How does the increase in renewable power generation in the US affect natural gas demand and prices?

A study examining the key determinants impacting electric and energy consumption across 25 OECD nations over the period from 1978 to 2004 peaked my interest into renewable power generation. The study leverages panel data analysis to unravel the relationship between rising income levels and energy consumption patterns. A pivotal finding of this study is the nuanced impact of energy conservation on economic growth, suggesting that while it may slow growth, it does not inherently damage it (Lee, Chien-Chiang, and Jun-De Lee, 2010). The analysis underscores the importance of sustained investment in alternative energy sources as a strategic measure to curtail wasteful practices, thereby fostering a steady trajectory of energy development in these countries.

Building on the understanding of energy dynamics in OECD countries, the scope of this research extends to analyzing energy expenditure patterns in the southern United States over a two-decade period. This particular analysis reveals a notable trend: despite a decrease in overall energy costs in 2006 compared to 1984, the most significant drop in energy spending occurred around 1985, coinciding with the global decline in petroleum prices. Intriguingly, the proportion of total expenses dedicated to energy in the South has remained relatively stable since 1984 (Abbot, 2008).

Complementing the analysis of energy consumption and expenditure trends in the southern United States and OECD countries, this study delves into the realm of renewable energy, specifically examining the development of wind energy systems for offshore applications outside the European Union. The research highlights the critical factors in offshore wind farm development, including wind conditions, potential for extreme weather, water depth, and proximity to land. The findings indicate that an expansive area of over 275,000 km^2 is suitable for offshore wind energy, with plans to install over 300,000 turbines (Siegfriedsen, 2003).

Furthering the discussion on renewable energy solutions, the focus shifts to Egypt's strategy for meeting its escalating energy demands. This study emphasizes the dual approach of enhancing energy efficiency in power plants and distribution networks, alongside diversifying renewable energy sources. The primary focus is on harnessing wind energy, with an eye towards future expansion through regional interconnection of electrical grids with neighboring countries. This approach not only addresses conservation and loss reduction but also mirrors the global trend towards sustainable energy practices, as seen in the offshore wind energy initiatives in Poland and the Baltic states (El-Kholy, Hosni, and Ragy Faried, 2011).

Culminating together to build on the themes of renewable energy adoption and efficiency from Egypt's national strategy to the innovative offshore wind projects in Europe, these global trends resonate on a more personal level through my grandfather's experience with solar energy. His decision to transition to solar power at his home not only reflected a microcosm of the larger shift towards sustainable energy globally (Garvey, 2015), but also sparked my deeper interest in renewable energy systems. This personal connection was further strengthened when I had the opportunity to work on a dataset from Schneider Electric. Engaging with this data allowed me to explore the practical applications of renewable energy technologies in a real-world context, bridging the gap between personal experience and professional interest. Working on this project, I set out to see firsthand how energy supply and demand contribute to broader energy trends and how companies like Schneider Electric are pivotal in driving these changes through innovation and data-driven solutions.

Initial specification:

(-) (-) (-) (-) (-) (-)

$$AVGPRICE = \beta_0 + \beta_1 COAL + \beta_2 GAS + \beta_3 NUCLEAR + \beta_4 HYDRO + \beta_5 WIND + \beta_6 SOLAR +$$

(+/-) (+/-)

$$\beta_7 OTHER + \beta_8 DEMAND + \varepsilon$$

Below are the independent variables, symbology, sign expectations, and notes (where applicable)

Ind. Variables	**Symbol**	**Expected Sign**	**Notes/Supporting**
Supply of Coal	COAL	-	Theoretically, coal could have an ambiguous sign, but I expect it to be negative.
Supply of Gas	GAS	-	Supply of gas could also be ambiguous, but here I am going to assume it to be negative.
Supply of Nuclear Energy	NUCLEAR	-	As supply of nuclear increases, Avg Price would theoretically decrease.
Supply of Hydro Energy	HYDRO	-	As supply of hydro increases, Avg Price would theoretically decrease.
Supply of Wind Energy	WIND	-	As supply of wind energy increases, Avg Price would theoretically decrease.
Supply of Solar Energy	SOLAR	-	As supply of supply increases, Avg Price would decrease.
Other Energy	OTHER	-	Excess sources of energy could make Ave Price decrease, but niche methods could see an increase to price.
Demand	DEMAND	+	Demand increases, that puts pressure on price to increase.

And below is the original regression output:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.22447657
R Square	0.05038973
Adjusted R Square	0.05007306
Standard Error	573.890031
Observations	23999

Y = Ave Price

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	376.952054	33.2167654	11.3482469	9.0125E-30
S-coal	-0.2596014	0.01397354	-18.578073	1.6656E-76
S-gas	-0.2187481	0.01427694	-15.32177	9.7265E-53
S-nuclear	-0.2753703	0.01484532	-18.549304	2.8241E-76
S-hydro	0.50239788	0.07865455	6.38739751	1.7183E-10
S-wind	-0.2327061	0.01401016	-16.609803	1.3079E-61
S-solar	-0.239869	0.01373693	-17.461615	7.3871E-68
S-other	-0.2258546	0.01478478	-15.276161	1.9469E-52
D	0.22793233	0.01398919	16.2934659	2.2894E-59

**** Highlighted coefficients indicate signs that do not match expectations****

Almost all of the independent variables match the expected signs according to common sense/theory. However, upon looking over the ANOVA results, hydro did not match my expectations. Hydro I am curious about because I would expect this to be negative. With all independent variable's P-value being very low, this strikes me as problematic. Initially, I am considering that there could be a multitude of factors/impacts that would lead to these results.

Before I go to run tests for multicollinearity, here are our **T-test** results:

Coal:
H_o: $\beta_{Coal}=0$ H_a: $\beta_{Coal}\neq 0$
$p = 1.6656E\text{-}76 < 0.01 = 1\%$
Therefore, we do not accept the null hypothesis and conclude that Coal is significant at less than 1% in determining average price.

Gas:
H_o: $\beta_{Gas}=0$ H_a: $\beta_{Gas}\neq 0$
$p = 9.7265E\text{-}53 < 0.01 = 1\%$
Therefore, we do not accept the null hypothesis and conclude that Gas is significant at less than 1% in determining average price.

Nuclear:
H_o: $\beta_{Gas}=0$ H_a: $\beta_{Gas}\neq 0$
$p = 2.8241E\text{-}76 < 0.01 = 1\%$
Therefore, we do not accept the null hypothesis and conclude that Nuclear is significant at less than 1% in determining average price.

Hydro:
H_o: $\beta_{Hydro}=0$ H_a: $\beta_{Hydro}\neq 0$
$p = 1.7183E\text{-}10 < 0.01 = 1\%$
Therefore, we do not accept the null hypothesis and conclude that Hydro is significant at less than 1% in determining average price.

Wind:
H_o: $\beta_{Wind}=0$ H_a: $\beta_{Wind}\neq 0$
$p = 1.31E\text{-}10 < 0.01 = 1\%$
Therefore, we do not accept the null hypothesis and conclude that Wind is significant at less than 1% in determining average price.

Solar:
H_o: $\beta_{Solar}=0$ H_a: $\beta_{Solar}\neq 0$
$p = 7.3871E\text{-}68 < 0.01 = 1\%$
Therefore, we do not accept the null hypothesis and conclude that Solar is significant at less than 1% in determining average price.

Other:
H_o: $\beta_{Other}=0$ H_a: $\beta_{Other}\neq 0$
$p = 1.9469E\text{-}52 < 0.01 = 1\%$
Therefore, we do not accept the null hypothesis and conclude that Coal is significant at less than 1% in determining average price.

Demand:
H_o: $\beta_{Demand}=0$ H_a: $\beta_{Demand}\neq 0$
p = 2.2894E-59 < 0.01 = 1%
Therefore, we do not accept the null hypothesis and conclude that Demand is significant at less than 1% in determining average price.

F-Test Results:

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	419260767	52407595.9	159.124435	1.68E-262
Residual	23990	7901100940	329349.768		
Total	23998	8320361707			

$\beta_k = \beta_0 + \beta_1 COAL + \beta_2 GAS + \beta_3 NUCLEAR + \beta_4 HYDRO + \beta_5 WIND + \beta_6 SOLAR + \beta_7 OTHER + \beta_8 DEMAND + \varepsilon$

H_o: All $\beta_k = \beta_k$

H_a: At-least one $\beta_k \neq 0$

With such a low p-value, much less than 1%, we can reject the H_o and accept the H_a that there is joint significance amongst the independent variables. (p(1.68E-262) < a=.05 or 5%).

<u>Multicollinearity Testing</u>

Checking for multicollinearity, I ran a correlation matrix to see what independent variables I should be testing to see how correlated they are/are not. The correlation matrix results are below:

	S-coal	*S-gas*	*S-nuclear*	*S-hydro*	*S-wind*	*S-solar*	*S-other*
S-coal	1						
S-gas	0.70637367	1					
S-nuclear	0.13480961	0.10379858	1				
S-hydro	0.25762499	0.32798843	0.00832856	1			
S-wind	-0.4297576	-0.5678394	-0.0727064	-0.1934282	1		
S-solar	0.25881385	0.37094832	0.00780092	0.2378784	-0.2646828	1	
S-other	-0.0137064	-0.0888971	0.00343984	0.0208667	0.0059148	-0.0069976	1

Now that I have an idea for which variables may be correlated, I will run VIF tests for each of them. The formula for this is $VIF(b_i) = \left(\frac{1}{1-R_i^2}\right)$. Here are the VIF results with included formula(s):

Y = Wind **VIF** = 492.8827728	
(1/(1-0.99797112))	
Multiple R	0.99898504
R Square	0.99797112
Adjusted R Square	0.99797053
Standard Error	264.460785
Observations	23999

Y = Gas **VIF** = 2.784168527	
(1/(1-0.64082634))	
Multiple R	0.8005163
R Square	0.64082634
Adjusted R Square	0.64073652
Standard Error	1355.73541
Observations	23999

Y = Coal **VIF** = 2.012817542	
(1/(1-0.50318398))	
Multiple R	0.70935462
R Square	0.50318398
Adjusted R Square	0.50305973
Standard Error	1594.48522
Observations	23999

Variables like demand, Wind, Gas, and Coal are most problematic in terms of multicollinearity. This could affect any regression model's ability to accurately estimate the effects of these individual variables. To address this issue, one may consider removing one or more of the highly collinear variables. While Gas and Coal are showing some signs of multicollinearity, Wind is problematic given such a high VIF. However, due to the significance of the Wind variable, I am not going to remove Wind yet. I am going to run an Omitted Variable Test to further test my assumptions.

Omitted Variable Tests

An Omitted Variable Test (OVT) is a test in which auxiliary regressions are performed to detect whether a key variable has been left out of a model. The purpose of this test is to determine if omitting a theoretically important variable leads to inconsistencies within our model. If there are inconsistencies, such as bias, it can lead to incorrect conclusions can Omitted Variable Bias. I highly suspect this to be the case due to such a low adjusted R^2 in our original regression. Because of this suspicion, I am going to run an OVT on every independent variable. Below are the results of my **informal** OVT tests:

SUMMARY OUTPUT
Demand Omitted

Regression Statistics	
Multiple R	0.19970277
R Square	0.0398812
Adjusted R Square	0.03960106
Standard Error	577.044646
Observations	23999

SUMMARY OUTPUT
Other Omitted

Regression Statistics	
Multiple R	0.20286069
R Square	0.04115246
Adjusted R Square	0.04087269
Standard Error	576.662496
Observations	23999

SUMMARY OUTPUT
Solar Omitted

Regression Statistics	
Multiple R	0.19575591
R Square	0.03832038
Adjusted R Square	0.03803978
Standard Error	577.513493
Observations	23999

SUMMARY OUTPUT
Wind Omitted

Regression Statistics	
Multiple R	0.19866854
R Square	0.03946919
Adjusted R Square	0.03918893
Standard Error	577.168444
Observations	23999

SUMMARY OUTPUT
Hydro Omitted

Regression Statistics	
Multiple R	0.22085009
R Square	0.04877476
Adjusted R Square	0.04849722
Standard Error	574.365849
Observations	23999

SUMMARY OUTPUT
Nuclear Omitted

Regression Statistics	
Multiple R	0.19175489
R Square	0.03676994
Adjusted R Square	0.03648889
Standard Error	577.978844
Observations	23999

SUMMARY OUTPUT

Gas Omitted

Regression Statistics	
Multiple R	0.20272449
R Square	0.04109722
Adjusted R Square	0.04081743
Standard Error	576.679107
Observations	23999

SUMMARY OUTPUT

Coal Omitted

Regression Statistics	
Multiple R	0.19164462
R Square	0.03672766
Adjusted R Square	0.0364466
Standard Error	577.991529
Observations	23999

Out of all the results of the informal omitted variable test, Hydro being omitted showed the least drop in the Adjusted R Square, I am going to choose Hydro for the formal omitted variable test.

Formal OVT:

SUMMARY OUTPUT

Hydro Included

Regression Statistics	
Multiple R	0.22447657
R Square	0.05038973
Adjusted R Square	0.05007306
Standard Error	573.890031
Observations	23999

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	376.952054	33.2167654	11.3482469	9.0125E-30
S-coal	-0.2596014	0.01397354	-18.578073	1.6656E-76
S-gas	-0.2187481	0.01427694	-15.32177	9.7265E-53
S-nuclear	-0.2753703	0.01484532	-18.549304	2.8241E-76
S-hydro	0.50239788	0.07865455	6.38739751	1.7183E-10
S-wind	-0.2327061	0.01401016	-16.609803	1.31E-61
S-solar	-0.239869	0.01373693	-17.461615	7.3871E-68
S-other	-0.2258546	0.01478478	-15.276161	1.9469E-52
D	0.22793233	0.01398919	16.2934659	2.2894E-59

SUMMARY OUTPUT

Hydro Omitted

Regression Statistics	
Multiple R	0.22085009
R Square	0.04877476
Adjusted R Square	0.04849722
Standard Error	574.365849
Observations	23999

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	394.179229	33.1345391	11.8963245	1.527E-32
S-coal	-0.2755248	0.01376076	-20.022492	1.8532E-88
S-gas	-0.235032	0.01405913	-16.717395	2.2083E-62
S-nuclear	-0.2937394	0.01457617	-20.152024	1.4252E-89
S-wind	-0.2493674	0.01377661	-18.100786	9.5833E-73
S-solar	-0.254518	0.01355536	-18.776192	4.2958E-78
S-other	-0.2416033	0.01458983	-16.559713	2.9821E-61
D	0.24454407	0.01375671	17.7763479	3.0432E-70

a) **Actual Bias**: -

b) **Expected Bias**: β_{Wind}*f(β_{Hydro},B_{Wind})

- - -

There is evidence to keep β_{Hydro} in the specification because the actual bias matches the results of the expected bias.

With the results of the multicollinearity testing suggesting that Wind was high multicollinear, it suggested that removing it from the PRL would theoretically improve the Adjusted R^2. Though, running an informal OVT against every independent variable showed no improvement in the Adjusted R^2 value. Because the Adjusted R^2 shows no improvement no improvement when omitting the included independent variables, I am left to conclude that there should be no changes made to the PRL. My PRL remains unchanged:

$$AVGPRICE = \beta_0 + \overset{(-)}{\beta_1 COAL} + \overset{(-)}{\beta_2 GAS} + \overset{(-)}{\beta_3 NUCLEAR} + \overset{(-)}{\beta_4 HYDRO} + \overset{(-)}{\beta_5 WIND} + \overset{(-)}{\beta_6 SOLAR} + \overset{(-)}{\beta_7 OTHER} + \overset{(+)}{\beta_8 DEMAND} + \varepsilon$$

<u>Heteroscedasticity Testing</u>

Heteroscedasticity, or hetsy for short, is a phenomenon in regression testing in which the variability of the dependent variable is not consistent across different levels of the independent variable(s). Ideally, a regression analysis will return results in which the error terms (residuals) have a constant variance, this is known as homoscedasticity. Hetsy violates a classical assumption (A5) in the Gauss Markov Theorem, and as a consequence would result in our model not being best linear unbiased estimates (BLUe).

Hetsy is typically found in cross-sectional data, but can also be present in time series data as well. I suspect that there could be evidence of hetsy here, and so I'm going to perform some auxiliary regressions. The first of two I'll perform is a Park Test, which is used to test for pure heteroscedasticity. I'll also use the White Test, which tests for impure heteroscedasticity. As the correlation matrix showed in the tests for multicollinearity, I am going to run a Park Test against Gas, Coal, and Wind. Here are the results for the Park Tests:

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.42512193	0.26620427	9.11000396	8.8645E-20
ln gas	0.62007887	0.02722435	22.7766312	1.253E-113

<u>H</u>

$H_o = \beta_{GAS} = 0$ $H_a = \beta_{GAS} \neq 0$

alpha(a) = .05 = 5%

p-value = 1.253E-113

1.253E-113 < 0.05

p-value < alpha at 5%, Hetsy is present.

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	6.58464553	0.47182883	13.9555813	4.33E-44
Ln of Coal	0.21092652	0.05252198	4.0159669	5.9383E-05

$H_o = \beta_{COAL} = 0$ $H_a = \beta_{COAL} \neq 0$

alpha(a) = .05 = 5%

p-value = 5.9383E-05

5.9383E-05 < 0.05

p-value < alpha at 5%, Hetsy is present.

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	12.0227202	0.19423521	61.8977377	0
Ln of Wind	-0.389175	0.02126234	-18.30349	2.4893E-74

H$_o$ = $\beta_{WIND} = 0$ **H$_a$** = $\beta_{WIND} \neq 0$

alpha(a) = .05 = 5%

p-value = 2.489E-74

2.489E-74 < 0.05

p-value < alpha at 5%, Hetsy is present.

With such low p-values for all of the independent variables I am testing, this leads me to conclude that there is heteroscedasticity in the model. I am also going to run a White Test and test for impure heteroscedasticity. If the results of the white test correlate with my results of the Park Test, I will consider removing Wind and Gas. I am going to perform a White Test using Wind and Gas due to the problems encountered with multicollinearity.

(23999*0.01450373 = 348.075016)

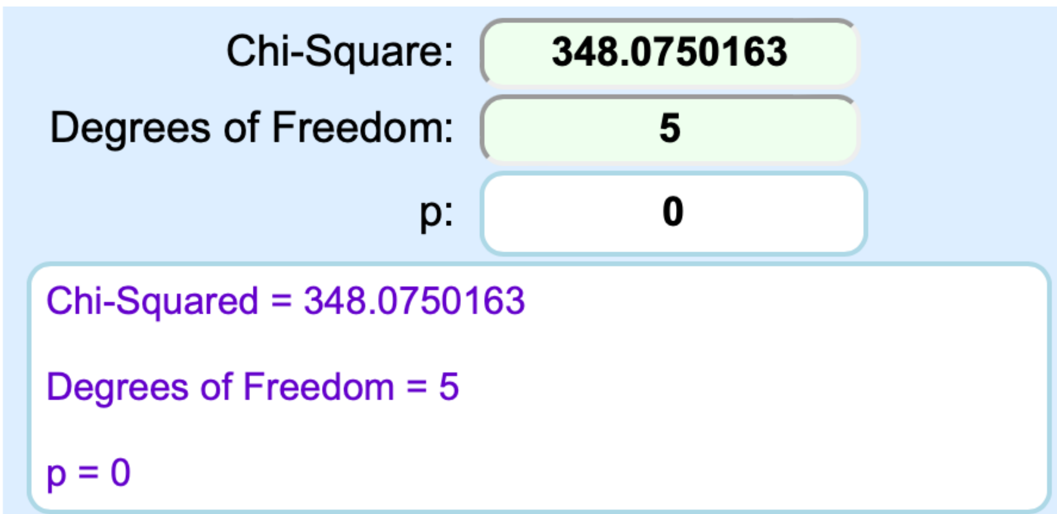
Regression Statistics	
Multiple R	0.12043143
R Square	0.01450373
Adjusted R Square	
Standard Error	4584288.95
Observations	23999

Chi Square	d.f.
348.075016	5

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	7.4208E+15	1.4842E+15	70.6218712	1.3272E-73
Residual	23993	5.0423E+17	2.1016E+13		
Total	23998	5.1165E+17			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-3058965.9	487031.917	-6.2808324	3.4254E-10
S-wind	154.772486	41.0073854	3.77425882	0.00016087
S-gas	226.690447	27.9388397	8.11381035	5.1389E-16
Wind Squared	-0.0003868	0.00108438	-0.35673	0.72129711
Gas Squared	-0.0021908	0.00041575	-5.2695502	1.3794E-07
Wind*Gas	-0.0090627	0.00097445	-9.3003186	1.5162E-20



H_0: Error terms and Variables are NOT related (homoscedastic)

H_a: Error terms and Variables ARE related (heteroscedastic)

With a p-value of 0, we reject the null hypothesis and accept the alternate hypothesis that the Error Terms and Variables are in fact related. As a result of this, it is worth considering removing gas and wind from the PRL. I decided to run a quick auxiliary regression omitting gas and wind, and the results were as I expected.

Gas and Wind Omitted	
Regression Statistics	
Multiple R	0.13928511
R Square	0.01940034
Adjusted R Sq	0.01915511
Standard Erro	583.154654
Observations	23999

Removing gas and wind decreases the robustness of the model in the Adjusted R Squared. There seems to, at-this point in the testing, be confirmed omitted variable bias, and heteroscedasticity. I'm going to run a Weighted Least Squares Test, as this test is required to potentially servev as a remedy for hetsy. The equation for my Weight Least Squares test is:

$$\frac{\text{AvePrice}}{\text{Wind}} = \frac{\text{Bo}}{\text{Wind}} + \beta_1 \frac{\text{Coal}}{\text{Wind}} + \beta_2 \frac{\text{Gas}}{\text{Wind}} + \beta_3 \frac{\text{Nuclear}}{\text{Wind}} + \beta_4 \frac{\text{Hydro}}{\text{Wind}} + \beta_5 \frac{\text{Wind}}{\text{Wind}} + \beta_6 \frac{\text{Solar}}{\text{Wind}} + \beta_6 \frac{\text{D}}{\text{Wind}} + \beta_6 \frac{\text{Other}}{\text{Wind}} + \varepsilon$$

<u>Weighted Least Squares Test</u>:

Original results:

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	376.952054	33.2167654	11.3482469	9.01246E-30	311.845105	442.059003	311.845105	442.059003
S-coal	-0.2596014	0.01397354	-18.578073	1.66557E-76	-0.2869904	-0.2322124	-0.2869904	-0.2322124
S-gas	-0.2187481	0.01427694	-15.32177	9.72648E-53	-0.2467318	-0.1907644	-0.2467318	-0.1907644
S-nuclear	-0.2753703	0.01484532	-18.549304	2.82408E-76	-0.3044681	-0.2462725	-0.3044681	-0.2462725
S-hydro	0.50239788	0.07865455	6.38739751	1.71826E-10	0.34823002	0.65656575	0.34823002	0.65656575
S-wind	-0.2327061	0.01401016	-16.609803	1.30785E-61	-0.2601669	-0.2052453	-0.2601669	-0.2052453
S-solar	-0.239869	0.01373693	-17.461615	7.38709E-68	-0.2667943	-0.2129438	-0.2667943	-0.2129438
D	0.22793233	0.01398919	16.2934659	2.2894E-59	0.20051265	0.25535201	0.20051265	0.25535201
S-other	-0.2258546	0.01478478	-15.276161	1.94688E-52	-0.2548337	-0.1968755	-0.2548337	-0.1968755

WLS results with Wind as scaling factor:

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.2514457	0.02087111	-12.047549	2.49385E-33	-0.2923544	-0.2105371	-0.2923544	-0.2105371
S-Coal/W	-0.3443447	0.02032461	-16.942256	5.17653E-64	-0.3841822	-0.3045072	-0.3841822	-0.3045072
S-Gas/W	-0.2387186	0.02081118	-11.47069	2.22247E-30	-0.2795098	-0.1979273	-0.2795098	-0.1979273
S-nuclear/W	-0.2480037	0.01987775	-12.476448	1.2947E-35	-0.2869654	-0.2090421	-0.2869654	-0.2090421
S-Hydro/W	0.65816944	0.10791245	6.09910598	1.08294E-09	0.44665426	0.86968462	0.44665426	0.86968462
1/Wind	-1.005E-06	3.1211E-07	-3.2193241	0.001286641	-1.617E-06	-3.93E-07	-1.617E-06	-3.93E-07
S-Solar/W	-0.3110348	0.01978232	-15.722863	1.99506E-55	-0.3498094	-0.2722602	-0.3498094	-0.2722602
D/W	0.26769432	0.02021678	13.2411941	7.00391E-40	0.22806816	0.30732048	0.22806816	0.30732048
S-Other/W	-0.2520219	0.02511667	-10.034046	1.20246E-23	-0.3012521	-0.2027916	-0.3012521	-0.2027916

Each of the above coefficients is color-coordinated to refer back and forth to as necessary. The slopes, standard error, t Stat, and P-value does not change drastically. The Weight Least Squares test shows that while there is heteroscedasticity is present, its affects are not very drastic. It would seem as omitting wind would not be necessary. Due to β_5WIND's significance in the model, and theoretical significance, it will remain in the model. Our PRL remains the same in the form of:

(-) (-) (-) (-) (-) (-)

AVGPRICE = β_0+ β_1COAL+ β_2GAS+ β_3NUCLEAR+ β_4HYDRO+ β_5WIND+ β_6SOLAR+

(-) (+)

β_7OTHER+ β_8DEMAND + Ɛ

Serial Correlation

Serial Correlation (or Autocorrelation) is what occurs when values in a time series dataset are related to their past values. Serial Correlation can be positive or negative. Positive Serial Correlation means that if a variable is above its mean in a period, it would be likely to also be above (positive) its mean in the following period. Equally, Negative Serial Correlation entails the opposite, a variable below its mean in a period would be likely to be below its mean in the following period. With Serial Correlation, minimum variance does not hold, and the independent variables are not BLUe. There are two tests for Serial Correlation that I will use, the Runs Test will be utilized first. The Runs Test is a simple test use to determine and investigate the presence of Serial Correlation. Because of the 24,000 observations in my dataset, I will only be observing random portions of my original residual list:

-124.5187
-4.9552278
3.60067657
51.295271
75.9538524
67.3420189
86.3571604
15.5683281
-17.203513
-36.427699
-6.1656366
-28.275054
-22.015024
-44.573251
71.194846
5.57336686
-20.297292
-81.868566
-20.020083
39.9134045
111.63827
80.3874392

**** Highlighted sings are indicators of sign changes ****

The residual list does not give a definite pattern that would lead me to believe there is Serial Correlation, and so I will employ the Durban Watson test. This test helps validate the results of regression models. The equation for the Durban Watson Test is:

$$d = \sum_{2}^{T} (e_t - e_{t-1})^2 \Big/ \sum_{1}^{T} e_t^2$$

Residuals	**(et-1)**	**diff**	**diff^2**	**res^2**
-46.394475				2152.44727
-37.172722	-46.394475	9.22175221	85.0407139	1381.81129
-24.499105	-37.172722	12.6736175	160.62058	600.206141
5.99848691	-24.499105	30.4975918	930.103106	35.9818453
-5.418547	5.99848691	-11.417034	130.348663	29.3606513
-23.342051	-5.418547	-17.923504	321.251985	544.85133
-28.853168	-23.342051	-5.511117	30.372411	832.505288
-37.988336	-28.853168	-9.1351679	83.4512917	1443.11364
-32.04619	-37.988336	5.94214539	35.3090918	1026.95831
-37.273622	-32.04619	-5.227432	27.326045	1389.32291
-38.973253	-37.273622	-1.6996313	2.8887465	1518.91448
-73.574186	-38.973253	-34.600933	1197.22453	5413.16084
-45.989227	-73.574186	27.5849588	760.92995	2115.00902

Numerator	**Denominator**
336536723	7901100940

d = 0.04259365

H₀: p = 0 (no serial correaltion)

Hₐ: p ≠ 0 (serial correlation)

dL: 1.592

dU: 1.757

(Link to DW tables - https://www3.nd.edu/~wevans1/econ30331/Durbin_Watson_tables.pdf)

With my d-stat being lower than my dL value of 1.592, much lower in fact. We would reject the null hypothesis and accept the alternate hypothesis that there is Serial Correlation in my data. There will need to be converted to rates, and I will need to re-run the DW test. However, it must be noted that while attempting to convert everything to rates in order to re-run the DW test, I was unable to correct the data to run a regression. Due to the amount of zeros in the data, this returned many errors. I employed several strategies to combat this, but was unsuccessful in the conversion. If this were to be successful, I would re-run my DW test and see if the conversion was successful in correcting the serial correlation.

Conclusion

My final PRL is as stated:

$$AVGPRICE = \beta_0 + \overset{(-)}{\beta_1} COAL + \overset{(-)}{\beta_2} GAS + \overset{(-)}{\beta_3} NUCLEAR + \overset{(-)}{\beta_4} HYDRO + \overset{(-)}{\beta_5} WIND + \overset{(-)}{\beta_6} SOLAR + \overset{(+)}{\beta_7} OTHER + \overset{(+)}{\beta_8} DEMAND + \varepsilon$$

Final Regression:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.22447657
R Square	0.05038973
Adjusted R Square	0.05007306
Standard Error	573.890031
Observations	23999

Y = Ave Price

ANOVA

	df	SS	MS	F	Significance F
Regression	8	419260767	52407595.9	159.124435	1.68E-262
Residual	23990	7901100940	329349.768		
Total	23998	8320361707			

	Coefficients	Standard Error	t Stat	P-value
Intercept	376.952054	33.2167654	11.3482469	9.0125E-30
S-coal	-0.2596014	0.01397354	-18.578073	1.6656E-76
S-gas	-0.2187481	0.01427694	-15.32177	9.7265E-53
S-nuclear	-0.2753703	0.01484532	-18.549304	2.8241E-76
S-hydro	0.50239788	0.07865455	6.38739751	1.7183E-10
S-wind	-0.2327061	0.01401016	-16.609803	1.31E-61
S-solar	-0.239869	0.01373693	-17.461615	7.3871E-68
S-other	-0.2258546	0.01478478	-15.276161	1.9469E-52
D	0.22793233	0.01398919	16.2934659	2.2894E-59

Overall, my model and specification thereof are not BLUe. Throughout all of the testing deployed in this model, we violate several classical assumptions. Going into this testing, I understood there would be omitted variable bias, however going through the testing showed that there are problems in pretty much every turn possible. One of the biggest issues that is present with this dataset is that it is hourly data. Hourly data, and being able to supplement it with relevant up-to-date data is extremely challenging. I would postulate this dataset with macroeconomic data containing daily (hourly would be necessary) stock trading data. Another idea would be to supplement this data with hourly weather data. Though my specification is not BLUe and not robust, I cannot conclude a definitive decision on my hypothesis, and must state that my results are inconclusive and require further postulation.

References

Lee, Chien-Chiang, and Jun-De Lee. "A Panel Data Analysis of the Demand for Total Energy and Electricity in OECD Countries." *The Energy Journal*, vol. 31, no. 1, 2010, pp. 1–23. *JSTOR*, http://www.jstor.org/stable/41323268.

El-Kholy, Hosni, and Ragy Faried. "MANAGING THE GROWING ENERGY DEMAND: THE CASE OF EGYPT." *Energy & Environment*, vol. 22, no. 5, 2011, pp. 553–63. *JSTOR*, http://www.jstor.org/stable/43735553. Accessed 20 Oct. 2023

Abbot, Cheryl. "An Analysis of Southern Energy Expenditures and Prices, 1984–2006." *Monthly Labor Review*, 2008, pp. 3–18. *JSTOR*, https://www.jstor.org/stable/monthlylaborrev.2008.04.003. Accessed 23 Oct. 2023.

Siegfriedsen, S., et al. "Primary Markets for Offshore Wind Energy Outside the European Union." *Wind Engineering*, vol. 27, no. 5, 2003, pp. 419–29. *JSTOR*, http://www.jstor.org/stable/43750352.

Garvey, Seamus D. "Integrating Energy Storage with Renewable Energy Generation." *Wind Engineering*, vol. 39, no. 2, 2015, pp. 129–40. *JSTOR*, https://www.jstor.org/stable/90007058.