The Relationship of Natural Gas to Oil Prices

Peter R. Hartley, Kenneth B Medlock III** and Jennifer E. Rosthal***

We investigate the relationship between the prices of natural gas and crude oil, and the factors that cause short run departures from the long run equilibrium price relationship. We find evidence that the link between natural gas and crude oil prices is indirect, acting through competition at the margin between natural gas and residual fuel oil. We also find that technology is critical to the long run relationship between fuel prices, and short run departures from long run equilibrium are influenced by product inventories, weather, other seasonal factors and supply shocks such as hurricanes.

1. INTRODUCTION

The relationship between natural gas and crude oil prices affects both energy consumers and providers, especially by influencing their incentives to invest in inventories or different types of energy using equipment. Energy market traders also are interested to know whether there is a tendency for the relative prices of different energy commodities to return to a particular value, since if such a tendency exists it might form the basis of a trading strategy. A historical "rule-of-thumb" of a ratio of WTI to the Henry Hub of 10:1, so that natural gas priced at one-tenth the price of a barrel of crude oil, seemed to disintegrate during the late 1990s and early 2000s, evolving to something closer to 6:1. Variability in the relative price relationship, which has in fact ranged from 4:1 to 12:1, has prompted questions as to whether or not the prices of natural gas and crude oil have decoupled, or if there is a stable relationship at all.

The Energy Journal, Vol. 29, No. 3.

* Corresponding author. Economics Department, MS22, and James A, Baker III Institute for Public Policy, Rice University, 6100 Main Street, Houston, TX, 77005. E-mail: hartley@rice.edu.

** James A, Baker III Institute for Public Policy and Economics Department, MS22, Rice University, 6100 Main Street, Houston, TX, 77005.

The authors thank the Energy Forum of the James A. Baker III Institute for Public Policy and McKinsey and Company for financial support and seminar participants at Rice University and the USAEE annual meetings, 2006, Ann Arbor Michigan for valuable comments.

In this paper, we use monthly data from February 1990 through October 2006 to demonstrate the existence of a long run cointegrating relationship between the residual fuel oil price, the natural gas price and technological change in electricity generation. We thus provide a technological explanation for the apparent change in the relationship between crude oil and natural gas prices in recent years. By estimating a vector error correction model (VECM) that includes some stationary exogenous variables, we also identify shocks that cause departures from that relationship. The VECM also allows us to identify a causal ordering in price adjustment, and how fast that adjustment occurs.

We begin from the premise that electricity generation plays a key role in influencing the relative prices of different energy commodities. In a recent paper, Hartley, Medlock and Rosthal (2007) show that substitution between natural gas and residual fuel oil is particularly strong in a few regions in the United States where there is sufficient system-wide switching capability.[1] Even in regions where individual plants cannot switch between fuels, different types of plant can be operated for different lengths of time as fuel prices change, extending the switching capability of the system.

Plant and grid level switching between different fuel types by electricity generators imposes strong pressure to limit deviations in the relative prices of competing fuels. Specifically, when possible, generators will arbitrage the cost of producing electricity in dollars per megawatt hour ($/MWh), which equals the price of fuel in $/Btu times the heat rate in Btu/MWh. Hence, changes in the heat rates of the plants using the different fuels will change their relative competitiveness. We therefore argue that the development of combined cycle gas turbines (CCGT) has raised the attractiveness of natural gas as a fuel for generating electricity. The resulting demand-side pressure, in turn, has contributed to an increase in the price of natural gas relative to fuel oil and hence also to crude.

2. PREVIOUS RESEARCH

Other authors have considered the cointegration of various energy prices. Of particular interest to us are papers that examine the cointegration of prices of different commodities.[2] One paper in particular considered the relationship between natural gas and residual fuel oil prices. Serletis and Herbert (1999) test for the existence of common trends in daily natural gas prices at Henry Hub and Transco Zone 6, the price of power in PJM, and the price of residual fuel oil at New York Harbor from October 1996 through November 1997. They find that the three fuel prices are cointegrated and that Transco Zone 6 prices adjust signifi-

1. They also found limited substitutability between natural gas and distillate-fired peaking plants. Natural gas and heating oil also compete in space heating applications. We find in this paper, however, that the relationships between distillate, natural gas and residual fuel oil were quite weak at the aggregate level. Eliminating the few marginally significant distillate variables did not materially affect any of the remaining coefficients and hence the variables have been omitted to simplify the exposition.

2. There is also a literature examining the cointegration of a single commodity across different locations (see, for example, DeVany and Walls (1993, 1999) and Siliverstovs et al. (2005)).

cantly faster than do Henry Hub prices to deviations in their long run relationship. Serletis and Herbert also find that fuel oil prices show no significant adjustment to deviations in the long run relationship with either Henry Hub or Transco Zone 6 natural gas prices. However, the Transco Zone 6 natural gas price does appear to adjust to movements in the fuel oil price at New York Harbor, indicating regional competition between the two fuels. Their results thus support weak exogeneity of fuel oil prices in the system of equations. Similarly, the fact that the Transco Zone 6 price adjusts most quickly to both long run price relationships suggests that it is in a sense the "most endogenous" price of the three, a result that is not surprising given that Transco Zone 6 reflects a local end-of-pipe market.

Building on this analysis, Serletis and Rangel-Ruiz (2002) examine the existence of common price cycles in North America energy commodities using the daily prices of natural gas at the Henry Hub and WTI from 1991 through 2001. In addition, they studied cointegration of U.S. and Canadian natural gas prices. They concluded that natural gas prices at Henry Hub and AECO (a liquid pricing point in Alberta) demonstrate common cycles, but Henry Hub and WTI do not have common price cycles. They claim this decoupling of U.S. energy prices is a result of deregulation.

Villar and Joutz (2006) examine the apparent decoupling of the prices of WTI crude oil and Henry Hub natural gas in more detail, finding a cointegrating relationship between the two prices that exhibits a positive time trend. This indicates that the prices have a long run relationship that is slowly evolving rather than constant. Villar and Joutz estimate an error correction model that includes exogenous variables such as natural gas storage levels, seasonal dummy variables, and dummy variables for a few other transitory shocks. Their analysis supports the findings of Serletis and Rangel-Ruiz (2002) that the price of WTI is weakly exogenous to the price of natural gas at the Henry Hub.

Brown and Yücel (2007) used an ECM to analyze weekly prices from January 1994 through July 2006. They found that the price series are cointegrated over this period, indicating a stable long run relationship.[3] They found that short run deviations from the estimated long run relationship could be explained by market fundamentals such as storage levels, weather, and the quantity of production shut-in due to hurricanes. They report that the price of natural gas at Henry Hub responds significantly to the deviation from the long run relationship, changes in the prices of natural gas for the preceding two weeks, and the change in the price of oil one week earlier. Furthermore, they report that weather and storage levels both have significant effects on the price of natural gas by moving it temporarily away from the long run relationship to crude oil prices. Similar to previous studies, Brown and Yücel found the direction of causality is from the price of WTI to the price of natural gas at Henry Hub, but not the other direction.

Bachmeier and Griffin (2006) also examine the evidence for cointegration within as well as across various commodity markets. Specifically, they find

3. However, they also found that a cointegrating relationship does *not* exist if they consider the shorter time period of June 1997 through July 2006.

that various global crude oils are strongly cointegrated, but that the cointegrating relationship between coals in the U.S. is not strong. Moreover, they report that cross-commodity cointegration in the U.S. is weak, and conclude that the market for energy can only be considered a single market for primary energy in the very long run. By contrast, Asche, Osmundsen and Sandsmark (2006), using data for the U.K., report that the prices of crude oil, natural gas and electricity are cointegrated. Moreover, they find that there is a single market for primary energy in the U.K. in which price is determined exogenously by the global market for crude oil. In addition, they conclude that changes in regulatory structures and capacity constraints can make prices appear to be more or less cointegrated. Neither of these studies, however, considers the influence of exogenous variables, such as weather and inventories, on short run price adjustment. In addition, none of the studies we reviewed considered the influence of technology on the long run price relationships.

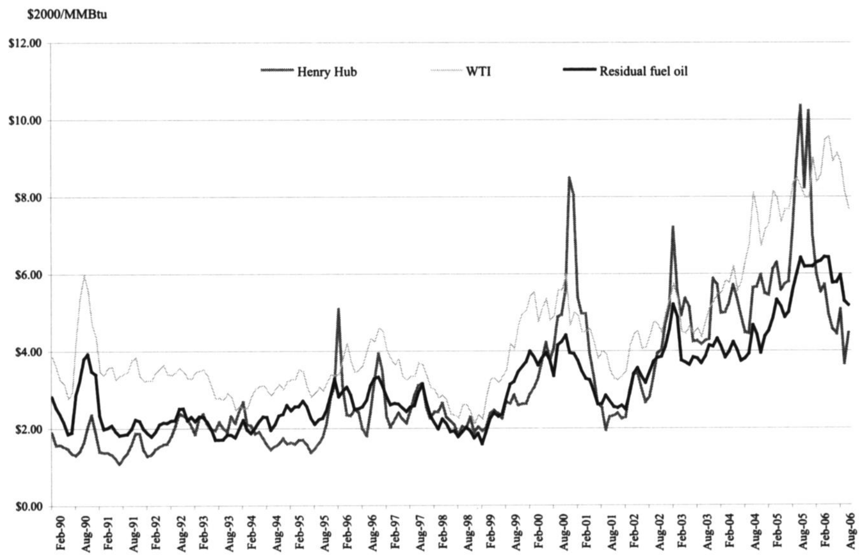
We also examine the relationship between oil and natural gas prices. Like Villar and Joutz, we use monthly data and attempt to find a stable cointegrating relationship between gas and oil prices by adding an additional variable, but we consider technology rather than a time trend. More specifically, we assume that an electricity producer chooses among alternative fuels to minimize costs in $/MWh given as the fuel price times the heat rate. The substantial increase in combined-cycle power generating capacity over the past decade has lowered the capacity-weighted average heat rate for natural gas plants, effectively lowering the cost of producing electricity with natural gas relative to other fuels. Since a substantial amount of fuel competition occurs in the power sector, we would expect this technological change to have affected the long run relationship between natural gas and crude oil prices. Thus, we hypothesize that the increased efficiency of producing electricity with natural gas is responsible for the increasing price differential observed by Villar and Joutz.

We also follow both Villar and Joutz and Brown and Yücel by allowing market fundamentals such as storage levels and weather to influence the short run dynamic relationship between the prices. Finally, while much of the recent literature focuses on the relationship between the prices of crude oil and natural gas, we follow the earlier papers by Serletis et. al. in relating gas prices not to the price of crude oil but rather to the prices of the main competitive oil product, namely residual fuel oil. However, we also allow crude prices to enter the system of equations that we estimate and thus to influence both of the other prices.

3. DATA

We examine a system of three fuel prices: the price of natural gas at the Henry Hub (compiled from *Natural Gas Weekly*), the wholesale price of residual fuel oil and the price of WTI crude (the latter two series were obtained from the EIA web site). We examine the price at Henry Hub rather than natural gas prices in other regions because variations in basis differentials primarily reflect transportation constraints, and hence the shadow value of scarce transportation capacity, rather than changes in the value of energy as such. Consistent with our theoretical framework, fuel prices are expressed in real 2000$/MMBtu,[4] and are deflated using industrial electricity retail prices, which most closely resemble a wholesale output price for the electricity sector.[5]

Figure 1. Real Energy Commodity Prices (February 1990 – August 2006)



Sources: Natural Gas Weekly and the Energy Information Administration

Figure 1 plots the real prices of WTI, residual fuel oil and natural gas at the Henry Hub in $2000/MMBtu. It shows that natural gas prices have tended to relate more closely to residual fuel oil than to crude prices. Specifically, gas prices have tended to fluctuate around residual fuel oil prices with alternating periods of several months to a year where they are persistently above or below the residual fuel oil price. In some brief episodes, however, the natural gas price spikes substantially above the residual fuel oil price, and even the West Texas Intermediate (WTI) price, in energy-equivalent terms.

4. The conversion factors for energy content, obtained from the EIA website, are 1.03 MMBtu per thousand cubic foot for natural gas, 6.287 MMBtu per barrel for residual fuel oil, and 5.8 MMBtu per barrel for WTI.

5. We related real rather than nominal prices since general inflation could make any nominal price non-stationary and the general inflation rate would need to be included in the cointegrating relationship. This may obscure the real relationship between the different energy commodities. From the perspective of a cost-minimizing electricity producer, the relevant real input price for each fuel is the nominal price times the heat rate divided by the price of electricity. We estimated the cointegrating relationship using the logs of these variables.

The heat rate data were constructed from two sources. The EPA NEEDS 2004 data provides the heat rates for many generating plants in the U.S., but very few capacities and no information about month of first use. To obtain the additional information, the EPA data were matched to the facilities listed in the EIA Form-860 (Annual Electric Generator Report) based on generator identification codes, in-service year, fuel, and prime mover type.[6] The formula used for calculating the capacity weighted heat rate for plants using fuel of type f in month t (HR^f_t) is then given as

$$HR^f_t = \frac{\sum_i (Capacity^f_{i,t} * HeatRate^f_{i,t})}{\sum_i \text{Capacity}^f_{i,t}}$$

for all plants i using fuel f that were available for use at any time during month t. The EIA database provides as many as six energy sources for any one generator. Only the primary energy source was considered for the heat rate calculations. The use of the heat rate variable in our analysis restricts us to using data at the monthly frequency. One advantage of using monthly data, however, is that we can cover quite a long time series from February 1990 to October 2006.

Other variables used in the dynamic adjustment process include beginning of period inventory levels, variables reflecting weather conditions, and a variable to capture disruptions to Gulf of Mexico production as a result of hurricanes. The inventory variables, which were obtained from the Energy Information Administration web site, allow for short-term supply availability to either mitigate or exacerbate the effects of shocks on price movements. For natural gas, we used working gas in storage at the end of the previous month (beginning of the current month), and, for residual fuel, we used monthly stocks at the end of the previous month measured in thousands of barrels.[7]

The weather variables are included to capture the effects of weather on demand and hence price. The variables were calculated using data on heating and cooling degree-days (HDD_t and CDD_t) from the National Oceanic and Atmospheric Administration (NOAA). Normal seasonal variations in weather were calculated by the 15-year average degree-days for each month ($HDDavg_t$ and $CDDavg_t$) over the period 1990-2005. These were not included in the dynamic adjustment equations, however, since we also included monthly indicator variables.[8] Instead, we included deviations in heating and cooling degree-days in each month, measured as the actual values minus the 15-year average:

$$HDDdev_t = HDD_t - HDDavg_t$$

$$CDDev_t = CDD_t - CDDavg_t$$

6. The specific matching methodology we used is available upon request.

7. Natural gas inventories are measured in units of trillion cubic feet. We converted the residual fuel oil stocks to *trillions* of barrels prior to the regression analysis.

8. An argument for including monthly effects rather than normal weather variables is that seasonal factors other than weather, such as the distribution of holidays or variations in the number of working days in a month, could influence demands for different types of energy commodities and hence prices.

We also included a measure of extreme winter weather events calculated as the top decile of the *HDD* distribution:[9]

$$HDDext_t = \begin{cases} 0 & \text{if } HDDev_t \text{ is not in the top 10\% of values} \\ HDDev_t & \text{if } HDDev_t \text{ is in the top 10\% of values} \end{cases}$$

The hurricane variable is included to capture the price impacts of short term supply disruptions. We derived the variable by regressing Federal Offshore Gulf of Mexico natural gas production on a cubic time trend and a set of dummy variables representing periods when major hurricanes, as reported by NOAA, affected Gulf producing areas

$$NG_t^{Gulf} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \sum_j \sum_{t_j} \delta_{jt_j} D_{t_j} + \varepsilon_t. \tag{1}$$

In equation (1), j indexes the hurricanes that tracked through producing areas in the Gulf of Mexico within the sample period, t_j indexes months for which hurricane j had a statistically significantly negative effect on production (relative to trend), and $D_{t_j} = 1$ for $t = t_j$ and 0 otherwise. The measure of shut-in production due to hurricanes is then

$$HurrShutIn_t = -\sum_j \sum_{t_j} \delta_{jt_j} D_{t_j}$$

We take this approach to capture the effect of hurricanes for multiple reasons. First, a number of hurricanes over this period affected production beyond the month in which they occurred. The method we use allows moderating effects over a number of months. Second, unlike a simple dummy variable for the months of major hurricane strikes, our method allows different hurricanes to affect production by different amounts. Third, we allow for slow moving changes in Gulf production for other reasons, such as depletion or new discoveries, by measuring shut-in production as statistically significant deviations from a smooth polynomial time trend. The method indicates production was lost due to hurricanes during August-September 1992, October 1995, September 1998, September-October 2002, September-October 2004, and September-December 2005.[10]

We also included an indicator variable (*Chicago*) for February 1996. The first week of that month was very cold in many parts of the U.S. and produced very high demand. In particular, when coupled with low storage levels, the pe-

9. A similar extreme cooling-degree day variable was neither numerically nor statistically significantly different from zero in any equation.

10. Using a dummy-variable instead of our measure of lost production made the coefficient on the hurricane shut-in variable less significant but did not materially affect any of the remaining coefficients. Another alternative involves using data from the MMS to measure shut-in production (www.gomr.mms.gov/homepg/whatsnew/hurricane/history.html). However, these statistics are not consistent over the entire sample period, and do not generally track shut-in production for an extended period. Our measure also has the virtue of measuring statistically significant variation in *total* production from the Gulf of Mexico, which is what is relevant as a supply shock.

riod witnessed unprecedented prices of natural gas in the Chicago market area that were transmitted back to Henry Hub. Although we have included variables to capture the effect of extreme weather on high prices, the 1996 incident had a peculiarly large effect on prices. This might be related to the then relatively new emergence of major market hubs so that hub services such as "parks and loans" were not widely used at that time.[11]

Finally, we allowed for seasonality in the adjustment process by including a set of monthly dummy variables. The natural gas price has a pronounced seasonal pattern, and although inventories rise and fall in an effort to arbitrage seasonal price movements, they do not eliminate them. Nevertheless, we might expect that inventory levels and normal weather conditions could explain seasonal price movements. However, other factors such as the number of days in a month, and normal seasonal demand patterns in fuel consumption may not necessarily be captured by seasonal changes in inventories and heating or cooling degree-days.[12]

4. ANALYSIS AND RESULTS

Engle and Granger (1987) showed that if two series, y_{1t} and y_{2t}, are cointegrated, then there must exist an error correction representation of the dynamic system governing the joint behavior of y_{1t} and y_{2t} over time. This system can be written as

$$\Delta y_{1t} = \alpha_{10} + \alpha_{11}\Omega_{t-1} + \sum_{i=1}^{p_1} \alpha_{12,i}\ \Delta y_{1,t-i} + \sum_{i=1}^{p_2} \alpha_{13,i}\ \Delta y_{2,t-i} + \varepsilon_{1t}$$

$$\Delta y_{2t} = \alpha_{20} + \alpha_{21}\Omega_{t-1} + \sum_{i=1}^{p_1} \alpha_{22,i}\ \Delta y_{1,t-i} + \sum_{i=1}^{p_2} \alpha_{23,i}\ \Delta y_{2,t-i} + \varepsilon_{2t}$$

where Ω_{t-1} is an error correction term representing the deviation from the equilibrium or cointegrating relationship between $y_{1,t-1}$ and $y_{2,t-1}$. The coefficients on Ω_{t-1} are *speed of adjustment parameters* measuring how fast y_{1t} and y_{2t} revert to their long run equilibrium relationship. Note that since y_{1t} and y_{2t} are cointegrated, the estimation of the relationship between y_{1t} and y_{2t} is superconsistent and the series Ω_t can be treated in the estimation of the ECM as if it were known. Each equation

11. The February 1996 episode is discussed in *Natural Gas 1996: Issues and Trends* available at www.eia.doe.gov/oil_gas/natural_gas/analysis_publications/natural_gas_issues_and_trends/it96.html. The EIA notes on page 21 that "some industrial gas consumers paid more than $45.00 per MMBtu in Chicago in order to avoid pipeline imbalance penalties of over $60.00 per MMBtu." On page 78 they claimed, "Other evidence that market centers are not being fully utilized is the size of the daily price spikes experienced this past winter." "Parks and loans" services can mitigate the impact of such combinations of severe weather and low storage levels because they allow consumers to meet contractual obligations while smoothing the profile of capacity utilization on market area pipelines. Brownfield and greenfield expansions of pipeline infrastructure (by Northern Border and Alliance) after the winter of 1996 increased access to Canadian supplies and storage and helped mitigate similar problems in subsequent years.

12. We also investigated the use of actual weather rather than deviations from normal, but the monthly dummies remain significant. Thus, we adopt the approach taken here.

in the system above has the desirable property that if we are at long run equilibrium ($\Omega_t = 0$) and there is no change in any of the other variables, there will be no change in y_{1t} and y_{2t}, provided the intercept terms (α_{10} and α_{20}) are equal to zero.

The premise of the ECM is that, although natural gas and residual fuel oil prices are each non-stationary (or more specifically integrated of order 1), there exists a stable long run relationship between them. Statistically, if two non-stationary variables are cointegrated, the residual after estimating their cointegrating relationship will be stationary. Phillips-Perron tests indicate that the levels of the logs of the three price variables and the relative heat rate variable are non-stationary and integrated of order one, $I(1)$. The remaining variables are all stationary, $I(0)$.[13]

To obtain a better understanding of the relationship among the prices and the relative heat rate variable, we estimated a VAR on a vector $\mathbf{Y}_t$ of natural gas price, residual fuel oil price, WTI, and the relative heat rate using Johansen's maximum likelihood method.[14] Since the elements of $\mathbf{Y}_t$ are each $I(1)$, the changes in the variables at time t, $\Delta\mathbf{Y}_t$, are estimated as a function of $\mathbf{Y}_{t-1}$ and n lags of $\Delta\mathbf{Y}_t$, where the optimal n is determined using the Akaike Information Criterion (AIC). The rank of the matrix multiplying $\mathbf{Y}_{t-1}$ is the number of cointegrating relationships in the system. The errors from the estimated cointegrating relationships are then used to construct an error correction model similar to that used in the Engle-Granger method.

The Johansen tests imply that there are two cointegrating relationships, and the AIC indicates that the optimal number of lags n is one. The two normalized cointegrating relationships are given as:[15]

$$ce_1 = \ln P_t^{NG} - 0.3327 - \underset{(0.2612)}{0.6540} \ln P_t^{rfo} + \underset{(1.7855)}{3.5045} \ln \frac{HR_t^{NG}}{HR_t^{rfo}}.$$

$$ce_2 = \ln P_t^{rfo} + 0.2053 - \underset{(0.0708)}{0.8914} \ln P_t^{WTI}.$$

Table 1 gives the corresponding estimated vector error correction model (VECM). The coefficients on the two cointegrating equations imply that only natural gas prices respond to divergences in the first long run relationship ce_1 and

13. The test statistics are $Z(\rho) = -8.725$ and $Z(\tau) = -2.082$ for $ln(P^{NG})$ and $Z(\rho) = -155.443$ and $Z(\tau) = -12.306$ for $\Delta ln(P^{NG})$. For $ln(P^{rfo})$, the test statistics are $Z(\rho) = -6.639$ and $Z(\tau) = -1.666$ compared with $Z(\rho) = -131.664$ and $Z(\tau) = -10.836$ for $\Delta ln(P^{rfo})$. For $ln(P^{WTI})$, the test statistics are $Z(\rho) = -4.839$ and $Z(\tau) = -1.345$ compared with $Z(\rho) = -143.120$ and $Z(\tau) = -11.493$ for $\Delta ln(P^{WTI})$. For *ln(HRrel)*, the test statistics are $Z(\rho) = 1.363$ and $Z(\tau) = 1.898$ compared with $Z(\rho) = -63.369$ and $Z(\tau) = -6.133$ for Δ*ln(HRrel)*. The interpolated 10% critical value for $Z(\rho)$ is -11.133, and for $Z(\tau)$ it is -2.573. For the weather, the statistics are $Z(\rho) = -161.882$ and $Z(\tau) = -11.074$ for *HDDdev*, and $Z(\rho) = -136.925$ and $Z(\tau) = -10.005$ for *CDDdev*. For the storage variables, they are $Z(\rho) = -59.174$ and $Z(\tau) = -5.426$ for *ngstor*, and $Z(\rho) = -23.467$ and $Z(\tau) = -3.774$ for *rfostor.*

14. For more information on maximum likelihood estimation in this context see Hamilton (1994).

15. The likelihood ratio test of the over identifying restrictions in this normalization yields a statistic χ_1^2 with a p-value of 0.378.

only residual fuel oil prices adjust in response to deviations in ce_2. Furthermore, the negative coefficients imply that the subsequent adjustments will tend to restore the long run relationships.

The fact that neither changes in the WTI price nor changes in the relative heat rate variable respond to deviations in the two cointegrating relationships implies that both of these variables are weakly exogenous.[16] The VECM also implies that the WTI price influences the remaining prices mainly through its effect on the residual fuel oil price. However, the estimated dynamic adjustment process in the VECM needs to be treated with some caution since the AIC test for lag length only considers uniform increments of all lags in all equations. In addition, the system may omit important exogenous variables such as the weather and storage variables and this could bias the estimated coefficients. To investigate more flexible dynamic adjustment models, we used the Engle-Granger two-step methodology.

Table 1. Estimated VECM model (without exogenous variables)

Variable	$\Delta\ln P_t^{NG}$	$\Delta\ln P_t^{rfo}$	$\Delta\ln P_t^{WTI}$	$\Delta\ln HRrel_t$
$ce_{1,t-1}$	−0.1699***	0.0010	−0.0128	−0.0004
	(0.0455)	(0.0259)	(0.0263)	(0.0004)
$ce_{2,t-1}$	0.0031	−0.1522***	−0.0092	−0.0011
	(0.1002)	(0.0570)	(0.0580)	(0.0008)
$\Delta\ln P_{t-1}^{NG}$	0.1478*	0.1465***	0.0536	−0.0006
	(0.0780)	(0.0444)	(0.0451)	(0.0007)
$\Delta\ln P_{t-1}^{rfo}$	0.1306	−0.1718*	−0.2086**	0.0027*
	(0.1835)	(0.1044)	(0.1062)	(0.0015)
$\Delta\ln P_{t-1}^{WTI}$	0.1316	0.4858***	0.3275***	−0.0015
	(0.1892)	(0.1077)	(0.1096)	(0.0016)
$\Delta\ln HRrel_{t-1}$	−0.0181	−5.3305	−3.6922	−0.6520***
	(6.5275)	(3.7142)	(3.7801)	(0.0545)
constant	−0.00001	−0.00001	−0.0002	−0.0003***
	(0.0113)	(0.0064)	(0.0065)	(0.0001)
R^2	0.1249	0.2442	0.0625	0.5701
joint significance	$\chi^2_7 = 27.40$	$\chi^2_7 = 62.03$	$\chi^2_7 = 12.81$	$\chi^2_7 = 254.57$

*** statistically significantly different from zero at the 1% level;
** statistically significantly different from zero at the 5% level;
* statistically significantly different from zero at the 10% level

16. A test that the coefficients on $ce_{1,t-1}$ and $ce_{2,t-1}$ are zero except for $ce_{1,t-1}$ in the natural gas price adjustment equation and $ce_{2,t-1}$ in the residual fuel oil price adjustment equation yields a test statistic χ^2_6 = 5.35 (p = 0.4994). A test that only the coefficients on $ce_{1,t-1}$ and $ce_{2,t-1}$ in the WTI equation are zero yields a test statistic χ^2_2 = 0.38 (p = 0.8257), while a test that both coefficients on $ce_{1,t-1}$ and $ce_{2,t-1}$ are zero in the relative heat rate equation yields a test statistic χ^2_2 = 4.72 (p = 0.0946). While this is just significant at the 10% level, neither coefficient individually is significantly different from zero (the corresponding z-statistics have p-values of 0.273 and 0.199). Finally, a joint test that the coefficient on $ce_{1,t-1}$ is zero in the residual fuel oil equation and the coefficient on $ce_{2,t-1}$ is zero in the natural gas equation yields a test statistic χ^2_2 = 0.001 (p = 0.9986).

Engle-Granger Methodology

The Engle-Granger method estimates each cointegrating relationship individually using ordinary least squares (OLS). Then, the errors from those cointegrating equations are included, along with the exogenous variables, in short run dynamic adjustment equations to explain adjustment to the long run equilibrium. Following the Johansen method results, we estimated equations (2) and (3) below by OLS[17]

$$\ln P_t^{NG} = \underset{(0.0913)}{0.0701} + \underset{(0.0849)}{0.8779} \ln P_t^{rfo} - \underset{(0.5785)}{3.0032} \ln \frac{HR_t^{NG}}{HR_t^{rfo}} + \varepsilon_t^{NG} \tag{2}$$

$$\ln P_t^{rfo} = \underset{(0.0339)}{-0.2931} + \underset{(0.0234)}{0.9637} \ln P_t^{WTI} + \varepsilon_t^{rfo} \tag{3}$$

Since the Phillips-Perron test indicates both residuals are stationary, in each case there is a stable long run relationship and the parameter estimates will be superconsistent.[18] The strong and statistically significant negative coefficient on the relative heat rate in (2) indicates that improvement in the heat rate of gas relative to oil fired generating plant has raised the price of natural gas relative to residual fuel oil as hypothesized. Furthermore, if we omit the relative heat rate from (2) the residual is closer to being non-stationary. In addition, if we use WTI rather than the residual fuel oil price in (2), the residual is non-stationary at the one percent level.

Next, we estimate the short run dynamic adjustment to the long run relationships including stationary variables X_t, such as storage levels and weather, which affect short run price adjustments. Using $\hat{\varepsilon}_t^{NG}$ to denote the predicted residual from (2), the ECM for the change in gas prices can be written as

$$\Delta \ln P_t^{NG} = \beta_{00} + \beta_{01}\, \hat{\varepsilon}_{t-1}^{NG} + \alpha_0(L) X_{0t} + \gamma_0(L)\, \Delta \ln P_t^{rfo} + \delta_0(L)\, \Delta \ln P_{t-1}^{NG} \tag{4}$$
$$+ \phi_0(L)\, \Delta \ln P_{t-1}^{WTI} + \omega_t^{NG}$$

Equation (4) reveals that if we are at long run equilibrium, so that $\hat{\varepsilon}_t^{NG} = 0$, and all other variables remain unchanged, then the price of natural gas will remain unchanged. Otherwise, if $\hat{\varepsilon}_t^{NG} > 0$ ($\hat{\varepsilon}_t^{NG} < 0$), the price of natural gas is above (below) its long run equilibrium value, and if $\beta_{01} < 0$ subsequent movements in the natural gas price will tend to restore the long run equilibrium relationship between fuel prices. The terms $\gamma_0(L)$, $\delta_0(L)$ and $\phi_0(L)$ in (4) are polynomials in the lag operator while, since X is a vector, $\alpha_0(L)$ is a matrix of polynomials in the lag operator.

17. The estimated standard errors are in parentheses below each estimated coefficient.

18. For ε_t^{NG}, the test statistics are $Z(\rho) = -30.793$ (10% critical value -11.133) and $Z(\tau) = -4.033$, which has a MacKinnon approximate p-value of 0.0012, while for ε_t^{rfo}, the test statistics are $Z(\rho) = -25.686$ and $Z(\tau) = -3.719$, which has a MacKinnon approximate p-value of 0.0039.

An equation similar to (4) can be written for the dynamic price adjustment of residual fuel oil as

$$\Delta \ln P_t^{rfo} = \beta_{10} + \beta_{11}\, \hat{\varepsilon}_{t-1}^{rfo} + \alpha_1(L) X_{1t} + \gamma_1(L)\, \Delta \ln P_t^{NG} + \delta_1(L)\, \Delta \ln P_{t-1}^{rfo} \qquad (5)$$
$$+ \phi_1(L)\, \Delta \ln P_t^{WTI} + \omega_t^{rfo}$$

where $\hat{\varepsilon}_t^{rfo}$, the predicted residual from (3), represents deviations from the long run equilibrium between the residual fuel oil price and WTI. The interpretation of the variables in (5) is analogous to that for (4).

To provide a baseline for our subsequent analysis, we estimated (4) and (5) using ordinary least squares (OLS). Variables that proved individually and jointly insignificant at the 10% level, apart from the full set of monthly dummy variables and the change in WTI prices, have been eliminated from the equations.[19] The results are reported in Table 2. The monthly dummy variables are not reported, but are available upon request.

Several features of these estimated equations are of interest. First, all variables have the expected signs, and diagnostic tests indicate the models fit the data reasonably well while leaving uncorrelated and homoskedastic residuals. Second, the change in the residual fuel oil price has a much larger effect in the natural gas price equation than vice versa. Third, while the contemporaneous (and lagged) change in WTI has a large effect in the residual fuel oil equation, its coefficient in the natural gas equation is much smaller and not statistically significantly different from zero. This suggests that crude oil prices influence natural gas prices mainly via competition with residual fuel oil as hypothesized.

A potential problem with the OLS estimates is that the prices of residual fuel oil and natural gas may be jointly determined. To examine this possibility, we re-estimated the equations using instrumental variables (IV) using the weather variables, own inventories, lagged values of own price, and current and lagged values of WTI as instruments. The weather variables are exogenous, and the OLS results suggest that only the most extreme winter weather directly affects residual fuel oil prices. We therefore used the contemporaneous and lagged heating and cooling degree-day deviations, as well as the hurricane shut-in variable, as instruments for the change in natural gas prices in (5). We also used beginning of month inventories as an instrumental variable in each equation. Although beginning of month inventories should influence the change in price over the subsequent period, they should not be influenced by that change in price.

19. The OLS analog of the equation for $\Delta \ln P_t^{WTI}$ from Table 1, and using the OLS estimates of the cointegrating relationships, is

$$\Delta \ln P_t^{WTI} = \underset{(0.007)}{0.001} - \underset{(0.026)}{0.001}\, \hat{\varepsilon}_{t-1}^{hhrfo} + \underset{(0.057)}{0.001}\, \hat{\varepsilon}_{t-1}^{rfoWTI} + \underset{(0.045)}{0.047}\, \Delta \ln P_{t-1}^{NG} - \underset{(0.106)}{0.210}\, \Delta \ln P_{t-1}^{rfo}$$
$$+ \underset{(0.110)}{0.336}\, \Delta \ln P_{t-1}^{WTI} - \underset{(3.758)}{3.241}\, \Delta \ln HRrel_{t-1}$$

Table 2. Error Correction Model Estimation Results

Variable	OLS $\Delta\ln P_t^{NG}$	OLS $\Delta\ln P_t^{rfo}$	IV $\Delta\ln P_t^{NG}$	IV $\Delta\ln P_t^{rfo}$
$\hat{\varepsilon}^j_{t-1}$	−0.2316*** (0.0427)	−0.1668*** (0.0407)	−0.2313*** (0.0427)	−0.1666*** (0.0432)
$\Delta\ln P_t^{NG}$		0.0746*** (0.0271)		0.0948 (0.0818)
$\Delta\ln P_{t-1}^{NG}$	0.2319*** (0.0660)	0.0656** (0.0266)	0.2251*** (0.0656)	0.0657** (0.0270)
$\Delta\ln P_{t-2}^{NG}$	−0.1104* (0.0594)		−0.1086* (0.0593)	
$\Delta\ln P_t^{rfo}$	0.5173*** (0.1415)		0.5909*** (0.1162)	
$\Delta\ln P_{t-1}^{rfo}$	−0.1720* (0.1019)		−0.1807* (0.1024)	
$\Delta\ln P_t^{WTI}$	0.0757 (0.1465)	0.7039*** (0.0474)		0.6996*** (0.0568)
$\Delta\ln P_{t-1}^{WTI}$		0.2184*** (0.0485)		0.2205*** (0.0486)
ngstor	−0.1196*** (0.0349)		−0.1193*** (0.0349)	
rfostor		−1.9730* (1.0230)		−2.0584** (1.0552)
$HDDdev_t$	0.00068*** (0.00024)		0.00067*** (0.00024)	
$HDDdev_{t-1}$	−0.00048*** (0.00018)		−0.00048*** (0.00018)	
$CDDdev_t$	0.00136*** (0.00043)		0.00138*** (0.00042)	
$CDDdev_{t-1}$	−0.00078* (0.00042)		−0.00081* (0.00042)	
HDDext	0.00080** (0.00040)	0.00038*** (0.00013)	0.00077* (0.00040)	0.00036** (0.00016)
HurrShutIn	0.000031*** (0.000011)		0.000031** (0.000011)	
$Chicago_t$	0.6164*** (0.1073)		0.6274*** (0.1060)	
$Chicago_{t-1}$	−0.6875*** (0.1164)		−0.6781*** (0.1149)	
N	197	198	197	197
R^2	0.6220	0.7453	0.6213	0.7452
Joint significance	$F_{26,170} = 10.76$	$F_{18,179} = 29.10$	$F_{25,171} = 10.69$	$F_{18,178} = 28.62$
Q-stat (12 lags)	$\chi^2_{12} = 17.019$	$\chi^2_{12} = 15.843$	$\chi^2_{12} = 16.842$	$\chi^2_{12} = 15.294$
Breusch-Pagan	$\chi^2_1 = 0.60$	$\chi^2_1 = 0.00$		
Hausman test			$\chi^2_{19} = 0.06$	$\chi^2_{17} = 1.10$

*** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level

The OLS results suggest the twice-lagged change in natural gas price does not directly affect the change in residual fuel oil price, but is correlated with the contemporaneous change in the natural gas price. Hence, it is also a reasonable instrument for the contemporaneous change in natural gas price in the residual fuel oil equation. Similarly, the OLS results imply that the twice-lagged change in residual fuel oil price does not directly affect the change in the natural gas price. Hence, we also use it as an instrument for the contemporaneous change in fuel oil prices in the natural gas equation.

Since the price of WTI is determined in the world oil markets, it is effectively exogenous with respect to changes in the U.S. markets for residual fuel oil and natural gas. The OLS results also suggest that although changes in the residual fuel oil price are highly dependent on changes in WTI, the influence of WTI on natural gas prices is not statistically significant once the effects of residual fuel oil prices have been taken into account. We therefore also used the contemporaneous and lagged change in the WTI price as instruments for the change in the residual fuel oil price in (4) (after dropping the change in WTI price as an exogenous regressor in that equation).

Table 2 also presents the IV estimates of equations (4) and (5). Hausman tests for exogeneity suggest that one can treat the change in residual fuel oil price as exogenous in the natural gas price adjustment equation and vice versa. The estimated OLS and IV coefficients are generally very similar. However, the coefficient on changes in natural gas prices in (5) becomes statistically insignificant, while the estimated effect of residual fuel oil prices in (4) increases in both magnitude and statistical significance. This suggests that the residual fuel oil price causes movements in the natural gas price, but that the converse is not true. In turn, while the WTI price influences the Henry Hub natural gas price, it does so indirectly by affecting the residual fuel oil price.

We conclude that the natural gas price adjustment equation can be estimated using OLS despite the inclusion of contemporaneous changes in residual fuel oil prices. It will differ from the OLS equation in Table 2, however, since the WTI price will be excluded. Similarly, the residual fuel oil price adjustment equation can be estimated using OLS, but will differ from the OLS equation in Table 2 since contemporaneous changes in natural gas prices will be excluded. Table 3 presents the resulting final regression equations.

The right panel of Table 3 presents a vector error correction estimate of the same model using the OLS estimates of the cointegrating equations. These estimates were obtained by three stage least squares using the program *JMulti* (Lütkepohl and M. Krätzig, 2004). The two sets of estimates in Table 3 are similar except for the coefficients on the current change in the WTI price in the residual fuel oil equation. A potential problem with estimating a system of equations is that if one of the equations is misspecified the estimates in the other equations can be affected. Since developing a model of the world oil market is beyond the scope of our analysis, the model for the change in the WTI real price is rudimentary. It includes only the variables found to be statistically significantly different from

Table 3. Final OLS and VECM Results

	OLS			VECM	
Variable	$\Delta\ln P_t^{NG}$	$\Delta\ln P_t^{rfo}$	$\Delta\ln P_t^{rfo}$	$\Delta\ln P_t^{rfo}$	$\Delta\ln P_t^{WTI}$
$\hat{\varepsilon}^j_{t-1}$	−0.2305*** (0.0426)	−0.1800*** (0.0411)	−0.2306*** (0.0407)	−0.1926*** (0.0396)	
$\Delta\ln P^{NG}_{t-1}$	0.2269*** (0.0652)	0.0692** (0.0271)	0.2259*** (0.0645)	0.0738*** (0.0263)	
$\Delta\ln P^{NG}_{t-2}$	−0.1095* (0.0592)		−0.1092** (0.0562)		
$\Delta\ln P_t^{rfo}$	0.55729*** (0.0916)		0.6060*** (0.2290)		−0.0528 (0.0865)
$\Delta\ln P^{rfo}_{t-1}$	−0.1768* (0.1012)		−0.1869* (0.1051)		
$\Delta\ln P_t^{WTI}$		0.7335*** (0.0470)		0.1726 (0.1844)	
$\Delta\ln P^{WTI}_{t-1}$		0.2193*** (0.0493)		0.3160*** (0.0591)	0.2158** (0.0977)
ngstor	−0.1184*** (0.0348)		−0.1198*** (0.0339)		
rfostor		−2.1658** (1.0391)		−2.5987** (1.0088)	
$HDDdev_t$	0.00067*** (0.00024)		0.00069*** (0.00022)		
$HDDdev_{t-1}$	−0.00048*** (0.00018)		−0.00048*** (0.00017)		
$CDDdev_t$	0.00138*** (0.00042)		0.00139*** (0.00040)		
$CDDdev_{t-1}$	−0.00081* (0.00042)		−0.00079* (0.00039)		
HDDext	0.00078* (0.00040)	0.00047*** (0.00013)	0.00075* (0.00038)	0.00045*** (0.00012)	
HurrShutIn	0.000031*** (0.000011)		0.000031** (0.000010)		
$Chicago_t$	0.6253*** (0.1056)		0.6180*** (0.1013)		
$Chicago_{t-1}$	−0.6786*** (0.1149)		−0.6901*** (0.1070)		
N	197	198	197	197	197
R^2	0.6220	0.7346			
Joint significance	$F_{25,171} = 11.23$	$F_{17,180} = 29.30$			
Q-stat (12 lags)	$\chi^2_{12} = 17.087$	$\chi^2_{12} = 17.287$		$\chi^2_{108} = 132.133$*	
Hetero-skedasticity or ARCH	$\chi^2_1 = 0.78$	$\chi^2_1 = 0.32$	$\chi^2_{12} = 12.018$	$\chi^2_{12} = 8.388$	$\chi^2_{12} = 14.525$

***- significant at the 1% level; **- significant at the 5% level; *- significant at the 10% level zero in Table 1. Even then, the lagged change in the fuel oil price is not statistically significant in the VECM formulation.

Despite the possible problems, we needed the VECM formulation to calculate impulse response functions. These are graphed in Figure 2, along with 95% Hall bootstrap confidence intervals obtained with 250 bootstrap replications. The impulse response functions suggest that both the residual fuel oil and natural gas prices adjust proportionately in the long run to movements in the WTI price, with the residual fuel price approaching full adjustment within a few months and the natural gas price taking a little longer. In all other cases, the shocks have a period of significantly positive effect on the residual fuel oil and natural gas prices, but the effects ultimately disappear. Since only lagged natural gas prices have an effect on the residual fuel oil price, the impulse response for residual fuel oil of a natural gas price shock is non-monotonic.

The price adjustments are stable in the long run because the estimated coefficients on deviations from the long run relationships, $\hat{\varepsilon}^{NG}_{t-1}$ and $\hat{\varepsilon}^{rfo}_{t-1}$, are negative. Deviations from the long run relationships thus cause natural gas and residual fuel oil prices to return to their respective long run equilibrium relationships with residual fuel oil and WTI.

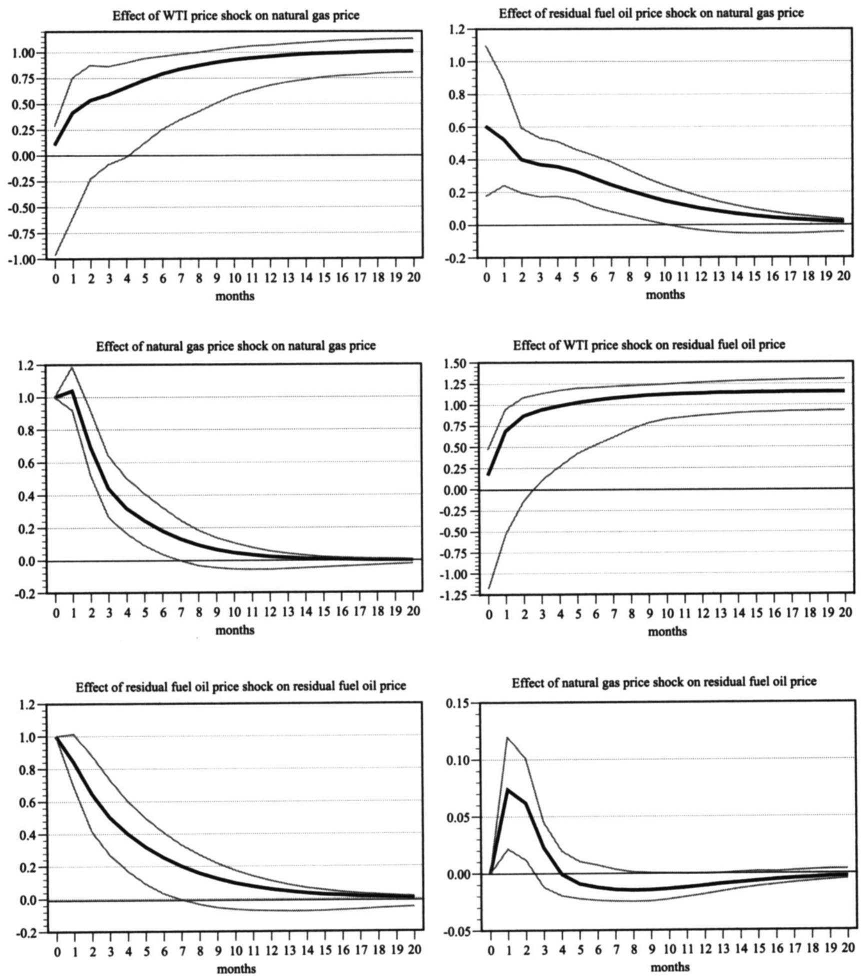
The estimated coefficients also show that weather significantly affects short run price movements. Deviations from normal weather tend to influence only the natural gas price, but extreme deviations in cold weather significantly affect the residual fuel oil price. Since the coefficients on *HDDdev* and *CDDdev* are almost completely reversed after one period, unusual weather evidently has short-lived impacts on natural gas prices. In addition, extremely cold weather in Chicago in February 1996 had an especially large effect on natural gas prices, but the change was reversed the following month.[20]

The estimated negative coefficients on the product inventory variables imply that higher beginning of month storage leads to lower prices over the month, holding all else equal. This follows from the fact that inventories represent readily available supply in any given month, and ample supply will tend to reduce prices. As might be expected, hurricanes tend to have a significant impact on natural gas prices as supply is reduced. The estimated coefficient implies that for each billion cubic feet of production that is shut in as a result of a hurricane in the Gulf of Mexico, natural gas prices at the Henry Hub increase by approximately \$1.03/MMBtu (= exp(0.000032*1000)).

Finally, the monthly dummies (not reported in Table 3) reveal some seasonal tendency in the price series that is not captured by the other variables. With

20. There is considerable evidence the Chicago incident is an outlier. Omitting $Chicago_t$ and $Chicago_{t-1}$ raises the standard deviation of the residuals from 0.0919 to 0.1090, reduces the minimum residual from -0.2528 to -0.4930 and the maximum from 0.2290 to 0.5370. The t-statistic for including $Chicago_t$ is 5.35, and for adding $Chicago_{t-1}$ it is –5.34. Omitting these variables also reduces the R^2 from 0.6214 to 0.4682, while the test for joint significance of the included variables becomes $F_{23,173} = 6.62$ instead of $F_{25,171} = 11.23$. Furthermore, omitting these variables substantially reduces the magnitude and statistical significance of the coefficients of $\Delta \ln P^{NG}_{t-1}$, $\Delta \ln P^{rfo}_{t-1}$ and *HDDext*, while raising the magnitude (and standard error) of the coefficient of *ngstor*.

Figure 2. Impulse Response Functions (VECM with exogenous variables)



regard to natural gas, holding all other variables constant, prices increase more than they do in January from August through December, and increase less than they do in January from February through July. However, only the differences from January in May, and October through December, are statistically significant. In the case of residual fuel oil, price tends to increase more in January than in any other month, but the differences from January are not statistically significantly different in July, October and November.

5. CONCLUDING REMARKS

This paper has demonstrated some important points regarding the relationship between crude oil prices and natural gas prices. First, our analysis suggests that the relationship between these two commodities is indirect, acting via competition between natural gas and residual fuel oil. Most of the previous literature has focused on a *direct* relationship between crude oil and natural gas prices.

The second point, which is closely related to the first, is that the results indicate that crude oil prices are weakly exogenous to a system that includes the natural gas and residual fuel oil prices. More specifically, the results suggest that U.S. natural gas and residual fuel oil prices tend to respond to movements in the international crude oil market, but the reverse is not true. Thus, disequilibria in the long run relationship between natural gas and residual fuel oil prices can be driven by random shocks to the international crude oil market, which themselves influence disequilibria in the long run relationship between the prices of residual fuel oil and crude oil.

Third, similar to Brown and Yücel (2007), we find that variables such as weather, inventories, hurricanes, and other seasonal factors have significant influence on the short run dynamic adjustment of prices. This is important because many other studies ignore these influences. In addition, prolonged periods of low product inventories or active hurricane seasons can extend periods of disequilibria by acting to counter the tendency of the system to return to long run equilibrium. This latter point is important for commercial considerations and short-term policy more generally.

Fourth, the analysis indicates that changes in electricity generating technology can explain the apparent drift in the long run relationship between residual fuel oil and natural gas prices in recent years. The time trend found to be important in previous literature might be serving as a proxy for the evolving relative *cost* of fuels, taking into account improvements in the heat rate of natural gas fired generation capacity. None of the previous literature has considered the influence of technology as an explanatory factor in the evolving relationship between crude oil and natural gas. This is important because future innovations will influence the long run relationship between crude oil and natural gas in a way that simple time trends cannot identify.

REFERENCES

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

Bachmeier Lance and James Griffin (2006). "Testing for Market Integration: Crude Oil, Coal and Natural Gas", *The Energy Journal*, 27(2): 55-72.

Brown, Stephen and Mine Yücel (2007). "What Drives Natural Gas Prices?" Federal Reserve Bank of Dallas. Working Paper 0703.

De Vany, Arthur and David Walls (1993). "Pipeline Access and Market Integration in the Natural Gas Industry: Evidence from Cointegration Tests," *The Energy Journal*, 14(4): 1–19.

De Vany, Arthur and David Walls (1999). "Cointegration analysis of spot electricity prices: Insights on transmission efficiency in the western US," *Energy Economics*, 21(5): 435-448.

Enders, Walter (1995). *Applied Econometric Time Series*. John Wiley and Sons, Inc.

Engle, Robert and Clive Granger (1987). "Cointegration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55(2): 251-276.

Granger, Clive and P. Newbold (1974). "Spurious Regressions in Econometrics." *Journal of Econometrics* 2(2): 111-120.

Hartley, Peter, Kenneth Medlock and Jennifer Rosthal (2007). "An Econometric Evaluation of the Demand for Natural Gas in the Power Generation and Industrial Sectors." James A Baker III Institute for Public Policy, working paper.

Hamilton, James D. (1994). *Time Series Analysis*. Princeton University Press.

Lütkepohl, H. and M. Krätzig (2004), *Applied Time Series Econometrics*, Cambridge: Cambridge University Press.

Serletis, Apostolos and John Herbert (1999). "The Message in North American Energy Prices." *Energy Economics*, 21(5): 471–483.

Serletis, Apostolos and Ricardo Rangel-Ruiz (2004). "Testing for Common Features in North American Energy Markets." *Energy Economics*, 26(3): 401–414.

Siliverstovs, Boriss, Guillaume L'Hegaret, Anne Nuemann, and Christian von Hirschhausen (2005). "International Market Integration for Gas? Cointegration Analysis of Gas Prices in Europe, North America and Japan." *Energy Economics*, 27: 603–615.

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