

# Forecasting Inter-related Energy Product Prices

M. E. Malliaris  
Loyola University Chicago, mmallia@luc.edu

S. G. Malliaris  
Yale University, steven.malliaris@yale.edu

## Abstract.

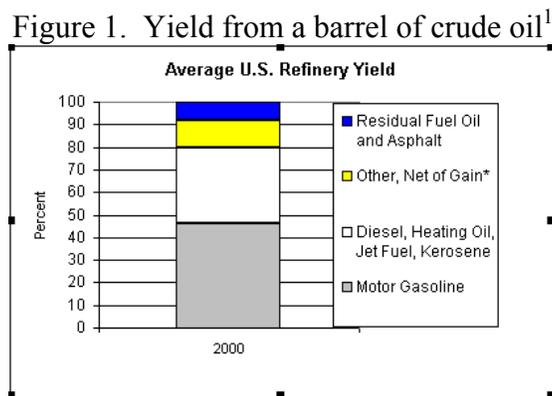
Five inter-related energy products are forecasted one month into the future using both linear and non-linear techniques. Both spot prices and data derived from those prices are used as input data in the models. The models are tested by running data from the following year through them. Results show that, even though all products are highly correlated, the prediction results are asymmetric. In forecasts for crude oil, heating oil, gasoline and natural gas, the nonlinear forecasts were best while for propane, the linear model gave the lowest error.

## Introduction.

Crude oil, heating oil, gasoline, natural gas, and propane are five energy products whose prices are inter-related. This relation might seem to derive logically from their relationship: no one product is independent of all others in either usage or production.

Crude oil and natural gas are both raw energy products; they come out of the ground. The former can be refined to produce heating oil and gasoline, and one component of gasoline is propane. The latter (natural gas) can also be processed to give propane, but only in limited amounts. Overall, propane is sourced in approximately equal amounts from 1) gasoline derived from crude oil refining and 2) natural gas processing [EIA4].

Crude oil, the most actively traded commodity in the world, breaks specifically into the following products (Figure 1, sourced from US Department of Energy, EIA1)



<sup>1</sup> Note: Processing gain is the volume increase that results as denser molecules (e.g., residual fuel oil) are split into less dense ones (e.g. gasoline). The processing gain in U.S. refineries is equal to about 6%. See: [http://www.eia.doe.gov/pub/oil\\_gas/petroleum/analysis\\_publications/oil\\_market\\_basics/graphs\\_and\\_charts.htm](http://www.eia.doe.gov/pub/oil_gas/petroleum/analysis_publications/oil_market_basics/graphs_and_charts.htm)

Similarly, natural gas breaks into the following (Figure 2)

Figure 2. Components of natural gas<sup>2</sup>

TYPICAL NATURAL GAS COMPONENTS	
COMPONENTS	PERCENT BY VOLUME
Methane	89.5 - 92.5
Ethane	5.1 - 2.0
Propane	2.1 - 0.7
Butanes	1.6 - 0.5
Nitrogen	0.6 - 1.9
Carbon Dioxide	1.1 - 2.4
<b>Total</b>	<b>100.0 - 100.0</b>

At present natural gas accounts for about a quarter of US energy use, but large consumers of energy (factories, power plants) often absorb an up-front cost to implement systems allowing, with minimal change-over cost, the use of either fuel oil or natural gas [NYMEX] as their main energy source as a hedge against fluctuating market prices and uncertain availability. Because of this substitute relationship, demand for either energy source is more elastic than one would otherwise predict—and, as one would expect, the NRG News states that one of the primary determinants of the price of natural gas is the price of oil [NRG News Energy Shop].

Table 1 shows the (as expected) high correlations between the above-mentioned energy markets.

Table 1. Correlation between energy market prices.

	CL	HO	PN	HU	NG
CL	1	-	-	-	-
HO	0.959721	1	-	-	-
PN	0.842248	0.881154	1	-	-
HU	0.964905	0.926191	0.847288	1	-
NG	0.669869	0.731288	0.677979	0.657551	1

Academic and industry research and position papers differ as to their explanations for the causes and effects of both absolute energy prices and the relationships therein. Asche et al found a long-run relationship between crude oil and gas oil prices, but no similar relationship between crude oil and fuel oil. Brown et al found that rising oil prices hurt economic activity and preceded nine of the ten recessions following World War II. Current measures indicate that the potential for an energy crisis is historically high [Williams and Alhajji]. OPEC, often a target of criticism about rising prices in energy products, acknowledges a relationship between gasoline, heating oil, and crude oil, but

<sup>2</sup> Roanoke Gas Company, <http://www.roanokegas.com/aboutus/origin.html>

believes the relationship is neither direct nor proportional, and claims that rising prices are due to taxes rather than OPEC policies [Heselbarth].

There is much uncertainty about energy prices in the future [Dept. of Energy]. This uncertainty in energy prices commands a great deal of foreign and domestic political attention, and facilitates an active market ranking second only to financial products<sup>3</sup> in amount of trading on futures contracts. As a result, energy commodity price risk has a dominant role in the energy industry [EIA 2, EIA4]. Herard and Taylor emphasize that a first step in minimizing earnings volatility in industry is to stabilize margins; a major reason for the margin volatility is the price change in natural gas. Therefore, by accurately forecasting prices into the future, we can create an effective way of managing this risk. This paper looks at a short-term forecast (one month) and compares the ability of two models to handle that forecast. There are about 248 trading days per year, or, around 21 per month. More specifically, we use information from five energy markets to forecast each of those markets 21 trading days from the day the forecast is made. Data was specifically kept simple as we wanted to focus on the effect of recent price change on pricing one month away.

### **Data and Variables.**

#### Data

The data consists of daily spot prices for Crude Oil (CL), Heating Oil (HO), Gasoline (HU), Natural Gas (NG), and Propane (PN). All the data are original data, provided by Barchart at [www.barchart.com](http://www.barchart.com), in absolute values. HO is in cents per gallon, e.g. for 1/3/1994 the price was \$ 0.453/gallon. PN is in cents per gallon, e.g. for 1/3/1994 the price was \$ 0.2500/gallon. NG was measured in dollars per MMBtu, e.g. the price for 1/3/1994 is \$2.050/MMBtu. HU was in cents per gallon, e.g. for 1/3/1994 the price is \$0.3970/gallon. Finally, CL is in dollars per barrel, e.g. for 1/3/1994 the price was \$14.54/barrel.

The initial data set contained daily spot prices for the period between Jan 3, 1994 and Dec 31, 2002. Inspection of the data over time showed that the relationship between the five markets changed significantly between 1994 and 2002. They became more correlated in movement [see Appendix 1 for graphs illustrating, for example, this relationship between gasoline and heating oil for the month of February in three years]. Anticipating that energy markets will continue to become more rather than less inter-related, and attempting to build a stable model, we used data from December 1997 through November 2002 for this study.

#### Input and Output Variables

The input variables for each model consisted of the daily closing spot price, percent change in daily closing spot price from the previous day, standard deviation over the previous 5 trading days, and standard deviation over the previous 21 trading days, for each of the five markets.

---

<sup>3</sup> US Department of Energy statistic

For the Crude Oil variables, for example, these variables were labeled as CL, CLchg, CLStdDv5, and CLStdDv21. This gave us a total of 20 initial input variables. In addition, the neural network model used a cluster indicator. This is a non-numeric variable indicating the cluster group to which each row belonged. The output variable was the daily spot price one month into the future (21 trading days). Again using Crude Oil as an example, this variable was labeled CLplus21. Models were built for each of the five markets. The table below shows the correlation of each energy commodity price today with each price 21 trading days into the future. These values range from a low of .6599 to a high of .9293.

Table 2. Correlation of today's spot prices with prices 21 trading days away.

	CL	HO	PN	HU	NG
CLplus21	0.9293	0.9012	0.8029	0.9021	0.6383
HOplus21	0.9242	0.9204	0.8208	0.8898	0.6546
PNplus21	0.8185	0.8329	0.8820	0.8115	0.5684
HUplus21	0.8869	0.8702	0.9265	0.8482	0.6440
NGplus21	0.7160	0.7708	0.7346	0.6810	0.6599

As the table indicates, these correlations remain high into the future with natural gas showing the lowest overall match. Notice that most of the correlations of one product's prices today with another product's prices 21 days away are slightly lower than those shown between those same products in Table 1, with the exception of natural gas. For example, the correlation between HO and CL is .9597, while the correlation between HO and CLplus21 is .9012 and between CL and HOplus21 is .9242. Though slightly lower, they are still good enough to encourage us to attempt a model.

#### Training and Validation Sets

The data set ran from December 1997 through November 2002, and contained 25 columns. This was split into disjoint training and validation sets.

Twenty of the columns were used for input. Of the remaining five columns, one was selected at a time as the output for each model. The remaining four were unused for that model. For the non-linear model, another input column was generated by the K-Means algorithm, a cluster identifier.

The training set contained four years of data, from December 1997 through November 2001, and had 1001 rows. The remaining year of data, December 2001 through November 2002 was used for the validation set. This set contained 247 rows.

The models were judged by their results on validation data that they had not seen during training and which occurred after the end of the training period.

### Models and Methodology

#### Models

Two model types, one linear, and one non-linear were built for the forecasts. A multiple regression model was built for each of the five energy markets with the spot price 21 trading days into the future as the dependent variable. These models were developed in

Microsoft Excel and run with the Excel Data Analysis option. As variables indicated lack of importance to the model, or multicollinearity, they were deleted and the model was re-run. A final model for each market was then used for the validation set forecasts. The variables used in each of the multiple regression models are shown in Table 3.

Table 3. Variables used in each final regression model.

<b>Variables</b>	<b>CL</b>	<b>HO</b>	<b>HU</b>	<b>NG</b>	<b>PN</b>
CL	x	x	x		x
CLchg	x	x			x
CLStdDv5		x			x
CLStdDv21	x	x	x	x	
HO		x		x	x
HOchg	x				
HOStdDv5	x		x	x	x
HOStdDv21		x	x		x
HU		x	x		
HUchg	x	x			x
HUStdDv5	x	x		x	x
HUStdDv21				x	
NG	x	x	x	x	x
NGchg			x		
NGStdDv5					x
NGStdDv21	x	x	x		x
PN	x	x		x	
PNchg			x		x
PNStdDv5				x	x
PNStdDv21		x	x	x	x

The number of variables used in each of the final regression models ranged from nine to fourteen. Only NG appeared in every model. The variables appearing in at least four of the five models were CL, CLStdDv21, HOStdDv5, HUStdDv5, NGStdDv21, and PNStdDv21. Five out of these six variables appearing in four models are based on movement within the market the past week or month.

Following the development of each of the regression models, the SAS data mining package Clementine was used to develop a second set of models. These models were non-linear and used two processes in sequence. The data was first run through a K-Means clustering algorithm. In a clustering algorithm, the numerical records are grouped by similarity and a center for the group is calculated. This center is the arithmetic mean of the records in the cluster. The cluster can be described not only by its mean, but also by its distance from other clusters.

The K-Means algorithm in Clementine generated 5 distinct clusters. Three of these clusters were fairly small, and the remaining two clusters were large. The number of rows of the training set per cluster and the center of each cluster (that is, the mean of the records in that cluster for each of the twenty input variables in the training set) are shown in Table 4.

The distance between cluster means, also called cluster proximity, is shown in Table 5. The greater the distance, the more distinct the clusters are. In addition to generating centers for each variable per cluster, the K-Means algorithm identifies the cluster to which each row belongs and generates a node to inspect data and assign it to a cluster. These cluster assignments are non-numeric. Other data sets can be fed through a trained K-Means node to generate a cluster assignment for each row in them.

Table 4. Cluster centers and population.

Variable	cluster-1	cluster-2	cluster-3	cluster-4	cluster-5
	482 Records	17 Records	384 Records	58 Records	60 Records
CL	15.992	28.231	27.873	28.213	31.167
CLchg	0	-0.01	0.001	0	0.002
CLStdDv5	0.342	0.82	0.59	0.6	0.827
CLStdDv21	0.676	2.382	1.196	0.863	1.304
HO	0.424	0.929	0.765	0.774	0.91
HOchg	0	-0.005	0.001	0	-0.002
HOStdDv5	0.009	0.032	0.02	0.017	0.024
HOStdDv21	0.017	0.06	0.041	0.027	0.041
HU	0.461	0.727	0.794	0.879	0.869
Huchg	0.001	0.001	0.001	-0.003	-0.002
HUStdDv5	0.011	0.021	0.023	0.027	0.026
HUStdDv21	0.023	0.085	0.044	0.057	0.038
NG	2.166	9.236	3.375	4.408	6.686
NGchg	0	0.02	0.002	-0.005	-0.001
NGStdDv5	0.072	0.764	0.102	0.127	0.391
NGStdDv21	0.144	1.313	0.213	0.246	0.821
PN	0.286	0.776	0.503	0.594	0.547
PNchg	0.001	0.016	0.001	0.013	-0.008
PNStdDv5	0.006	0.031	0.013	0.042	0.022
PNStdDv21	0.013	0.059	0.03	0.13	0.062

Table 5. Distance between clusters.

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	-				
Cluster2	1.996	-			
Cluster3	0.916	1.421	-		
Cluster4	1.253	1.402	0.638	-	
Cluster5	1.376	0.961	0.691	0.744	-

The resulting cluster assignment, Cluster1 through Cluster5, became an additional input into the neural network model. This input was symbolic data. Neural networks can accept both numeric and non-numeric data. A neural network model was developed for each of the five markets. Neural networks have been successful in forecasting because of

their ability to forecast difficult types of relationships well, given the correct inputs and training set. See, for example, Trippi and Turban or Smith and Gupta for collections of papers using neural networks in financial problems. Olson and Mossman found, in financial models, that they outperform regression models and thus yield greater profitability when trading.

Each neural network model used twenty-one inputs (the 20 original fields, plus the non-numeric cluster identifier), one hidden layer with twenty nodes, and one output node. The neural network model generates a sensitivity analysis that lists the model variables in order of their effect on the model. Table 6 contains this output for each of the five models. Thus, the variables used in the neural network models are listed in their order of importance to the model in this table. Note that in four of the models, heating oil price was a prominent variable.

Table 6. Order of Variable Significance in Neural Network Models

<b>CL</b>	<b>HO</b>	<b>HU</b>	<b>NG</b>	<b>PN</b>
Cluster	HO	CL	HO	HO
PNStdDv5	NGStdDv5	HO	PN	CL
Huchg	NGStdDv21	NGStdDv21	CL	NGStdDv5
NGchg	PN	CLStdDv5	NG	PN
HU	PNStdDv5	Cluster	HUStdDv5	NGStdDv21
HUStdDv5	CL	NG	PNStdDv5	PNStdDv5
NGStdDv21	Cluster	PNStdDv5	PNStdDv21	HUStdDv21
CL	PNStdDv21	NGStdDv5	NGStdDv21	CLStdDv5
NG	HU	PNStdDv21	Cluster	HU
HOSStdDv21	CLStdDv5	HUStdDv21	HUStdDv21	Cluster
PNStdDv21	Huchg	HU	CLStdDv21	Huchg
CLStdDv21	HOSStdDv21	HOSStdDv21	CLStdDv5	HUStdDv5
CLchg	HUStdDv5	PNchg	NGStdDv5	NGchg
HUStdDv21	CLStdDv21	NGchg	Huchg	Hochg
Hochg	HUStdDv21	PN	NGchg	PNchg
HOSStdDv5	Hochg	Huchg	HOSStdDv5	PNStdDv21
NGStdDv5	HOSStdDv5	CLchg	HU	HOSStdDv21
HO	NG	CLStdDv21	Hochg	CLStdDv21
CLStdDv5	CLchg	Hochg	PNchg	HOSStdDv5
PN	PNchg	HUStdDv5	CLchg	CLchg
PNchg	NGchg	HOSStdDv5	HOSStdDv21	NG

If we look at the top five variables in each of the models, we find little agreement among them. In particular, the CL model has no variable in the top five in common with any other model. HO is in the top five of each of the other models. PN and CL are in the top five of three of the models. Thus, even though the products are inter-related, the price forecasts emphasize different variables.

## Methodology

A regression model and neural network model was developed for each of the five energy commodities to forecast the price one month into the future. Data was used, for the training models, from December 1997 through November 2001, for a total of 1001 row of input data.

The regression models were refined until a model using only variables showing no multicollinearity were used. The non-linear approach used first a K-Means clustering algorithm, then a neural network to develop a forecasting model. The networks were trained using a random 50% of the data, with the other 50% used as a testing set to prevent the network from overtraining. When a network is allowed to develop with all of the training data, it frequently does very well on the training data, but does less well on data it has never seen (in our case, we will use the validation set to finally judge this). To prevent this, data is randomly selected from the training set, held out from the training process, and used to periodically check the performance of the network. When the network begins to do less well on the testing set than the training set, Clementine stops training the network. This results in a network that is able to generalize better when predicting new data.

Next, a validation set, comprised of a year's worth of data that occurred after the training data, was run through each of the models. This validation set data, for the nonlinear model, ran first through the clustering algorithm and each row was assigned a value indicating the cluster to which it belongs. The data then flowed into the trained neural network and a forecast was generated.

Models were compared, for each of the five energy products, by looking at the validation set mean absolute error and the mean squared error of each regression and neural network model. These statistics from the validation sets were used as the measures of ability of the linear and non-linear models to forecast the five energy markets. In addition, a simple forecast was generated by assuming no change between today and 21 days from now. The same comparison statistics were calculated for this simplistic view of prediction.

### Models Results on Training Sets

The regression statistics generated by Excel on the training data for each of the models are shown in Table 7. As can be seen in the table, all models did well on the training set, with the lowest R Square going to Propane.

Table 7. Regression Statistics

	<b>CL</b>	<b>HO</b>	<b>HU</b>	<b>NG</b>	<b>PN</b>
Multiple R	0.948154	0.949655	0.919218	0.913553	0.884675
R Square	0.898997	0.901845	0.844962	0.83458	0.78265
Adjusted R Square	0.897977	0.900552	0.843396	0.833077	0.779564
Standard Error	2.199953	0.066021	0.079985	0.6553	0.070904

Table 8 gives the coefficients of each of the regression equations. Blank cells indicate unused variables for that model. The coefficients from model to model emphasize the difference placed on each variable between models.

Table 9 shows the relative importance of each of the variables in the non-linear neural network model. With a neural network, the sensitivity analysis values vary from 0 to 1. The higher the number attached to an input variable, the greater the importance of that variable in determining the value of the output variable.

Table 8. Regression Coefficients.

<b>Variable</b>	<b>CL</b>	<b>HO</b>	<b>HU</b>	<b>NG</b>	<b>PN</b>
Intercept	1.2802	-0.0265	0.0574	-0.4755	0.0049
CL	0.7676	0.0217	0.0120		0.0076
CLChg	-6.2214	-0.2938			-0.1580
CLStdDv5		0.0253			-0.0146
CLStdDv21	1.1483	0.0192	0.0216	-0.3455	
HO		0.3671		3.8124	0.4179
HOChg	5.2500				
HOStdDv5	29.2962		0.7437	-4.0060	0.9232
HOStdDv21		-0.5076	0.9226		-0.5630
HU		-0.2205	0.2709		
HUChg	7.8516	0.2477			0.1479
HUStdDv5	-32.4310	-0.7473		5.3742	-1.0396
HUStdDv21				9.9400	
NG	0.6215	0.0185	0.0460	0.7399	0.0191
NGChg			-0.0773		
NGStdDv5					0.0946
NGStdDv21	-4.3548	-0.1398	-0.2250		-0.2425
PN	5.0062	0.1433		-1.6273	
PNChg			0.1381		0.0881
PNStdDv5				-2.8196	0.2912
PNStdDv21		-0.4951	-0.2878	-13.0017	-0.2469

Variables whose importance are greatest in the neural network forecast (over .1) are shown in bold-face type in the table. Notice that NG, while important in every regression model, is more important in only two of the neural network models. Also, in the CL model, the cluster affinity variable is much greater in significance than any other variable in the model. Comparison of these two tables indicates a very different approach taken by the two model types to forecasting each product.

Table 9. Relative importance of inputs in neural network models.

Input	CL	HO	HU	NG	PN
CL	0.0329	<b>0.1606</b>	<b>0.4019</b>	<b>0.1462</b>	<b>0.2507</b>
CLchg	0.0218	0.0172	0.0338	0.0034	0.0131
CLStdDv21	0.0245	0.0487	0.0321	0.0641	0.0309
CLStdDv5	0.0087	0.0890	<b>0.1709</b>	0.0572	<b>0.1193</b>
HO	0.0131	<b>0.3612</b>	<b>0.3828</b>	<b>0.2424</b>	<b>0.3594</b>
Hochg	0.0183	0.0246	0.0217	0.0156	0.0577
HOStdDv21	0.0265	0.0717	0.0775	0.0016	0.0477
HOStdDv5	0.0162	0.0241	0.0123	0.0435	0.0200
HU	0.0386	<b>0.1369</b>	0.0822	0.0170	<b>0.1158</b>
Huchg	0.0568	0.0820	0.0421	0.0512	0.0986
HUStdDv21	0.0213	0.0433	0.0884	0.0826	<b>0.1374</b>
HUStdDv5	0.0383	0.0686	0.0211	<b>0.1249</b>	0.0925
NG	0.0289	0.0228	<b>0.1445</b>	<b>0.1365</b>	0.0032
NGchg	0.0406	0.0057	0.0675	0.0436	0.0690
NGStdDv21	0.0374	<b>0.2743</b>	<b>0.3026</b>	0.0904	<b>0.1712</b>
NGStdDv5	0.0145	<b>0.2779</b>	<b>0.1159</b>	0.0562	<b>0.2345</b>
PN	0.0028	<b>0.2301</b>	0.0570	<b>0.1940</b>	<b>0.2315</b>
PNchg	0.0023	0.0101	0.0739	0.0081	0.0517
PNStdDv21	0.0247	<b>0.1405</b>	0.0989	<b>0.1097</b>	0.0485
PNStdDv5	0.0593	<b>0.1843</b>	<b>0.1392</b>	<b>0.1098</b>	<b>0.1406</b>
Cluster	<b>0.1992</b>	<b>0.1446</b>	<b>0.1546</b>	0.0830	<b>0.1130</b>

The neural network results on the training data are shown in Table 10. This table gives the estimated accuracy of the model along with typical statistics generated by the network. The estimated accuracy for each model compares well with the regression results. On the training data results, shown in Tables 7 and 10, both models appear to do rather well.

Table 10. Training Data Results with Neural Network

	CL	HO	HU	NG	PN
Minimum Error	-11.524	-0.199	-0.317	-2.461	-0.215
Maximum Error	10.403	0.399	0.480	4.746	0.389
Mean Error	-2.152	-0.036	-0.024	-0.157	0.01
Mean Absolute Error	4.732	0.099	0.096	0.641	0.061
Standard Deviation	5.183	0.111	0.124	0.98	0.087
Linear Correlation	0.774	0.859	0.789	0.792	0.817
Estimated Accuracy	93.761	95.436	96.080	97.064	95.675

### Model Results on Validation Data

After the development of the models on the training set data, the next step was to run the validation set data through each of the models. The validation set data occurred in time after the end of the training data, so it is a good measure of the ability of each of the models to forecast into the future. As a measure for each regression and neural network model, we used the mean absolute error and the mean squared error. For comparison

purposes, a “Simple Model Forecast” was made by assuming that the market in 21 days would be the same as it was today. The same error quantities were calculated for this model.

Table 11 shows the mean absolute error and mean squared error for each of the models developed. The best model, that is, with the lowest error, for each product is shown in bold type. With the exception of the models for Propane, the neural network models gave a forecast with smaller error over the year. For forecasting propane, the best results came from the Simple Model. So, neither the regression nor neural network approaches were able to do well in forecasting Propane with the given variables.

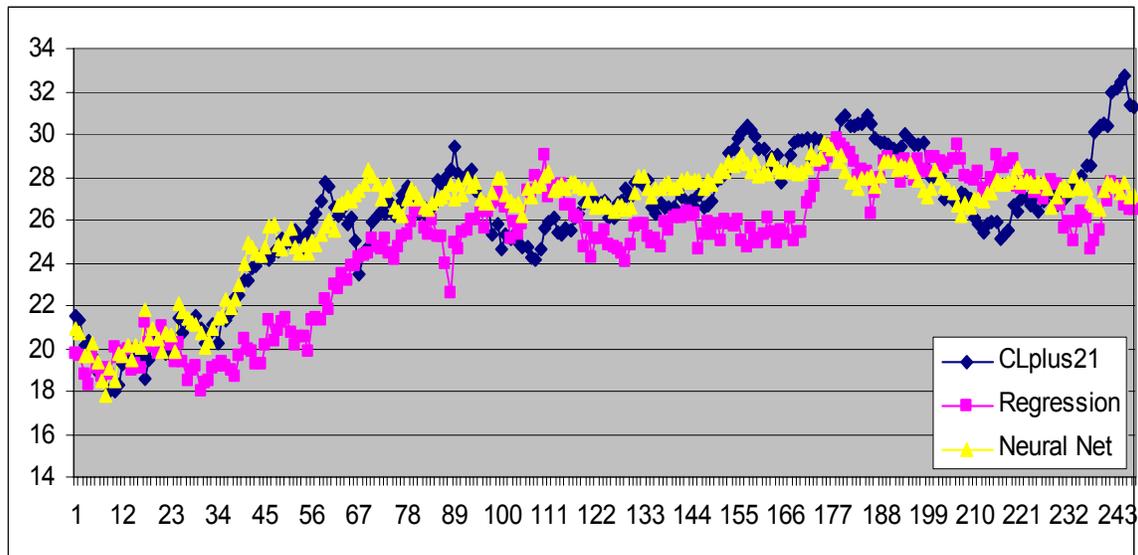
Table 11. Comparison of Simple Model, Regression and Neural Network Errors

	Avg. Absolute Error			Mean Squared Error		
	Simple	Regression	Neural Net	Simple	Regression	Neural Net
CL	1.973	2.126	<b>1.120</b>	6.013	6.653	<b>2.269</b>
HO	0.051	0.055	<b>0.035</b>	0.004	0.005	<b>0.002</b>
HU	0.057	0.053	<b>0.029</b>	0.006	0.004	<b>0.001</b>
NG	0.388	0.414	<b>0.218</b>	0.240	0.242	<b>0.075</b>
PN	<b>0.041</b>	0.061	0.080	<b>0.003</b>	0.006	0.009

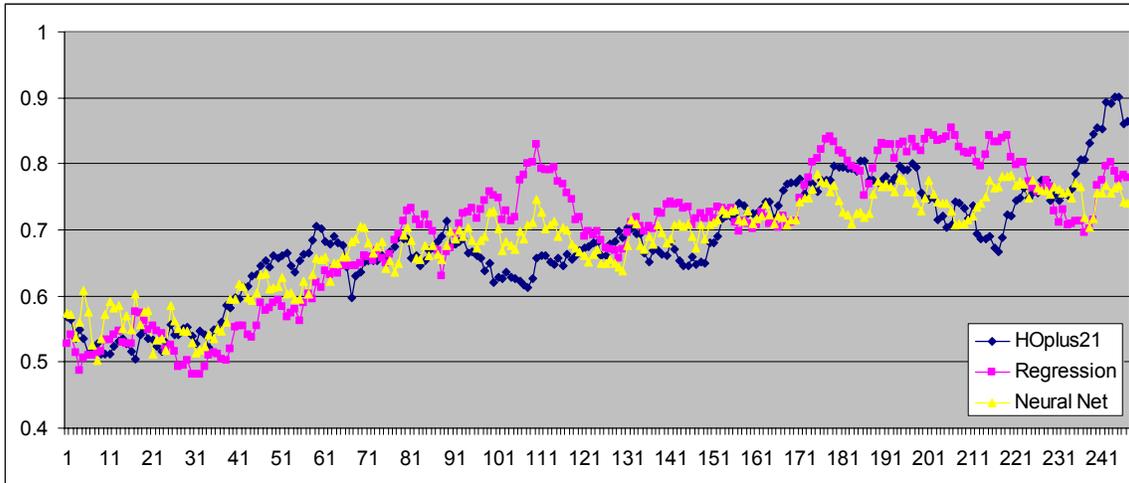
Graphs 1 through 5 show the relationship of the actual prices, forecasted regression prices and forecasted network prices. As can be seen, other than for Propane, the neural network forecast is closer to the actual price than that of the regression model.

Graphs 1 - 5. Regression and Neural Network Forecasts compared with Actual Prices for the Validation Set time period.

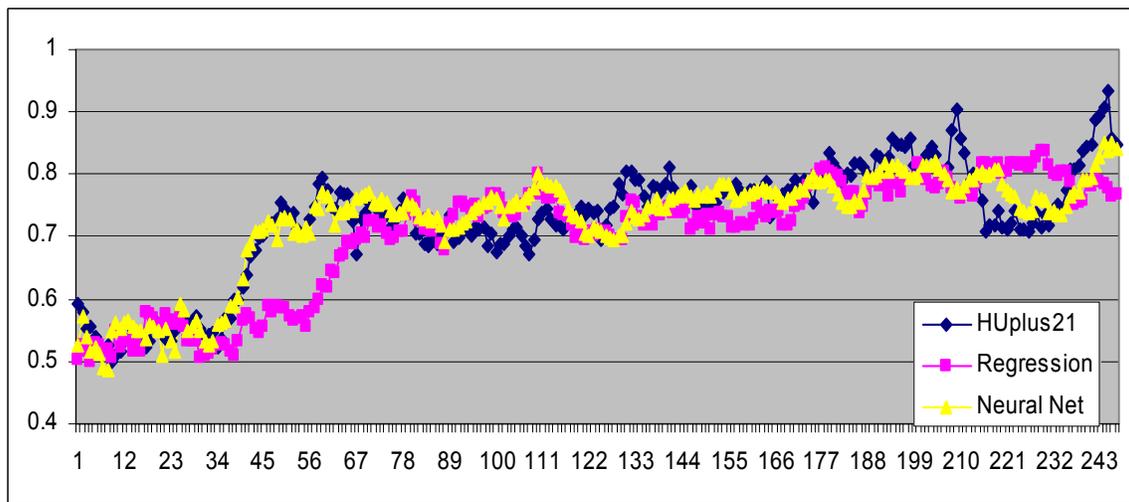
Graph 1. Crude Oil



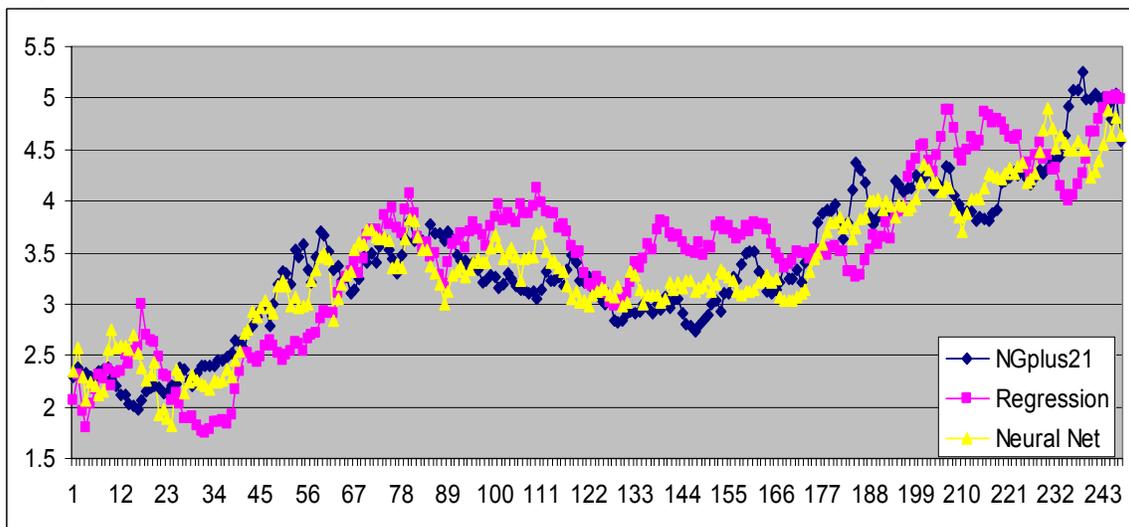
Graph 2. Heating Oil



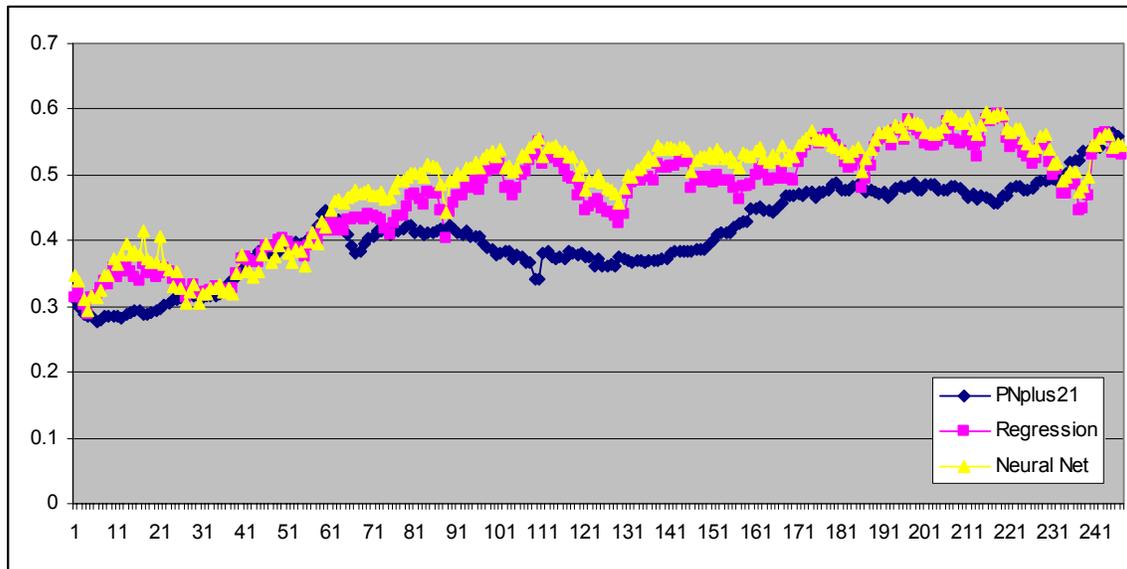
Graph 3. Gasoline



Graph 4. NaturalGas



Graph 5. Propane



### Conclusion and Suggestions for Future Research

While the nonlinear models derived by the neural network provided superior forecasting in a majority of cases, there remain undeniable asymmetries in the predictive abilities of the three models examined. For crude oil, heating oil, gasoline, and natural gas, the neural network gave the best results, consistently boasting a mean squared error less than half that of the regression or simple predictions. However, with propane, the neural network gave the least accurate prediction. Also surprising were the results of the simple model, which predicted that the commodity price 21 trading days into the future would hold unchanged from the present day's price. Except in the case of gasoline, the mean squared error of the simple prediction was lower than that of the regression model.

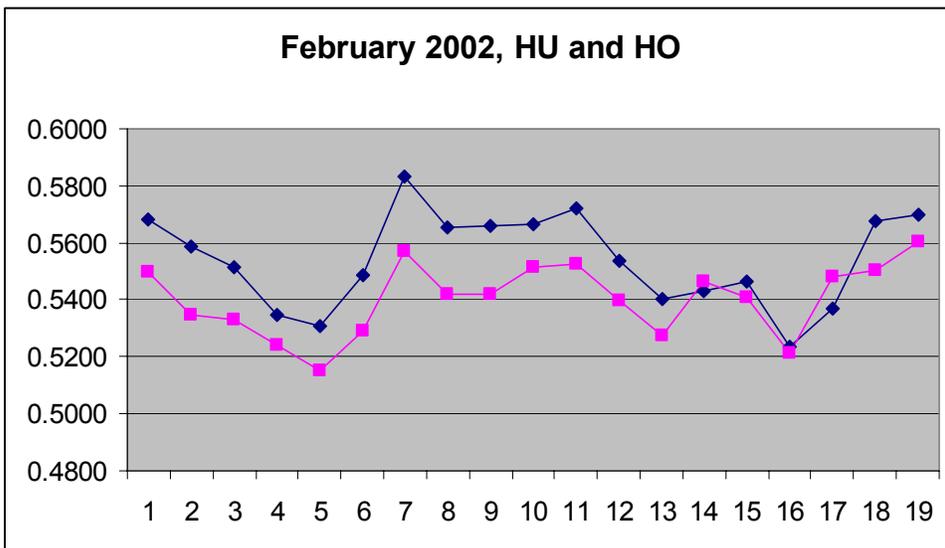
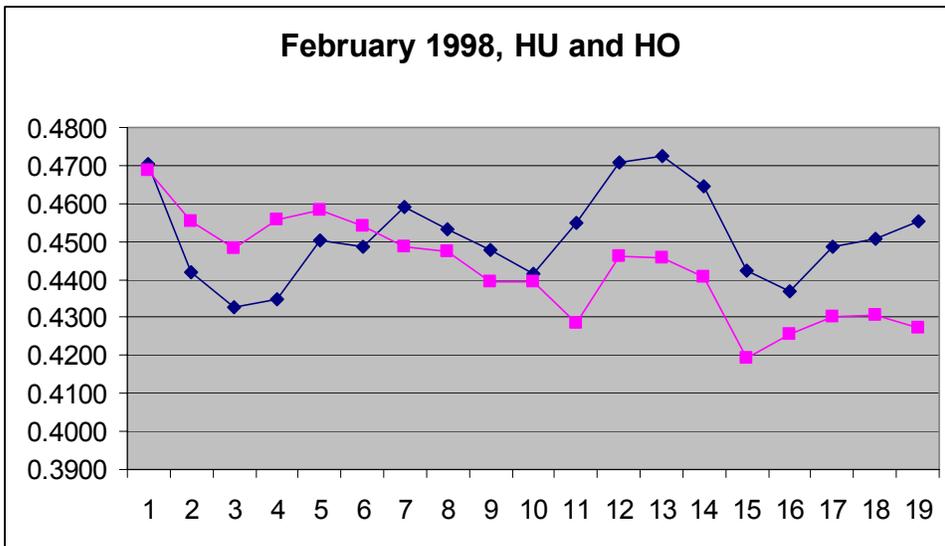
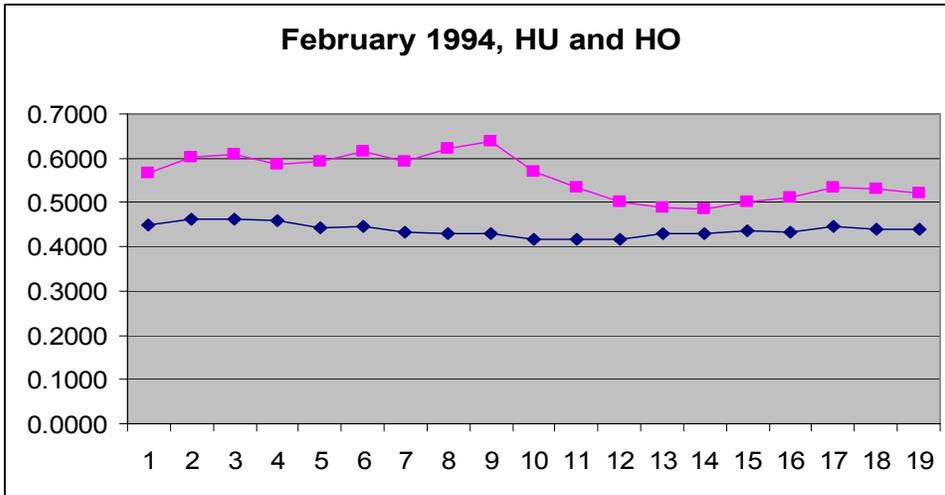
One can draw a number of conclusions from these results. First, it is clear that in a number of—though by no means in all—cases, there is enough information contained in a simple set of price data to allow effective forecasting. That is to say, while the neural network had no extraneous knowledge (say, of news items regarding energy consumption, or foreign wars, or what season it might be), it was nonetheless able to make a reasonably accurate prediction for four out of five energy products.

Second, an ability to predict the price of a given source good does not necessarily imply an ability to predict the price of such a good's byproducts. This asymmetry is exhibited by the neural network model's inability to effectively price propane futures. Propane, as we recall from the introduction, is produced by processing or refining natural gas or crude oil; the model is generally capable of predicting prices for both these raw materials. But when we try to turn the same model to predicting the price of propane itself, not only does the model produce less-than-stellar results, but it would be bested both by a standard regression and by a trader following the simplistic idea that prices would remain constant.

Lastly, we observe that traditional statistical techniques for understanding and extracting information about trends are often less than ideal in market situations, despite their otherwise broad acceptance by the profession. Regression analysis might be a common tool in analysis of market movements, but the results show that in energy markets, regressions are not appreciably better at predicting future movements than models making the blanket assumption of a flat market.

This paper has concerned itself with an examination of the performance of models simulating the actions of a trader concerned with only the most basic of information: present and historical commodity prices. As we have seen, while such simple information can often provide comparatively excellent predictions, some surprising asymmetries remain. Future work along these same lines might begin to explain the root causes of these asymmetries by examining more closely the relationships between these various energy products, considering specifics such as time-to-market, refining and processing costs, change-over costs for energy users switching from one fuel to another, and volume of usage and cyclical or seasonal factors.

Appendix 1. Prices of Gasoline and Heating Oil, February 1994, 1998, and 2002.



## **Bibliography**

- Asche, F., O. Gjolberg, and R. Volker, “Price Relationships in the Petroleum Market: An analysis of crude oil and refined product prices”, Discussion Paper #D-04/2001, Department of Economics and Social Sciences, Agricultural University of Norway.
- Brown, S., M. Yucel, and J. Thompson, “Business Cycles: The Role of Energy Prices”, *Federal Reserve Bank of Dallas*, Research Department Working Paper 0304.
- Department of Energy, “Annual Energy Outlook 2004 with Projections to 2025”, [www.eia.doe.gov/oiaf/aeo/economic.html](http://www.eia.doe.gov/oiaf/aeo/economic.html)
- Energy Information Administration [EIA1], “A Primer on Gasoline Prices”, [www.eia.doe.gov/pub/oil\\_gas/petroleum/analysis\\_publications](http://www.eia.doe.gov/pub/oil_gas/petroleum/analysis_publications)
- Energy Information Administration [EIA2], “Derivatives and Risk Management in the Petroleum, Natural Gas, and Electricity Industries”, U.S. Department of Energy, October 2002, SR/SMG/2002-01.
- Energy Information Administration [EIA3], “Energy Price Impacts on the U.S. Economy”, April 2001, [www.eia.doe.gov/oiaf/economy/energy\\_price.html](http://www.eia.doe.gov/oiaf/economy/energy_price.html)
- Energy Information Administration [EIA4], “Propane Prices, What Consumers Should Know”, <http://tonto.eia.doe.gov/oog/info/hopu/hopu.asp>.
- Herard, Bruce and D. Taylor, “Reducing earning volatility: a short course in energy cost risk management”, *Chemical Market Reporter*, Sept 8, 2003.
- Heselbarth, R., “Oil prices hitting homes”, *Contractor*, July 2000.
- NYMEX [New York Mercantile Exchange], “Risk Management with Natural Gas Futures and Options, [www.nymex.com](http://www.nymex.com).
- NRG News Energy Shop, “What Drives Prices?”, May 2004, [www.energystore.com/energystore/pricearticles.cfm](http://www.energystore.com/energystore/pricearticles.cfm)
- Olson, Dennis, and C. Mossman, “Neural network forecasts of Canadian stock returns using accounting ratios”, *International Journal of Forecasting*, Vol. 19, Issue 3, July-September 2003, pages 453-465.
- Smith, K, and J. Gupta, Neural Networks in Business: Techniques and applications, Idea Group Publishing, 2002.

Trippi, R. and E. Turban, Neural Networks in Finance and Investing, revised edition, Irwin, 1996.

Williams, J. and A. Alhajji, "The Coming Energy Crisis?", *Energy Economics Newsletter*, 2003, [www.wtrg.com/EnergyCrisis/index.html](http://www.wtrg.com/EnergyCrisis/index.html).