obtained from an EIA Internet website that reported the average daily temperature for four eastern cities and compared it to the historical average for that day. My independent variable, **TDEVAL4**, is the average of these daily temperature deviations for the last four weeks, but lags the natural gas spot price by four weeks. The EIA used Heating Degree Days per month and divided the value by the number of days for the month.

e) My equation model uses the Henry Hub spot price for natural gas. The EIA model for the Spot Natural Gas Wellhead Price, NGSPUUS, is a composite spot price of which the Henry Hub is a component.

A complete list of the acronyms and a brief description for the variables which are used in alternative model regressions are shown in Table 2. A more detailed description for these variables is described below.

SP = Natural Gas Spot Price, ($ per million Btu (MMBtu))

SPLAG4 = SP[-4] = average weekly natural gas Spot Price of 4 weeks previous, which is roughly the same as the previous months spot price, ($ per million Btu (MMBtu))

TEMPDEV = the weekly average of the deviation of the average daily temperature from the normal average temperature of four major east coast cities over the last 35 years.

TDEVAL4 = Average of four, five, six and seven weeks past Weekly Temperature Deviations (TEMPDEV) for Average of Chicago, Kansas City, New York & Pittsburgh for that time of Year, (deg F)

WTTEDL4 = a weighted average of Temperature Deviation which is (0.4 * four weeks past) + (0.3 * five weeks past) + (0.2 * (six weeks past)) + (0.1 * (seven weeks past))

STORDEV = Deviation of Natural Gas Storage from Average for that time of year (BCF), lagged by 4 weeks, so that Observation 5 of STORDEV was observed at the same time as Observation 1 for the Natural Gas Spot Price.

F1LAG4 = average weekly natural gas 1 month NYMEX Futures Price, 4 weeks prior to the natural gas Spot Price

F2LAG4 = Friday closing natural gas 2 month NYMEX Futures Price, 4 weeks prior to the natural gas Spot Price

F3LAG4 = Friday closing natural gas 3 month NYMEX Futures Price, 4 weeks prior to the natural gas Spot Price

SPOILL4 = #2 Distillate Oil in NY Harbor Spot Price, 4 weeks prior to natural gas Spot Price

Historically when natural gas usage is higher than normal the local natural gas distribution company reaps higher profits. This is attributed to the fact that most of the cost to operate a local gas company is fixed and the variable costs are very low. The cost of gas does not affect the profit of a local distribution natural gas company because this cost is passed directly to the customer without any mark-up. If market analysts anticipate higher usage (demand) for natural gas, then they anticipate greater profits and will value the gas companies stock price higher. A market value stock price index was constructed from the stock prices for eleven companies whose primary business is the local distribution of natural gas. Many of these companies are now holding companies but with only one exception, each obtains 90% or more of its gross revenues from its local gas distribution utility. These gas utilities serve the eastern portion of the U.S, from basically the same area from which the temperature data was taken. A list of these companies and a description of their service area can be seen in the Appendix Table T-2. This index was constructed from stock prices taken during the weeks of the same four winter heating seasons that the other data is taken. This index is abbreviated GASCOI.

$$\text{GASCOI} = \frac{\Sigma\ (\text{\# of shares} * \text{Avg Weekly Price of Stock})}{\Sigma\ (\text{\# of shares})}$$

The four week lagged value of this index is called **GASCOIL4.**

The effect of national economic prosperity is also suspected to have an effect on the usage of natural gas. To examine this relationship the macro-economic variables of M-2 money supply and weekly initial claims for unemployment insurance were added. I have taken the six week lagged values of these variables since it takes two weeks for the values of these variables to be published and we are trying to use them to predict the spot price of natural gas at a point of time four weeks in the future. These variables are published in both a seasonally adjusted figure and as the actual figure. I have chosen to look at both forms of these variables. The M2 value that is seasonally adjusted is known as **M2ADJL6.** The unadjusted value is abbreviated as **M2NTAJL6.** The variable acronym for weekly new claims for unemployment insurance, seasonally adjusted is **UNEMAJL6.** The variable for actual new claims for unemployment insurance, not seasonally adjusted is **UNEMATL6.**

A variable, **TDVSUSWA,** was created that is a weighted average of the TEMPDEV variable when this variable is negative for four or more consecutive weeks. This **TDVSUSWA** variable is zero at all other times.

Table 3 displays the Correlation matrix for all of the independent variables which were used in the regressions that were run. As would be expected this Correlation Matrix depicts close correlation between Spot and Futures Prices. High correlation values are seen for the weekly values for the money supply M2, adjusted and not adjusted. There was also a fairly high degree of correlation between the Spot Price for #2 Diesel Oil and the spot and futures prices for natural gas. The index of eleven gas utility companies had a fair degree of correlation with the variables for M2, money supply,(**M2NTAJL6** and **M2ADJL6**). The correlation between the values of adjusted and unadjusted Initial Claims for Unemployment Insurance, (**UNEMATL6** and **UNEMAJL6**), was rather low. This would appear to be due to the large effect that the seasonal adjustment creates.

Regression Results

A complete summary of the statistics for most of the 29 regressions run using these different independent and lagged dependent variables can be found in Table T-1 of the Appendix. The list of sources for the data from which these Regressions were estimated can be seen in Table T-5 in the Appendix.

Equations 2 through 4 add the one, two or three month Futures price for natural gas to the core **NGSPUUS** model respectively. Futures prices are representative of the knowledge available to traders regarding anticipated changes in the market. Adding the Futures price for natural gas as an independent variable will increase the forecasting accuracy of my alternative model. The next such variable to be added is that of #2 Distillate Heating Oil, **SPOILL4.** Distillate oil is sometimes used by commercial, industrial and electric generating facilities as a substitute for natural gas. There is no short term substitute for natural gas for residential customers. The presence of partial substitutes such as #2 distillate oil and to a lesser extent propane provide some substitute possibilities during periods of high demand and high price of natural gas. Therefore the prices of these partial substitute commodities should have some bearing on the price of natural gas.

Similar to Futures prices, the stock price of investor owned gas companies should carry with it some indication of how the market forces anticipate the demand for natural gas and its corresponding price. As previously described, an index of eleven gas companies was created and the weekly value of that index was incorporated as a variable. Table T-2 in the Appendix lists the eleven gas companies and describes their service areas.

The state of the economy would also seem an important determinant of the demand for natural gas in the commercial, industrial and electric generation markets. The macro-economic variables M2 money supply and initial claims for unemployment benefits were introduced as independent variables in some of the regressions to take account of the overall level of economic activity.

These different variables were used in various combinations to try and determine which would yield more accurate forecasts of the spot price of natural gas four weeks in the future. All of the equations were found to have a good deal of autocorrelation as evidenced by the Durbin-Watson and Durbin H statistics. Many of the equations also appear to have a good deal of multicollinearity as well. The multicollinearity was measured by calculating the Condition number as described by Belsley, Kuh and Welsch (1980). Condition numbers which exceed 20 are said to demonstrate that a regression equation has serious multicollinearity.

Each of these regressions was used to make out-of-sample forecasts for the most recent heating season, 2000 to 2001. None of the regressions accurately predicted the sustained high levels of natural gas prices during late November 2000 and throughout December 2000. After examining the regression forecasts and also a similar but smaller deviation from forecast spot prices during the fall of 1996, it was noticed that these periods corresponded with colder than normal temperatures. A dummy variable was created, TDVSUSWA, to account for this price behavior. This variable is assigned a value of zero, unless there are four consecutive weeks of colder than normal temperatures. On the fourth week

TDVSUSWA = (0.5 * four weeks past deviation of weekly average temperature from normal weekly average temperature) + (0.3 * five weeks past deviation of weekly average temperature from normal weekly average temperature) + (0.2 * six weeks past deviation of weekly average temperature from normal weekly average temperature).

This TDVSUSWA variable continues to be calculated in this manner until a full week of above average weekly temperatures has taken place. The two equations which use this variable, Equations #28 and #29, show markedly improved regression statistics. However their out-of-sample forecasts still do not reach the high values that were actually encountered during the 2000 - 2001 heating season.

Each of the 29 linear regression results were ranked with respect to the degree with which they fit the actual natural gas spot prices for the four heating seasons 1996 to 1997, 1997 to 1998, 1998 to 1999, and 1999 to 2000. Table 4 shows the statistics for the five linear regressions that had the highest ranking along with the EIA core model, Equation #1. Two ranking procedures were used. The complete list of rankings for all of the regressions is shown on Table T-3, in the Appendix. The first method of ranking shown as Rank Method A and is calculated as

(0.5 * **RANK** R sq Adj.) + (0.25 * **RANK** Multicollinearity with Constant) + (0.25 * **RANK** H Statistic)

The second method of ranking is shown as Rank Method B and is calculated as

(0.4 * **RANK** R sq Adj.) + (0.2 * RANK Minimum Amemiya Prediction) + (0.25 * **RANK** Multicollinearity w Constant) + (0.25 * RANK H Statistic)

Both ranking methods were arbitrarily selected. The adjusted R squared (R sq Adj.) and Amemiya Prediction criteria were selected because they are both statistics that measure how well the regression equation fits the data. Both statistics penalize an equation model which adds more independent variables which don't add any explanatory power. Each of the alternate equation