Projecting Energy Industry Trends Using Econometric Models

Ryan Reiser

Colorado School of Mines

EBGN 303: Econometrics

Golden, CO USA

Instructed by Dr. Steven Piper

May 5th, 2025

I. Introduction

As artificial intelligence innovation continues to carry greater promises to transform all forms of production, the world's energy demands are expected to increase dramatically over the coming years. According to the International Monetary Fund's scholarly paper *Power Hungry: How AI Will Drive Energy Demand*, "As AI technologies continue to evolve and proliferate, the demand for computational power and electricity is poised for a significant surge. AI-related electricity consumption could reach up to 1,500 TWh by 2030, possibly outpacing other emerging sources of demand, like electric vehicles, and becoming comparable to India's total electricity consumption" [1]. It is clear that the energy industry will need to substantially increase output to meet higher electricity demands, but how can our world built on a foundation of fossil fuel burning afford to drive energy production when the consequential greenhouse gas emissions will undoubtedly cause permanent environmental destruction? On an individualistic perspective, should modern investors continue to fund traditional energy stocks, or should they put their faith in the renewable energy sector as the solution to this crisis?

This paper attempts to use econometric methods to understand and predict the growth of both the fossil fuel and renewable energy industries. The remainder of this paper examines which variables are used in the model, where the data for each variable was sourced from, what transformations that were made to each variable to ensure model accuracy, and how the model overall performed as an explanation for trends within the energy industry.

II. Theoretical Model

To most accurately measure growth within each section of the energy industry, this project utilizes two dependent variables: Yearly Global Fossil Fuel-Based Energy Consumption and Yearly Global Renewable Energy Consumption, both measured in terawatt hours. It is important that we view the industry from a global perspective because climate change, AI development, and energy production are all world-wide issues that depend on the collective capabilities of all nations.

The first independent explanatory variable used in the model is Yearly World GDP, measured in USD. This variable acts as a measure of global economic status overtime, allowing analysis of how economic growth influences the expansion of each sector. The variable is expected to positively influence both the fossil fuel and renewable sectors because both rely on a strong economy to fund growth.

The second independent explanatory variable used in the model is Yearly Crude Oil Prices, measured in USD. This variable specifically examines the operating costs of the fossil fuel sector as an explanation of why energy companies may want to drive petroleum refinement under low oil prices or switch to renewable alternatives under high oil prices. Therefore, this variable is expected to negatively influence the fossil fuel sector but positively influence the renewable energy sector.

The third independent explanatory variable used in the model is Global Tertiary Education Enrollments. The energy industry is a highly technical one which was built and is maintained off the work of scientists and engineers. Furthermore, innovation in complex renewable energy forms like nuclear energy require extreme amounts of education, and educated citizens tend to support renewable energy alternatives more as they learn the dangers that climate change presents. Therefore, we expect tertiary educational enrollments to have a slightly positive influence on the fossil fuel sector and a substantially positive influence on the renewable energy sector.

The fourth independent explanatory variable used in the model is Global Nuclear Powerplants Built. Many countries, such as France, look to nuclear fission as the cleanest way to mass produce energy. Nuclear power particularly has the potential for high production levels once enough plants are built and operated. This variable measures the expansion of the nuclear sector through the quantity of nuclear power plants built globally each year, which is expected to take demand away from the fossil fuel sector and thus negatively influence fossil fuel consumption while positively influencing renewable energy consumption.

The fifth and final independent explanatory variable used in the model is a dummy variable for the Environmentalism Movement. 2007 is a rough estimate for when the public opinion shifted support away from fossil fuel sector and in factor of cleaner solution due to a mix of the release of Al Gore's *An Inconvenient Truth*, climate change beginning to be taught in school curriculums, and the near election of President Obama. This variable is expected to negatively influence fossil fuel consumption and positively influence renewable energy consumption.

A few years from now, it would be interesting to use a metric like global AI queries made yearly to see how much AI usage drives the energy industry. However, today there is simply not nearly a big enough sample size of such data to include as an explanatory variable.

III. Data Sources

Variable	Source (Hyperlinked)	Years	Strengths	Weaknesses	Descriptive Analysis
Yearly Global Fossil Fuel Energy Consumption (TWh)	Our World in Data	1965-2023	Large dataset of independent variable in appropriate form of measurement	None of note	Fossil fuel energy consumption is steadily increasing over time
Yearly Global Renewable Energy Consumption (TWh)	Our World in Data	1965-2023	Large dataset of independent variable in appropriate form of measurement	None of note	Renewable energy consumption is steadily increasing over time
Yearly World GDP (USD)	WorldBank.org	1965-2023	Large dataset of a relatively accurate way to measure global economic trends	There are many factors that influence and measure global economic trends, not just World GDP	World GDP is steadily increasing over time
Yearly Crude Oil Prices (USD)	InflationData.com	1965-2023	Large dataset of the most influential price that affects fossil fuel operation	Does not explain operating costs of natural gas and coal-based energy sources	Crude oil prices are steadily increasing over time, notably jumping very high between 2010-2014
Yearly Global Tertiary Education Enrollments	Our World in Data	1970-2023	Accurate way of measuring global educational trends	Slightly reduces overall datapoints for regression since the data starts in 1970	Global education is steadily increasing over time
Yearly Global Nuclear Powerplants Built	Global Energy Monitor	1965-2023	Accurate way of measuring the expansion of the nuclear industry	The nuclear industry is highly complex and is frequently closing plants, which is not considered	There does not appear to be any clear trends for when nuclear power plants are built
Environmentalism Movement Dummy Variable	Self-Made	1965-2023	Attempts to explain the social impact on each sector	2007 is a very rough estimate of when the variable became significant	The environmentalism movement began to see notable impact in 2007

IV. Empirical Model

When looking at the Descriptive Analysis section of our Data Sources table, we can identify that most variables used in our model tend to steadily increase over time. This is a problem because, if unedited, these variables will cause spurious correlations that will appear to explain the dependent variable very well but are actually mostly corrected due to mutual upwards trending. Therefore, we will transform our data into log differences of variables for the most accurate results. Log difference of variables explain how much a variable changes in percent from one year to the next, which will have the benefit of reducing heteroskedasticity, reducing multicollinearity, and preventing trending relationships from inflating results. Importantly, we do not transform the dummy variable because you cannot take logarithms of zeros. The model uses a linear ordinary least squares regression because most of the variables trend linearly.

Empirical model for Fossil Fuel Energy Consumption:

$$\text{dln\_FossilFuelTotal}_t = \beta_0 + \beta_1 * \text{dln\_WorldGDP}_t + \beta_2 * \text{dln\_CrudeOilPrice} + \beta_3 * \text{dln\_TeriaryEnrollment}_t + \beta_4 * \text{dln\_NuclearPlants} + \beta_5 * \text{EnviroDummy} + \varepsilon_t$$

Empirical model for Renewable Energy Consumption:

$$\text{dln\_RenewableTotal}_t = \beta_0 + \beta_1 * \text{dln\_WorldGDP}_t + \beta_2 * \text{dln\_CruddeOilPrice} + \beta_3 * \text{dln\_TeriaryEnrollment}_t + \beta_4 * \text{dln\_NuclearPlants} + \beta_5 * \text{EnviroDummy} + \varepsilon_t$$

As can be seen, each empirical model indicates that each of the noted independent variables are explained by the noted dependent variables, in addition to an error term ε.

V. Empirical Results

Regression Results (Fossil Fuel Consumption):

R^2 = 0.4094		**Adjusted R^2 = 0.3438**		**Root MSE = 0.0161**		
Variable	Coefficient	Stan Error	t-statistic	P-statistic	95% confidence interval	
Log Difference World GDP	0.1765	0.0487	3.63	0.001	0.0785	0.2746
Log Difference Crude Oil Price	0.0046	0.0105	0.44	0.663	-0.0165	0.0257
Log Difference Tertiary Enrollments	0.0903	0.1269	0.71	0.481	-0.1653	0.3457
Log Difference Nuclear Plants Built	0.0073	0.0034	2.11	0.041	0.0003	0.0142
Environmentalism Movement Dummy	-0.0023	0.0054	-0.42	0.678	-0.0132	0.0086
Constant	0.0047	0.0059	0.80	0.428	-0.0072	0.0166

After running our regressions, we can conclude that each model does a mediocre job of explaining trends within our dependent variables. Looking at our fossil fuel regression, an adjusted R^2 value of 0.3438 means that our model accurately explains 34.38% of the variation within fossil fuel-based energy consumption. The only statistically significant variables within the model that had both a t-statistic greater than 2 and an P statistic less than 0.05 were World GDP and nuclear power plants built. However, it is important to note that nuclear power plants built was statistically significant in the opposite direction than intended, suggesting that the quantity of nuclear power plants built has a statistically significant positive influence on fossil fuel consumption. I was particularly surprised to see crude oil prices have such an insignificant influence on fossil fuel consumption, since I would assume that as the finite resource of oil becomes scarcer, more consumers would see their prices increase and thus switch to fossil fuel-based energy sources. It would be interesting to see if this variable becomes more statistically significant in the future as global finite oil reserves deplete further.

Further Testing Results (Fossil Fuel Consumption):

Diagnostic Test	Test Statistic	p-value	Conclusion
Variance Inflation Factor Test	Mean VIF = 1.27	N/A	No evidence of multicollinearity
Heteroskedasticity Test	$X^2(1) = 0.82$	0.364	No significant heteroskedasticity
Omitted Variable Test	$F(3, 42) = 1.92$	0.1414	No significant omitted variables

Upon further diagnostic testing of our regression, we can see that there is no evidence of multicollinearity, no significant heteroskedasticity, and no significant omitted variables. This means that, while taking the log differences of each variable did reduce their overall R^2 results, they also make our results more accurate and reliable.

Regression Results (Renewable Energy Consumption):

R^2 = 0.4409		**Adjusted R^2 = 0.3788**		**Root MSE = 0.0097**		
Variable	Coefficient	Stan Error	t-statistic	P-statistic	95% confidence interval	
Log Difference World GDP	0.2485	0.0293	0.85	0.402	-0.3426	0.0840
Log Difference Crude Oil Price	0.0019	0.0063	0.31	0.759	-0.0108	0.0147
Log Difference Tertiary Enrollments	-0.3845	0.0764	-5.03	0.000	-0.5385	-0.2305
Log Difference Nuclear Plants Built	0.0042	0.0033	2.25	0.029	0.0005	0.0088
Log Difference Environmental Dummy	0.0042	0.0033	1.28	0.208	-0.0024	0.0107
Constant	0.0299	0.0036	8.41	0.000	0.0227	0.0370

The renewable energy regression shows marginally better results than the fossil fuel regression, with an adjusted R^2 value of 0.3788 meaning that our model accurately explains 37.88% of the variation within renewable energy consumption. The only statistically significant variables within the model that have both a t-statistic greater than 2 and a P statistic less than 0.05 were tertiary education enrollments and nuclear power plants built. However, it is important to note that tertiary education enrollments was statistically significant in the opposite direction than intended. There is no logical reason why this would be the case, so it is likely that this is a result of coincidence, and that tertiary education does not actually have any notable impact on renewable energy consumption. It was particularly interesting to see that World GDP was not nearly as statistically significant as it was in the fossil fuel regression, suggesting that fossil fuel consumption relies much more on global economic state than renewable energy consumption does.

Further Testing Results (Renewable Energy Consumption):

Diagnostic Test	Test Statistic	p-value	Conclusion
Variance Inflation Factor Test	Mean VIF = 1.27	N/A	No evidence of multicollinearity
Heteroskedasticity Test	$X^2(1) = 1.10$	0.750	No significant heteroskedasticity
Omitted Variable Test	$F(3, 42) = 3.49$	0.024	No significant omitted variables

Upon further diagnostic testing of our regression, we can see that this regression also has no evidence of multicollinearity, no significant heteroskedasticity, and no significant omitted variables.

VI. Summary and Conclusions

In conclusion, the econometric model developed through this project serves to moderately explain variance within both the fossil fuel and renewable energy sectors, marginally performing better explanations for renewable energy consumption. This model uses five independent variables that logically influence energy consumption, though the results of our model show that only a few of those variables actually perform statistically significant explanations. When running OLS regressions with many variables that trend similarly over time and potentially have multicollinearity or heteroskedasticity, we can generate the log differences of each variable to gain a more accurate understanding of how each independent variable explains the dependent variable. It is important to run diagnostic tests of regressions to highlight any potential inaccuracies, though this model did not display any. I would predict that as more data for each variable is collected into the future, this model would increase in accuracy and R^2 results as issues like climate change and the finiteness of fossil fuels reserves become more socially prominent.

This model was also briefly used to attempt to explain variation within stock prices for large energy companies like Chevron and NextEra Energy, though those regressions performed extremely poorly and proved to not be statistically significant methods of explaining company stock prices over time. Further collection of different variables, such as company-specific variables, would likely do a greater job at explaining energy company stock prices.

VII. References

[1] C. Bogmans et al., “Power hungry: How AI Will Drive Energy Demand,” International Monetary Fund, https://www.imf.org/en/Publications/WP/Issues/2025/04/21/Power-Hungry-How-AI-Will-Drive-Energy-Demand-566304 (accessed May 5, 2025).

[2] Hannah Ritchie and Pablo Rosado (2017) - “Fossil fuels” Published online at OurWorldinData.org. Retrieved from: 'https://ourworldindata.org/fossil-fuels' [Online Resource] (accessed May 5, 2025).

[3] Hannah Ritchie, Max Roser, and Pablo Rosado (2020) - “Renewable Energy” Published online at OurWorldinData.org. Retrieved from: 'https://ourworldindata.org/renewable-energy' [Online Resource] (accessed May 5, 2025).

[4] “GDP (current US$),” World Bank Open Data, https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2023&start=1960 (accessed May 5, 2025).

[5] T. McMahon, “Historical Crude Oil Prices,” InflationData.com, https://inflationdata.com/articles/inflation-adjusted-prices/historical-crude-oil-prices-table/ (accessed May 5, 2025).

[6] Hannah Ritchie, Veronika Samborska, Natasha Ahuja, Esteban Ortiz-Ospina, and Max Roser (2023) - “Global Education” Published online at OurWorldinData.org. Retrieved from: 'https://ourworldindata.org/global-education' [Online Resource] (accessed May 5, 2025).

[7] “Global Nuclear Power Tracker,” Global Energy Monitor, https://globalenergymonitor.org/projects/global-nuclear-power-tracker/ (accessed May 5, 2025).