Industry Trends Can Help Predict Future Generation Mix

By: Mary Graham, Project Manager
Utility Engineering Corporation

Using historical data can provide a basis for estimating future electrical generation market trends. This paper will examine the historical statistics in the fossil fuel electrical power generation sector, specifically historical coal and natural gas generation, fuel prices and CO2 emissions data. This analysis offers only a glimpse of the total historical electric power picture as only the following historical electric power variables are examined.

- Coal Electrical Net Generation[1]
- Natural Gas Electrical Net Generation[2]
- Coal Prices (Nominal)[3]
- Natural Gas Prices (Nominal)[4]
- Total CO2 Emissions[5]

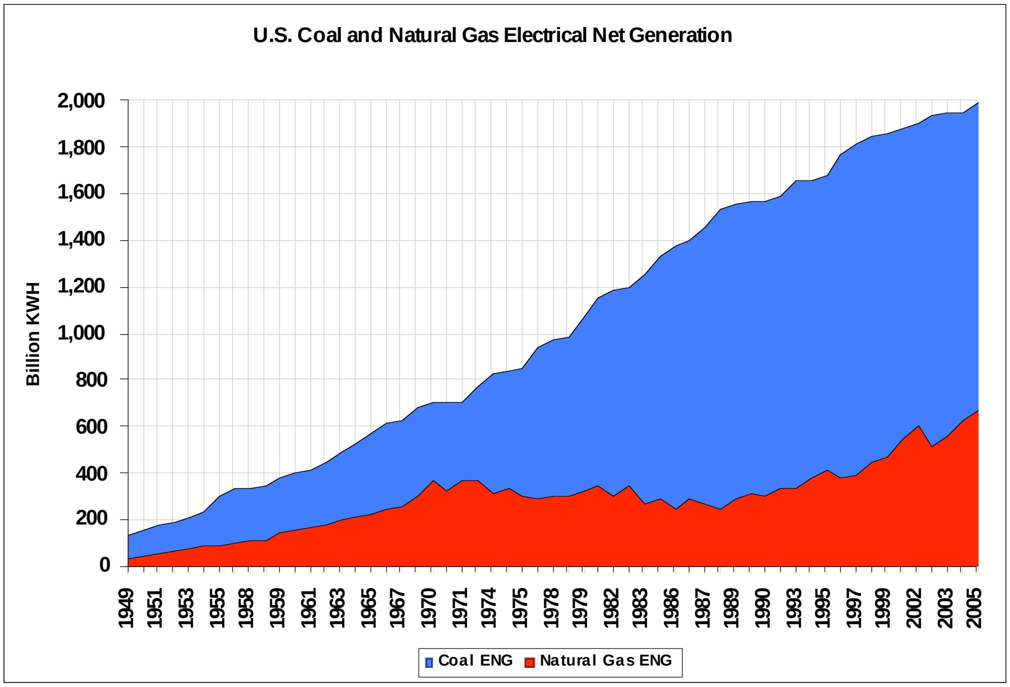
The data for this paper was collected from the Department of Energy, Energy Information Administration website (EIA)[6]. The data spans a 56-year period from 1949 to 2005 and contains several different independent variables that are ratio-level in nature; that is, the value of zero has meaning. The variables' relationships with economic market drivers such as population growth, inflation, interest rates and oil prices have not been considered.

The analysis outlined below shows that these model results are statistically significant and indicate that future natural gas generation can be predicted by Coal Electrical Net Generation, Coal Price, Natural Gas Price and Total CO2 Emissions. However, because some relationships are uncertain (for example Coal ENG and Natural Gas ENG) it is recommended to include other historical database variables related to electric power generation such as nuclear generation and renewable generation, which were not considered. Also, analyzing the relationships of variables that might not have obvious or historical relationships with electric generation fossil fuel data such as population growth and weather data is recommended.

Descriptive Statistics

The first step in examining the historical statistical trends is to examine the raw data of each variable and determine the distribution of the data.

As shown by the net generation graph below, Coal Electric Net Generation (Coal ENG) steadily increases annually while the Natural Gas Electric Net Generation (Natural Gas ENG) increase is sporadic year to year.

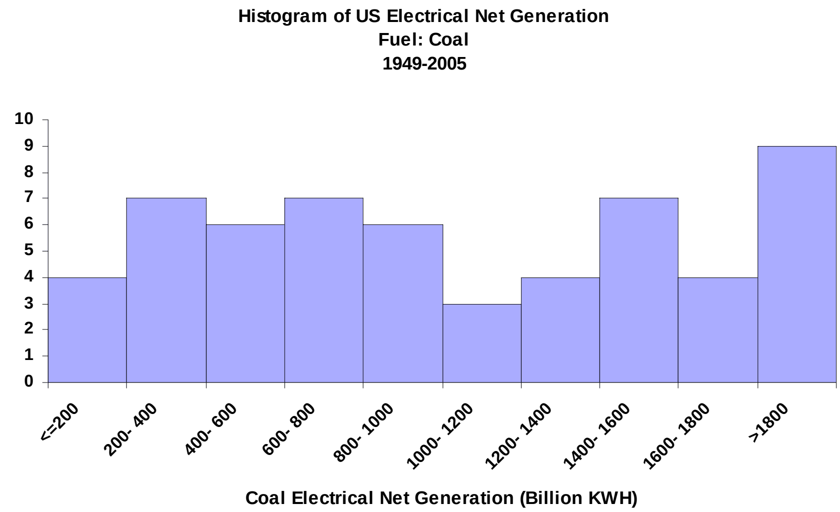


To determine whether trends exist in the Coal ENG and Natural Gas ENG data, I evaluated the distribution (normality) of the data. Descriptive statistics and histograms are utilized for this evaluation.

Coal ENG

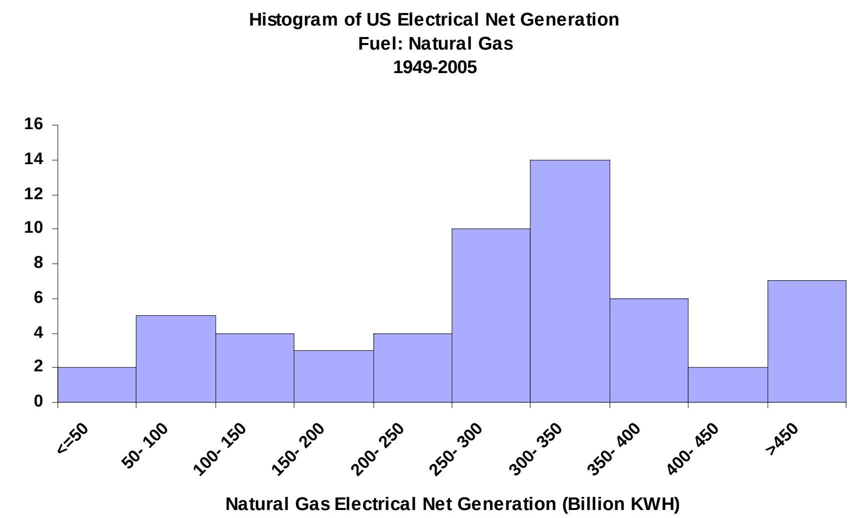
According to the EIA a total of 59,246.30 billion KWH of Coal ENG was measured over the 56-year period. The Coal ENG data has a negative kurtosis value of -1.41, which indicates the data distribution is flatter than that of a normal distribution. This is due to the steady increase of coal generation throughout history (about 4.9 percent a year) as shown by the above graph. The skewness value of the coal data, 0.10, indicates that the value of the distribution of the coal electric generation is slightly positively skewed; that is, the centroid of the data is skewed toward the higher values of the generation values. This data distribution is indicative that Coal ENG data can be compared to the other historical variables and captured in a model.

Coal ENG	
Count	57
Sum	59,246.30
Mean	1,039.41
Median	975.74
Standard deviation	602.12
Maximum	1,857.08
Minimum	135.45
Range	1,992.53
Variance	362,554.34
Skewness	0.10
Kurtosis	(1.41)



Natural Gas ENG
A total of 16,878.49 billion KWH of Natural Gas ENG was measured over the same 56-year period. The skewness value of 0.39 indicates that the value of the distribution of the ENG Natural Gas variable is more positively skewed than that of the Coal ENG data which is evident in the histogram. The positive kurtosis value of 0.15 indicates that the distribution is more peaked than that of a normal distribution and more peaked than that of the coal generation data. This is most likely due to sporadic increase of gas generation over the years as shown by the above net generation graph. The several peaks in the data lead to the possibility that there are distinct periods in the data range that can be captured in a model.

Natural Gas ENG	
Count	57
Sum	16,878.49
Mean	296.11
Median	304.43
Standard deviation	148.56
Maximum	638.10
Minimum	36.97
Range	675.07
Variance	22,070.89
Skewness	0.39
Kurtosis	0.15



For brevity purposes, the descriptive statistics and histograms for the other variables studied in this paper, Coal Prices, Natural Gas Prices, and Total CO2 Emissions are located in Appendix A of this report.

Correlation and Scatterplot Analysis

The correlation between two variables indicates the strength and direction of their linear relationship. Using historical coal and natural gas generation, fuel prices and CO2 emissions data, a correlation table was generated to observe these relationships. It is interesting to note that there are no inverse (or negative) correlations between the variables. Also, the majority of the variables have strong relationships (values close to 1.000) indicating their probability of a fit within a predictive model.

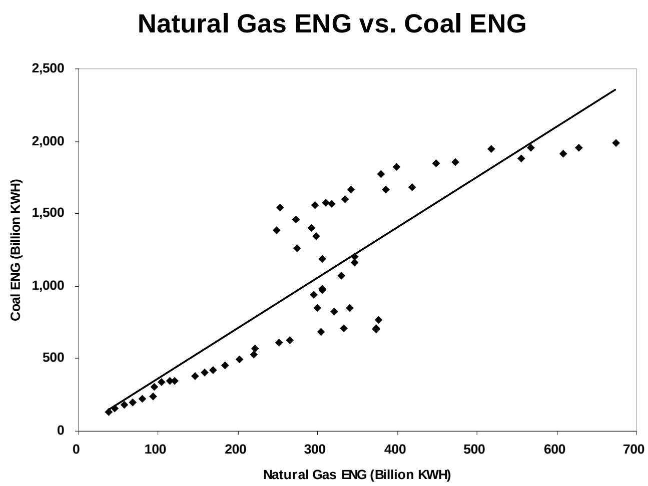
A correlation table containing other relevant historical electric power variables such as nuclear generation and renewable generation[7] is included in Appendix B of this paper for information purposes.

Table of Correlations	Coal ENG	Natural Gas ENG	Coal $ (Nom)	Nat. Gas $ (Nom)	Total CO2
Coal ENG	1.000	0.857	0.788	0.791	0.991
Natural Gas ENG		1.000	0.560	0.689	0.894
Coal $ (Nom)			1.000	0.559	0.802
Natural Gas $ (Nom)				1.000	0.805
Total Fossil Fuel CO2					1.000

To supplement the investigation between the relationships of the variables, scatterplot diagrams are used to evaluate linear relationships between the variables.

Coal ENG vs. Natural Gas ENG

Because, Coal ENG and Natural Gas ENG have a relatively high correlation factor of 0.857, these variables will be examined first. As graphically shown on the scatterplot diagram, a linear relationship is revealed. The plot shows a slight deviation from linear in the center of the data where Natural Gas ENG peaks, however the coefficient of determination (R^2) indicates a 73.39 percent explanation of variation, which points toward only a modest linear relevance. Clearly, there exists a relationship between Coal ENG and Natural Gas ENG. When modeling future fossil fuel electrical generation, it is obvious that both of these variables must be taken into account.

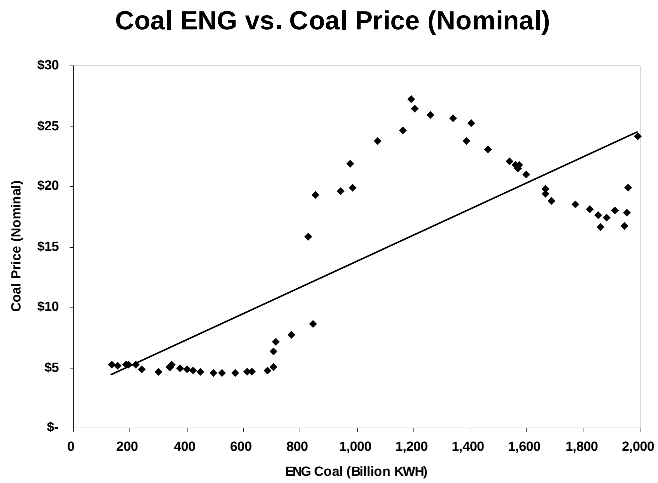


Trend Line Equation: $y = 3.4721x + 11.274$
R^2 Value = 0.7339

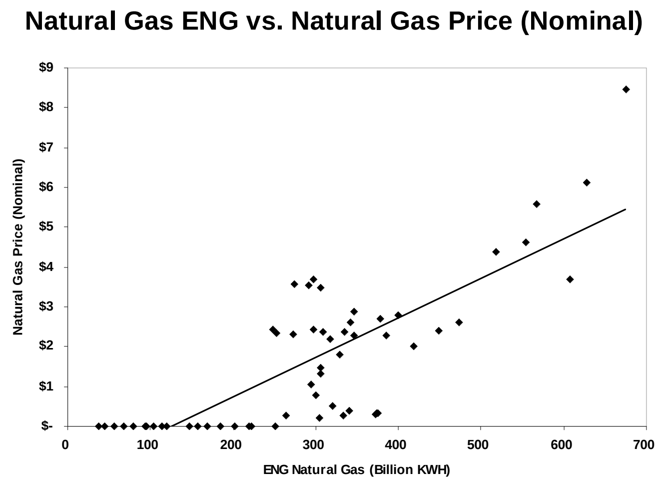
Coal ENG vs. Coal Price and Natural Gas ENG vs. Natural Gas Price

With Coal ENG and Coal Prices, the correlation value of 0.788 reveals a linear relationship to some extent between the two variables. However, when you examine the scatterplot, the relationship appears to be curvilinear. The linear relationship between Coal ENG and Coal Prices also appears to be in question with an R^2 value of 0.6215. When modeling future coal electrical generation, it can be assumed that these variables have some impact on each but it is also evident that other variables need to be considered.

The Natural Gas ENG and Natural Gas Prices relationship has a correlation value of 0.689, therefore the linear relationship between these two variables is relatively unclear. This uncertainty is evident in the scatterplot diagram below where the data appears to be co-located at the Natural Gas ENG peaks. The diagram also contains several data points that are outliers (anomalies) which may be due to the data inaccuracy or the distinct peaks in the generation data. Regardless, with the R^2 value of 0.6341, only a little over half of the variation in the data is explained. It would be difficult to model future natural gas without including other significant variables.



Trend Line Equation: y = 0.0108x + 2.9142
R^2 Value = 0.6215

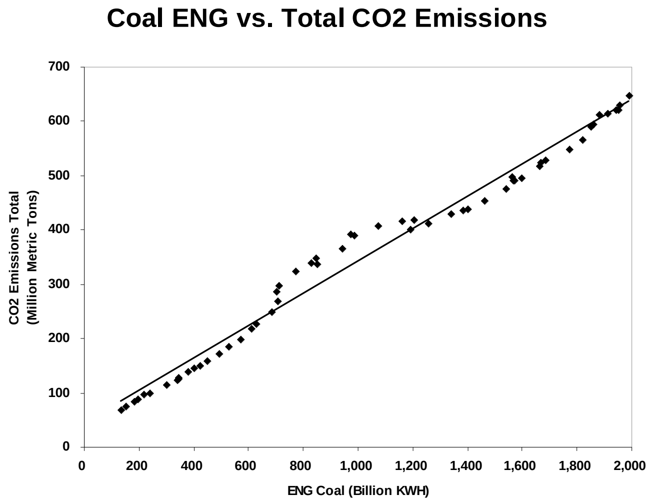


Trend Line Equation: 0.0099x - 1.2677
R^2 Value = 0.6341

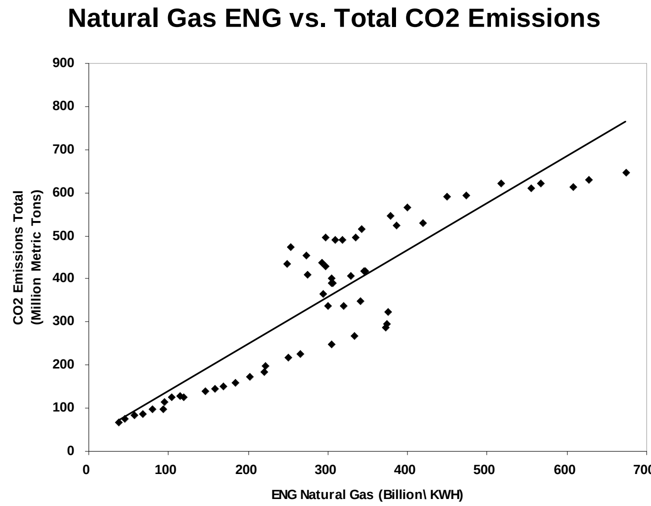
<u>*Coal ENG vs. Coal CO2 Emissions and Natural Gas ENG vs. Natural Gas CO2 Emissions*</u>

The linear relationship between Coal ENG and Total CO2 Emissions is not a surprise. As observed in the correlation table, the linear relationship between Coal ENG and Total CO2 is close to a perfect value at 0.991. The scatter plot diagram confirms this observation as the data is aligned in a straight line with 98.14 percent of the variance in the data is explained. It is obvious that when modeling future coal CO2 emissions, coal generation is the significant variable.

Interestingly enough, a linear relationship between Natural Gas ENG and Total CO2 Emissions is evident but not as significant as coal generation. The correlation table gives a value of 0.894. However, graphically the data is similar to the Natural Gas ENG vs. Coal ENG scatterplot diagram where there is a slight deviation from linear in the center of the data where Natural Gas ENG peaks. The resultant R^2 value indicates a 79.88 percent explanation of variation which indicates linear relevance. Not a significant relevance, but a relationship nonetheless.



Trend Line Equation: y = 0.2974x + 42.62
R^2 Value = 0.9814



Trend Line Equation: y = 1.0875x + 29.722
R^2 Value = 0.7988

Multiple Regression Model

Trend lines in the scatterplots diagrams can be simply approximated through a set of data points, but to properly calculate the slope of the trend line and create a model to simulate future values, the course of action is linear regression. Linear regression will model the relationship by varying independent variable values such as fuel prices and emissions data with a dependent variable such as Natural Gas ENG.

After analyzing the Natural Gas ENG through descriptive statistics and scatterplot diagrams, it will be interesting to see if the other variables impact future natural gas generation trends. Natural Gas ENG will be the dependent variable while Coal ENG, Natural Gas Price (nominal), Coal Price (nominal) and Total CO2 Emissions will be the independent variables. Additionally, based on the scatterplots above, Natural Gas ENG and Coal ENG possibly have a curvilinear relationship so a pseudo-variable was created (Coal ENG ^{2}), hopefully to capture this association.

With a p-value of 0.2 used as used as a cut-off point for the multiple regression outcomes, the results were found to be significant. Complete multiple regression results shown in Appendix C.

- The multiple regression models were run twice: The first time with the Coal ENG ^{2} variable and the second time without. Even though the p-value for the Coal ENG ^{2} coefficient is less than 0.2 for the first regression, the resultant coefficient had a 0 zero value, therefore this variable was dropped.
- The first regression model resulted in an adjusted R^2 of 93.00%.
- The second regression model resulted in an adjusted R^2 92.36%. Removing the Coal ENG ^{2} proved to have little significance.
- In both regression models the p-values of the dependent variables (regression coefficients) were all less than 0.2. The second regression coefficients p-values were substantially smaller.

The Regression equation for the final model is below. The resulting equation can be utilized to simulate future natural gas electrical generation (Natural Gas ENG) by varying the independent variable values.

ENG Natural Gas = 112.1 + [-0.4*Coal ENG] + [-10.2*Coal Price] + [- 39.4*Natural Gas Price] + [2.1*Total CO2]

For example, when the independent variables are held at their mean values, the predicted value of Natural Gas ENG (billion KWH) is as follows:

ENG Natural Gas	C	Coal ENG	Coal Price	NG Price	Total CO2
342.48	112.10	1,039.41	$ 14.15	$ 2.44	351.76

Practical Conclusions

By examining the historical descriptive statistics for fossil fuel variables independently, it was found that there is an opportunity to capture the data in a model. To further reveal the data characteristics, the relationships between the variables were looked at through their correlation values and scatterplot diagrams. There are several results that are particularly noteworthy:

- The correlation value between Coal ENG and Natural Gas ENG is a reasonably strong value at 0.857 indicating a strong relationship. This is validated by the scatterplot diagram which indicates a linear relationship with a modest R^2 value of 73.39%.
- Although the correlation value between Coal ENG and Coal Price showed is 0.788, the scatterplot indicates basically no linear relationship. In fact, it may be interpreted that the relationship is curvilinear.
- The linear relationship between Natural Gas ENG and Natural Gas Price is relatively unclear. This uncertainty could be a result of the sporadic increase and decrease in natural gas generation or imprecise data.
- The linear relationship between Coal ENG and CO2 Emissions is perfectly linear which, is a contrast to the relationship between Natural Gas ENG and CO2 Emissions which has a modest linear relationship at best.

In summary, it is obvious that the relationships between the data variables have some impact on each other. To capture and model these relationships, two multiple regressions were completed. Choosing to look at the Natural Gas ENG as the dependent variable, one regression was run with all of the data plus a pseudo-variable created to capture the possible curvilinear relationship between Coal ENG and Coal Prices; and another regression was completed with all of the data minus pseudo-variable.

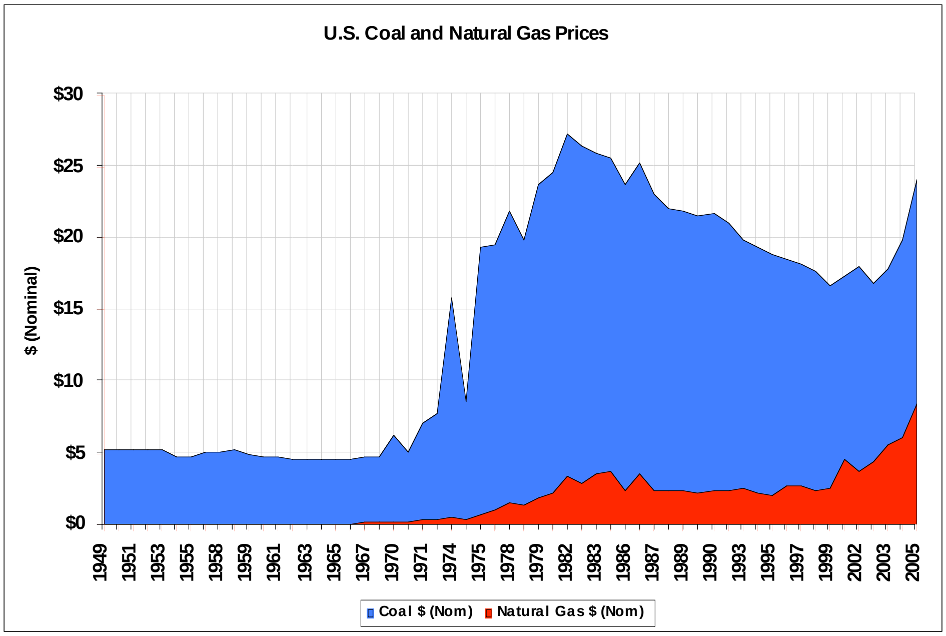
The results of the multiple regressions were favorable. Using a p-value of 0.2 for a cut-off point for the for the resulting multiple regression coefficients, the resulting adjusted R^2 values were noticeably high values:

- Regression model one (including the pseudo-variable) resulted in an adjusted R^2 explaining 93.00% of the variance in the Natural Gas ENG variable.
- Regression model two (excluding the the pseudo-variable) resulted in an adjusted R^2 that explained 92.36% of the variance in the Natural Gas ENG variable. Removing the Coal ENG ^{2} proved to have little significance which possibly indicates that a further examination between the Coal ENG and Natural Gas ENG variable is warranted.

These model results are statistically significant and indicate that future natural gas generation can be predicted by Coal ENG, Coal Price, Natural Gas Price, and Total CO2 Emissions. However, because some relationships are uncertain (i.e. Coal ENG and Natural Gas ENG) it is recommended in to include other historical database variables related to electric power generation such as nuclear generation and renewable generation which, were not considered. Also, analyzing the relationships of variables that might not have obvious or historical relationships with electric generation fossil fuel data such as population growth and weather data is recommended.

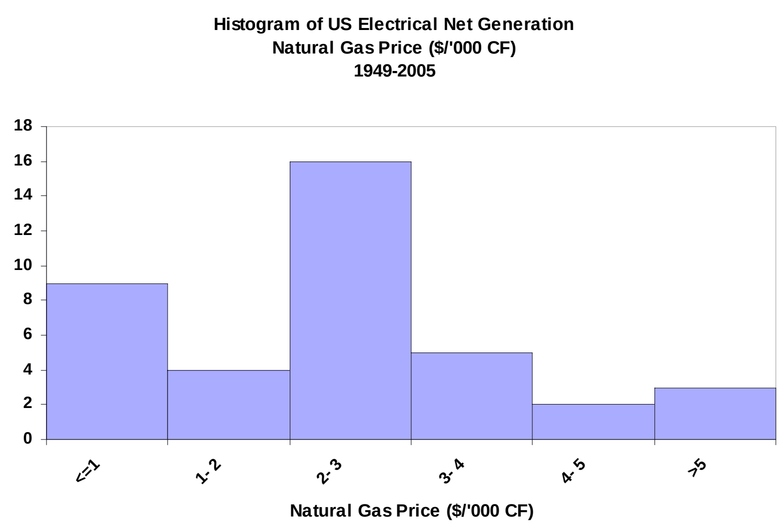
Appendix A - Descriptive Statistics (expanded)

US Coal and Natural Gas Prices, Raw Data

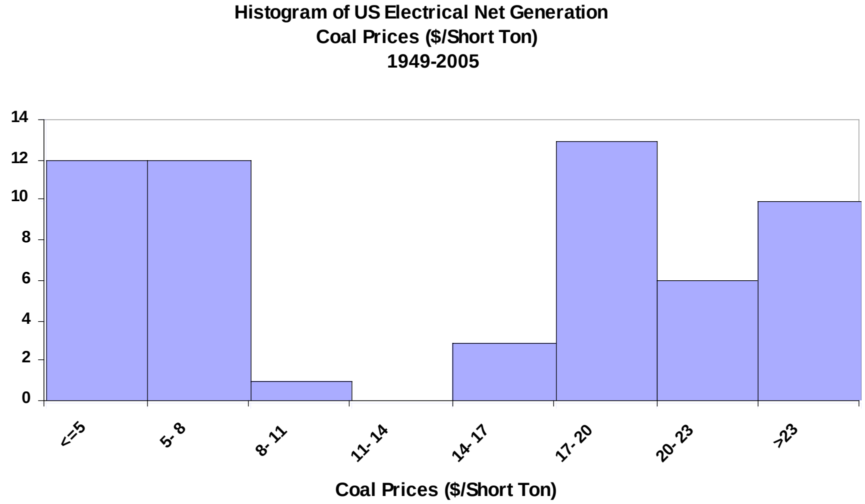


US Coal and Natural Gas Prices, Descriptive Statistics

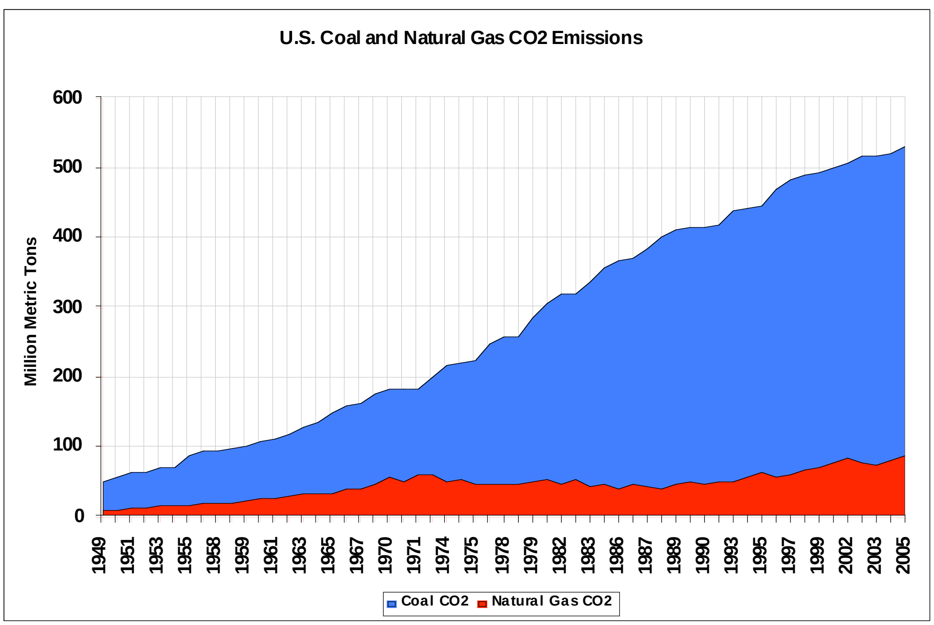
Natural Gas Prices		
Count	$	39.00
Sum		
Mean	$	2.44
Median	$	2.36
Standard deviation	$	1.77
Maximum	$	8.23
Minimum	$	0.22
Range		8.45
Variance		3.12
Skewness		1.19
Kurtosis		2.47



Coal Prices		
Count	$	57.00
Sum		
Mean	$	14.15
Median	$	17.38
Standard deviation	$	8.25
Maximum	$	22.70
Minimum	$	4.55
Range		27.25
Variance		68.12
Skewness		(0.02)
Kurtosis		(1.72)

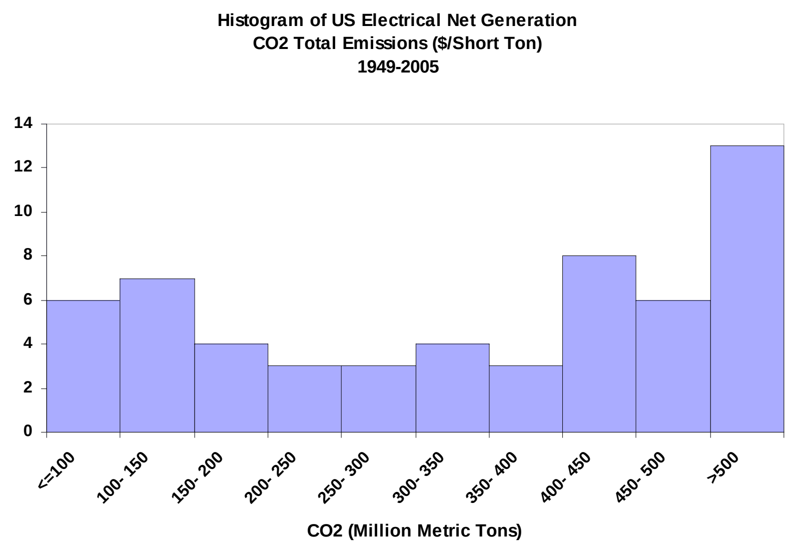


US Coal and Natural CO2 Emissions, Raw Data

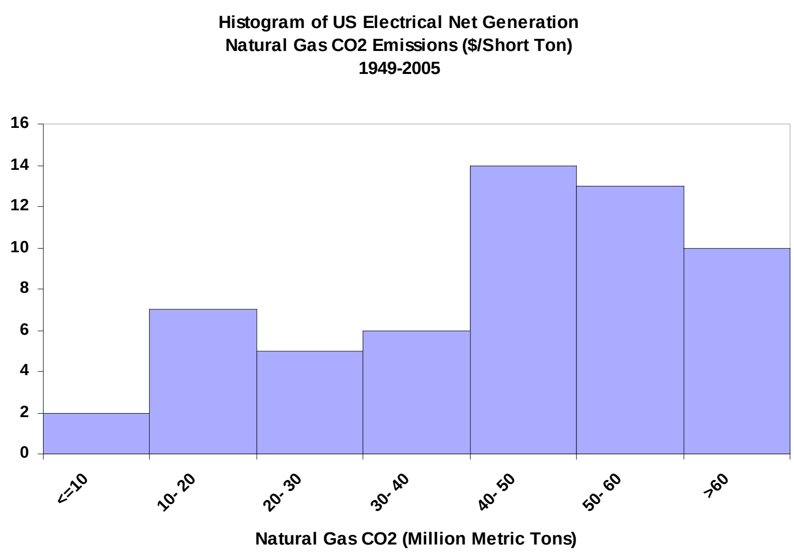


US Coal and Natural CO2 Emissions, Descriptive Statistics

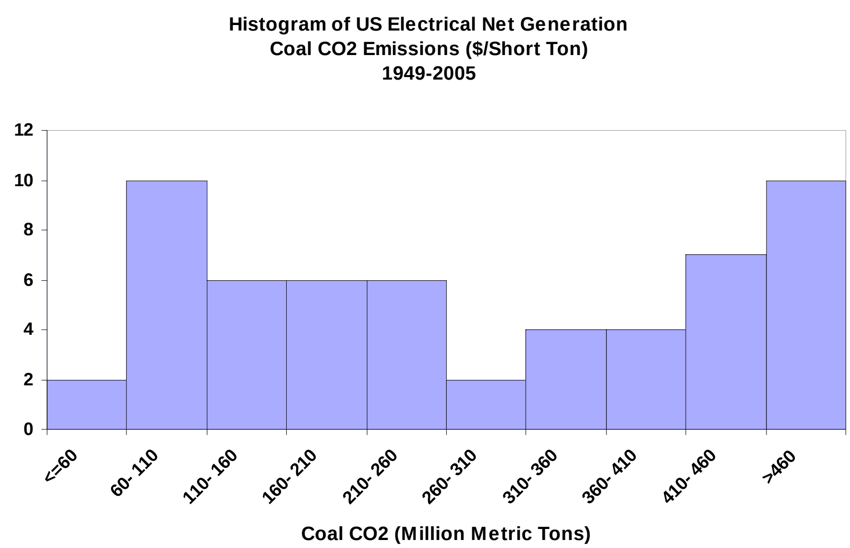
CO2 Total	
Count	57
Sum	20,050.33
Mean	351.76
Median	389.55
Standard deviation	180.77
Maximum	580.14
Minimum	67.58
Range	647.72
Variance	32,679.27
Skewness	(0.06)
Kurtosis	(1.30)



Natural Gas CO2	
Count	57
Sum	2,573.54
Mean	45.15
Median	46.98
Standard deviation	19.70
Maximum	78.78
Minimum	8.21
Range	86.98
Variance	387.93
Skewness	0.02
Kurtosis	(0.43)



Coal CO2	
Count	57
Sum	15,730.84
Mean	275.98
Median	259.60
Standard deviation	157.53
Maximum	479.56
Minimum	50.68
Range	530.24
Variance	24,816.72
Skewness	0.15
Kurtosis	(1.44)



Appendix B – Table of Correlations (expanded)

Table of Correlations	Coal ENG	NG ENG	Fossil Fuel ENG Total	Nuclear ENG	Renew ENG Total	ENG Total	Coal $ (Nom)	Coal $ (Real)	NG $ (Nom)	NG $ (Real)	NG CO2	Coal CO2	Total CO2
Coal ENG	1.000	0.857	0.984	0.982	0.886	0.993	0.788	(0.159)	0.791	0.570	0.857	1.000	0.991
NG ENG		1.000	0.916	0.756	0.812	0.894	0.560	(0.225)	0.689	0.406	0.993	0.851	0.894
Fossil Fuel ENG Total			1.000	0.948	0.914	0.996	0.788	(0.092)	0.820	0.588	0.921	0.981	0.998
Nuclear ENG				1.000	0.759	0.970	0.672	(0.232)	0.752	0.512	0.743	0.983	0.958
Renew ENG Total					1.000	0.905	0.777	0.040	0.443	0.320	0.845	0.879	0.912
ENG Total						1.000	0.789	(0.117)	0.797	0.563	0.896	0.992	0.999
Coal $ (Nom)							1.000	0.442	0.559	0.755	0.576	0.789	0.802
Coal $ (Real)								1.000	(0.246)	0.142	(0.189)	(0.160)	(0.089)
NG $ (Nom)									1.000	0.901	0.615	0.796	0.805
NG $ (Real)										1.000	0.336	0.577	0.585
NG CO2											1.000	0.850	0.899
Coal CO2												1.000	0.989
Total CO2													1.000

Appendix C – Multiple Regression Results

Regression 1

Results of multiple regression for ENG_NaturalGas_2

Summary measures

Multiple R	0.9691
R-Square	0.9392
Adj R-Square	0.9300
StErr of Est	28.77

ANOVA Table

Source	df	SS	MS	F	p-value
Explained	5.00	422,230.95	84,446.19	102.00	0.00
Unexplained	33.00	27,320.34	827.89		

Regression coefficients

	Coefficient	Std Err	t-value	p-value	Lower limit	Upper limit
Constant	389.19	140.37	2.77	0.0091	103.60	674.77
ENG_Coal_2	(0.91)	0.24	(3.82)	0.0006	(1.40)	(0.43)
SQ_ENG_Coal_2	0.00	0.00	2.03	0.0503	(0.00)	0.00
TotalCoal_Nominal	(4.82)	2.79	(1.73)	0.0929	(10.49)	0.85
NaturalGas_Nominal	27.96	7.35	3.81	0.0006	13.02	42.91
Total_CO2	1.82	0.27	6.74	0.0000	1.27	2.37

Regression 2

Results of multiple regression for ENG_NaturalGas_2

Summary measures

Multiple R	0.9652
R-Square	0.9316
Adj R-Square	0.9236
StErr of Est	30.07

ANOVA Table

Source	df	SS	MS	F	p-value
Explained	4.00	418,813.57	104,703.39	115.82	0.00
Unexplained	34.00	30,737.73	904.05		

Regression coefficients

	Coefficient	Std Err	t-value	p-value	Lower limit	Upper limit
Constant	112.10	34.72	3.23	0.0028	41.54	182.66
ENG_Coal_2	(0.44)	0.06	(7.12)	0.0000	(0.57)	(0.32)
TotalCoal_Nominal	(10.19)	0.92	(11.05)	0.0000	(12.07)	(8.32)
NaturalGas_Nominal	39.38	4.94	7.97	0.0000	29.33	49.43
Total_CO2	2.10	0.24	8.67	0.0000	1.61	2.59

Notes

[1] Energy Information Administration Annual Energy Review 2006, Table 8.2b Electricity Net Generation: Electrical Power Sector, 1949-2005, Subset of Table 8.2a (Thousand Kilowatthours). Coal is defined as anthracite, bituminous coal, subbituminous coal, lignite, waste coal, and coal synfuel.

[2] Energy Information Administration Annual Energy Review 2006, Table 8.2b Electricity Net Generation: Electrical Power Sector, 1949-2005, Subset of Table 8.2a (Thousand Kilowatthours). Natural gas is defined as natural gas, plus a small amount of supplemental gaseous fuels that cannot be identified separately.

[3] Energy Information Administration Annual Energy Review 2006, Table 7.8 Coal Prices, 1949-2005 (Dollars per Short Ton).

[4] Energy Information Administration Annual Energy Review 2006, Table 6.8 Natural Gas Wellhead, City Gate, and Import Prices, 1949-2005 (Dollars per Thousand Cubic Feet)

[5] Energy Information Administration Energy Related Emissions Data and Environmental Analysis, *http://www.eia.doe.gov/environment.html*. Total Energy-Related Carbon Dioxide Emissions by End-Use Sector, and the Electric Power Sector, by Fuel Type, 1949-2005.

[6] Energy Information Administration website, *http://www.eia.doe.gov/*. Historical energy data dating back to 1949, *http://www.eia.doe.gov/emeu/aer/elect.html*.