Predicting American Energy Trends During Decarbonisation

Matthew Dyer

University of Colorado Boulder

1. Introduction

In January 2020, President Joe Biden signed an executive order calling for the United States to produce emission-free electricity by 2035 and to reduce net greenhouse-gas emissions to zero by 2050 [1]. Net zero emissions is a state where the amount of greenhouse gas pollution being produced is equivalent or less than the amount of greenhouse gases being sequestered from the environment. In April, this target was updated to a 50% reduction in U.S. greenhouse gas pollution from 2005 levels in 2030. These lofty goals coincide with President Biden re-joining the Paris Agreement. This international treaty on climate change has the overarching goal of limiting a global rise in temperature to under 2 degrees Celsius this century [2]. America's energy sector is due to evolve significantly. Despite these concrete targets, America does not have a clear strategy yet. Even with a clear federal strategy, the diversity of America and decentralization of the government among the States makes a rigid plan unlikely. In this analysis, I aim to predict how America's energy sector will look in the future now that we have clear goals in sight. I will specifically predict the energy production makeup of America by predicting the output of various energy sources including non-renewables like coal and renewable energy such as solar and wind.

1.1 Purpose

Energy production is only part of the picture. Greenhouse gas emissions come from several other industries including shipping, agriculture and transportation. Therefore, curtailing energy production pollution is only a piece of the puzzle. President Biden made an earlier deadline for electricity emissions partly because of its prime importance in emissions. Limiting electric emissions will have a rippling effect across other industries. For example, the burgeoning electric car market will directly have reduced emissions with diminished electricity emissions. However, certain industries like concrete production and cargo shipping do not have electric analogues and reducing their emissions will prove intractable [3]. While electricity production is not the entire solution, its importance and influence on other emissions is prominent and deserving to be the center of this study. A successful strategy in electricity is indicative of a successful strategy overall.

In order to understand how the electric sector may evolve, it is important to know its current state. America has experienced reductions in reliance on carbon despite recent federal negligence on climate change mitigation. This promising trend is partly due to the shift from coal to natural gas. Coal dominated electric production in 2010 producing 45% of all electricity. In 2020, that share dropped to 19% [3]. Clean energy has also experienced a boost. Wind and solar have become remarkably cheaper. The cost of energy from wind has dropped 70% in the past decade while the cost of energy from solar has dropped 90% [3][4]. These two renewable energies appear to be the most promising tools in curtailing emissions significantly. Nuclear energy has remained stagnant with no new nuclear power plants built in the last decade. The recent tragedy of Fukushima has largely hindered confidence in nuclear power.

In order to assess the trajectory of America's energy production, I will analyze the current trends. If the government keeps a laissez-faire attitude, it may be possible that current trends in energy may be enough to reach President Biden's goals. After all, minimal federal action and litigation has occurred during America's current drop in a carbon footprint. I will first analyze America's energy trends alone and then attempt to incorporate trends in similar countries using partial pooling. I will determine what constitutes as a "similar country" by assessing what factors are elementary in explaining a country's energy production and then incorporating countries with similar key factors.

1.2 Data

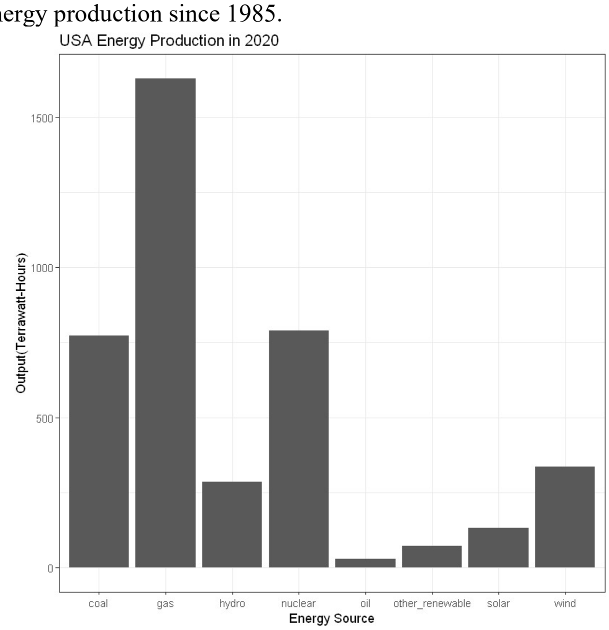
The dataset used for analysis is the Energy Dataset from Our World in Data [5]. This dataset contains electricity consumption figures extracted from the BP Statistical Review of World Energy, Shift Data Portal, and Ember. The energy data includes electricity generated from oil, natural gas, coal, biofuel, wind, solar, hydro, nuclear, and other renewables. Many of the variables in the dataset are variations of the same data expressed in different ways such as percent shares, percent change, annual change, etc. Therefore, to avoid issues with multicollinearity, I only extracted the electricity generation metrics. In addition to these metrics, I used the population figures from this dataset which come from the UN population estimates and I also used the GDP figures which come from the Maddison Project Database.

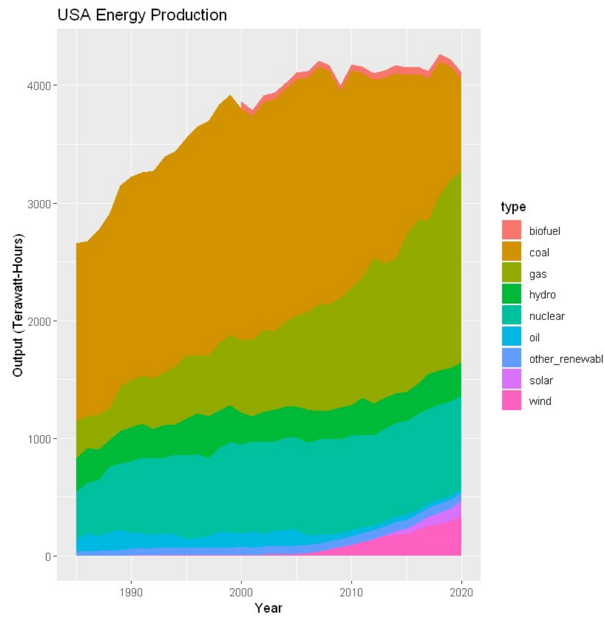
The electric consumption variables are measured in terawatt-hours. This is a unit of energy equal to outputting one trillion watts for one hour. GDP is inflation-adjusted in this dataset. Missing values was a significant issue with these data. In particular, electric output in the US was missing before 1985 and GDP figures were missing after 2016. Biofuel figures for the US are also missing up until 2000.

2. Exploratory Data Analysis

2.1 U.S. Energy Trends

The first step is to take a look at the current US energy production figures (as of 2020) and the overall trend in energy production since 1985.



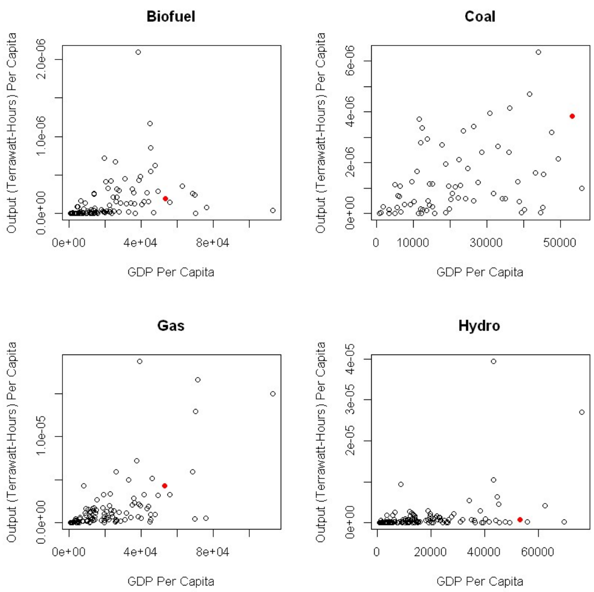


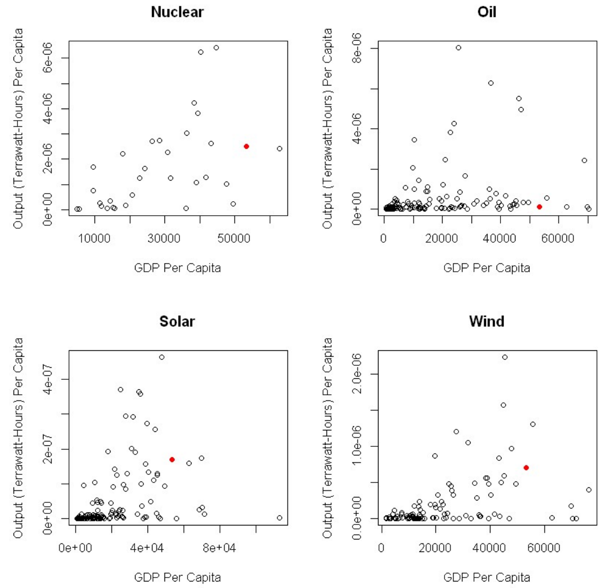
Natural gas is the main energy source today in the U.S. It accounted for over 1500 terawatt-hours in 2020 while no other energy source exceeded 1000 terawatt-hours. Coal and nuclear are the next largest producers accounting for roughly 750 terawatt-hours each. Wind and hydro accounted for roughly 250 terrawatt-hours in 2020 with the rest accounting for less than 250 terawatt-hours. While better than coal, natural gas still produces greenhouse gases. Unfortunately two of the top 3 energy sources come from fossil fuels. Natural gas appears to have more than 5 times the output of the highest producing renewable, wind. In addition, coal appears to have more than double the output of wind. Current output is, unfortunately, nowhere close to the targets. It is important to look at the trends to assess if we are heading in the right direction.

The stacked area plot seems to agree with the current energy landscape outlined in the introduction. Natural gas energy production has experienced the biggest increase in the past 35 years. Coal has experienced a contraction since around the turn of the millennium. Oil, which never had a large share of energy production, has seemed to diminish drastically around 2005. Biofuel and hydropower have both remained small but steady portions of American energy production. Nuclear power growth seemed to have thrived prior to 1990. It was the second biggest energy provider by a significant margin behind coal in 1990. Since its apex, nuclear power has not changed but still remains a significant energy producer in the US. Finally, solar and wind have had large increases in power generation in the past 10 years. Solar has become a viable and practical energy source since around 2013 and has risen in output ever since. Unfortunately, the magnitude in output between solar and wind energy does not appear to be adequate to reach President Biden's targets. The current hopes in carbon-free electricity by 2035 appear to reside in the rate of growth of solar and wind.

2.2 Energy Trends and GDP

The Paris Agreement distinguishes between developing and developed countries [2]. There are different requirements and goals based on this characterization. The idea is that developed countries are in a better position to decarbonize with their established industrialization and carbonization. As such, developing countries are not expected to curtail emissions as drastically as developed countries to ensure that developing countries can progress to enjoy the same quality of life as developed countries. I am interested in how energy production is impacted by a country's economic output. The following plots are the output per capita of each energy source plotted against the GDP per capita of each country. Only countries that had an output greater than zero for each respective energy source were included in each plot. The US data are incorporated as the red dot in each plot.





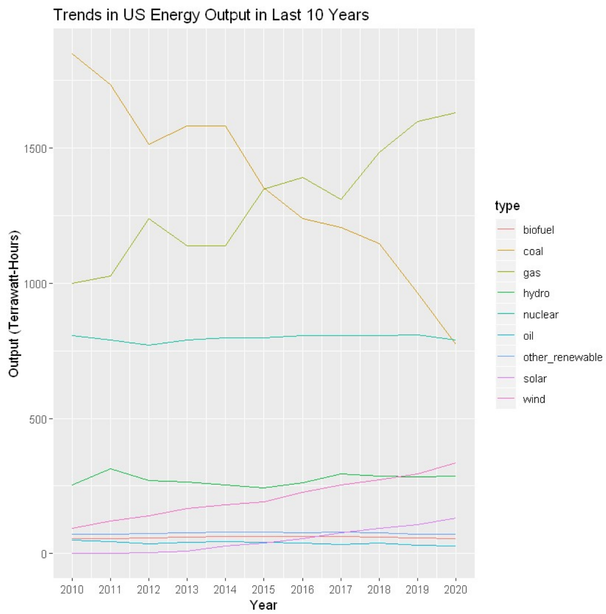
The correlation in each plot is positive indicating that there is a positive relationship between each energy output and GDP. The data corroborate the Paris Agreement treating developed countries differently than developing countries. Countries with higher GDPs produce more electric output per capita in each energy source on average than countries with lower GDPs. The highest correlations exist between GDP and nuclear energy as well as GDP and natural gas. These correlations are 0.528 and 0.607 respectively. This indicates that nuclear energy and natural gas are luxuries enjoyed by more developed nations. Intuitively, nuclear power requires substantial upfront costs and resources to build a reactor. In addition, nuclear proliferation is a consistent worry of developed nations that often stifle nuclear progress in developing countries. Natural gas has benefited from large advances in fracking technology. This technology may also require significant resources and infrastructure that developing countries lack.

3. Methods

3.1 Linear Model

In order to analyze the trends in energy production, a statistical model must be developed to accurately forecast energy production for the next 30 years. A linear model is a universally well-understood model that is readily interpretable. Therefore, it is the natural first attempt to predict energy production trends. However, the issues with a linear model for these data are apparent. Most importantly, these data are a time series. The data are not independent observations and are autocorrelated. Therefore, any linear analysis should be taken with caution and skepticism. Furthermore, the trends in the data are not regular and linear. For example, solar and wind power has not been a significant energy contributor until the past decade. The increases in output were negligible and then increased exponentially.

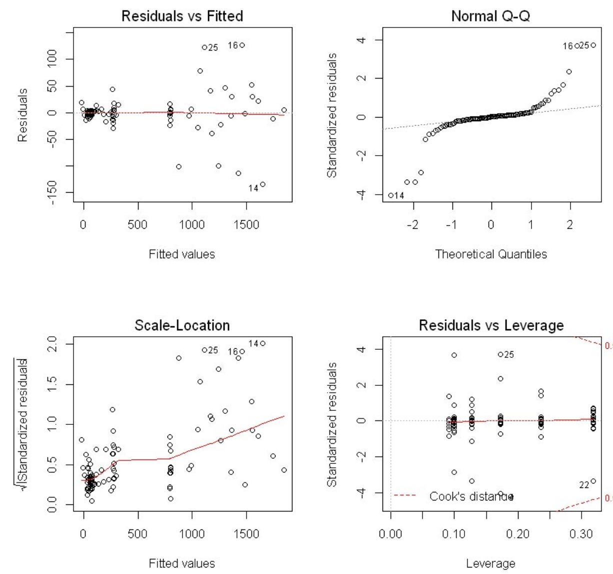
In order to account for these drastic changes in trends, I will only use data from the U.S. over the past 10 years where the trends are relatively consistent. The following is a visualization of the trends of the past decade in the U.S.



The trends appear roughly linear in this most-recent 10 year window. The most problematic trend is solar power which appears more exponential. The linear analysis will most likely underpredict the trend in solar power. This linear model will one-hot-encode for a factor describing the type of energy output. The linear model will also include interaction terms for each energy category and time in years since I am interested in how each category changes individually over time.

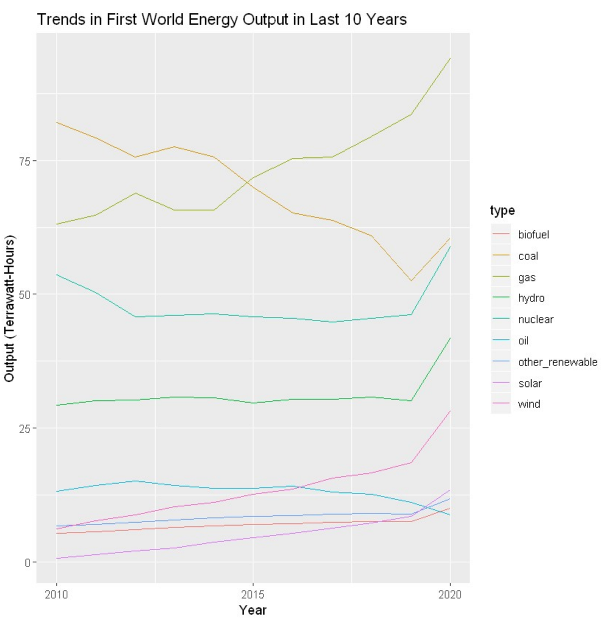
```
Residual standard error: 36.44 on 81
degrees of freedom
Multiple R-squared:  0.9961, Adjusted
R-squared:  0.9953
F-statistic:  1226 on 17 and 81 DF,  p-
value: < 2.2e-16
```

The overall model has a p-value of machine zero signifying that the overarching model is statistically significant. The individual t-tests also indicate that the slope interaction coefficients for coal, gas, solar, and wind are all statistically significant. This indicates that the energy production from these resources has been changing over time. The model suggests that hydro, nuclear, oil, biofuel, and other renewables have been stagnant over the past decade. This quantitative assessment supports the qualitative assessment of the corresponding almost flat lines on the trends graph. This linear model suggests that there is insignificant evidence that hydro, nuclear, oil, and biofuel will change production capacity over time. Therefore, in order to understand the changing energy production makeup, I can focus my attention on coal, gas, solar, and wind. The linear model automatically violates the independence assumption but the other assumptions must be checked.

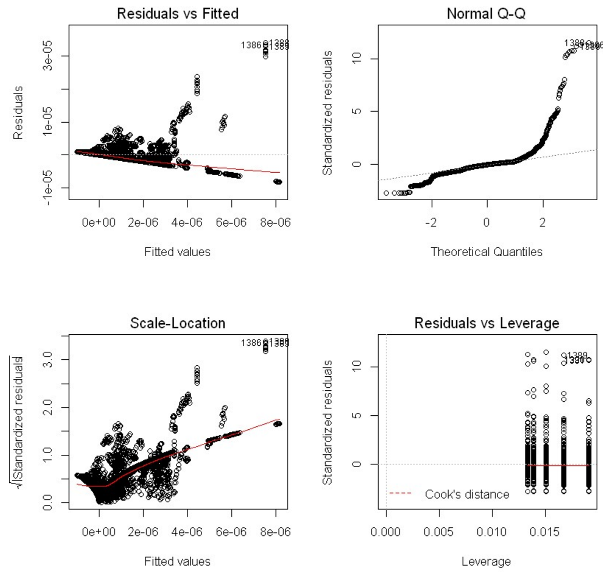


The residuals vs. fitted plot indicates no overall trend in the residuals. The linearity assumption appears valid. However, there is a higher spread in residuals at higher fitted values. Homoskedasticity appears to be violated. The scale-location plot corroborates this apparent nonconstant variation showing a positive overall trend. The variance of residuals increases as the fitted values increase. The normal QQ plot shows overdispersion of the residuals beyond the 1 and -1 theoretical quantiles. The residual histogram also shows heavy tails. Large residuals can be a result of heteroskedasticity and not necessarily a violation in normality. In this regard, the robust normality assumption appears possibly valid and should be reassessed after the heteroskedasticity is remedied. This model could be further amended through a possible transformation or weighted least squares. Finally, the residuals vs. leverage plot does not indicate any influential points. Observations 14, 16, and 25 are flagged as outliers with standardized residuals around 4. These observations correspond to coal output in 2012 and 2014 as well as gas output in 2010. Even though these data do not have alarmingly high Cook's distances, these observations will need to be carefully considered in future models to assess whether they affect model fit. Due to the clear violation of the independence assumption, I will only use this model as a guideline to illuminate possible trends. Actual predictions will need a different valid model.

One issue with the linear model is the lack of data. I truncated the data to the last 10 years in order to include linear trends in the data. The data vary widely across the 35 years of observations and the latest 10 year window is the largest window without drastic changes in the behavior of energy production. 10 years of data for one country does not provide a robust analysis. One possible solution is a hierarchical model that allows for the flow of information between data. The above linear model can be thought of as an unpooled model for an individual country. Using the data for the rest of the countries, a pooled model can be created using the plethora of data provided in the dataset. However, the earlier analysis of GDP and energy production output indicated that the pooled data includes developing and developed countries that require different models. Therefore, I will only include developed countries in the pooled model in order to accurately describe the class of energy production that includes the U.S. A common metric for a developed country is a country with a GDP per capita over $25,000 [6]. The pooled model will only include these countries. Since GDP figures were not included past 2016, the pooled model will only include countries with a GDP per capita above $25,000 in the most recent year of GDP data, 2016. First, a plot of the first-world global trends will indicate if the trends are linear from 2010-2020.



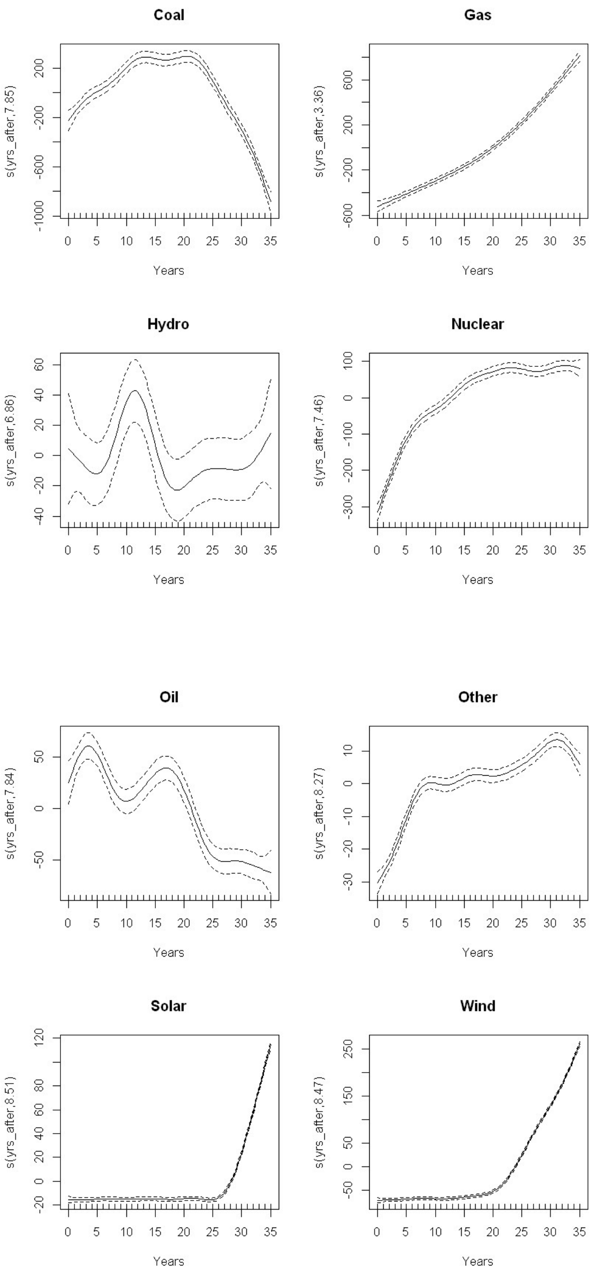
The first apparent anomaly is the strange jump across all energy sources from 2019-2020. Aside from this year, the trends appear consistent. Therefore, I will truncate the data from 2010-2019 to avoid the irregularities of 2020 affecting the data. The following linear model is the pooled data of energy production trends in first world countries adjusting for each country.



Unfortunately, the diagnostic plots indicate major problems. There is a clear downward linear trend at lower fitted values and a linear model does not appear appropriate for this data. The normal QQ plot shows an overdispersion in the data of up to 15 standard residuals. The scale-location plot shows clear heteroskedasticity. Similar to the previous time series, these data are not independent. This model is not valid. There are, most likely, missing predictors that explain a country's energy production. It is not reasonable to assume that the trends in energy output are similar enough to be modelled across countries. Unfortunately, this dataset does not include any other metrics besides various measures of energy production, population and GDP. Therefore, improving model selection is not an option with this dataset alone. A hierarchical model cannot be produced with an invalid pooled model.

3.2 Non-parametric Model

A major issue with the linear model here is its dependence on several assumptions. Independence is not valid for autocorrelated data. Output is not necessarily linearly related to time. Even the homoskedasticity of the linear US model appeared violated. Therefore, the flexibility of a nonparametric model may be of use. Using the smoothed spline method, I will fit a simple generalized additive model that incorporates smoothed splines using each energy source as a predictor. A GAM cannot be fitted to the above linear model as the interaction terms are not additive. The GAMs may not be effective models by themselves which such few data but they will be effective tools to assess the trends of each predictor. The GAMs will reveal whether a relationship between time and output exists and whether that relationship can reasonably be modelled as linear.



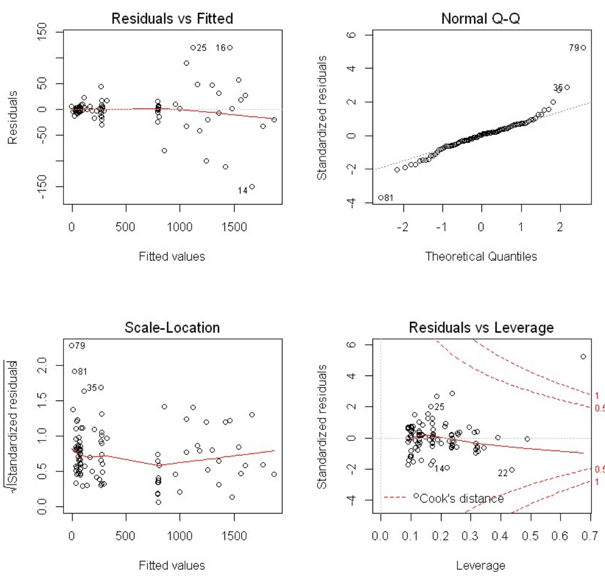
The smoothed spline GAM plots reveal interesting results. First, some of the plots yield nonsensical results. Each plot shows negative output on at least some portion of the graph. It is clear that these smoothed splines do not give accurate predictions by themselves. However, the overall trends are useful. The confidence bands of oil, hydro, and nuclear all indicate a stagnant trend in output. These results agree with the earlier p-values of the slope coefficients of the linear model. Coal appears to be exponentially decreasing while natural gas is exponentially increasing. Wind and solar appear to have changed their behaviors drastically around a decade ago and are now increasing at a sharp linear or possibly slight exponential rate.

3.3 Weighted Least Squares

The diagnostic plots of the linear model and smoothed spline plots have indicated that a linear relationship is appropriate for the data when limited to the past decade. The original model produced an R-squared of 0.9961. This model exhibits that 99.61% of the variation of output can be attributed to the variation of type of energy source and years elapsed since 1990. The two outstanding issues of this model were the violations of independence and homoskedasticity. The independence assumption can be remedied with a time series model that takes autocorrelation into account. However, the interpretability of a linear model is ideal and the main trends that exist in the solar and wind output appear to be appropriate for linear modeling. The original linear model also consisted of a high r-squared. Although this does not mean the model is correct, it is indicative of accurate predictions, which is the main focus of this paper.

The violation of homoskedasticity can be remedied by weighted least squares. The weights were selected by calculating the inverse of the fitted values of the original linear model residuals regressed on the original linear model's fitted values. The resulted weighted least squares model produces an r-squared of 0.9939 and an overall p-value of machine zero. Even though this r-squared is slightly less than the original value, the diagnostic plots show adequate homoskedasticity. It is important to note, however, that the weighted least squares model contains a possible influential point flagged as observation 79. This observation contains a Cook's distance of beyond 1 and is most likely influential. Removing this point increased the model fit.

Acknowledgements



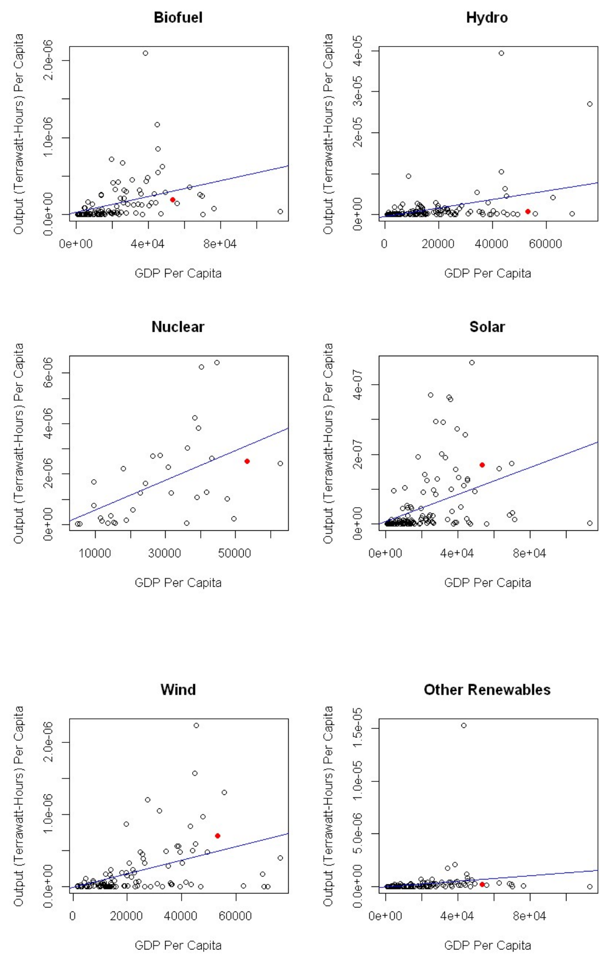
4. Conclusion

The original purpose of this research was to predict if America would rise to President Biden's goal of the electricity sector being emissions free by 2035. The data above has indicated that, out of all the low-carbon energy sources, wind and solar are the only nonstationary energy sources. Nuclear and hydro are stagnant while biofuel and other renewables are small producers and relatively unchanging. By the principle of Occam's razor, the simplest explanation is the best. Since zero emissions would mean no fossil fuels, the original research question can be simplified to: Will the low-carbon energy sources produce enough output to cover America's energy needs by 2035? If making the assumption that energy trends will continue, then natural gas production will increase exponentially until 2035 and emission-free energy will be impossible. In addition, coal and natural gas are not renewable and, therefore, as supplies exhaust in the upcoming decades, energy production will drop dramatically. Therefore, the assumption that fossil fuel trends will continue the same unimpeded is unreasonable.

The fundamental question then is "Will current production in the stationary (and roughly stationary) nuclear, hydro, biofuel and other renewable power sources in conjunction with the rapidly increasing solar and wind power sources cover America's energy quota in 2035?"

The final weighted least squares model indicates that, adjusting for all other predictors, a 1 year increase is associated with a 12.5962 terawatt-hour increase in output from solar power. Equivalently, adjusting for all other predictors, a 1 year increase is associated with a 22.402 terawatt-hour increase in output from wind power. Assuming all fossil fuel electricity sources fall to zero output by 2035 and assuming all other energy sources maintain constant output, the increases in wind and solar output would not even be enough to cover today's energy needs.

The next question is where can America improve in order to meet these energy needs? Some suggestions can be discovered in the output vs. GDP plots. The plots where America falls below the fitted line shows America's output is below expected for that energy source based on its GDP. In the following plots, the blue line is the linear regression line and the red dot is America's output.



America's output is below expected for every renewable energy source except for wind and solar. Perhaps the answer to America's energy problem is not with how fast wind and solar can rise to the renewable energy needs, but how the other low-carbon energy sources can increase output. In particular, nuclear energy is the largest low-carbon energy source but has remained stagnant for decades. In fact, plants may start to shut down as they reach a certain age and nuclear energy production could even contract. The untapped potential in nuclear power could be the solution President Biden is seeking

For further analysis, I would investigate other countries' energy production and natural factors to determine where America has untapped potential. For example, Stephen Pacala of Princeton University posits that the South's abundance of sunshine provides an opportunity for more solar power. He also claims that the plains and long coasts provide potential for more wind power. Untapped uranium deposits could provide more opportunity for nuclear power. I would look at countries with similar factors and assess if America has untapped potential in each sector.

In conclusion, America has benefited in a reduction of a carbon electrical footprint despite federal inaction. However, these natural trends in energy production are not enough to accomplish President Biden's goal to have its electrical sector emission-free by 2035. The overall output of low-carbon sources is not expected to be enough to accommodate America's energy needs.

References

[1] "The Bumps Ahead for Joe Biden's Plan to Decarbonise America." *The Economist*, 2021.
[2] "The Paris Agreement." *United Nations Climate Change*, https://unfccc.int/. Accessed 30 Apr. 2021.
[3] "Decarbonising America." *The Economist*, Feb. 2021.
[4] Rollett, Catherine. "Solar Costs Have Fallen 82% since 2010." *PV Magazine*, June 2020.
[5] Ritchie, Hannah, et al. "GitHub - Owid/Energy-Data: Data on Energy by Our World in Data." *Our World in Data*, https://github.com/owid/energy-data. Accessed 30 Apr. 2021.
[6] Majaski, Christina. "Developed Economy." *Investopedia*, 21 Nov. 2020.
[7] Storrow, Benjamin. "U.S. Wind and Solar Installations Are Smashing Records, but the Trend May Not Last." *Scientific American*, Springer Nature America, Inc., Oct. 2020.