THE JAMES A. BAKER III INSTITUTE FOR PUBLIC POLICY
RICE UNIVERSITY

AN ECONOMETRIC EVALUATION OF THE DEMAND FOR NATURAL GAS IN THE POWER GENERATION AND INDUSTRIAL SECTORS

BY

PETER HARTLEY
PROFESSOR, ECONOMICS DEPARTMENT

KENNETH B MEDLOCK III
FELLOW IN ENERGY STUDIES, THE JAMES A. BAKER III INSTITUTE FOR PUBLIC POLICY
ADJUNCT ASSISTANT PROFESSOR, ECONOMICS DEPARTMENT

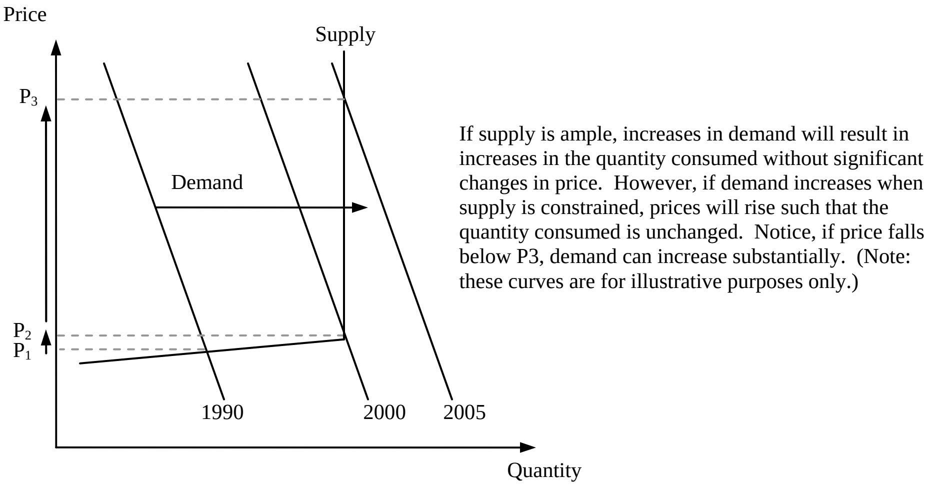
JENNIFER ROSTHAL
GRADUATE STUDENT, ECONOMICS DEPARTMENT

UNPUBLISHED

I. Introduction

Natural gas prices have recently increased to unprecedented levels due to a rapid increase in natural gas demand relative to maturing domestic production. While higher prices have tended to mitigate growth in actual consumption, the expansion of the gas-fired power generation fleet, in particular, has created a tremendous amount of latent demand that, if prices were to fall, could trigger a rapid increase in consumption. In fact, in the summer of 2006, a heat wave coupled with natural gas prices below parity with residual fuel oil spurred two weeks of withdrawals from natural gas storage. Such an event had never previously occurred and highlights the reality that the demand potential (the position of the demand curve, see Figure 1) for natural gas is enormous.

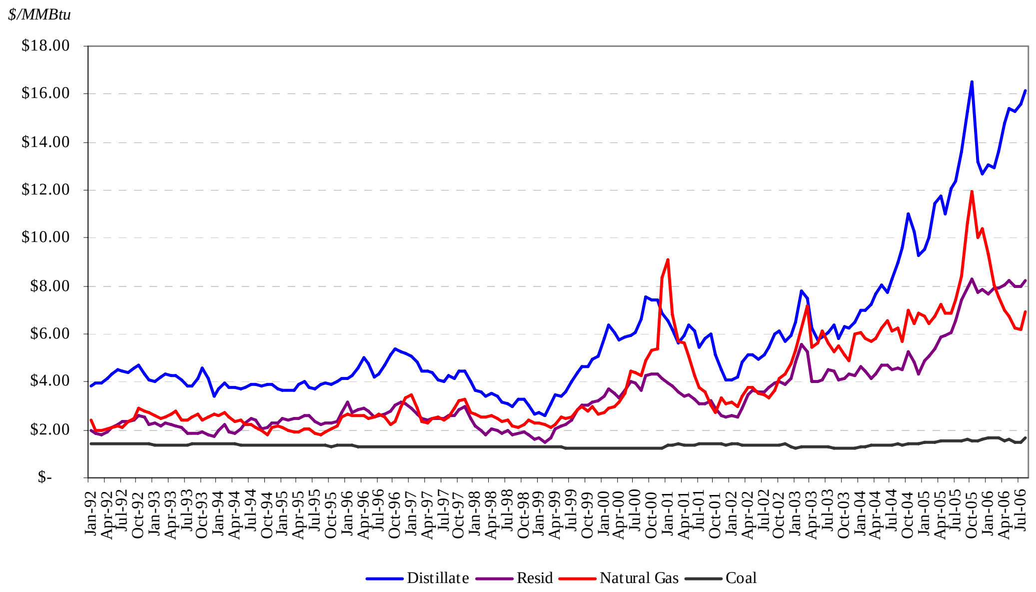
Figure 1: Natural gas demand and actual consumption



If supply is ample, increases in demand will result in increases in the quantity consumed without significant changes in price. However, if demand increases when supply is constrained, prices will rise such that the quantity consumed is unchanged. Notice, if price falls below P3, demand can increase substantially. (Note: these curves are for illustrative purposes only.)

As this paper will discuss, the nature of the relationship between natural gas prices and crude oil prices, in general, is critical in understanding the outlook for the future. Although the price of natural gas remained above $5 per million British thermal unit (MMBtu), which is high by historical standards, consumption increased dramatically. Of note, however, is the price of natural gas *relative* to the price of competing fuels. In particular, as noted above, natural gas prices dipped well below parity with the price of residual fuel oil (see Figure 2). This encouraged gas use in power generation, rather than residual fuel oil, when electricity demand spiked due to higher than average temperatures. This substitution in end use is a critical determinant in understanding the long and short run relationship between natural gas prices and crude oil prices.

Figure 2: Competing fuel prices

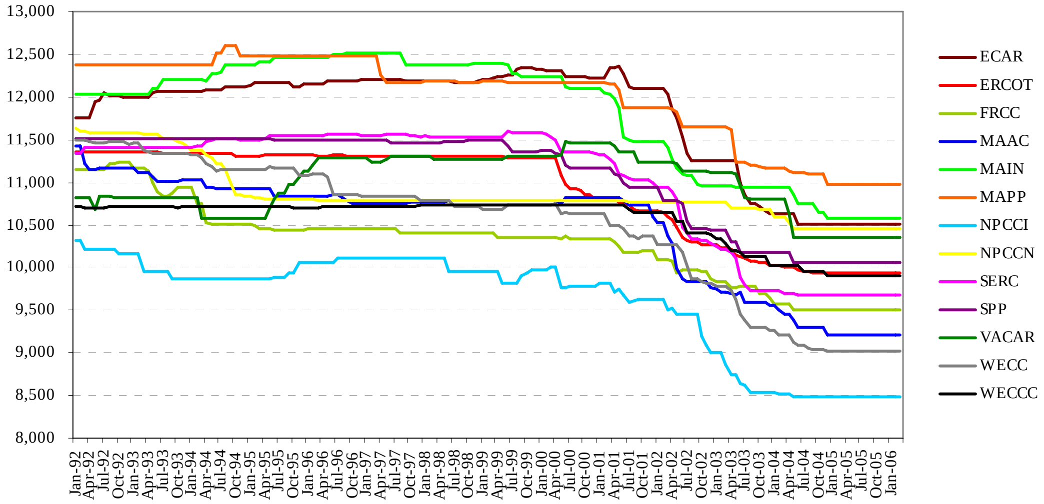


II. Natural gas demand for electricity generation

There are several reasons why substitutability between different fuels is higher in the electricity sector than in most end-uses in the industrial sector. In some cases, electricity generating plants can substitute fuel oil for natural gas at relatively low cost. More importantly, however, the relative position of different types of plants in the supply stack will change as fuel prices vary. Specifically, when natural gas costs are high relative to the cost of oil-fired electricity generation, natural gas generation will shift up in the supply stack. As a result of differing marginal costs, different plants will be used for different lengths of time within a day, which, in turn, can greatly affect the demands for different types of fuel. In particular, we would expect the demand for natural gas for electricity generation to depend on the price of natural gas and the price of fuels it most competes with on the supply stack. For combined-cycle plants, the competing fuel will likely be residual fuel oil, while for gas turbines, the competing plants will be fueled by diesel.

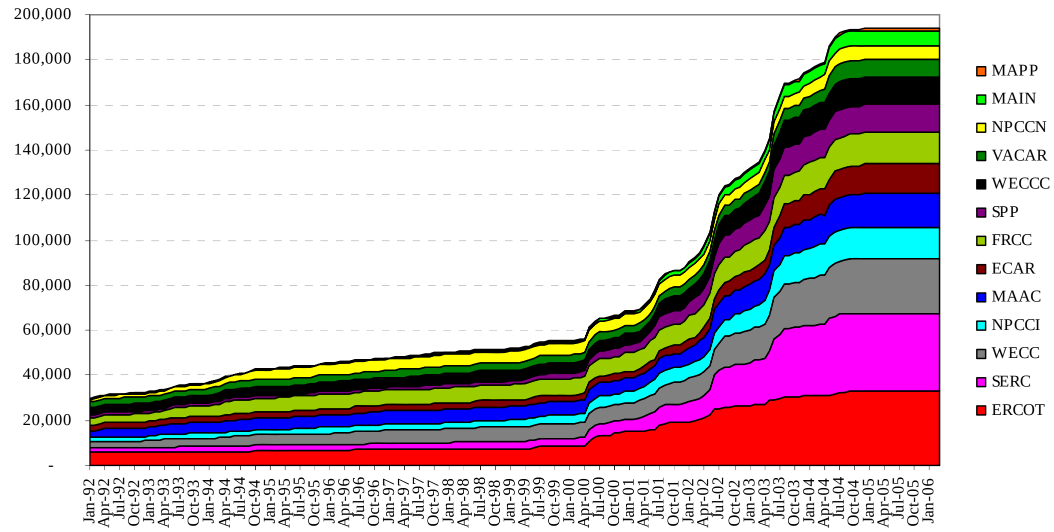
The aggregate time-series analysis shows that the technology available in the electricity sector plays a central role in maintaining the long run relationship between natural gas and crude oil prices (see “VECM.ppt” for more details). In particular, we find evidence that oil and natural gas prices tend to return to a long run relationship that depends on the relative heat rates for different types of electricity generating plants. Furthermore, historical data shows deviations in natural gas prices from that long run relationship induced changes in demand that tended to bring prices back into long run equilibrium.

Figure 3: Capacity-weighted average natural gas heat rates (Btu/kWh)



Given the aggregate results, a natural hypothesis is that the demand for natural gas to generate electricity responds negatively to divergences from the long run relationship between gas and oil prices. The aggregate results further suggest that the relative price differences need to be adjusted for changes in relative heat rates since changes in heat rates reflect the adoption of new technologies that change the relative costs of using gas or oil to generate electricity. The capacity-weighted average heat rate for natural gas fired generation capacity in each NERC region is depicted in Figure 3. One notable feature is decrease in gas heat rates, which is the result of rapid expansion in high efficiency natural gas fired combined cycle (NGCC) generation capacity (see Figure 4). Note that no such improvement in heat rates occurred for the oil-fired generation capacity (not pictured).

Figure 4: Natural gas combined cycle capacity (MW)



In the disaggregated analysis of the demand for natural gas for electricity generation we allow the relative cost of generating electricity using either gas or oil to affect gas demand. Specifically, for each NERC region, we form a capacity-weighted real cost of natural gas using the average electricity price as deflator

$$NGRCost_{it} = \frac{\sum_{j=1}^{P_i} K_{ij} HR_{ij} NGPrice_j}{ElecP_i \sum_{j=1}^{P_j} K_{ij}}$$

where P_i equals the number of gas-fired plants on line in NERC region or sub-region i in period t. The capacity of plant j is K_{ij} and its heat rate (obtained from the EPA NEEDS 2004 data) is HR_{ij} (see appendix 2 for more details). The formula above also allows for the possibility that the natural gas price is different for each plant. We use the state-specific city gate price reported by EIA for plants located in a given state.[1] This procedure allows electricity generation to adjust to persistent basis differentials between states with deviations from those differentials driving changes in demand. Similarly, the electricity price for region i is a weighted average of state electricity prices with the weights given by the proportion of overall generating capacity within the NERC region that is located in a given state.

The NERC region petroleum product prices were constructed in much the same way as the natural gas price. However, the same level of disaggregation was not available. Rather than using state-specific prices, the product prices are reported at the PADD level. The United States is divided into 5 PADD districts.[2] NERC region oil generation costs were then formed in a similar manner to the natural gas costs by multiplying product prices by plant heat rates and then forming a weighted average of the results in each region using generating capacities as the weighting variable.

We estimate a long run relationship between real natural gas and oil generation costs in each of the thirteen NERC sub-regions by regressing the logarithm of the real natural gas cost on the logarithm of the real oil cost

$$\ln NGRCost_{it} = \beta_0 + \beta_1 \ln OilRCost_{it} + \omega_{it} .$$

Because the terms are cointegrated, the resulting parameter estimates are superconsistent and the error term, ω_t, can be constructed as if it were known. Moreover, the error term is interpreted as the deviation from the long run equilibrium between the heat rate adjusted oil and gas prices. The deviation from the long run relationship is used in the regression analysis for estimating the demand for natural gas in each NERC sub-region.

---

[1] In some instances city gate price data were missing due to confidentiality restrictions. A regression analysis of the relationship between the average US city gate price and the non-missing values of the state city gate price was used to forecast the missing observations.

[2] As with for natural gas prices, missing values were interpolated using a regression of non-missing values on the US average price.

In effect, deviations in the long run relationship ought to affect the electricity generation fuel mix in such a way that subsequent price adjustments tend to bring the relative costs back into line.

Table 1: Cointegration of the real cost variables

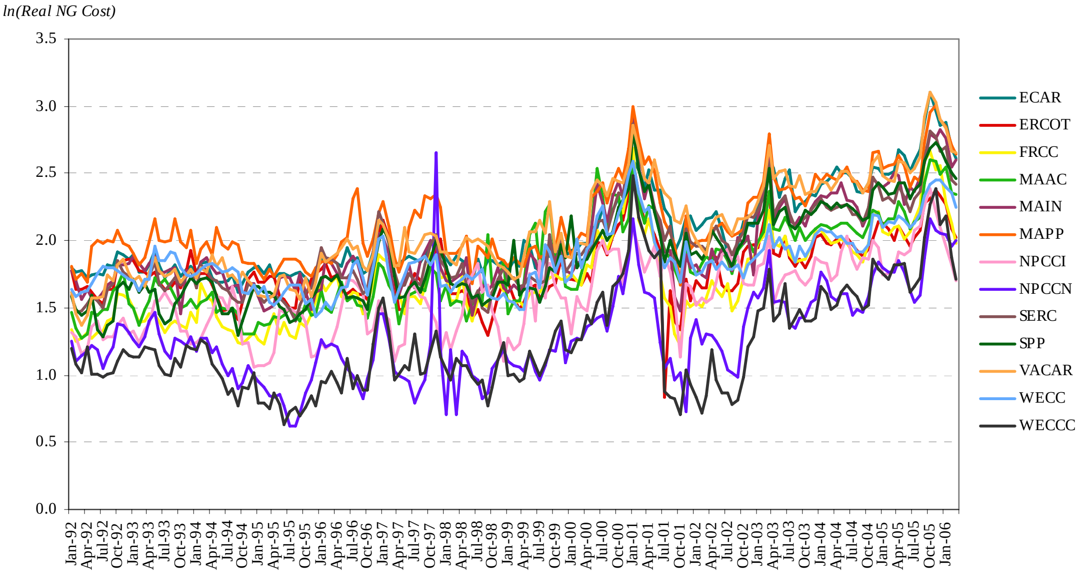
NERC sub-region	β_0	β_1	Test for *NGRCost* non-stationarity[a]	Test for *OilRCost* non-stationarity[a]	Test for error non-stationarity[a]
FRCC	0.050	0.896	0.081	0.708	0.000
VACAR	-0.075	0.961	0.427	0.906	0.001
MAAC	0.346	0.757	0.138	0.885	0.000
MAIN	-0.212	0.893	0.168	0.859	0.000
MAPP	-0.146	0.864	0.093	0.888	0.000
NPCCN	-0.054	0.918	0.003	0.790	0.000
ECAR	-0.252	0.973	0.587	0.913	0.000
SPP	-0.923	1.102	0.298	0.778	0.000
SERC	-0.523	0.967	0.222	0.849	0.000
WECC	-0.070	0.724	0.091	0.818	0.000
WECCC	-0.902	1.012	0.072	0.732	0.000
ERCOT	0.050	0.741	0.000	0.738	0.000
NPCCI	0.710	0.628	0.003	0.637	0.000

[a] MacKinnon approximate p-value for the null hypothesis that the variable is non-stationary.

There is evidence that the real natural gas cost variable is stationary in a few regions, especially ERCOT, NPCCI and NPCCN and, to a lesser extent, WECCC, WECC and MAPP. It may be the case that, in these regions, natural gas plants often determine the marginal cost of electricity. Thus, once we have corrected for heat rate changes, the relative price of natural gas to electricity does not appear to follow a trend. As the graph of the real cost variables below shows, however, the ratio of input cost to output price generally follows a similar trend in all regions. In the regions where we can reject the hypothesis of non-stationarity, however, the variability is much higher, making it much harder to detect any change in trend before and after 1999.

In addition, the evidence for stationarity of the real gas cost variable is less conclusive than might first appear to be the case. Certainly the p-values for the test of non-stationarity suggest the hypothesis can be rejected in three and perhaps five or six regions. On the other hand, contrary evidence is provided by the fact that the real oil cost variable appears non-stationary in all regions, while a linear function of the real natural gas and real oil costs is stationary in every region.

Figure 5: Real natural gas costs, Jan 1992-May 2006



Evidently, the real gas cost variable in each NERC sub-region contains a non-stationary component that cancels with a similar non-stationary component in oil costs. In addition, however, the real gas cost variable in some regions must contain some other (stationary) high variance component that makes it difficult to discern the non-stationarity that must be present in real gas costs.

The estimated equation for each NERC sub-region is given as[3]

$$\ln\left(-\ln NGConFrac_t\right) = b_0 + b_1\omega_t + b_2 \ln FossEgen_t + b_3 CDD + b_4 HDD + \sum_i Month_i + \varepsilon_t$$

---

[3] In WECCC (California), the equation included an indicator variable set equal to 1 for the period January through June of 2001 and 0 elsewhere. This allowed for departures from the estimated relationship during the California electricity crisis of that year. The period was characterized by large deviations in ω_i and the use of natural gas to generate electricity that do not fit the estimated patterns for remaining time periods.

where $\varepsilon_t = \rho\varepsilon_{t-1} + u_t + \theta u_{t-1}$, which allows for autocorrelation in the error terms.[4] Autocorrelation could arise for a number of reasons, including slow adjustment to changes in factors that affect natural gas demand that continue to alter demand in subsequent periods, such as contracting behavior. In addition, any important influences on the demand for natural gas that have been omitted from the equation would appear in the error term, and these influences could themselves be autocorrelated over time.

The dependent variable, the double log of the capacity factor for natural gas plant usage in a month ($\ln(-\ln NGConFrac_t)$), was so-specified for the following reasons.

- The minimum natural gas usage in a month is obviously zero, while the maximum usage is limited by the total natural gas-fired generating capacity in the NERC sub-region in a given month. Thus, we defined a maximum level of natural gas consumption for the month by calculating how much gas would be consumed if all available gas capacity were run for all hours of any given month. The ratio of actual natural gas consumed to generate power to this theoretical maximum level would then be a number constrained to lie in the [0, 1] interval.[5] Hence, the term $\ln NGConFrac$ will be negative and the logarithm of the negative logarithm will be well-defined and can take any real value.
- An advantage of this transformation is that it allows for an error term with classical properties as assumed by the statistical theory underlying the estimation of the equation and the hypothesis tests for statistical significance. If the dependent variable were constrained to lie in the unit interval, for example, the error terms in the equation would need to be bounded.
- A somewhat less technical justification for the functional form is that it ensures that the amount of natural gas input is bounded by the physical constraints of the system. No matter what values the independent variables on the right hand side of the equation take, natural gas usage cannot be predicted to lie outside the bounds of what is feasible.
- Finally, the functional form allows a non-linear response to changes in the determinants of natural gas demand that make intuitive sense given the way that the electricity system is operated in practice. Since combined-cycle electricity generation, conventional gas-fired steam generation and gas turbines each have different heat rates, they are used to supply power at different points on the load curve and thus for different amounts of time during the month. As total gas-fired generation increases, the most efficient plants are used first and the least efficient

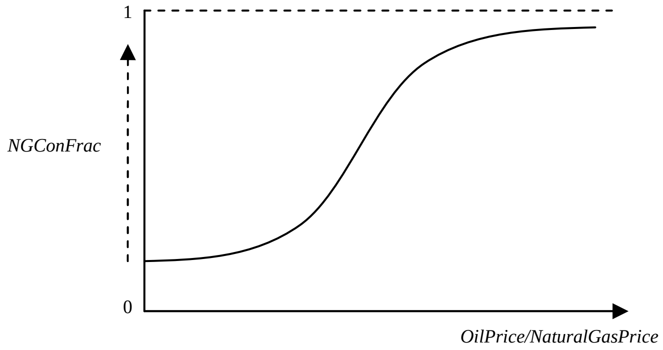
---

[4] We examined some other models for the error term, including second-order autoregressions and non-stationary specifications. However, allowing for first order autoregressive and first order moving average components appeared to be most satisfactory. We also examined models that included a lagged dependent variable as an alternative, or supplement to autoregressive and moving average structure in the error term, but again the model as written above proved most satisfactory.

[5] In practice, some natural gas is used to generate power in every NERC sub-region in every month, so the ratio is bounded above zero, ensuring that the logarithm of the ratio remains finite.

ones last. Gas demand can rise rapidly as many of the more efficient plants are brought on-line, but then will level off as the remaining smaller plants are added more gradually. This type of response is illustrated in the following Figure 6.

Figure 6: Response of *NGConFrac* to changes in oil price for a natural gas price



We turn next to the different determinants of demand specified on the right-hand side of the estimated equation. The first term, ω_i, by construction will be positive when real natural gas costs are above their long run relationship with real oil costs. If this were the case, we would expect a reduction in the demand for natural gas as oil-fired capacity is dispatched more extensively. Since $\ln(-\ln NGConFrac_t)$ decreases as $NGConFrac_t$ increases we should find $b_1 > 0$.

In general, we would also expect natural gas consumption to increase as total electricity generation from fossil fuels (ln*FossEgen*) increases as gas-fired plants would be part of the mix of plants called upon to meet peak demands. We use fossil fuel generation rather than *total* electricity generation as the determining variable because dispatch of a substantial amount of the non-fossil fuel generating capacity is essentially passive. Generation from wind turbines or run-of-river hydroelectric plants is determined by natural factors independent of the level of overall power demand or the cost of competing sources of power. Also, while the generation output from nuclear plants could in principle be varied in response to demand or cost variations, in practice such plants typically are operated at full capacity or not at all. Hydroelectric plants based on stored water (or pumped storage facilities) are dispatched on an economic basis and would compete with gas-fired plants. However, the key determinant of the dispatch decision in that case is the shadow value of the stored water (the marginal value of that same water in its next best alternative period of use) and we did not have sufficient data to estimate such costs. Hence, we treated all non-fossil generation as exogenous and looked at total demand net of such generation output. Obviously, we would expect gas demand for electricity generation to rise as ln*FossEgen* rises, implying that $b_2 < 0$.

The two weather variables are included for different reasons. A month that has a larger number of cooling degree days (*CDD*) will also have a higher demand for electricity to run air conditioning equipment. ln*FossEgen* will measure the higher monthly levels of electricity demand in such months; however, more extensive use of air conditioning will also change the shape of the load curve, emphasizing peaks compared to months with equivalent total overall demand for electricity but with a more temperate climate. Since gas turbines are called upon to provide peak power, we therefore expect a larger value of *CDD* to be associated with higher gas demand ($b_3 < 0$).

Months with a larger number of heating degree days (*HDD*) might also be associated with an elevated demand for electricity for heating purposes. This effect is not likely to be large, however, since providing space heating is not a significant factor in electricity demand. The motivation for including *HDD* is therefore somewhat different from the motivation for including *CDD.* Natural gas itself is a major source of space heating services on cold days. Local gas prices therefore are likely to be driven higher in months when *HDD* is large. Such higher prices will be reflected in the cost differential term ω_i. However, electric generating companies might also have gas supply contracts with interruptibility provisions that allow some quantitative reductions in gas demands for electricity generation when gas demand for heating purposes is high. If so, a large *HDD* value would be associated with lower gas use for generating power independently of any effects operating via higher prices. The two effects discussed here are offsetting in sign, so it is not clear *a priori* whether *HDD* would have a positive or negative coefficient, or even whether it would differ significantly from zero.

The monthly indicator variables (*Month*) capture multiple influences on demand. For one, there are different numbers of days in each month, so *all else equal* a month with 31 days should see greater gas demand than a month with 30 days. Second, the variable *Month* is also correlated with variations in weather. Hence, the effects of *CDD* and *HDD* should be interpreted as the marginal effects of departures of cooling or heating degree days from their normal monthly averages. Third, the monthly indicator variables will also capture seasonal regularities in gas price movements relative to oil. For example, since there are seasonal effects in gas price basis differentials, the cost differential term ω_i will vary by season. Any response of gas demand to these normal seasonal price fluctuations will be captured by the monthly indicator variables rather than the ω_i terms. Last, to the extent that certain generating facilities are taken off-line for maintenance, the monthly indicator variable will capture the resulting impact on gas demand.

In order to measure the sensitivity of natural gas demand to changes in each the individual variables, we can calculate the elasticity based on the estimated coefficient. For illustrative purposes, suppose we have a right hand side variable x measured in logarithmic form with estimated coefficient α. The relationship given as

$$\ln(-\ln y) = \alpha \ln x$$

implies $y = e^{-x^{\alpha}}$ so the estimated elasticity of response becomes

$$\frac{x}{y}\frac{dy}{dx} = -xe^{x^{\alpha}}\alpha x^{\alpha-1}e^{-x^{\alpha}} \;= -\alpha x^{\alpha}$$

Then, $\alpha < 0$ indicates a positive effect of variable x on the consumption of natural gas, but the elasticity decreases as x increases. When $\alpha > 0$, variable x has a negative effect on the consumption of natural gas that becomes more negative, but at a decreasing rate, as x increases.

The weather variables are measured in levels rather than logs. In that case

$$\ln(-\ln y) = \alpha x$$

implies $y = e^{-e^{\alpha x}}$ so the estimated elasticity of response becomes

$$\frac{x}{y}\frac{dy}{dx} = -xe^{e^{\alpha x}}\alpha e^{\alpha x}e^{-e^{\alpha x}} \;= -\alpha x e^{\alpha x}.$$

For the indicator variables, it makes little sense to measure an elasticity of response since the variable can only be either 0 or 1. Instead, if

$$\ln(-\ln y) = k + \alpha D\,,$$

where k is a constant and D = 0 or 1, we measure the ratio of natural gas consumption between the two cases (D = 0 or 1) as

$$\frac{y_{D=1}}{y_{D=0}} = e^{e^{k}(1-e^{\alpha})}\,.$$

A positive value of α makes $y_{D=1}$ smaller than $y_{D=0}$ and vice versa.

As an example to illustrate the elasticities, consider the estimated equation for the NERC sub-region MAIN (Mid-America):

$$\begin{aligned}\ln\left(-\ln NGConFrac_t\right) = 10.061 + 0.1358\omega_t - 0.5167\ln FossEgen_t \\ - 0.0020CDD_t - 0.0002HDD_t + \sum_i \gamma_i Month_i + \varepsilon_t\end{aligned}$$

$$\varepsilon_t = 0.9452\varepsilon_{t-1} + u_t - 0.3617u_{t-1}$$

The interpretation in terms of elasticity implies that when fossil fuel generation increases by one percent, the fraction of potential natural gas output that is actually used increases by $0.5617\ln FossEgen_t^{-0.5167}$ percent, holding all other influences fixed. Meanwhile, cooling and, to a lesser extent, heating degree days have positive effects on the consumption of natural gas.

Although they are not presented above, the coefficients on the monthly indicator variables imply that natural gas demand for electricity generation in MAIN is significantly higher in February through May and again in September through December than it is in January. Demand in June and August (but not July) is also estimated to be higher than in January, but the difference is not statistically significantly different from zero.

Finally, for ω positive, natural gas cost is high relative to the long run relationship with the oil cost of production. Therefore, this variable ought to have a negative impact on natural gas consumption, or positive coefficient, as is the case with MAIN. The estimated equation for MAIN thus implies that substitution between gas and oil-fired generation contributes to bringing gas and oil costs back into line when they deviate from their long run relationship.

The magnitude of the consumption response to ω varies greatly across regions with FRCC being the most sensitive to the deviations from the long run relationship. Using the cointegrating relationship that defines ω

$$\omega = \ln\left(\frac{NGRCost}{OilRCost^{\beta_1}} e^{-\beta_0}\right)$$

the elasticity in this case becomes

$$-\alpha\left(\frac{NGRCost}{OilRCost^{\beta_1}} e^{-\beta_0}\right)^{\alpha} = -0.1358\left(\frac{NGRCost}{OilRCost^{0.893}} e^{-0.212}\right)^{0.1358}.$$

Figure 7: Estimated response of gas demand to cost variations

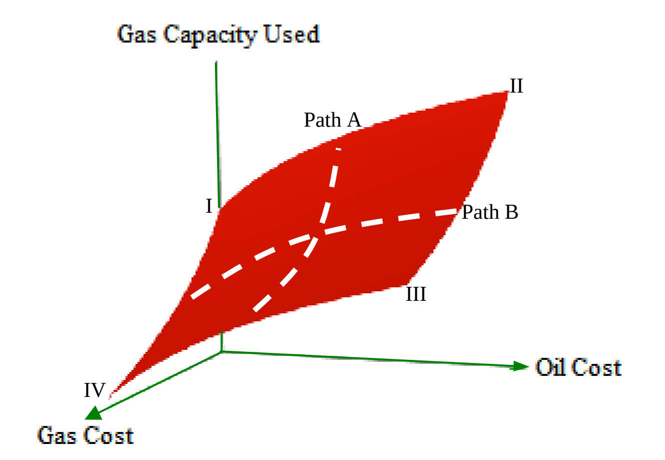


Figure 7 indicates the estimated response surface to variations in costs in the case of the MAIN sub-region (taking into account also the estimated cointegrating relationship between costs for that region). The graph has been drawn only for the range of cost variations actually observed in the MAIN region over the sample period. A decline in gas costs, holding oil costs fixed, leads to an increase in the use of gas capacity at an increasing rate (Path A). On the other hand, an increase in oil costs holding gas costs fixed leads to an increase in the use of gas capacity use at a decreasing rate (Path B). A consequence is that if prices moved from a region of high gas and low oil costs to one of low gas and high oil costs, there would be an S-shaped response of gas capacity use (along the diagonal connecting IV to II). The use of gas capacity would tend to rise quickly at first, then more slowly until we move toward the opposite corner of the region where gas capacity use increases more rapidly once again. This may reflect the ability to substitute different types of gas-fired capacity for oil-fired capacity at different relative costs. It must be stressed, however, that in practice most of the data lies in the vicinity of the other diagonal in Figure 7 (along the diagonal connecting I to III), which is precisely because gas and oil prices tend to return to a long run equilibrium, where prices move together.

Table 2: NERC sub-region estimation results

NERC sub-region	Variables							
	Constant	ω	ln*FossEgen*	*CDD*	*HDD*	AR(1)	MA(1)	*CAcrisis*
FRCC	5.8676*** (1.2289)	0.2997*** (0.0812)	-0.3179*** (0.0773)	-0.0006*** (0.0002)	-0.0006*** (0.0002)	0.9240*** (0.0377)	-0.2878*** (0.0937)	
VACAR	4.2145*** (1.5755)	0.2186** (0.1067)	-0.1802* (0.0977)	-0.0019*** (0.0004)		0.6172*** (0.0555)		
MAAC	5.0230*** (0.6536)	0.1727** (0.0719)	-0.2320*** (0.0404)	-0.0020*** (0.0004)		0.6863*** (0.0624)		
MAIN	10.0609*** (2.0583)	0.1358** (0.0551)	-0.5167*** (0.1251)	-0.0020*** (0.0003)	-0.0002*** (0.0001)	0.9542*** (0.0296)	-0.3617*** (0.0878)	
MAPP	1.5686*** (0.0657)	0.1079** (0.0505)		-0.0020*** (0.0002)	-0.0001** (0.00004)	0.5049*** (0.0726)		
NPCCN	4.0712*** (1.0092)	0.0934*** (0.0320)	-0.2002*** (0.0664)	-0.0008** (0.0003)		0.8403*** (0.0457)		
ECAR	11.0475*** (3.5495)	0.0477 (0.0750)	-0.5512*** (0.2023)	-0.0017*** (0.0002)		0.9613*** (0.0302)	-0.4453*** (0.0980)	
SERC	11.6386*** (1.3072)	0.0382 (0.0515)	-0.6116*** (0.0768)	-0.0006*** (0.0002)		0.9380*** (0.0285)		
SPP	13.7826*** (2.3196)	0.0313 (0.0721)	-0.7969*** (0.1400)	-0.0012*** (0.0002)		0.9487*** (0.0237)		
WECC	12.8011*** (1.8125)	0.0125 (0.0649)	-0.7045*** (0.1078)	-0.0013*** (0.0004)	-0.0002*** (0.0001)	0.8908*** (0.0355)		
WECCC	6.9219*** (0.1234)	0.0091 (0.0231)	-0.4153*** (0.0080)	-0.0003*** (0.0001)		0.9302*** (0.0412)		-0.1042*** (0.0249)
ERCOT	7.3650*** (0.5759)	0.0009 (0.0316)	-0.4000*** (0.0345)	-0.0007*** (0.0001)	-0.0002*** (0.0001)	0.9663*** (0.0262)		
NPCCI	8.7014*** (0.9621)	-0.0390 (0.1432)	-0.5146*** (0.0627)			0.8515*** (0.0332)		

*** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level. Statistically insignificant variables are reported in grayed font.

In addition to MAIN, five other regions had significant responses to deviations from the long run relationship between gas and oil costs relevant to their region: FRCC, VACAR, MAAC, MAPP, and NPCCN. These regions encompass the east coast from Florida north to New York and Pennsylvania and the Mid-Western states centered on Illinois, Wisconsin, Iowa and Minnesota. Three additional regions (ECAR, SERC and SPP) have positive and reasonably large responses to deviations in costs, although the coefficients are not statistically significantly different from zero. The coefficients in the remaining four regions (WECC, WECCC, ERCOT and NPCCI) are so small relative to their estimated standard errors that no meaning can be attached to the estimated values.

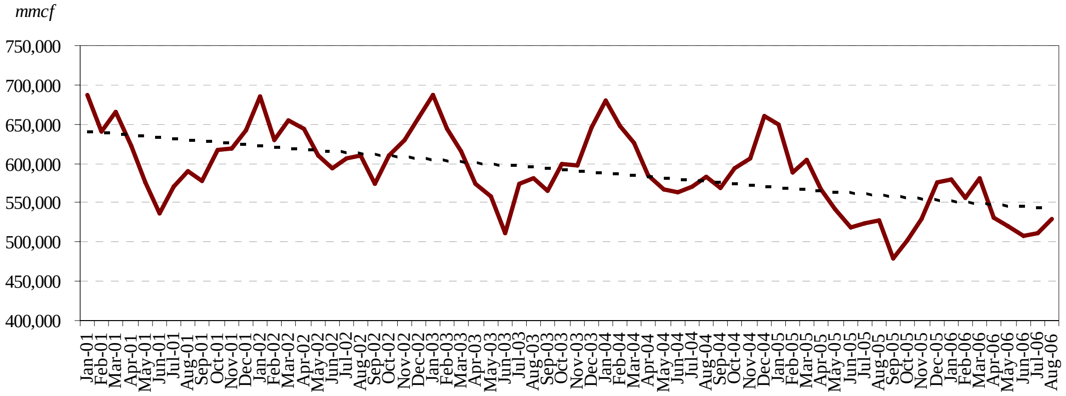
The ability to respond to the cost differences by varying the position of plants on the supply stack varies from one region to the next, which could explain why some regions exhibit stronger responses than others to movements in oil and gas costs from their long run equilibrium. It should be emphasized, however, that part of the estimated monthly effects could be a response of gas demand to seasonal and predictable relative price fluctuations, so the significant coefficients on ω in the above six regions may not be the only response directly aimed at maintaining relativity between gas and oil prices (adjusting for variations in heat rates).

All regions except MAPP had a strong and statistically significant response to changes in the quantity of fossil fuel powered electricity generation ($\ln FossEgen$), with SPP being the most responsive. MAPP does, however, show a significant response to departures of gas costs from the gas/oil cost long run equilibrium. All regions except NPCCI are also responsive to cooling degree days. The effect of heating degree days, when it is statistically significant, is of the same sign as the effect of cooling degree days but usually an order of magnitude smaller. Finally, all sub-regions had significant autocorrelation in the error term. This may indicate a lagged adjustment of demand to changes in driving factors, but it could also indicate that significant omitted explanatory variables are themselves autocorrelated. In three of the sub-regions, the error terms also displayed a significant moving average structure, which could reflect omitted explanatory variables that are correlated only over neighboring months.

III. Natural gas demand in the industrial sector

US industrial natural gas demand has been declining steadily in recent years (see Figure 8). While there is some ability to switch fuels in the face of short term relative price movements (such as to residual fuel oil for boiler fuel), anecdotal evidence suggests that much of the recent decline in natural gas demand is due to longer term forces. Higher prices have encouraged some industries to relocate offshore, where input costs are lower, and the final/intermediate product is then imported to the US.

Figure 8: US industrial natural gas demand, Jan 2001 – Aug 2006



The analysis of industrial natural gas demand is composed of two parts: (i) panel data analysis of monthly total industrial natural gas consumption by state, and (ii) analysis of monthly and daily natural gas deliveries to industrial end-users for a subset of interstate pipelines across the United States.

A. Monthly Total Industrial Consumption by State

The data, which was collected from the Energy Information Administration and the Federal Reserve Bank, covers a panel of all states except Alaska and Hawaii for the period June 2001 through April 2006. (The raw data, layout of the data file, and data sources are found in “EIA Industrial.xls”.) Many econometric specifications were tested, but the results indicated herein posit natural gas consumption (*NG*) to be a function of the real price of natural gas to industrial consumers (P^{NG}), the real price of residual fuel oil (P^{RESID}), the real price of electricity to industrial consumers (P^{ELEC}), an index of industrial production in natural gas intensive industries (IP^{NGINT}), heating and cooling degree days (*HDD* and *CDD*) monthly dummies (*Month*), and lagged natural gas consumption. The estimated equation was

$$\begin{aligned}\ln NG_{i,t} &= \alpha_i + \beta_1 \ln P_{i,t}^{NG} + \beta_2 \ln P_{i,t}^{RESID} + \beta_3 \ln P_{i,t}^{ELEC} + \beta_4 \ln IP_{i,t}^{NGINT} \\ &\quad + \beta_5 \ln HDD_{i,t} + \beta_6 \ln CDD_{i,t} + \gamma \ln NG_{i,t-1} + \sum_{n=2}^{12} \sigma_j Month_{j,t} + \varepsilon_{i,t}\end{aligned}.$$

In the estimation, we allow for the presence of panel-specific autocorrelation in the residuals. Demand is expressed million cubic feet and prices are expressed in real terms (2005$) where the PPI for the natural gas intensive industries was used to adjust for inflation in output prices.

The natural gas intensive industries are defined as in the National Petroleum Council’s 2003 study of long term North American natural gas markets:

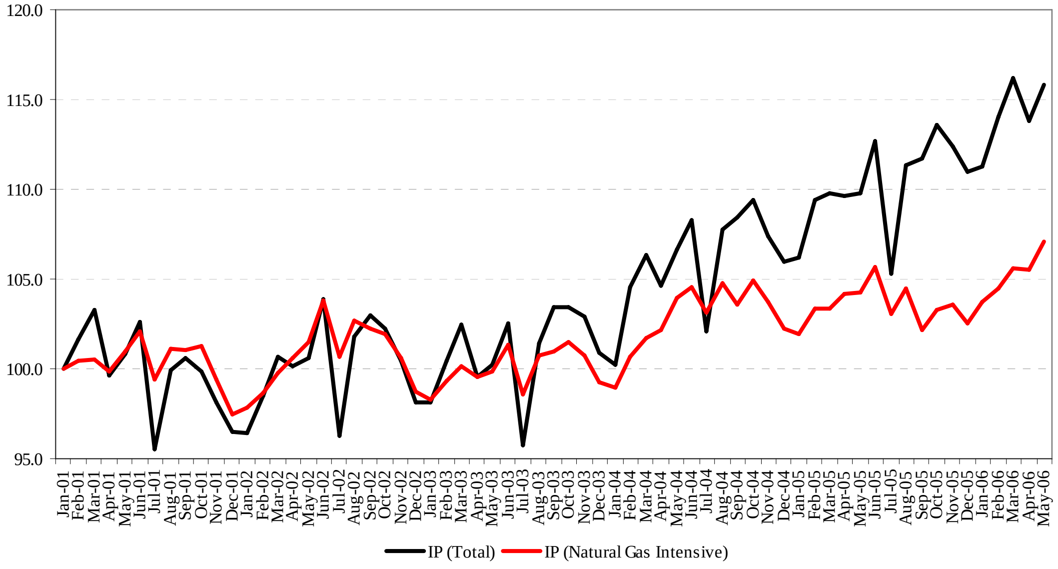
- Mining (NAICS 21),
- Food, Beverage, and Tobacco (NAICS 311,2)
- Paper (NAICS 322)
- Petroleum and Coal Products (NAICS 324)
- Chemicals (NAICS 325)
- Plastics and Rubber Products (NAICS 326)
- Non-metallic Mineral Products (NAICS 327)
- Primary Metals (NAICS 331)
- Fabricated Metals (NAICS 332)

Collectively, these industries account for over 80% of all natural gas consumption in industry. The index of industrial production in natural gas intensive industries and the PPI for the natural gas intensive industries are both constructed from the individual series associated with each of the above industries. The data were collected from the US Bureau of Labor Statistics and the US Census.

It is important to distinguish these industries from all other industry due to the fact that much of the growth in the general IP index over the past few years has been due to growth in industries that are not natural gas intensive, such as the computer industries. To highlight this point, Figure 9 shows the general IP index and the IP index for the natural gas intensive industries, each indexed to 100 in the initial time point. As can be seen, the time series are different, with the IP index for the natural gas intensive industries exhibiting less growth.

Theory suggests that a higher price of natural gas should reduce demand, but that a higher price of a competing fuel, such as residual fuel oil, should increase demand. The price of electricity could have a positive impact on natural gas consumption in industry due to switching for process heat and/or due to cogeneration of electricity. In addition, the level of industrial activity should increase demand, while the degree day variables could have a positive influence to the extent that space heating/cooling matters. The monthly dummy variable are defined for months 2 through 12, so they measure the monthly effect on average natural gas consumption for each of the remaining months relative to average January consumption.

Figure 9: Industrial production indices – total and natural gas intensive only



The lagged dependent variable permits a gradual adjustment of demand to changes in the included variables (due to investments in energy using capital or influences such as contracting behavior) and produces different long and short run effects. In particular, the estimated coefficients β_k are short run elasticities, while the long run elasticities can be calculated as $\beta_k/(1-\gamma)$.

Table 3: Parameter estimates – EIA state-level industrial demand

Variable	Parameter	Parameter estimate *(std. err.)*
P^{NG}	β_1	-0.08858^{***} *(0.01695)*
P^{RESID}	β_2	-0.03575 *(0.02484)*
P^{ELEC}	β_3	0.07733^{**} *(0.03287)*
IP^{NGINT}	β_4	0.56572^{**} *(0.26589)*
HDD	β_5	0.00018^{***} *(0.00003)*
CDD	β_6	0.00007 *(0.00006)*
NG_{t-1}	γ	0.52192^{***} *(0.01496)*
Feb	σ_2	-0.05120^{***} *(0.01533)*
Mar	σ_3	-0.01714 *(0.01691)*
Apr	σ_4	-0.03045 *(0.02240)*
May	σ_5	-0.04589^{*} *(0.02761)*
Jun	σ_6	-0.06140^{*} *(0.03511)*
Jul	σ_7	-0.02871 *(0.04000)*
Aug	σ_8	-0.00128 *(0.03929)*
Sep	σ_9	-0.02964 *(0.03208)*
Oct	σ_{10}	0.04343^{*} *(0.02413)*
Nov	σ_{11}	0.02043 *(0.01882)*
Dec	σ_{12}	0.02105 *(0.01555)*

*** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level. Statistically insignificant variables are reported in grayed font.

A Hausman specification test revealed that this fixed effects specification was preferable to alternatives that effectively require additional assumptions about the determinants of state differences in demand. In addition, because industrial activity is a function of the price natural gas, instrumental variables estimation was used. The instruments chosen for IP^{NGINT} were the current and lagged total IP index and current and lagged values of the other regressors. The parameter estimates (standard errors in parentheses) are given in Table 3.

The regression indicates that natural gas prices, electricity prices, industrial production, and heating degree days all have statistically significant effects on industrial natural gas consumption. Moreover, the sign of each parameter estimate is consistent with expectations:

- an increase in price of natural gas causes a decline gas consumption,
- an increase in the price of electricity causes an increase in gas consumption,
- an increase in output is associated with an increase in gas consumption, and
- cold weather tends to increase gas consumption in the industrial sector.

The negative own price elasticity indicates that industrial consumers will curtail their use of gas when gas prices increase, which is consistent with the idea that industrial consumers of natural gas are cost-minimizing. Thus, if gas prices were to increase enough, *holding all else constant*, we could expect industrial gas consumption drop to very low levels. Of course, in reality this would likely result in a decline in industrial production as well, which is precisely why we chose to use an instrumental variables approach in the estimation – to account for any simultaneity in the right hand side variables.

The positive elasticity of demand with respect to the electricity price can be attributed to substitution in process, but another likely explanation involves cogeneration. In particular, if electricity prices rise relative to gas prices, cogeneration facilities may consume more gas in order to generate electricity. This could be either for sale to the grid or for own use in order to reduce electricity purchases. Plant level data is currently being collected to investigate this conjecture.

Rising output also results in increased gas consumption. This is not surprising given that the measure of industrial production is for the gas-intensive industries. Thus, if these industries were to expand (shrink) their output, a high gas intensity should push them to increase their consumption of natural gas. Cold weather also tends to increase consumption, which is a result of space-heating requirements at industrial facilities.

The coefficient on the price of residual fuel oil is of the wrong sign, but is not significantly different from zero. The statistical insignificance of the cross price effects is indicative of a minimal response by industrial users of gas to changes in residual fuel prices. This is consistent with the notion that fuel switching from natural gas to residual fuel oil is not very prevalent in the industrial sector in the aggregate, at least at a monthly time increment.

B. Interstate Pipeline Deliveries to Industrial End-users

While the analysis of the EIA state-level industrial demand data is informative of broader trends in natural gas consumption, the data is not granular enough to provide insights into the nature of the trends seen in industrial demand over the past several years. In particular, a different set of data is necessary to determine if the declining trends in US industrial demand (see Figure 8) is endemic of all sectors within industry, or if it due only a few sectors. This is important when considering the future of natural gas consumption in industry.

We, therefore, used interstate pipeline delivery data (from *Energy Velocity*) to draw conclusions about industrial natural gas demand in the United States. The pipeline delivery data was grouped by 3-digit NAICS code, where the 3-digit code assigns each metering point to a major industry group. Four specific industries were chosen for analysis: Paper Manufacturing (NAICS 322), Petroleum and Coal Products Manufacturing (NAICS 324), Chemical Manufacturing (NAICS 325), and Primary Metal Manufacturing (NAICS 331). (See “Raw Pipeline Data.xls” for the data used in this analysis.)

After compiling the pipeline delivery data, we determined that interstate deliveries represent less than 10% of total U.S. industrial gas demand. Thus, we chose to analyze a subset of these data that appeared reasonably indicative of the trends observed in less granular data available from the US Energy Information Administration (EIA).

The data was aggregated by pipeline for both daily and monthly consumption for all industrial facilities as well as by industry group. While some of the pipeline data extend back to 2000, other data only extend back to 2004, effectively limiting our analysis to two years of daily data. These two years were found to be an inadequate representation of the total consumption in the United States both in the aggregate and by industry group. In order to have a longer, more representative time series, we narrowed the 21 pipelines down to four.

Two primary criteria were used to determine which pipelines were most representative of industrial natural gas consumption.

- The first criterion was whether the aggregate industrial consumption on a given pipeline mimicked the aggregate trends seen in the United States for total industrial demand for natural gas. The aggregate monthly data was collected from the EIA and spans the time period January 2000 to June 2006.
- The second criterion focused on the sectoral distribution of industrial natural gas consumption. Specifically, we identified a subset of the pipeline data that matched, as closely as possible, data from the Manufacturing Energy Consumption Survey (MECS) published by the EIA in 2002.

By satisfying both criteria, we hoped to produce a set of econometric estimates that would reflect the trends likely to be exhibited in a full nationwide sample.

Table 4: Regression results for examination of pipeline representativeness

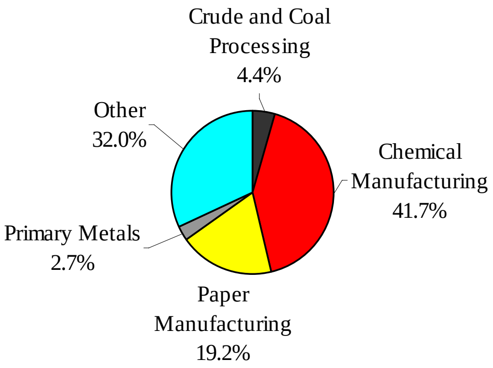
	Gulf South	Trunkline	NWPL	PNGTS	GTN	CGT
β_1	0.0075^{***} (0.0023)	0.0109^{**} (0.0051)	0.0753^{**} (0.0051)	0.1643^{***} (0.0457)	1.6955^{***} (0.5930)	0.0182^{**} (0.0087)
R^2	0.6905	0.6435	0.6466	0.6841	0.6620	0.6422

*** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level

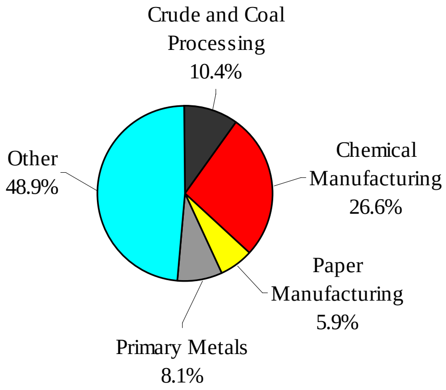
To determine the time series fit to the aggregate EIA data, the EIA data was regressed on the aggregate industrial demand from each pipeline individually. Only pipelines with a statistically significant (at the 20% level) and positive coefficient on the aggregate pipeline consumption were retained in the sample. Six pipelines qualified under this condition: Gulf South, Trunkline, Northwest Pipeline (NWPL), Portland Natural Gas Transmission System (PNGTS), Gas Transmission Northwest (GTN), and Columbia Gas Transmission (CGT).

Figure 10: Comparison of pipeline data to MECS 2002 data

Pipeline Data



MECS 2002 Data



The second step compared the industry composition data on different aggregates of these six pipelines with the sector breakdown from MECS. The sum of Gulf South, Trunkline, NWPL, and PNGTS was revealed to have sector level data that best represented the nationwide end-use sector composition of demand in 2002. One caveat is that the sector compositions are not an exact match. For example, the chemicals sector is slightly over-represented in the pipeline data while the paper and refining sectors are somewhat under-represented. This is illustrated in Figure 10. Another potential problem of unknown importance is that the pipeline data may not be representative of the distribution of demand within the sectors we have examined. For example, the entire US chemical sector may have more fertilizer production than indicated in the chemical sector represented by our subset of pipeline data.

Regression analysis was again used to determine if a simple sum of the four pipelines rather than differential weightings on each pipeline was a more appropriate proxy for total industrial demand. The EIA monthly data was regressed on all four pipelines simultaneously. We found that we could not reject the hypothesis that the sum of the pipeline data is an appropriate metric for representing the time series aspects of the aggregate industrial data. Thus, we used the sum of industrial demand on Trunkline, Gulf South, PNGT, and NWPL in each of the four sectors to examine how the price of natural gas and alternative fuels affected industrial demand in each sector.

i. Pipeline delivery analysis – Monthly industrial demand

In all cases, we begin with a specification similar to that used in the analysis of the EIA state-level data with all variables similarly defined. However, because demand is specific to a few industrial end-users, the issue of simultaneity bias is muted. Specifically, the aggregate measure of industrial production turns out to have no explanatory power in the regression analyses for the pipeline data. A more suitable indicator of activity would be desirable, but output data for specific industrial plants was not collected. In general, we estimate equations of the form:

$$\ln NG_t = \alpha + \beta_1 \ln P_t^{NG} + \beta_2 \ln P_t^{RESID} + \beta_3 \ln P_t^{ELEC} + \beta_4 \ln P_t^{COAL} + \beta_5 \ln IP_t^k$$

$$+ \gamma \ln NG_{t-1} + \sum_{n=2}^{12} \sigma_j Month_{j,t} + \varepsilon_t$$

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \mu$$

where the error structure allows for autocorrelation in the residuals.

The industrial production index used in each sector regression is the measure specific to that industrial sector, *k*. For total demand, we use the weighted average IP index for the natural gas intensive industries (as in the EIA state-level industrial demand analysis). Prices are expressed in real 2005$ with the inflation adjustment made using the PPI specific to sector *k*. Coal prices are included in the regressions for total demand and for primary metals, in order to capture the potential for substitution through increased use of metallurgical coal in metals manufacturing. The estimation results for each sector are summarized in Table 5.

For *Total* demand, we see that the results are similar to the results obtained in the analysis of the EIA state-level industrial demand data with regard to the influences of the price variables, but different with regard to the influence of industrial production. Specifically, there is evidence of a strong negative own price elasticity, a strong positive influence of the electricity price, and little evidence for fuel switching to residual fuel oil. Moreover, the industrial production index is negative but not statistically different from zero, which could be due to the fact that the pipeline data are regional whereas the industrial production data are national.

The econometric results for the four sectors identified for analysis are interesting in the sense that they may inform us more about what is NOT in the data than what is. For example, the lack of a significant response to the sector-specific industrial production index (or significant and negative in the case of chemicals) could be the result of the fact that those indices are national level indicators of output but the pipeline data represent regional consumption. With regard to the chemical sector, if industrial gas demand has been falling in the Gulf Coast region due to a large number of chemical manufacturers closing operations, then the chemicals IP index for the Gulf Coast states should be falling as well. However, the national level chemicals IP index has not generally been falling (see Figure 11), so the appropriateness of this measure for regional activity, and hence explanatory power, is questionable.

Table 5: Pipeline regression results (monthly)

		Sectors				
Variable	Parameter	Total	Primary metals	Refining	Paper	Chemicals
P^{NG}	β_1	-0.1805** (0.0902)	0.1184 (0.0913)	-0.5841* (0.3501)	-0.1937*** (0.0760)	-0.4100*** (0.1653)
P^{RESID}	β_2	0.0039 (0.1077)	-0.0897 (0.1185)	0.2930 (0.5902)	0.0817 (0.0920)	0.0269 (0.2192)
P^{ELEC}	β_3	0.5947** (0.3019)	0.1016 (0.2115)	0.0596 (0.2086)	-0.3860 (0.2854)	0.2130 (0.6229)
P^{COAL}	β_4	-0.1590 (0.1124)	0.1329 (0.1304)	---	---	---
IP	β_5	0.1483 (1.3895)	0.3553 (0.4325)	-0.2113 (1.6353)	-0.6162 (0.5346)	-2.1414** (1.0976)
NG_{t-1}	γ	0.7577*** (0.0888)	0.5663*** (0.1251)	0.3803*** (0.1471)	0.8527*** (0.0618)	0.7075*** (0.0889)
Feb	σ_2	-0.1272** (0.0650)	-0.0726 (0.0690)	-0.1566 (0.1742)	-0.0260 (0.0513)	-0.1543 (0.1515)
Mar	σ_3	0.0071 (0.0572)	0.0329 (0.0665)	-0.0665 (0.1862)	0.0088 (0.0467)	-0.0508 (0.1272)
Apr	σ_4	-0.1071 (0.0653)	-0.1154 (0.0675)	0.0263 (0.2054)	-0.0495 (0.0516)	-0.1334 (0.1400)
May	σ_5	-0.0801 (0.0709)	-0.0306 (0.0658)	-0.1052 (0.2273)	-0.0396 (0.0530)	-0.0877 (0.1427)
Jun	σ_6	-0.1933** (0.0907)	-0.1733** (0.0741)	-0.2484 (0.2603)	-0.0807 (0.0625)	-0.2146 (0.1606)
Jul	σ_7	-0.1307* (0.0759)	-0.1325* (0.0706)	-0.3123 (0.2505)	0.0473 (0.0658)	-0.1476 (0.1670)
Aug	σ_8	-0.0913 (0.0870)	-0.0485 (0.0726)	-0.2812 (0.2511)	0.0178 (0.0657)	-0.0620 (0.1640)
Sep	σ_9	-0.1272* (0.0751)	-0.0323 (0.0698)	-0.2366 (0.2252)	-0.0432 (0.0611)	-0.1846 (0.1501)
Oct	σ_{10}	-0.1241* (0.0711)	-0.0730 (0.0677)	-0.2511 (0.2131)	0.0886 (0.0592)	-0.1425 (0.1425)
Nov	σ_{11}	-0.0228 (0.0633)	-0.0821 (0.0626)	-0.0743 (0.2110)	0.0234 (0.0487)	-0.1344 (0.1270)
Dec	σ_{12}	0.0053 (0.0652)	-0.0449 (0.0681)	0.0550 (0.2058)	0.0517 (0.0560)	-0.0232 (0.1539)
Constant	α	4.5149 (6.3442)	4.6089** (2.1987)	7.8287 (6.9191)	3.2044 (2.2391)	13.8093*** (5.0663)
	ρ	-0.2793	-0.1296	0.0608	-0.1540	-0.3179
Adj. R^2		0.9969	0.8648	0.2159	0.9701	0.9569

*** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level. Statistically insignificant variables are reported in grayed font.

All sectors exhibited a statistically significant and negative own price elasticity, with the exception of the primary metal manufacturing sector. However, none of the sectors exhibited any significant response to the other prices. Among other things, this indicates that cogeneration facilities, while included in the total industrial demand data, have been separated from the sector-specific data, even if the facility is at a chemical plant, for instance. This is important because it masks the influence of electricity prices on plant level decisions regarding the consumption of natural gas for power generation for own use versus purchasing power from the grid. The data, so reported, is a by-product of the manner in which the data was coded by *Energy Velocity*.

The lack of a response to the prices may also be attributable to the use of national level prices when demands are regional. In particular, if the prices in a particular region are more volatile on a monthly basis than the national average price, then demand fluctuations in response to such movements will not be accounted for in the regression. As such, a large portion of the unexplained error could be due to a misspecification. One way to test this would be to incorporate data for plant-specific pricing into the analysis.

ii. Pipeline delivery analysis – Daily industrial demand

The pipeline delivery data that was provided by *Energy Velocity* is in daily time increments. This gives us a fairly large set of time series data that can be used to provide some insight into the manner in which industrial consumers respond to daily price movements. However, we are also limited in that we cannot include the same set of variables that are included in the monthly time series analysis because those data are not available on a daily basis. Thus, to control for variables that are not observable in daily increments, we examine time series models with an error structure that allows for unobservable influences. We estimate equations of the general form:

$$\ln NG_t = \beta_0 + \sum_k \beta_k \ln P_t^k + \sum_{j=1}^{n} \gamma_j \ln NG_{t-j} + \varepsilon_t$$

$$\text{such that } \varepsilon_t = \rho_1 \varepsilon_{t-1} + \ldots + \rho_p \varepsilon_{t-p} + \mu_t + \theta_1 \mu_{t-1} + \ldots + \theta_q \mu_{t-q}$$

where the error term, ε, is an ARMA(p,q) process. Natural gas demand is estimated to be a function of prices for k fuels and n lags. The estimation results are given in Table 6.

Table 6: Pipeline regression results (daily)

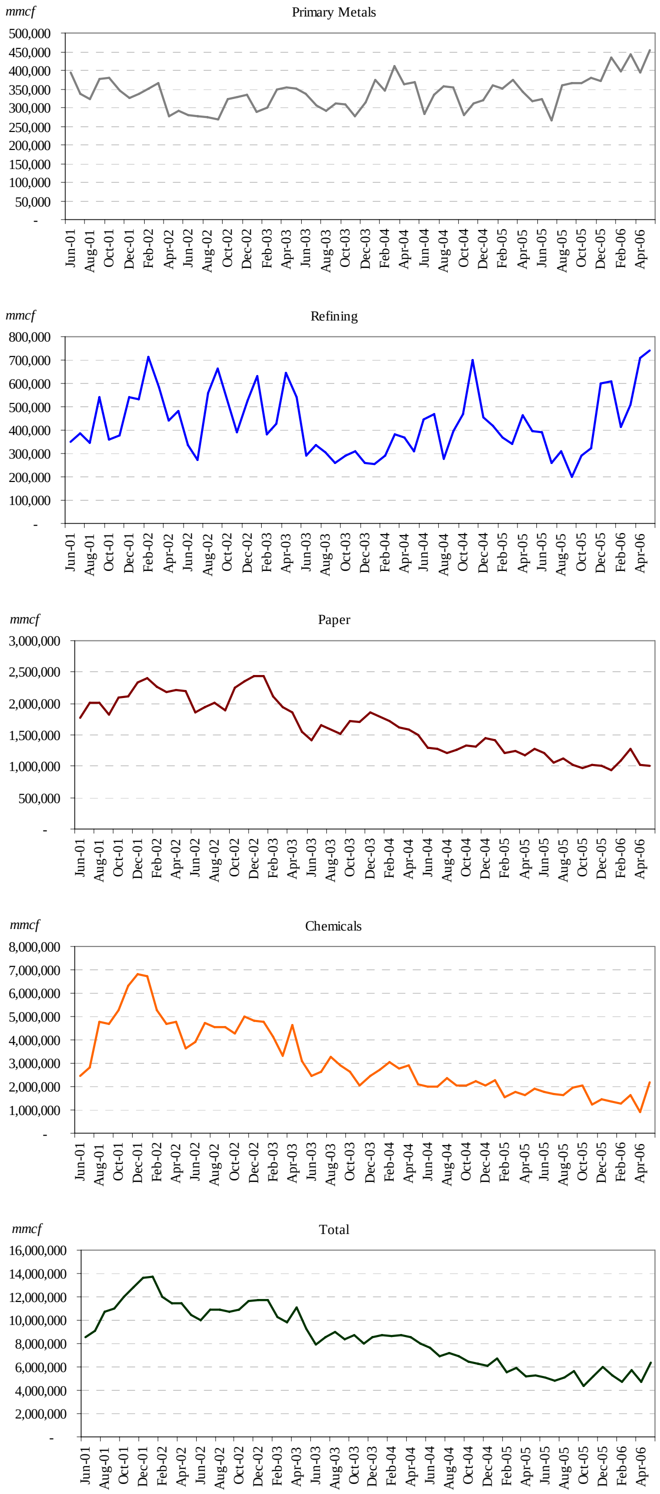
		Sectors				
Variable	Parameter	Total	Primary metals[a]	Refining	Paper	Chemicals
P^{NGHH}	β_1				-0.0040^{**} (0.0019)	-0.0134^{***} (0.0038)
P^{RESID}	β_2		-0.0099 (0.0076)			
P^{WTI}	β_3				-0.0057^{*} (0.0036)	
P^{NGHH}/P^{WTI}	β_4			-0.0621^{***} (0.0205)		
P^{NGHH}/P^{RESID}	β_5	-0.0020^{***} (0.0008)				
NG_{t-1}	γ_1	1.4229^{***} (0.0191)	1.3167^{***} (0.0729)	0.8850^{***} (0.0130)	0.8828^{***} (0.0215)	0.9198^{***} (0.0253)
NG_{t-2}	γ_2	-0.2714^{***} (0.0255)	-0.3715^{***} (0.0593)		0.1045^{***} (0.0227)	0.0674^{***} (0.0251)
NG_{t-3}	γ_3	-0.1527^{***} (0.0192)				
NG_{t-4}	γ_4					
Constant	β_0	0.0142^{**} (0.0074)	0.00001 (0.00001)	1.0739^{***} (0.1208)	0.1541^{***} (0.0619)	0.1672^{***} (0.0495)
AR1	ρ_1			-0.0753^{***} (0.0239)	0.6763^{***} (0.0344)	0.7454^{***} (0.0279)
AR2	ρ_2			-0.1041^{***} (0.0250)		
AR3	ρ_3			-0.0657^{***} (0.0270)		
MA1	θ_1	-0.9337^{***} (0.0119)	-1.6336^{***} (0.0705)		-0.9167^{***} (0.0210)	-0.9352^{***} (0.0166)
MA2	θ_2		0.6380^{***} (0.0691)			
	σ	0.0882^{***} (0.0008)	0.0939^{***} (0.0007)	0.2561^{***} (0.0030)	0.0690^{***} (0.0005)	0.1379^{***} (0.0013)

a – variables included as first differences (ARIMA model); Statistically insignificant variables are reported in grayed font.
*** indicates significance at the 1% level, ** indicates significance at the 5% level, and * indicates significance at the 10% level.

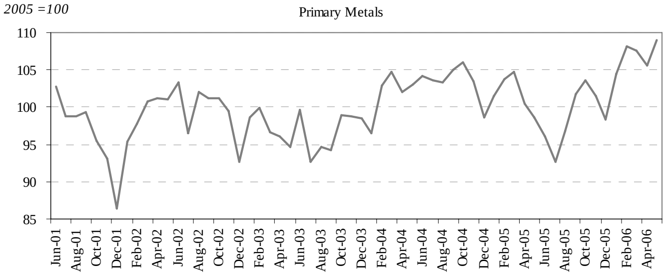
The chemicals sector appears to respond the most to changes in the price of natural gas on a daily basis while refining has the most significant response to the ratio of prices. However, with the exception of refining, all other industries appear to be more responsive to lagged natural gas consumption, perhaps due to considerable delays in adjustments to daily factors, which could be due to contracting behavior.

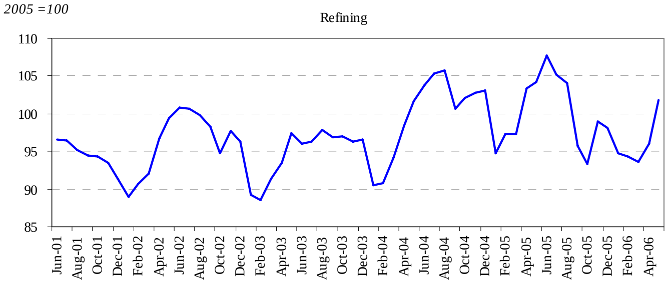
Figure 9: Data plots for monthly pipeline analysis

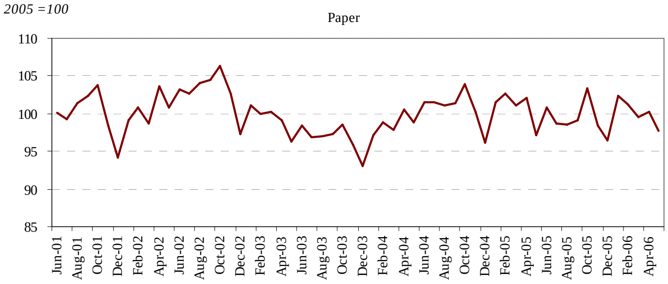
Natural gas demand

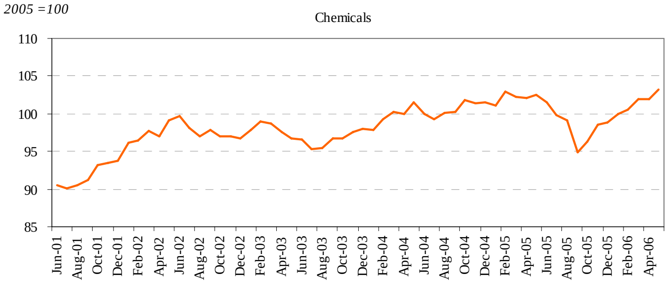


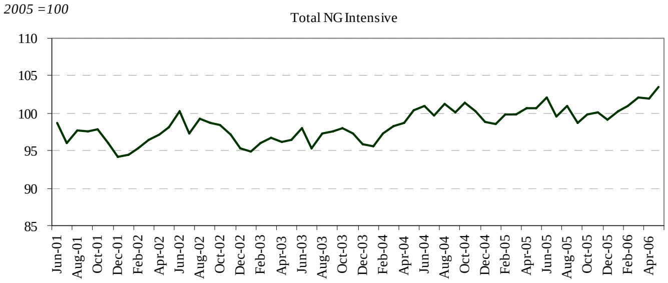
Industrial production



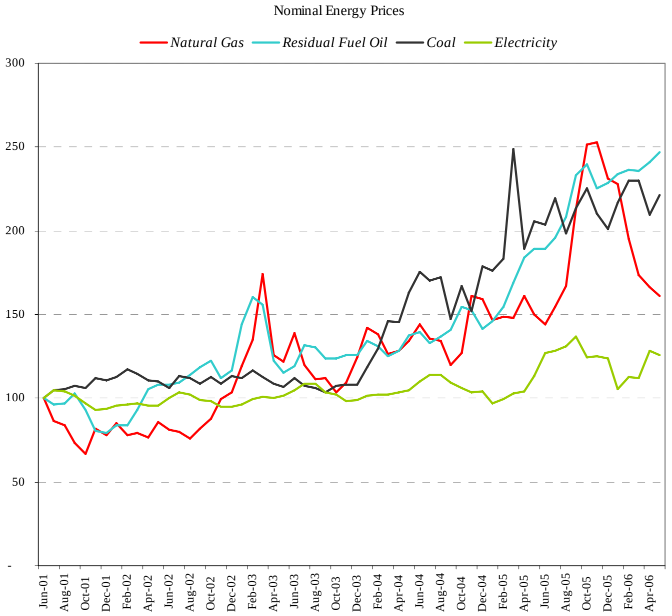


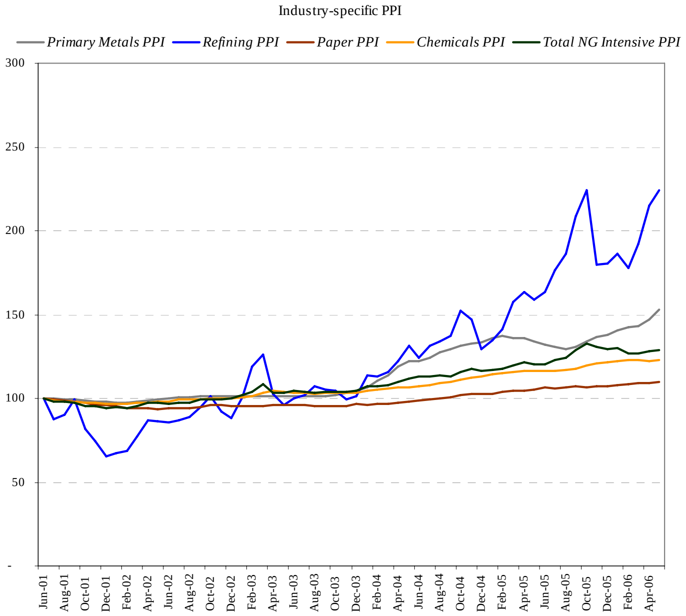




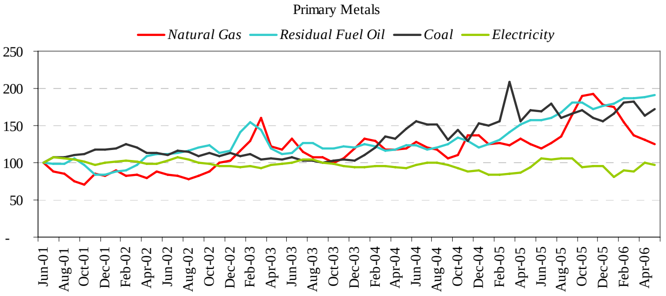


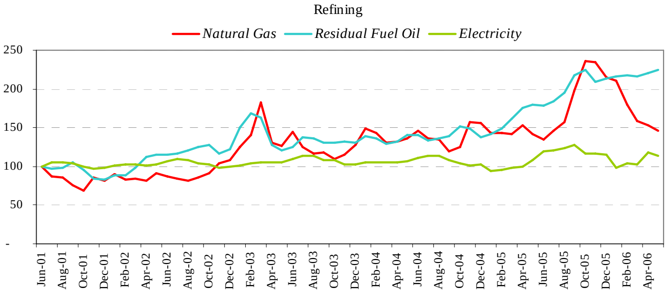
Energy prices and industry-specific PPI (indexed to Jun 2001 = 100)

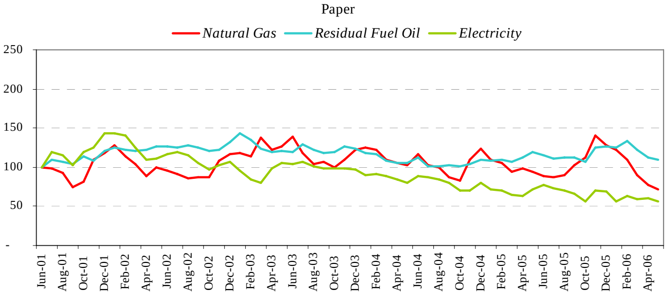

Nominal Energy Prices

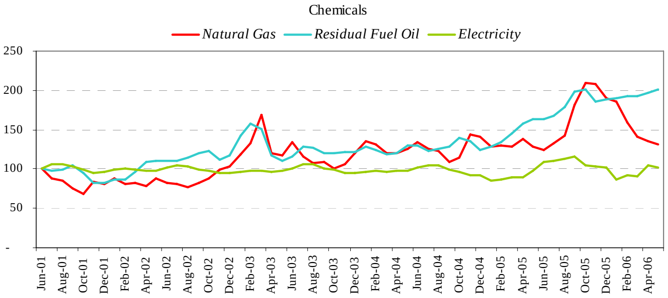

Industry-specific PPI

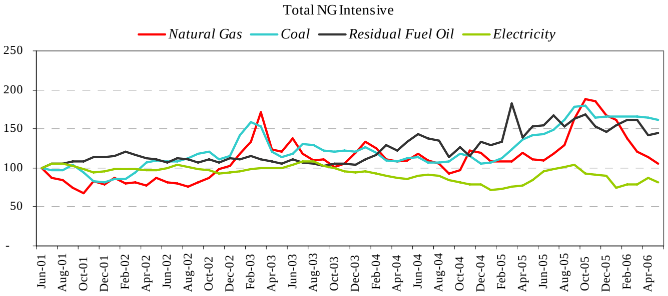
Industry-specific real energy prices (indexed to Jun 2001 = 100)











III. Concluding remarks

As stated at the outset, a primary reason for analyzing natural gas demand in the industrial and power generation sectors is to better understand the demand-side factors that drive a relationship between crude oil prices and natural gas prices. It is apparent that there is considerable influence coming from the power generation sector. We found that positive deviations from the long run relationship of the cost of using natural gas to generate electricity relative to the cost of using petroleum products exert a significant negative effect on natural gas demand in power generation. Moreover, the effect is generally largest in NERC regions where switching capability within the transmission grid is largest. This is important as it establishes a force that drives crude oil prices and natural gas prices back to a long run equilibrium that is dependent upon technology.

The industrial sector, as such, does not appear to *directly* contribute to the long run relationship between crude oil prices and natural gas prices because there is little evidence of fuel switching. We did, however, find evidence in the industrial sector that higher gas prices generally reduce demand. This effect would serve to alter the economy-wide responsiveness of demand to high gas prices, but, alone, would not restore any long run relativity between commodity prices.

There is, however, evidence that electricity from cogeneration serves an arbitrage role. If the electricity price rises, we found that this exerts a positive influence on natural gas consumption in industry. This could be due to increased generation from cogeneration facilities for either own use or sales to the grid. In either case, the effect would be to reinforce the relationship between natural gas prices and other competing fuels through the power generation sector. Thus, the linkage between fuel prices appears to be established mainly through the power market.

Appendix 1: Description of the Electricity Data

Capacity Weighted Heat Rates

The capacity weighted heat rates were determined based on the heat rates at each facility as given in the EPA NEEDS 2004 data. The heat rates in the EPA data were matched to the facilities listed in the EIA Form-860 (Annual Electric Generator Report) in four steps.

Step 1: A Unique ID number was created for each generator at each facility. This ID number consisted of the Facility ID and the Generator number. These two components were available in both the EIA and EPA datasets and for any plant where there was an exact match of facility and generator number, the reported heat rate was matched to the EIA data.
Step 2: The plant in the EIA database was matched to the plant in the EPA database that had the same Facility ID number, year of first use, prime mover, and fuel type.
Step 3: The average heat rate of similar facilities (based on prime mover type and fuel type and year of initial use) was used for the facility.
Step 4: The average heat rate of all plants with the same fuel type and prime mover.

If matching was accomplished in Step 1, only the remaining plants in the database were subjected to Step 2. As plants were matched in each step, the number of remaining unmatched plants dwindled until the Step 4, which is the least precise metric.

The capacity weighted heat rates were calculated each month based on the capacity that was online during that month. Thus, if a plant began operations in a particular month it was included in that month's heat rate calculation. The formula used for calculating the capacity weighted heat rate (*CapWtHR*) is:

$$CapWtHR_t = \frac{\sum_i (Capacity_{i,t} * HeatRate_{i,t})}{\sum_i Capacity_{i,t}}$$

where i = any plant in the specified NERC region at time t.

Capacity weighted heat rates are included for five groups – Coal, DFO, RFO, Total Oil, and Natural Gas. The RFO and DFO calculations were done separately by NERC sub-region and then a weighted average of them was calculated based on the capacity of RFO and DFO in the region. The EIA database was used to perform the calculations once the heat rates were determined using the EPA data.

It is important to note that heat rates are not available for all facilities. Those that have no heat rate published in the EPA and EIA data were not used in the heat rate calculations. Specifically, those plants that are powered by geothermal, hydro, or other fuel sources that are not necessary for the electricity demand analysis are not included in the heat rate calculation.

The EIA database provides as many as six energy sources for any one generator. For the heat rate calculations only the primary energy source was considered. However, the formatting of the file allows the user to include secondary energy sources in the heat rate calculations (see “HR Capacity Layout.xls”).

Natural Gas Consumption

EIA Forms 906 and 920 spanning the years 1986 through 2006 (found in file “Generation Data.xls”) report the total energy consumed by fuel type for electricity generators. Some modifications to the raw data were necessary in order to combine the data over the time period due to structural and formatting changes in the reports over the years.

1) Pre-2001 data include only the physical quantity of fuel consumed (bbl, mcf, tons), but do include neither the heat content of the fuel consumed nor the total energy content of fuel consumed (MMBtu). This problem was resolved by using the average heat content for each specific fuel type (‘Reported AER Fuel Type’) by state in 2001, and applying that heat content at each plant in that state that used that fuel type. This was then used to calculate the total energy consumed for electricity generation for that plant.
2) Prior to 1997 FRCC was not a separate NERC Region and thus did not appear in the dataset. Based on the Facility ID number, which remains constant over time, plants before 1997 were matched to facilities in later years to determine if they were in FRCC after its creation. Any plant that appeared prior to 1997, but not after 1997, and was located in Florida was assumed to be in FRCC. This allowed the construction of a longer time series for FRCC and SERC that was consistent throughout the time horizon.
3) The data were organized by reported NERC region with the exception of the aforementioned FRCC/SERC modification and the sub-regions specified for the analysis, the sub-regionalization of NPCC into NPCCN (any plant in NPCC that is located in NY) and NPCCI (any plant in NPCC not in NY), the distinction of

VACAR (a sub-region of SERC located in VA, SC, NC), and the distinction of California from the rest of the WECC.

4) Facilities that reported negative fuel consumption or electricity generation were not included in the dataset to eliminate those facilities that are either purchasing electricity or selling their fuel supply.

After the above modifications were made to the dataset, natural gas consumption (defined as MMBtu/month) was summed by month in each NERC region/sub-region. The data were not adjusted for the number of days in the month.

Natural Gas Price

Natural gas prices for each NERC region were constructed using state-specific city gate prices reported by EIA. The NERC region natural gas prices are a capacity-weighted city gate price, determined as in the following equation:

$$NGPrice_{i,t} = \sum_j \alpha_{j,t} NG\,\mathrm{Pr}\,ice_{j,t}$$

where:

$\alpha_{i,t}$ = Percent of total capacity in NERC region i that is in state j at time t

$NG\,\mathrm{Pr}\,ice_{j,t}$ = City gate price in state j at time t

In some instances, data were missing. Thus, the missing state city gate price was constructed based on a regression analysis of the relationship between the average US city gate price and the non-missing values of the state city gate price.

Residual Fuel Oil and Distillate Prices

The NERC region petroleum product prices were constructed in much the same way as the natural gas price. However, the same level of disaggregation was not available. Rather than using state-specific prices, the product prices are reported at the PADD level. The United States is divided into 5 PADD districts. The formula used to determine the NERC region prices is:

$$Price_{i,t} = \sum_j \alpha_{j,t}\,\mathrm{Pr}\,ice_{j,t}$$

where:

$\alpha_{i,t}$ = Percent of total capacity in NERC region i that is in PADD j at time t

$\mathrm{Pr}\,ice_{j,t}$ = PADD j price at time t

Any missing data was constructed in the same manner as described above for natural gas prices – missing values were interpolated using the regression of non-missing values on the on the US average price.

Natural Gas Combined Cycle Capacity

The capacity of natural gas combined cycle (NGCC) facilities is based on the prime mover characterization given in EIA Form 860, which results in any generator marked CA, CT, CS, or CC (see "HR Capacity Layout.xls" for descriptions) being characterized as NGCC. In addition, only if natural gas is reported to be the primary energy source is the facility considered to be part of the total natural gas combined cycle capacity.

Natural Gas Steam and Gas Turbine Capacity

Any natural gas generator that is not considered combined cycle is included in the steam and gas turbine categorization for natural gas capacity.

Heating and Cooling Degree Days

Heating and cooling degree days are population-weighted state-specific degree day averages where 2000 Census data on state population is used for the weightings. Each state is assigned to only one NERC region, even if the state lies in more than one NERC region.

Generation Cost

Generation cost is defined here as the fuel component of the variable cost of producing electricity. It is a function of the price of the fuel as well as the technology employed (heat rate), and is calculated as follows:

$$Cost\left(\frac{\$}{kWh}\right) = \left(\frac{\$}{MMBtu}\right) * \left(\frac{Btu}{kWh}\right) * \left(\frac{1}{1000}\right)$$

The capacity-weighted heat rates (*Btu/kWh*) are used as the technology measure. The oil generation cost is calculated as the capacity weighted average of residual fuel cost and distillate fuel cost.

Maximum Natural Gas Consumption

Maximum natural gas consumption is a measurement of the total amount of natural gas (MMBtu) that could theoretically be used in a given NERC sub-region if all gas-fired facilities operated 24 hours per day for an entire month. It is calculated based on the natural gas capacity in the region, the total number of hours in the month, and the capacity-weighted heat rate of the plants:

$$NGmax_t = \frac{MWCap * hours * HeatRate}{1000}.$$

This theoretical maximum is then used to create the variable, *NG Consumption Fraction,* which is a measure of the percentage of natural gas capacity that is actually used.

California Crisis Dummy Variable

A dummy variable was used to allow for the market peculiarities present at the time of the California energy crisis. This resulted in a dummy variable for the months January to June 2001. This period was indicated by an exceedingly large value of the cointegrating error term.

Appendix 2: Pipeline Data Compilation

Monthly and daily compilation pages in the "Pipeline Data" were used to sum natural gas deliveries by pipeline and by consumer NAICS code along each pipeline. A description of the layout of these compilation pages is included in the file "Pipeline Data Layout". The sum of the pipeline consumption for a specified period (monthly and daily) is calculated by summing the scheduled quantity delivered (Dth) to any industrial consumer included in the data set for that given period.

Monthly Data

The file is configured to calculate the total consumption for each pipeline by month (Dth/month). In order to compute these consumption figures, the user must fill cells E2 through W2 for the whole time period and calculate the page. To sum consumption for a given NAICS code only, follow these steps:

1) Enter the NAICS code in cells C2:C81 for the desired industry sector.
2) Clear all but the first line (E2:W2).
3) Highlight E2:W2.
4) Find/Replace $W with $Y.
5) Drag the formulas down for the whole array (E2:W81) and calculate cells.

Daily Data

The file is configured to calculate the total consumption for each pipeline by day (Dth/day). In order to compute these consumption figures, the user must fill cells E2 through W2 down for the whole time period and calculate the page. To sum consumption for a given NAICS code only, follow these steps:

1) Enter the NAICS code in cells C2:C2472 for the desired industry sector.
2) Clear all but the first line (E2:W2).
3) Highlight E2:W2.
4) Find/Replace $X with $Z.
5) Drag the formulas down for the whole array (E2:W2472) and calculate cells.