patterns with the shock and without it. However, these predictions assume a known oil price that partially determines the gas price. In the modeling exercises here, the oil price was held constant. In the real world, this would not be the case. This is an important weakness of the simulations. Some adaptations could be used to resolve this issue, however, such as using oil futures prices as rough proxies of predicted future WTI oil prices.

Given these shortcomings, perhaps the simplest lesson to take away from the exercise is the fact that a single shock can move the Henry Hub natural gas price quite far from the long-run relationship with WTI crude oil prices, and the effect can persist for quite some time. However, from 1997-2009 at least, gas prices have always returned to some kind of stable equilibrium with oil prices. One must simply be patient and watch it unfold.

4.11. The possibility of a changing relationship

Figure 36 provides another look at the long-run cointegrating relationship estimated by the VECM and used again in the Conditional ECM. The difference here is that the actual Henry Hub gas price series has been broken into segments. The first segment marks prices from January 25, 1991 through June 6, 1997 and is plotted in orange stars. This is the segment for which no data was available on HDD or CDD. As a result, the VECM did not include these dates in its analysis. The second segment spans from June 13, 1997 to December 16, 2005 and is plotted in dark green diamonds. This period marks the beginning of the point at which weekly HDD and CDD data became available and continues until the last week at which WTI crude oil traded below $60/Bbl. Note that WTI crude oil prices were in the $60/Bbl range in the summer and fall of 2005, but fell back into the mid-$50/Bbl range until the end of 2005. The third segment spans December 23, 2005 through February 20, 2009, and each weekly price is plotted with a bright blue X. This final period encompasses the entire crude oil price run-up of 2007-2008 and its collapse. The cointegrating relationship, which is the same as in Figure 28, is plotted in black.

Figure 36 - Oil-Gas Cointegrating Relationship Against Actual Oil and Gas Prices



This segmentation exercise illustrates an important point: the axes in the graph do not allow for gas prices to be plotted in time series, but only according to the oil price. However, by segmenting the data, it is clear that the gas prices are nonetheless clustered according to date. The 1991-1997 segment is clustered around an oil price under $25/Bbl and a gas price largely between $1 and $5/mmBtu. The 1997-2005 segment spans a wider price range in both oil and gas, with oil prices ranging from $11 to $66/Bbl and gas prices spanning the $1 to $15/mmBtu range. The high gas prices are outliers reflecting the effects of Hurricanes Katrina and Rita in September of 2005. The 2005-2009 segment, as mentioned above, captures the oil price run-up, with oil prices ranging from about $37 to $145/Bbl and gas prices ranging from about $3.60 to over \$13/mmBtu. The oil prices below \$60/Bbl represented by blue X's are all from the period after WTI oil prices crashed in the summer of 2008.

Considering that the data are clustered, one is led to ask: could the cointegrating relationship simply be bending as it does in an attempt to capture the mean of the data points in the middle period, and then bend down to capture the mean of the data points in the later period? Is the relationship found in the VECM actually more than one relationship, or a shifting relationship? One way to test this is to break the estimation into separate periods and evaluate whether or not the R^2 improves (signaling a better fit with the data). This exercise is conducted in the next section. The results will provide evidence as to whether the oil-gas relationship may actually shift across equilibria over time.

4.12. The cointegrating relationship estimated by segments

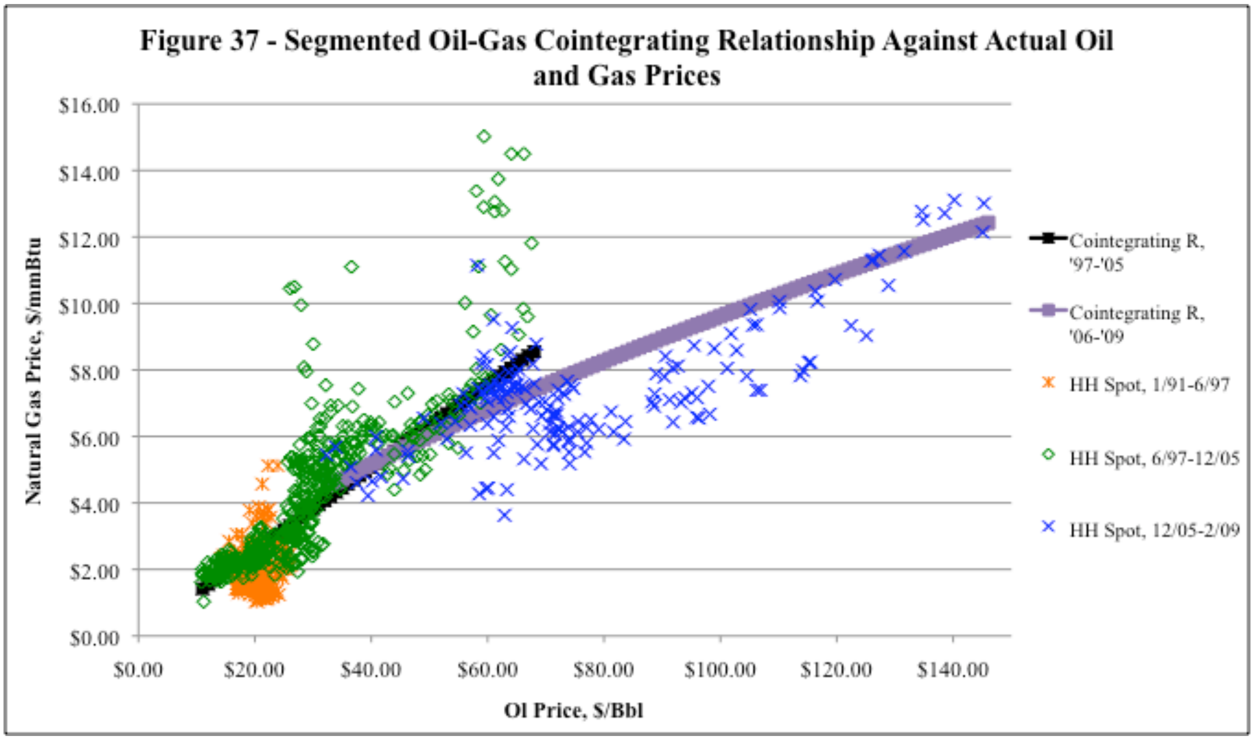
If the estimation is re-run to segment the 1997-2009 period into two blocks that correspond to the color-coded gas price periods, the cointegrating relationship changes for each period. The cointegrating relationship from June 13, 1997 through December 16, 2005 is:

$$logP_{HH,t} = -2.014 + 0.986logP_{WTI,t}$$

Meanwhile, the period covering the oil-price run-up, December 23, 2005 through February 20, 2009, is characterized by the following long-run relationship between natural gas and oil prices:

$$logP_{HH,t} = -1.049 + 0.734logP_{WTI,t}$$

The new relationships are plotted in Figure 37. The 1997-2005 VECM long-run relationship is plotted in black and is the steeper of the two segments. It is plotted between the \$11 and \$63/Bbl oil price range. The 2006-2009 regression's long-run relationship is plotted in purple and is the



longer of the two, since it spans a greater price range in oil. The oil prices to which this regression corresponds ranged from \$33 to nearly \$146/Bbl.

The two cointegrating relationships are clearly distinct from each other, and also from the 1997-2009 long-run predicted cointegrating relationship of the original VECM. But are they a better fit? The original 1997-2009 Conditional ECM has an R^2 of 0.16 and a Chi^2 statistic of 114.8. The segmented regressions produce much higher R^2 statistics, suggesting that the predicted relationship is a closer fit to the actual dataset. The 1997-2005 regression, when followed through to the Conditional ECM, has an R^2 of 0.21 and a Chi^2 statistic of 113.5. The 2006-2009 long-run relationship, using Conditional ECM statistics, has an R^2 of 0.31 and a Chi^2 statistic of 73.5.

What this means is that by segmenting the data, the model improves its ability to account for the volatility in Henry Hub natural gas prices from 16% to 21% for the pre-2006 regression. The 2006-2009 regression model accounts for 31% of gas volatility. The segments present credible evidence that there was a distinct relationship between oil and gas prices between the 1997-2005 period and the 2006-2009 period. Whether these are the only distinct periods, or whether there was a gradual shift over this time period, are not measured by this simple experiment. The fact that the Chi^2 statistics do not show the same improvement as the R^2 statistics should trouble no one. In all three cases, the Chi^2 statistic corresponds to a p-value of 0.0000, so all are highly significant from a statistical perspective. Furthermore, the segments manage to establish statistical significance with a much smaller number of observations, and cannot generally be expected to have Chi^2 statistics higher than the longer data set due to their smaller number of degrees of freedom.

It is a safe assumption that the 1997-2009 period can be interpreted as one of shifting relationships. The breakpoint set here was completely arbitrary, so whether there are exactly two stable equilibria or a gradual shift between the fuels is not discernible. A deeper study might be able to identify structural shifts in technology, policy, or demand/supply profiles that would justify exact points where the shifts are likely to occur. The fact that the segmented models fit the data better than the 1997-2009 model is a reflection of the hypothesis that the long model was attempting to split the difference between distinct or shifting equilibria. To provide readers with additional relevant information, Tables 11 and 12 provide the statistics for the VECM and

Conditional ECM for the 1997-2005 period, while Tables 13 and 14 present the same statistics for the December 23, 2005 through February 20, 2009 period.

Table 11 - VECM Model including Exogenous Variables (6/13/97-12/16/05)

Equation	Parms	RMSE	R-sq	chi2	P>chi2
D_lnhh	20	0.0912	0.1821	93.05402	0.0000
D-lnwti	20	0.0538	0.1128	53.1217	0.0001

Long-Term Variables:	Values	P-Values	
β	0.9860	0.0000	**
γ	-2.0142		

Henry Hub Effects (D_lnhh):

Short-Term Variables:	Values	P-Values	
constant (*a*)	-0.0015	0.925	
cointegrating term (t-1) (α)	-0.1375	0.000	**
$\Delta P_{HH(t-1)}$	-0.0242	0.623	
$\Delta P_{HH(t-2)}$	-0.1456	0.003	**
$\Delta P_{HH(t-3)}$	-0.0211	0.659	
$\Delta P_{HH(t-4)}$	-0.0911	0.058	+
$\Delta P_{HH(t-5)}$	0.0141	0.769	
$\Delta P_{HH(t-6)}$	-0.0089	0.852	
$\Delta P_{WTI(t-1)}$	-0.0975	0.249	
$\Delta P_{WTI(t-2)}$	-0.0041	0.962	
$\Delta P_{WTI(t-3)}$	-0.0703	0.406	
$\Delta P_{WTI(t-4)}$	-0.0610	0.473	
$\Delta P_{WTI(t-5)}$	-0.1299	0.122	
$\Delta P_{WTI(t-6)}$	0.0703	0.397	
$HDD_{(t)}$	2.49E-04	0.014	*
$HDDDEV_{(t)}$	1.41E-03	0.000	**
$CDD_{(t)}$	1.22E-05	0.968	
$CDDDEV_{(t)}$	3.03E-03	0.000	**
STORAGE $DIFF_{(t)}$	-2.83E-05	0.146	
SHUT $IN_{(t)}$	1.15E-05	0.014	*

+ = 0.1, * = 0.05, ** = 0.01 significance levels

Joint Significance:

Variable	Chi^2 Stat	P-Value	
Lagged HH	11.44	0.0757	+
Lagged WTI	4.86	0.5622	
Lagged HH & WTI	19.43	0.0787	+
Exogenous Vars.	58.91	0.0000	**
Exog + HH Lag	70.12	0.0000	**
Exog + WTI Lag	61.71	0.0000	**
Lagged + Exogs	74.66	0.0000	**

WTI Effects (D_lnwti):

Short-Term Variables:	Values	P-Values	
constant (*a*)	-0.0158	0.102	
cointegrating term (t-1) (α)	0.0135	0.385	
$\Delta P_{HH(t-1)}$	0.0576	0.048	*
$\Delta P_{HH(t-2)}$	0.0207	0.473	
$\Delta P_{HH(t-3)}$	-0.0363	0.199	
$\Delta P_{HH(t-4)}$	0.0295	0.299	
$\Delta P_{HH(t-5)}$	-0.0117	0.679	
$\Delta P_{HH(t-6)}$	0.0444	0.114	
$\Delta P_{WTI(t-1)}$	-0.0530	0.289	
$\Delta P_{WTI(t-2)}$	-0.1824	0.000	**
$\Delta P_{WTI(t-3)}$	0.0905	0.070	+
$\Delta P_{WTI(t-4)}$	-0.1201	0.017	*
$\Delta P_{WTI(t-5)}$	0.0720	0.147	
$\Delta P_{WTI(t-6)}$	-0.1329	0.007	**
$HDD_{(t)}$	8.08E-05	0.179	
$HDDDEV_{(t)}$	-6.88E-05	0.634	
$CDD_{(t)}$	3.35E-04	0.065	+
$CDDDEV_{(t)}$	4.57E-04	0.338	
STORAGE $DIFF_{(t)}$	1.69E-05	0.143	
SHUT $IN_{(t)}$	-2.22E-06	0.423	

+ = 0.1, * = 0.05, ** = 0.01 significance levels

Joint Significance:

Variable	Chi^2 Stat	P-Value	
Lagged HH	10.3	0.1125	
Lagged WTI	29.62	0.0000	**
Lagged HH & WTI	37.47	0.0002	**
Exogenous Vars.	10.56	0.1029	
Lagged + Exogs	50.43	0.0001	*

Table 12 - Conditional ECM Including Exogenous Variables (6/13/97 - 12/16/05)

Equation	Parms	RMSE	R-sq	chi2	P>chi2
dlnhh	15	0.089252	0.2058	113.504	0.0000

Long-Term Variables:	Values	P-Values	
β	0.9860	0.0000	**
γ	-2.0142		

Henry Hub Effects (dlnhh)

Short-Term Variables:	Values	P-Values	
constant (a)	0.0051	0.7480	
cointegrating term (t-1) (α)	-0.1361	0.0000	**
$\Delta P_{HH(t-1)}$	-0.0550	0.2340	
$\Delta P_{HH(t-2)}$	-0.1515	0.0010	**
$\Delta P_{HH(t-3)}$	-0.0216	0.6260	
$\Delta P_{HH(t-4)}$	-0.1082	0.0150	*
$\Delta P_{HH(t-5)}$	-0.0068	0.8770	
$\Delta P_{HH(t-6)}$	-0.0199	0.6520	
ΔPWTI	0.3367	0.0000	
$HDD_{(t)}$	2.087E-04	0.0320	
$HDDDEV_{(t)}$	1.399E-03	0.0000	
$CDD_{(t)}$	-1.421E-04	0.6290	**
$CDDDEV_{(t)}$	2.816E-03	0.0000	*
STORAGE $DIFF_{(t)}$	-3.370E-05	0.0720	**
SHUT $IN_{(t)}$	1.220E-05	0.0070	

+ = 0.1, * = 0.05, ** = 0.01 significance levels

Joint Significance:

Variable	Chi^2 Stat	P-Value	
Lagged HH	15.85	0.0146	*
Exogenous Vars.	62.80	0.0000	**
Exog + WTI	80.58	0.0000	**
Lagged + Exogs	76.56	0.0000	**
Lag + Exog + WTI	94.57	0.0000	**

Table 13 - VECM Model including Exogenous Variables (12/23/05-2/20/09)

Equation	Parms	RMSE	R-sq	chi2	P>chi2
D_lnhh	14	0.0797	0.2986	64.27627	0.0000
D-lnwti	14	0.0662	0.1586	28.4681	0.0123

Long-Term Variables:	Values	P-Values	
β	0.6782	0.0000	**
γ	-0.8598		

Henry Hub Effects (D_lnhh):

Short-Term Variables:	Values	P-Values	
constant (*a*)	0.0078	0.777	
cointegrating term (t-1) (α)	-0.1495	0.000	**
$\Delta P_{HH(t-1)}$	-0.0323	0.679	
$\Delta P_{HH(t-2)}$	0.2540	0.001	**
$\Delta P_{HH(t-3)}$	-0.1963	0.009	**
$\Delta P_{WTI(t-1)}$	0.0106	0.912	
$\Delta P_{WTI(t-2)}$	-0.0599	0.541	
$\Delta P_{WTI(t-3)}$	0.0471	0.627	
$HDD_{(t)}$	-1.75E-05	0.905	
$HDDDEV_{(t)}$	1.64E-04	0.630	
$CDD_{(t)}$	-1.06E-03	0.024	*
$CDDDEV_{(t)}$	4.34E-03	0.002	**
STORAGE $DIFF_{(t)}$	-1.03E-05	0.734	
SHUT $IN_{(t)}$	-9.72E-06	0.116	

+ = 0.1, * = 0.05, ** = 0.01 significance levels

Joint Significance:

Variable	Chi^2 Stat	P-Value	
Lagged HH	25.45	0.0000	**
Lagged WTI	0.86	0.8353	
Lagged HH & WTI	26.06	0.0002	**
Exogenous Vars.	16.24	0.0125	*
Exog + HH Lag	42.71	0.0000	**
Exog + WTI Lag	17.56	0.0406	*
Lagged + Exogs	43.93	0.0000	**

WTI Effects (D_lnwti):

Short-Term Variables:	Values	P-Values	
constant (*a*)	0.0164	0.476	
cointegrating term (t-1) (α)	0.0714	0.030	*
$\Delta P_{HH(t-1)}$	0.0602	0.353	
$\Delta P_{HH(t-2)}$	0.0547	0.377	
$\Delta P_{HH(t-3)}$	0.0165	0.792	
$\Delta P_{WTI(t-1)}$	-0.2486	0.002	**
$\Delta P_{WTI(t-2)}$	-0.0908	0.265	
$\Delta P_{WTI(t-3)}$	0.0927	0.249	
$HDD_{(t)}$	-9.96E-05	0.413	
$HDDDEV_{(t)}$	-1.89E-04	0.505	
$CDD_{(t)}$	-1.35E-04	0.729	
$CDDDEV_{(t)}$	1.99E-04	0.861	
STORAGE $DIFF_{(t)}$	3.12E-05	0.215	
SHUT $IN_{(t)}$	-1.44E-05	0.005	**

+ = 0.1, * = 0.05, ** = 0.01 significance levels

Joint Significance:

Variable	Chi^2 Stat	P-Value	
Lagged HH	1.41	0.7035	
Lagged WTI	12.43	0.0060	**
Lagged HH & WTI	12.91	0.0445	*
Exogenous Vars.	10.97	0.0892	+
Lagged + Exogs	22.49	0.0324	

Table 14 - Conditional ECM Including Exogenous Variables (12/23/05-2/20/09)

Equation	Parms	RMSE	R-sq	chi2	P>chi2
dlnhh	12	0.078379	0.3081	73.4754	0.0000

Long-Term Variables:	Values	P-Values	
β	0.6782	0.0000	**
γ	-0.8598		

Henry Hub Effects (dlnhh)

Short-Term Variables:	Values	P-Values	
constant (a)	0.0046	0.8610	
cointegrating term (t-1) (α)	-0.1642	0.0000	**
$\Delta P_{HH(t-1)}$	-0.0413	0.5730	
$\Delta P_{HH(t-2)}$	0.2458	0.0000	**
$\Delta P_{HH(t-3)}$	-0.1959	0.0050	**
ΔPWTI	0.1851	0.0380	*
$HDD_{(t)}$	8.270E-07	0.9950	
$HDDDEV_{(t)}$	1.775E-04	0.5770	
$CDD_{(t)}$	-1.038E-03	0.0200	*
$CDDDEV_{(t)}$	4.354E-03	0.0010	**
STORAGE $DIFF_{(t)}$	-1.530E-05	0.5900	
SHUT $IN_{(t)}$	-7.530E-06	0.1840	

Joint Significance:

Variable	Chi^2 Stat	P-Value	
Lagged HH	27.78	0.0000	**
Exogenous Vars.	17.80	0.0067	**
Exog + WTI	22.92	0.0018	**
Lagged + Exogs	46.57	0.0000	**
Lag + Exog + WTI	52.32	0.0000	**

+ = 0.1, * = 0.05, ** = 0.01 significance levels

5. Conclusions

This thesis has assessed the conventional wisdom that there is a stable link between oil and natural gas prices. The comparative history, structure and development of the oil and gas industries were scrutinized to identify points of overlap between the industries, including where the commodities complement each other and where they compete. The amount of current natural gas usage that could be displaced by shifting to petroleum products (and vice versa) was projected. A Vector Error Correction Model (VECM) was estimated to identify the long-run behavior of the oil and gas prices in relation to each other. A number of exogenous variables were included in an attempt to explain the volatility in natural gas prices. The model was refined through the Conditional Error Correction Model (Conditional ECM). The conclusion is that there is strong evidence of a historical relationship between the two commodities that is stable over time.

Through the use of the model, the inherent seasonality of natural gas prices was identified as a function of the heating- and cooling degree days (HDD and CDD). The model was then utilized to describe how weather and supply-related shocks to the natural gas industry, which are the cause of natural gas volatility, impact the gas market and then dissipate as gas marches back to the long-run relationship.

Finally, the question of whether the relationship could shift was tested by segmenting the 1997-2009 period into two pieces. Each piece was able to account for a greater proportion of natural gas volatility than the regression model over the entire time period. This suggests that the relationship had indeed changed over the twelve years studied.

In sum, this thesis confirms that a crude oil-natural gas price relationship existed over the 1997-2009 period. Its methodology ensured that any relationship discovered would be due to actual contemporary factors in the market of each commodity. First, spot prices instead of long-term contract prices were used. WTI and Henry Hub were chosen as the pricing points, since they are both highly liquid and also of similar geographic location (so as to avoid distortions due to transportation differentials). Furthermore, hubs and price series in the United States were chosen. This is because a spot market has been active in the U.S. for many years and the markets for each commodity have been deregulated for over 20 years.

The thesis also suggests that the price relationship can periodically shift to a new equilibrium relationship between the two commodities. It has shown that despite the fact that natural gas prices may stray from this relationship for long periods of time, it has always returned to a price range in line with a long-run cointegrating relationship with oil prices. The simulation exercise showed that a single shock can sometimes take over a year to completely dissipate – a period in which other shocks are constantly bombarding the natural gas market. This is also why the relationship does not hold to a very high correlation over the years.

5.1. Weaknesses in the current approach

The approach used here resolved many of the criticisms of earlier models that neglected to account for any factor other than the oil price as a determinant of Henry Hub prices. Nonetheless, there are still a number of issues that warrant further examination. There are some systemic issues that may be relevant to address.

Data on the exogenous variables are not available at the time prices are set

The prices in the market are set by traders acting on the information available to them at the time that they make their trade, which would incorporate forward-looking forecasts for the exogenous variables acting on Henry Hub prices. This model uses data that was not available until after the fact. These differences may be minor: by using weekly data at least some of the information could be assumed to have been available to traders at the time the natural gas price was set in the market. However, using this model with week-ahead forecasts of HDDDEV and CDDDEV and forecasts of HDD and CDD may prove to be more realistic. The difficulty is that the forecast data are not collected in databases provided by the NOAA (as is the actual data), so it would be necessary to cull the weekly press releases over the entire period to be studied in order to re-create the temperature forecast databases. Another possible alternative that avoids this difficulty would be to use the previous week's climate data to determine the current week's prices, since that data would have been available to traders at the time they were executing trades.

The model does not explain why a linkage exists

The model illustrates a likely link between crude oil and natural gas prices, but has a very poor fit when actually attempting to predict Henry Hub natural gas prices given the WTI crude oil price. The model has identified a statistically significant relationship between the fuels, but does not explain how oil price movements affect gas price movements directly. Other authors have claimed to identify the physical mechanisms by which the prices of these two fuel classes interact, but apart from Hartley's group, who included average heat rates of gas-fired units in their model, none have incorporated this explanation into their modeling exercise. Perhaps if a quantifiable mechanism could be identified and measured, a better picture of how the prices should interact could be developed. Perhaps a framework could be developed that could, for example, model a trigger point that shifts the two price series from one stable relationship to another.

Correlation is not causality

Despite the fact that the model seems to capture price movements between WTI crude oil and Henry Hub natural gas, no mechanism by which the price transfers from WTI to Henry Hub is identified. It is just as plausible that there is some variable affecting both commodities that simply affects crude oil first – such as economic output (GDP growth). Since crude oil is the more integrated, more global and more liquid market, it makes sense that it is the commodity that would react first to such stimuli. Much more is to be done before the relationship between crude oil and natural gas can be definitively characterized.

5.2. The current relationship

As noted earlier, there are those who believe that the oil-gas price relationship has broken since the crash of oil prices in the latter half of 2008. These people speculate that the new, stronger commodity price relationship is actually between coal and natural gas. While this thesis only briefly touched on a possible coal-natural gas linkage, that question remains unanswered here. There is not enough post-crash price data on the oil and gas markets to reliably establish whether the prices are still cointegrated or not. A possible further avenue of research could both test whether an oil-gas linkage still exists and also whether a coal-gas linkage is stronger. Perhaps there is a cointegrating relationship between all three commodities.

What should be kept in mind are two important points. (1) Natural gas prices can stray from the oil-gas cointegrating relationship for considerable periods of time. (2) There is evidence that the oil-gas cointegrating relationship can shift to a new equilibrium, so when oil and gas prices do not seem to share a prior relationship, it does not necessarily mean that they have become de-linked. It could simply be the case that a new relationship has developed. This relationship would require some time to become measurable using the techniques of this thesis, since longer data sets are required to establish statistical significance.

5.3. Future study

As mentioned in Section 4.12, the segmentation exercise used arbitrary date divisions. Further research could help identify points where structural changes in the relevant markets would be expected to provoke a realignment of the oil-gas price relationship. Such changes might come in the form of technological advances, shifts in policy, or changes in the demand or supply profiles of either fuel. Gathering a relevant dataset to test any of these effects is not a trivial pursuit, but could be quite helpful in determining *why* the oil-gas price relationship might change.

One way to see if participants in the market itself actually assume that a price relationship between oil and natural gas prices exists is to conduct similar analyses using futures prices. Futures prices are also appealing because they add another dimension to the data. Spot prices provide a single price for a single delivery date on a single trade date. Futures markets, on the other hand, provide a series of prices for a chain of future deliveries for each single trade date. This allows researchers to filter out the noise in the spot markets and examine whether the market pricing patterns reflect an oil-gas (or, for that matter, coal-gas) price linkage. This work has already begun and is expected to be ongoing.

As mentioned above, another interesting aspect worth investigating is whether a coal-natural gas pricing relationship is actually a more illuminating and useful linkage to study. Though work has not begun in this vein, the data are available for such work to proceed.

6. Bibliography

Asche, Frank, Peter Osmundsen and Maria Sandsmark (2006), "The UK Market for Natural Gas, Oil and Electricity: Are the Prices Decoupled?" *The Energy Journal* 27(2): 27-40.

Bachmeier, Lance J. and James M. Griffin (2006), "Testing for Market Integration: Crude Oil, Coal and Natural Gas," *The Energy Journal* 27(2): 55-71.

Barron, Michael J. and Brown, Stephen P. A. (1986), "Assessing the Market for Natural Gas," *Texas A&M Business Forum*, Fall 1986.

Blas, Javier and Hoyos, Carola (2009), "Oil-gas price link to weaken," *Financial Times*, No. 37151, p. 28, June 11, 2009.

Bloomberg Data Terminal, WTI and Henry Hub Weekly Price Series, Canadian gas prices (AECO and Empress), Sloan Business School, Massachusetts Institute of Technology.

Borenstein, Severin (2002), "The Trouble with Electricity Markets: Understanding California's Restructuring Disaster", *Journal of Economic Perspectives*, 16(1): 191-211.

Borenstein, Severin, James Bushnell and Frank Wolak (2002), "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market," *American Economic Review*, 92(5): 1376-1405.

Breyer, Stephen and Paul W. MacAvoy (1973), "The Natural Gas Shortage and the Regulation of Natural Gas Producers," *Harvard Law Review*, 86(6): 941-987.

Brown, Stephen P.A. and Yücel, Mine K. (2008), "What Drives Natural Gas Prices?" *The Energy Journal*, IAEE, 29(2): 45-60.

Brown, Stephen P.A. and Yücel, Mine K. (2007), "What Drives Natural Gas Prices?" Federal Reserve Bank of Dallas, Research Dept, Working Paper 0703, Feb. 2007, 19 pp.

Canada, National Energy Board of Canada (2007), *Canada's Energy Future – Reference Case and Scenarios to 2030, and Appendices*, November 2007, Calgary, Alberta, Canada.

Canadian Association of Petroleum Producers (CAPP) (2009), *Statistical Handbook for Canada's Upstream Petroleum Industry*, March 2009, Calgary, Alberta, Canada.

Costello, Ken, H.G.Huntington, and J.F. Wilson (2005), "After the Natural Gas Bubble: An Economic Evaluation of the Recent U.S. National Petroleum Council Study." *The Energy Journal*, 26(2): 9-110

Davis, Lucas W. and Lutz Kilian (2008), "The Allocative Cost of Price Ceilings: Lessons from the U.S. Residential Market for Natural Gas," working paper.

Dincerler, Cantekin, Khokler, Zeigham, and Simin, Timothy (2004), "The Convenience Yield and Risk Premia of Storage," University of Western Ontario Working Paper, Nov 2004, 30 pp., Available at SSRN: http://ssrn.com/abstract=687154

Enders, W., (2004), *Applied Econometric Time Series*, Second Edition. John Wiley & Sons: United States.

Engle, R.F., and Granger, C.W.J., (1987), "Cointegration and error correction: Representation, estimation, and testing", Econometrica, 55: 251–276. Accessed through jstor 4/25/2009, http://www.jstor.org/stable/1913236.

Engle, R.F., Kozicki, S., (1993), "Testing for common features," *Journal of Business and Economic Statistics*, 11: 369– 380.

Finon, Dominique (2008), "Why Would Oil-Indexation In Gas Contracts Survive In Europe?" EU Energy Policy Blog, http://www.energypolicyblog.com/?p=171

Hartley, Peter, Kenneth Medlock and Jennifer Rosthal (2008), "The Relationship of Natural Gas to Oil Prices," *The Energy Journal*, 29(3): 47-66.

Huntington, Hillard G. (2007), "Industrial Natural Gas Consumption in the United States: An Empirical Model for Evaluating Future Trends," *Energy Economics* 29(4): 743-759.

Jensen, J.T. (2003), "The LNG Revolution," *The Energy Journal*, 24(3): 1-45.

Joskow, Paul (2001), "California's Energy Crisis." *Oxford Review of Economic Policy*, 17: 365-388.

Joskow, Paul and E. Kahn (2002), "A Quantitative Analysis of Pricing Behavior in California's Wholesale Electricity Market During Summer 2000: The Final Word," *The Energy Journal*, 23: 1-35.

Lacombe, Romain H., and Parsons, John E. (2007), "Technologies, Markets and Challenges for Development of the Canadian Oil Sands Industry," 07-006, MIT Center for Energy and Environmental Policy Research, Massachusetts Institute of Technology, June 2007.

Lacombe, Romain H. and Parsons, John E. (2008), "Dynamics of the Heavy Light Spread in the North American Oil Market," Draft Paper, MIT Center for Energy and Environmental Policy Research, Massachusetts Institute of Technology, February 2008.

National Petroleum Council (2003, 2004), *Balancing Natural Gas Policy – Fueling the Demands of a Growing Economy, Volume I: Summary of Findings and Recommendations*, September 2003, and *Volume II: Integrated Report*, March 2004.

O'Neil, Patrick, ed. (1999-2001), *The Energy Market Report*, Economic Insight, Inc., Vol. 4-6, December 1999 through September 2001, Portland, Oregon, http://www.econ.com.

Organization for Economic Cooperation and Development (OECD), International Energy Agency (IEA) (2009), *World Energy Outlook 2009 – Prospects for Natural Gas*, November 2009.

Pyrdol, John and Bob Baron (2003), "Fuel Switching Potential of Electric Generators: A Case Study," Energy Modeling Forum Working Paper 20.3, Stanford University (July).

Serletis, Apostolos, and Herbert, John (1999), "The message in North American energy prices," *Energy Economics* 21 (1999): 471-483.

Serletis, Apostolos, and Rangel-Ruiz, Ricardo (2004), "Testing for Common Features in North American Energy Markets," *Energy Economics* 26 (2004): 401-414.

Stata Corporation, (2003), *Stata Time-Series Reference Manual Release 8*, Stata Press, Stata Corporation, College Station, Texas, USA.

United States, Bureau of the Census (2002), data files for the "2002 Economic Census, Manufacturing - Industry Series", Washington, D.C., USA.

United States, Department of Energy (DOE) (2010), Energy Information Administration (EIA), Office of Coal, Nuclear, Electric and Alternate Fuels, *Electric Power Annual 2008*, Washington, D.C., USA, January 2010.

United States, Department of Energy (DOE) (2009a), Energy Information Administration (EIA), "EIA Monthly Energy Review May 2009," Washington, D.C., USA.

United States, Department of Energy (DOE) (2009b), Energy Information Administration (EIA), Office of Coal, Nuclear, Electric and Alternate Fuels, *Electric Power Annual 2007*, Washington, D.C., USA, January 2009.

United States, Department of Energy (DOE) (2008), Energy Information Administration (EIA), Form EIA-860, *Annual Electric Generator Report 2008*, "Existing Capacity by Energy Source, 2008, (Megawatts)", Washington, D.C., USA, http://www.eia.doe.gov/cneaf/electricity/epa/epat1p2.html.

United States, Department of Energy (DOE) (2002), Energy Information Administration (EIA), *2002 Manufacturing Energy Consumption Survey,* Office of Energy Markets and End Use, Energy Consumption Division, Form EIA-846, Washington, D.C., USA.

United States, Department of Energy (DOE) (website(a)), Energy Information Administration (EIA), "U.S. Field Production of Crude Oil," Washington, D.C., USA http://tonto.eia.doe.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MCRFPUS1&f=M

United States, Department of Energy (DOE) (website(b)), Energy Information Administration (EIA), "U.S. Imports from All Countries," Washington, D.C., USA. http://tonto.eia.doe.gov/dnav/pet/pet_move_impcus_d_nus_z00_mbblpd_m.htm.

United States, Department of Energy (DOE) (website(c)), Energy Information Administration (EIA), "U.S. Natural Gas Consumption by End Use," Washington, D.C., USA. http://tonto.eia.doe.gov/dnav/ng/ng_cons_sum_dcu_nus_a.htm.

United States, Department of Energy (DOE) (website(d)), Energy Information Administration (EIA), "U.S. Natural Gas Imports by Country," Washington, D.C., USA. http://tonto.eia.doe.gov/dnav/ng/ng_move_impc_s1_a.htm.

United States, Department of Energy (DOE) (website(e)), Energy Information Administration (EIA), "U.S. Product Supplied for Crude Oil and Petroleum Products," Washington, D.C., USA. http://tonto.eia.doe.gov/dnav/pet/pet_cons_psup_dc_nus_mbbl_m.htm.

United States, Department of Energy (DOE) (website(f)), Energy Information Administration (EIA), "U.S. Total Crude Oil and Products Imports," Washington, D.C., USA. http://tonto.eia.doe.gov/dnav/pet/pet_move_impcus_a2_nis_ep00_im0_mbblpd_m.htm.

United States, Department of Energy (DOE) (website(g)), Energy Information Administration (EIA), "Weekly Working Gas in Underground Storage Statistics," Washington, D.C., USA. http://tonto.eia.doe.gov/dnav/ng/ng_stor_wkly_s1_w.htm.

United States, Department of Energy (DOE) (website(h)), Energy Information Administration (EIA), "Weekly Working Gas in Underground Storage Statistics Methodology," Washington, D.C., USA. http://www.eia.doe.gov/oil_gas/natural_gas/ngs/methodology.html.

United States, Department of Energy (DOE) (website(i)), Energy Information Administration (EIA), "Natural Gas Used as Feedstock for Hydrogen Production," Washington, D.C., USA. http://tonto.eia.doe.gov/dnav/pet/pet_pnp_feedng_k_a.htm.

United States, Minerals Management Service (MMS) (website), Gulf of Mexico Outer Continental Shelf Region, "Shut-in oil and Natural Gas Production Statistics," http://www.gomr.mms.gov/.

United States, National Oceanic and Atmospheric Administration (NOAA), National Weather Service (NWS), Climate Prediction Center (CPC) (website), Degree Days Statistics, http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/cdus/degree_days/.

Van Vactor, Samuel A. (1982), "Fuel Prices in the Northwest," Energy Analysis and Planning, Inc., Chapter 2, report to the Northwest Power Planning Council, 14 August 1982, 20 pp.

Villar, Jose A. and Joutz, Frederick L. (2006), "The Relationship Between Crude Oil and Natural Gas Prices," Energy Information Administration, Office of Oil and Gas, October 2006, 43 pp.

Yergin, Daniel (1991), *The Prize:The Epic Quest for Oil, Money & Power*, Free Press, New York, NY, USA, 906 pp.