

A Comparison of Natural Gas Spot Price Linear Regression Forecasting Models

highly integrated transmission piping network. The inter-connecting nodes of this transmission network are used to transfer quantities of gas between producers, traders, marketers, transmission companies and local distribution companies, (LDC). There are approximately 40 such trading points which are known as hubs. The most well known hub is the "Henry hub", located in Louisiana. This hub has become the node at which spot and futures prices are referenced on the NYMEX markets. This integrated national natural gas piping network has allowed competition to cause the pricing of natural gas at all of the major hubs to be inter-related. No supplier can charge an exorbitant fee at any point in the U.S. because other suppliers will be able to provide gas at competitive prices.

De-regulation now appears to be entering its final phase. Small customers, both residential and commercial are now gradually being allowed the ability to choose their source of gas supply. This customer choice is occurring at different rates on a state by state basis.

Plan for Analysis

This paper will focus on the forecasting of natural gas spot price four weeks in the future at the Henry Hub during the heating season. The term "heating season" is that period of the year when natural gas consumption is greatest and when natural gas prices are also generally at their peak. This period of time runs from November 1, through March 31. Almost all of the U.S. domestically produced natural gas and most natural gas that is released from storage that serves what is known as the East Consuming Region is priced at a level that is close to the value of prices for the Henry Hub. The East Consuming Region which includes all of the states east of the Mississippi River, except for Mississippi itself. Iowa, Nebraska and Missouri are also included in this East Consuming region. Historical data is only readily available for the Henry Hub, therefore the regressions that I have performed use only spot and futures natural gas prices at the Henry Hub. This data is taken from the heating seasons: 1996 - 1997, 1997 - 1998, 1998 - 1999, 1999 - 2000, and 2000 - 2001. This limited number of heating seasons was chosen for two reasons. The first was due to the fact that the data was more readily available. The second stems from the analysis performed by Arthur De Vany and David W. Walls. They performed a variety of natural gas price network analyses' that

A Comparison of Natural Gas Spot Price Linear Regression Forecasting Models

tested to see which of the marketing hubs were isolated from the other hubs. These analyses' showed that by the mid 1990's all of the natural gas marketing hubs were closely integrated so that competition caused all of these prices to be interrelated.

The Energy Information Administration of the U.S. Dept of Energy performs a short term analysis known as the Short-Term Integrated Forecasting (STIFS) model. This model generates predictions of demand, supply, storage stock levels, imports and prices for natural gas in the future. The predictions for natural gas price cover categories that EIA describes as Natural Gas Spot Wellhead Price, Natural Gas Composite Wellhead Price, Natural Gas Price to Electric Utilities, Residential Natural Gas Price, Commercial Natural Gas Price and Industrial Natural Gas Price. The Natural Gas Spot Wellhead Price (**NGSPUUS**) is the composite spot price from, Natural Gas Week, published by Energy Intelligence Group. This price is a weighted average of the spot prices at the major hubs one of which is the Henry Hub. The Natural Gas Composite Wellhead Price (**NGWPUUS**) is an average of both spot and long-term contracts. Since the focus of this paper is on spot price prediction the model for Natural Gas Spot Wellhead Price (**NGSPUUS**) is of greatest interest.

The **NGSPUUS** model has two independent variables, the one month lagged value of the dependent variable, and fourteen dummy variables that are only used during specific "shock" events or during a specific month.

The EIA defines the **NGSPUUS** model as:

$$\begin{aligned} \mathbf{NGSPUUS} = & \beta_0 + \beta_1(\mathbf{GASVAR}) + \beta_2(\mathbf{ZGHDPUS} - \mathbf{ZGHNPUS})/(\mathbf{ZSAJQUS}) + \\ & \beta_3(\mathbf{D_HAND}) + \beta_4(\mathbf{D_294}) + \beta_5(\mathbf{DUM9602}) + \\ & \beta_6[\mathbf{LAG(NGSPUUS)}] + \beta_7(\mathbf{DUM JAN}) \dots + \beta_{17}(\mathbf{DUM NOV}) \end{aligned}$$

where \mathbf{GASVAR} = the difference between the actual amount of natural gas that is present in storage in the United States and the average amount of natural gas that has been historically present in storage over the last ten years during that particular month of the year.

$\mathbf{ZGHDPUS}$ = Gas-weighted heating degree days during a month

$\mathbf{ZGHNPUS}$ = Normal (1965 - 1999 average with trend) gas-weighted heating degree days during a month for the average of the four major gas consuming

A Comparison of Natural Gas Spot Price Linear Regression Forecasting Models

eastern cities of Chicago, Kansas City, New York and Pittsburgh

ZSAJQUS = Number of Days in month

D_HAND = a dummy variable which only has a value during the months of September and October 1992, when Hurricane Andrew interrupted the production of natural gas

D_294 = a dummy variable which only has a value during the month of February 1994, when cold weather and ice storms interrupted the flow of gas to distribution areas, due to pipeline failures.

DUM9602 = a dummy variable to account for unusually high spot prices which are attributed to panic resulting from low inventories and expectations of cold weather in February 1996

LAG(NGSPUUS) = the one month lagged variable for NGSPUUS

DUM_JAN ...DUM_NOV = eleven dummy variables, each of which are held to be zero except for one month out of the year.

The EIA regression was estimated using spot hub prices from January 1992 to the present. Refer to Table 1, which displays the regression results for this EIA model.

This paper proposes that other variables can be added to the core EIA model for **NGSPUUS** and improve upon its predictive powers. Due to the high degree of volatility of heating season gas prices, the accurate prediction of the natural gas spot price four weeks in the future during the winter heating season months at an important hub such as the Henry Hub would be very beneficial to a variety of entities.

An assumption inherent to the models I have proposed is that there are no shocks to the supply side of the natural gas market. Events such as hurricanes in the Gulf of Mexico or ice storms in the Texas-Oklahoma area have caused temporary interruptions to the production of natural gas. Unusual price spikes in natural gas prices during those events are generally attributed to such supply side interruptions. The EIA model accounts for two such supply side price spikes during the period of 1992 to 1996 by the use of dummy variables. My alternative models seek to forecast natural gas spot price in the near

A Comparison of Natural Gas Spot Price Linear Regression Forecasting Models

Table 1 Regression Results for Natural Gas Spot Wellhead Price, (NGSPUUS) (dollars per million Btu)

$$\text{NGSPUUS} = \beta_0 + \beta_1(\text{GASVAR}) + \beta_2(\text{ZGHDPUS} - \text{ZGHNPUS})/(\text{ZSAJQUS}) + \beta_3(\text{D_HAND}) + \beta_4(\text{D_294}) + \beta_5(\text{DUM9602}) + \beta_6[\text{LAG}(\text{NGSPUUS})] + \beta_7(\text{DUM JAN}) \dots + \beta_{17}(\text{DUM NOV})$$

Variable	Coefficient (β)	Standard Error	T Ratio
Constant Coefficient	0.557987	0.1799	3.1
Lag(NGSPUUS)	0.760221	0.0673	11.29
GASVAR	-8.0 * E-05	0.000166	-0.49
(ZGHDPUS - ZGHNPUS)/ZSAJQUS	0.028218	0.0188	1.5
D_HAND	0.160223	0.1644	0.97
D_294	0.383219	0.2959	1.3
DUM_9602	1.103517	0.29	3.81
DUM_JAN	-0.09587	0.1294	-0.74
DUM_FEB	-0.20553	0.1397	-1.47
DUM_MAR	-0.24633	0.1363	-1.81
DUM_APR	-0.06895	0.1415	-0.49
DUM_MAY	-0.12619	0.1392	-0.91
DUM_JUN	-0.13369	0.1382	-0.97
DUM_JUL	-0.09228	0.138	-0.67
DUM_AUG	-0.07841	0.1383	-0.57
DUM_SEP	-0.11014	0.1379	-0.8
DUM_OCT	-0.00618	0.1363	-0.05
DUM_NOV	0.100599	0.135	0.75

Mean Squared Error = 0.0704

Root Mean Squared Error = 0.2653

R-Square = 0.747

Adj. R-Square = 0.693

Durbin-Watson = 1.30

A Comparison of Natural Gas Spot Price Linear Regression Forecasting Models

future. The natural gas industry supply curve characteristics allow me to assume that the industry supply curve is fixed in the short term. The EIA models use data over a whole year. The regressions that I ran include only the Natural Gas Spot prices during the heating season from November through March. The alternative independent variables that I propose are taken at periods from four or six weeks prior to the spot prices, depending on when the data is available. The reasons I have chosen only the heating season months are twofold. The first is that those are the months which see the greatest volatility in gas prices. The factors which drive this price volatility are different than the factors that drive the summer time price and consumption for natural gas. The second reason is that these are the months when most consumers actually consume the bulk of their annual gas requirements. A model that more accurately forecasts spot price four weeks in the future during the winter months has a much greater value than a model for the whole year which was less accurate during the winter months. The data I have used is based on weekly values for each of the variables, as opposed to the EIA model which uses average monthly values.

Equation #1 is the model which is meant to replicate the core of the EIA model for **NGSPUUS**, the composite of natural gas spot wellhead price. Refer to Table T-4 in the Appendix which displays Equation #1 and all of the alternative model regression equations. The differences between my Equation #1 regression and that of the EIA model are that:

- a) My regression is run on data from the weekly data that was compiled from average daily data for the five heating period months of November thru March. The EIA model uses full years data.
- b) The data that was used in my model is based on only the last four heating seasons of 1996-1997 to 1999-2000. The EIA model is based on data from January 1992 to the present.
- c) My equation model does not make use of any dummy variables. The EIA uses monthly dummy variables to account for monthly disparities between the regression results and the data. The EIA model also uses dummy variables to account for the two supply side disruptions in gas supply and for an unusually high demand side spike in price during February 1996.
- d) My independent variable for temperature deviation from normal was based on the same time period as the EIA model, from 1965 to 1999. My temperature deviation data was