models were also ranked with respect to multicollinearity and autocorrelation. The multicollinearity was measured by the value of the condition number, Belsley et al. (1980). The autocorrelation that is present with lagged dependent variables was measured by the H statistic, Durbin (1970). The lowest values for the condition number and the H statistic being the most desirable since they indicate the least multicollinearity and lowest autocorrelation.

The five equations which provided the least deviation from actual spot prices and the best regression statistics according to the two Ranking methods for the four In Sample heating seasons were Equations 28, 29, 19, 16, and 22. The regression statistics for these five equations were markedly better than the EIA core model, Equation #1. These results can be seen in Table 4.

The regression of the complete EIA model was estimated using spot hub prices from January 1992 to the present. Table 1 displays the regression results for the complete EIA model which includes the fourteen dummy variables. The regression statistics: R-squared, Adjusted R-squared and root mean squared for the complete EIA model are superior to those of any of the alternative equation models that I propose. The regression statistics for my alternative equations are shown in Table T-1 in the Appendix.

The T-ratios for the complete EIA model are poor compared to those of my better alternative models. The complete EIA model was generated from eight years of monthly data and uses eighteen coefficients. The 5 percent critical value for a t-distribution with 80 degrees of freedom is +/- 1.99. Only two of the seventeen variables in the complete EIA model are significant at this 5 percent critical value. My best five alternative equations have three or four variables that are significant at this 5 percent value and there are only four to seven variables in my alternative equations.

Intriligator (1978) discusses a method of depicting forecast values versus actual values as functions of their relative changes. The weekly change in forecast value is divided by the previous weeks forecast value. The weekly change in actual value is divided by the previous weeks actual value. These weekly relative changes are then graphed by letting the forecast relative change equal the y axis value and the actual relative change equal the x axis value. An accurate forecast model should have data points which fall along the line described by the equation, y = x. The graphing of all of the 29 equation regressions that are created in this paper are best described as being uniformly distributed about the origin, where y = x = 0. The data points tend to be dispersed in all four quadrants and not primarily Quadrants I and III, which Intriligator would anticipate for an accurate model. These graphical displays are shown for the best fit equations 28, 29, 19, 16, and 22 on Figures F-2 through F-6 in the Appendix. The EIA core model Equation #1 is shown as Figure F-1 in the Appendix. Figures F-13 through F-17 show the forecast values for spot price from these five best fit equations versus the acutal spot price for the four in-sample heating seasons. These graphical displays tend to question the validity of all of the regression equations.

Forecasting Ability of Regressions

The heating season for 2000 - 2001 had record breaking high natural gas prices which appear to be directly related to the abnormally low temperatures that have occurred this winter. Figure 4 shows a graph of the natural gas spot price for the last five heating seasons and reveals how much higher this past heating season prices have been compared to the previous four.

Each of the 29 linear regression results were also ranked with respect to the degree with which they accurately forecast the actual natural gas spot prices for the heating season of 2000 to 2001. Table 5 shows the statistics for the five linear regressions that had the highest ranking for this out-of-sample testing. 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 for this out-of-sample heating season is shown as Rank Method C and is calculated from :

(0.5 * **RANK** RMSE) + (0.5 * **RANK** Sum of Absolute Values of Predicted minus Actual Spot price).

The second method of ranking is shown as Rank Method D and is calculated from:

(0.3 * **RANK** RMSE) + (0.3 * **RANK** Sum of Absolute Values of Predicted minus Actual Spot price) + (0.2 * RANK Multicollinearity with Constant) + (0.2 * RANK H Statistic).

The two ranking methods yielded similar results.

The five equations which provided the least deviation for the single out-of-sample heating season 2000 to 2001 were Equations 23, 4, 3, 9 and 7. Figure 6 shows the spot prices forecast by the best five of these out-of-sample equations, as well as the EIA core model equation and the actual spot price for this most recent heating season.

The five equations which provided the least deviation from actual spot prices for the four in-sample heating seasons were Equations 28, 29, 19, 16, and 22. Figure 5 shows the spot prices forecast by the best five of these in-sample equations versus the EIA core model equation and the actual spot price for this most recent heating season, 2000 to 2001.

The Intriligator (1978) method of depicting forecast values versus actual values as functions of their relative changes are shown for the best fit equations 28, 29, 19, 16, and 22 on Figures F-7 through F-11 in the Appendix. Figures F-18 through F-22 show the forecast values for spot price from these five best fit out-of-sample equations versus the actual spot price for the four in-sample heating seasons.

The five best fit out-of-sample equations are diffeent from the five best fit in-sample equations. The fact that there is no single equation in both best-of-five rankings comes as no surprise after considering the history for natural gas spot prices over the last five heating seasons as shown by Figure 4. The four heating seasons from the fall of 1996 to the spring of 2000, for which the regressions were estimated have similar patterns and levels of natural gas prices. The pattern and level of natural gas price for the most recent heating season 2000 to 2001 is markedly different than that of the previous four. Figure 7 depicts the average weekly deviation from normal temperatures for the five heating seasons. There is no discernible pattern, such as exists for natural gas prices during this same period. However, the most recent heating season, shown with a bolder line, had below normal temperatures for an extended period from November 8^{th} through January 7^{th}. It is generally acknowledged in the natural gas industry that this extended cold period was a major factor in the unusual price pattern. Once this extended cold period broke, the price of natural gas started to return to normal, as is seen in Figure 4.

Overall the results suggest that the EIA core model equation can be improved by adding some of the variables discussed in this paper. The graphical displays of relative change in forecast versus actual prices suggest that none of the equations can be said to give a best representation of which variables best explain natural gas spot price. However the results of this paper do indicate that incremental improvements in the accuracy of existing forecast equations, such as the EIA model, can be made.