Evaluating the financial decision in renewable energy production investments.

Section 1 - Background

As world governments become more concerned with their greenhouse gas emissions, investment in dirtier energy sources has declined on a global scale. Recently, the president of Equinor was interviewed by the FT on this topic:

"We're currently spending about $400mn or $500mn a year globally in oil and gas. That's exactly what we have to do if we're on our way to the net zero emissions scenario," said Wærness.

But slashing oil supply only works if an increase in clean energy infrastructure drives down oil demand at the same time — and that is not happening, says Wærness: transition investment is nowhere near the level needed to wean global economies off oil.[1]"

This problem exists even as the United States and the European Union have introduced substantial subsidy plans to encourage investment in solar and wind power. This leads me to believe that there is a lack of returns on renewables projects relative to "dirty" energy projects (or other investments). The Economist covered this issue their recent article: *The world won't decarbonize fast enough unless renewables make real money[2]*. We often hear about the moral reasons for promoting renewable energies, but we are less often confronted with the financial decision. Reading these articles, I asked myself: How important is the financial decision when it comes to domestic investment in renewable energy?

The equity risk premium (ERP, the additional compensation above the risk-free rate required to hold a given asset) varies by country and is a function of various risk factors. Given that the ERP for a country considers the systemic risk involved in investing in the country, it can operate as a proxy for "willingness to invest" in a country, where higher ERPs represent lower willingness to invest.

My hypothesis is: the larger a country's equity risk premium and the more endowed the country is with fossil fuels, the higher returns required from a given clean infrastructure investment, and therefore, countries with higher ERPs and fossil fuel endowments will produce lower amounts of renewable energy per capita. If this is true, it implies that we would need to increase the financial incentive to build renewable energy projects in countries that have access to domestic fossil fuels.

This report will investigate this hypothesis by **exploring the relationship between the domestic production of renewable energy, and a country's equity risk premium, it's domestic extraction of fossil fuels (coal, natural gas, and petroleum), domestic production of nuclear energy, and area in square kilometers for the year 2019.**

The year 2019 was chosen since it is the most recent pre-covid year for which we have all required data. All energy production and population data used for was sourced from the U.S. Energy Information Administration (EIA). Equity risk premium data is harder to come by and not "standardized". I decided to

---

[1] https://ep.ft.com/permalink/emails/eyJlbWFpbCI6ImEyMGZmZGU5ZWU5OTgyNzljYmE2YmJhZmFlOWI0M2VjZGI4YzI3OTJhNCIsInRyYW5zYWN0aW9uSWQiOiIwODQ2MjNiOS1hZjFiLTQ4NGEtYjhhYS0wNWFjOTMzYjJkOTkifQ%3D%3D

[2] https://www.economist.com/leaders/2023/02/16/the-world-wont-decarbonise-fast-enough-unless-renewables-make-real-money

source equity risk premium data from NYU Professor Aswath Damodaran, who releases his international ERP estimates each year on his website[3]. There are degrees of accuracy with which you can estimate ERP. ERP is estimated most accurately if a country has a sovereign credit default swap market, therefore only counties with such a market have been included in the dataset for this paper.

Section 2 – Data Collection & Exploration

Section 2.1 – Data Structure and Variable Deep-Dive

Table 1 – Data Structure

Variable	Description	Transformation Applied	Source
ISO	ISO code for each country	none	EIA[4]
ERP in %	Equity risk premium is the excess return demanded by investors over the risk free rate in order to invest into a given asset.	Further explained here	Aswath Damodaran[5]
Population	Population in millions of people		EIA
CoalProduction/ Capita	Domestic coal production in BTU per person	$\frac{coal_{production}}{Pop} * 1{,}000{,}000$	EIA
GasProduction/ Capita	Domestic gas production in BTU per person	$\frac{natural\ gas_{production}}{Pop} * 1{,}000{,}000$	EIA
CrudeOilProduction/ Capita	Domestic crude oil production in BTU per person	$\frac{crude\ oil_{production}}{Pop} * 1{,}000{,}000$	EIA
NuclearProduction/Capita	Domestic nuclear energy production in BTU per person	$\frac{nuclear_{production}}{Pop} * 1{,}000{,}000$	EIA
RenewableProduction/ Capita	Domestic renewable energy production in BTU per person	$\frac{renewable_{production}}{Pop} * 1{,}000{,}000$	EIA
Area	Area of a country measure in Km^2		World Bank

Table 1 describes the variables that will be used to investigate the relationship of interest. The reasons for including all domestic production of non-renewable energy are two-fold, and can be explained using an example. In a country endowed with a lot of fossil fuels (Saudi Arabia), there is potential downward

[3] https://pages.stern.nyu.edu/~adamodar/pc/archives/ctryprem19.xls. For more information on how ERP is estimated, visit: https://raw.githubusercontent.com/AlexNever/multi-reg/main/assignment2_f/erp_calculation

[4] https://www.eia.gov/international/data/world

[5] https://pages.stern.nyu.edu/~adamodar/pc/archives/ctryprem19.xls

pressure on the production of renewable energy through two forces. The first force comes as the result of the abundance of domestic fossil fuels on the investment decision. One could argue that the returns on the marginal dollar invested in domestic renewable energy will need to cover the opportunity cost of having not invested in a domestic fossil fuel project. This opportunity cost may be higher in countries endowed with larger fossil fuel deposits. The second force is energy policy. If Saudi Arabia's oil production is a large percentage of its output as an economy, there may be political pressure to keep supporting the industry through fiscal and regulatory policy. For these reasons, it is important to understand the "profile" of a country's domestic energy endowments and include that context when explaining the country's production of renewable energy. All energy production variables are expressed in per capita terms to control for the population size of the given country.

I decided to include the area of each country in the sample as an additional variable. It would be unfair to compare the renewable energy capacity of, for example, Hong Kong, with the renewable energy capacity of China, purely on a per capita basis since there are limitations on the amount of renewable energy infrastructure that could physically fit into Hong Kong.

The reasons for including ERP were discussed in Section 1.

Section 2.2 – Data Exploration

Section 2.2.1 – Summary Statistics

Table 1 – Summary Statistics of Data

VARIABLE	**N**	MEAN	STD. DEV.	MIN	PCTL. 25	PCTL. 75	MAX
COAL_POP_M	70	18025	61678	0	0	8834	479131
ERP	70	6.8	1.7	5.2	5.6	7.5	12
GAS_POP_M	70	74069	298787	0	111	19163	2340479
NUC_POP_M	70	6703	13968	0	0	4428	62871
OIL_POP_M	70	76416	224210	0	1	26415	1372591
REN_POP_M	70	17159	29338	33	1935	18195	216234
AREA_SQK	70	1167329	2733112	783	64188	763106	16376870

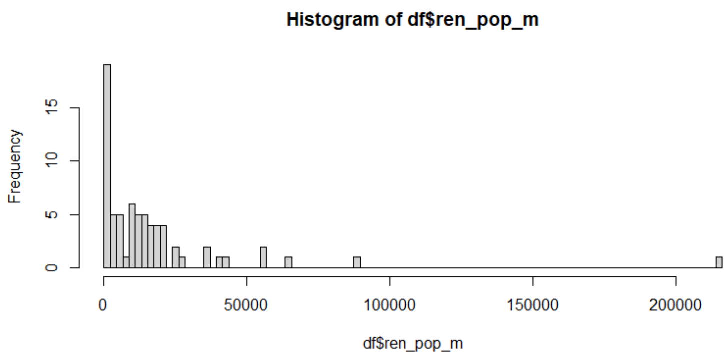
Let's begin by evaluating the summary statistics for our predictor variables. All predictor variables related to energy production (COAL_POP_M, GAS_POP_M, NUC_POP_M, OIL_POP_M), have an interesting issue. They are all extremely right tailed[6] but have minimum values of zero. Usually, taking the natural logarithm is a good way of dealing with extremely right tailed variables. However, it is not possible to take the natural logarithm of zero (or rather, the value of ln(0) in infinite). To solve this issue, I have decided to take the natural logarithm of the values in each of these variables plus one (referred to as a log-plus-one transformation). This is not a perfect solution, but I thought it could be appropriate given that none of the values for any of these variables that were larger than zero were less than or equal to 1. Therefore, taking their natural logarithm would not make the resulting logged variables negative or zero. Although not a perfect solution, I will be treating coefficients for these variables as if they were ordinary logged variables for the purposes of this report, since the approximation becomes

[6] See Appendix

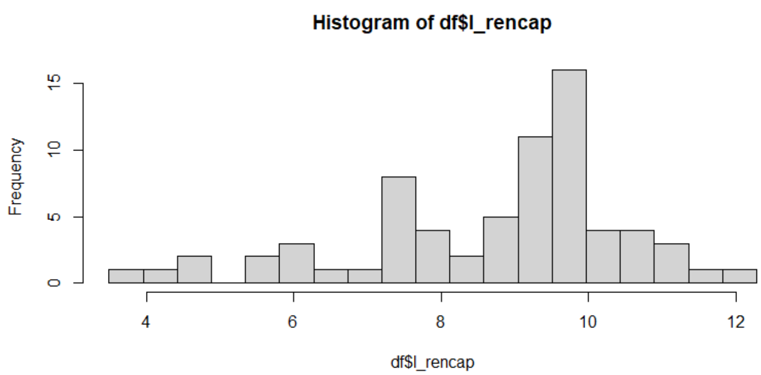
more like a standard log transformation, the larger the value of x[7]. This issue did not exist for the renewable energy production per capita, ERP or area variables, and therefore both were transformed using a standard log transformation.

Taking a closer look at the distribution of our dependent variable (renewable energy production per capita), we see that the log transformation helps suppress the right-tailed nature of the variable:

Graph 1 – Histogram of renewable energy production per capita



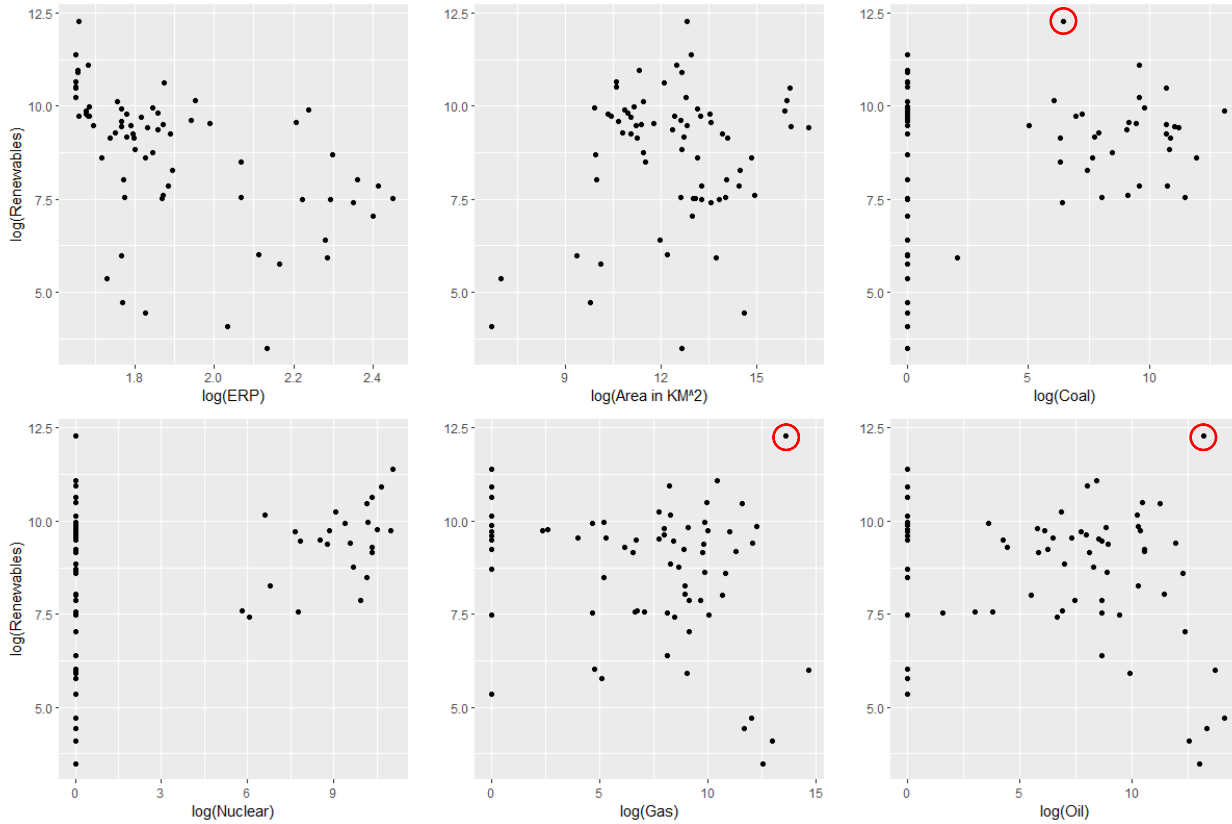
Graph 2 – Histogram of log(renewable energy production per capita)



[7] https://stats.stackexchange.com/questions/576504/interpreting-log-log-regression-with-log1x-as-independent-variable - There is an interesting comparison quoted here.

Section 2.2.2 – Scatterplots

Graphs 3a – 3f – Scatterplots of log(renewable energy production per capita) vs. predictors



Starting with ERP, there seems to be a weak inverse relationship between renewables and ERP as hypothesized in Section 1. Similar inverse relationships are observed between renewables and oil and gas production, which seems to support the existence of the "forces" outlined in Section 2.1. Looking toward coal, it is unclear if any relationship exists between renewables and coal production. Finally, although there are many zero values for nuclear energy production, there seems to be a strong positive relationship between renewables and nuclear production. It could be that the renewable energy policy push in some countries with large nuclear capacity includes provisions for nuclear power plant subsidies, since some countries classify nuclear energy as renewable or "sustainable".

One country, Norway, has been highlighted in the scatterplots above. Norway is uniquely endowed with vast amounts of fossil fuels, and uses the proceeds from these industries to push for more profitable domestic production of renewable energy projects[8]. However, I will not be removing Norway from the analysis, since the country is only an outlier for three of our variables of interest.

[8] https://energifaktanorge.no/en/om-energisektoren/verdt-a-vite-om-norsk-energipolitikk/

Section 3 – Regression Models

The first regression model in this paper will take the following shape:

$$\ln(ren_{pop_m}) = \alpha + \beta_1 \log(ERP) + \beta_2 ln(area) + \beta_3 ln(oil_{pop_m}) + \beta_4 ln(gas_{pop_m}) + \beta_5 ln(coal_{pop_m}) + \beta_6 ln(nuclear_{pop_m})$$

Output 1 – Linear model 1

```
MODEL FIT:
F(6,63) = 8.2212, p = 0.0000
R² = 0.4391
Adj. R² = 0.3857

Standard errors: OLS
------------------------------------------------------
                        Est.      S.E.     t val.        p
----------------- ---------- --------- ---------- --------
(Intercept)          13.7283    1.8044     7.6083   0.0000
erp                  -3.7343    0.7882    -4.7380   0.0000
l_area                0.2326    0.1091     2.1327   0.0369
l_oilcap             -0.0234    0.0755    -0.3097   0.7578
l_gascap             -0.1465    0.0819    -1.7888   0.0785
l_coalcap             0.0601    0.0463     1.2981   0.1990
l_nuccap              0.0278    0.0425     0.6537   0.5157
------------------------------------------------------
```

Linear model 1 contains some important results. The F-Statistic for the overall regression implies that the model is statistically significant[9]. Additionally, the adjusted R^2 of this model implies that around 39% of the variation in the logged renewable energy production can be explained by the predictors. The standard error of the regression is 1.417, which indicates that the dependent variable could roughly be predicted within $\pm 2\hat{\sigma} =$ 2.834 ln(renewable energy production per capita) using linear model 1. The coefficients can be interpreted as follows (each of these interpretations assume holding all else fixed and the dependent variable refers to renewable energy production per capita):

- The intercept of this model is difficult to interpret given we used a log-plus-one transformation for some of our predictors. Usually, the intercept of a log-log model will estimate the expected log(dependent) when all predictors are equal to 1. Given the shape of our data (as explained in Section 2.2.1), the intercept estimates the expected log(dependent) given x values take the value of 1 for ERP and l_area, and the value of 0 or 1 for all energy production per capita predictors. I would argue that the intercept in linear model 1 (and linear model 2) is less important to understand than the coefficients of the models' predictors, since the minimum non-zero values of all the non-logged predictors are > 1 (4.03 is the smallest observation from any predictor variable and is observed in oil production per capita).
- A one percent increase in ERP is associated with a -3.73% change in the dependent variable. This relationship is statistically significant.
- A one percent increase in area is associated with a 0.23% change in the dependent variable. This relationship is statistically significant.
- A one percent increase in oil production per capita is associated with a -0.02% change in the dependent variable. This relationship is **not** statistically significant.

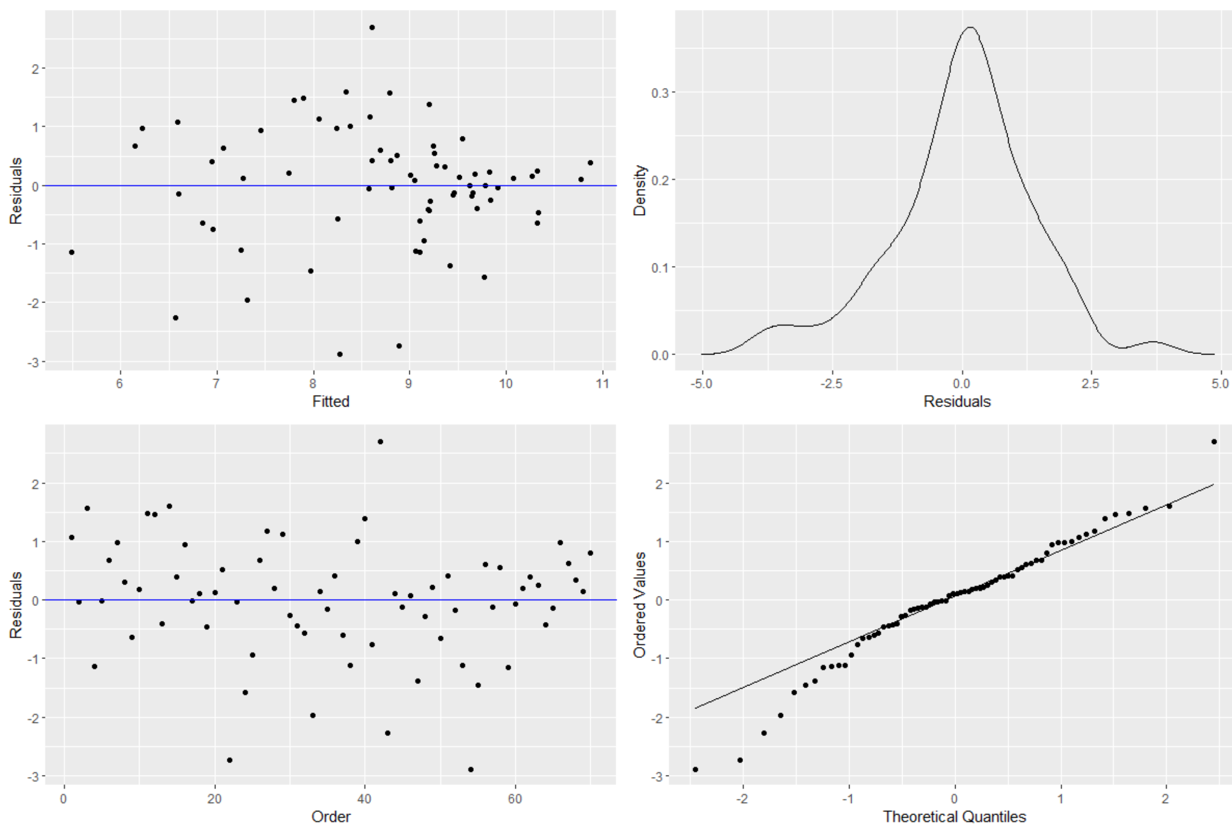
---

[9] Anytime anything in this paper is said to be "statistically significant", it implies that it is significant at the 5% level at minimum.

- A one percent increase in gas production per capita is associated with a -0.15% change in the dependent variable. This relationship is **not** statistically significant.
- A one percent increase in oil production per capita is associated with a 0.06% change in the dependent variable. This relationship is **not** statistically significant.
- A one percent increase in oil production per capita is associated with a 0.03% change in the dependent variable. This relationship is **not** statistically significant.

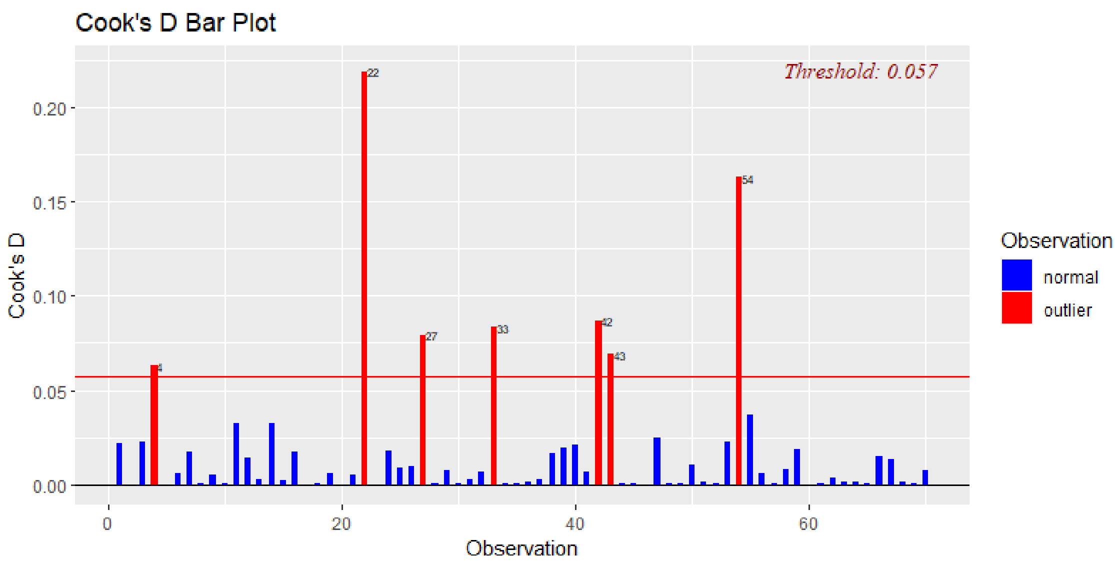
Linear model 1 seems to support part of the hypothesis. High ERP values for countries are associated with lower renewable energy production. The effect of ERP is also highest, indicating that this relationship is not only statistically significant, but also large. In addition, there is a positive relationship between renewable energy production per capita and the size of a country, supporting the hypothesis that larger countries are able to physically fit more renewable energy infrastructure within their borders. This model **does not** seem to suggest that there is any significant relationship between renewable energy production and the production of any other energy source.

Graphs 4a – 4d – <u>Standardized</u> Residual Plots for Linear Model 1



Graphs 4 do not point to any patterns in the residuals vs. order and vs. fitted plots. However, there is a strong indication of left skewness in our standardized residuals from our qq-plot and density plot.

Graph 5 – Cook's D Bar Plot for Linear Model 1



Running a Cook's D for each observation in the model flags seven points as potential outliers using a threshold of 0.057. No individual observation has a Cook's D greater than 1, indicating that (according to this diagnostic) we have no leverage points that could have an outsized effect on the regression in linear model 1.

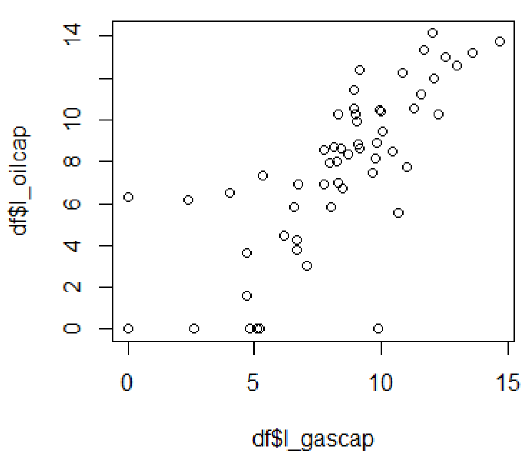
However, there are issues with collinearity in this model. Running a VIF analysis yields the following result.

Output 2 - VIF Analysis

ERP	L_area	L_oilcap	L_gascap	L_coalcap	L_nuccap	1/(1-R^2)
1.13	1.53	4.08	3.99	1.63	1.24	1.78

Variables l_oilcap and l_gascap both have VIFs that are larger than 1/(1-R^2) of the model. This makes sense since oil and gas are often found in the same place. Therefore, **if** a country has domestic oil operations, they are more likely to also have domestic gas operations:

Graph 6 – ln(oil production per capita) vs. ln(gas production per capita)



Therefore, it could make sense to generalize this model further by running a simplified model where oil and gas are summed to create the variable "oil and gas production per capita" (or sum_oil_gas"). The form and results of this model are presented below:

$$\ln\left(ren_{pop_m}\right) = \alpha + \beta_1 \log\left(ERP\right) + \beta_2 ln(area) + \beta_3 ln\left(oil_{pop_m} + gas_{pop_m}\right) + \beta_4 ln\left(coal_{pop_m}\right) + \beta_5 ln\left(nuclear_{pop_m}\right)$$

Output 3 – Linear Model 2

```
MODEL FIT:
F(5,64) = 9.9492, p = 0.0000
R² = 0.4373
Adj. R² = 0.3934

Standard errors: OLS
```

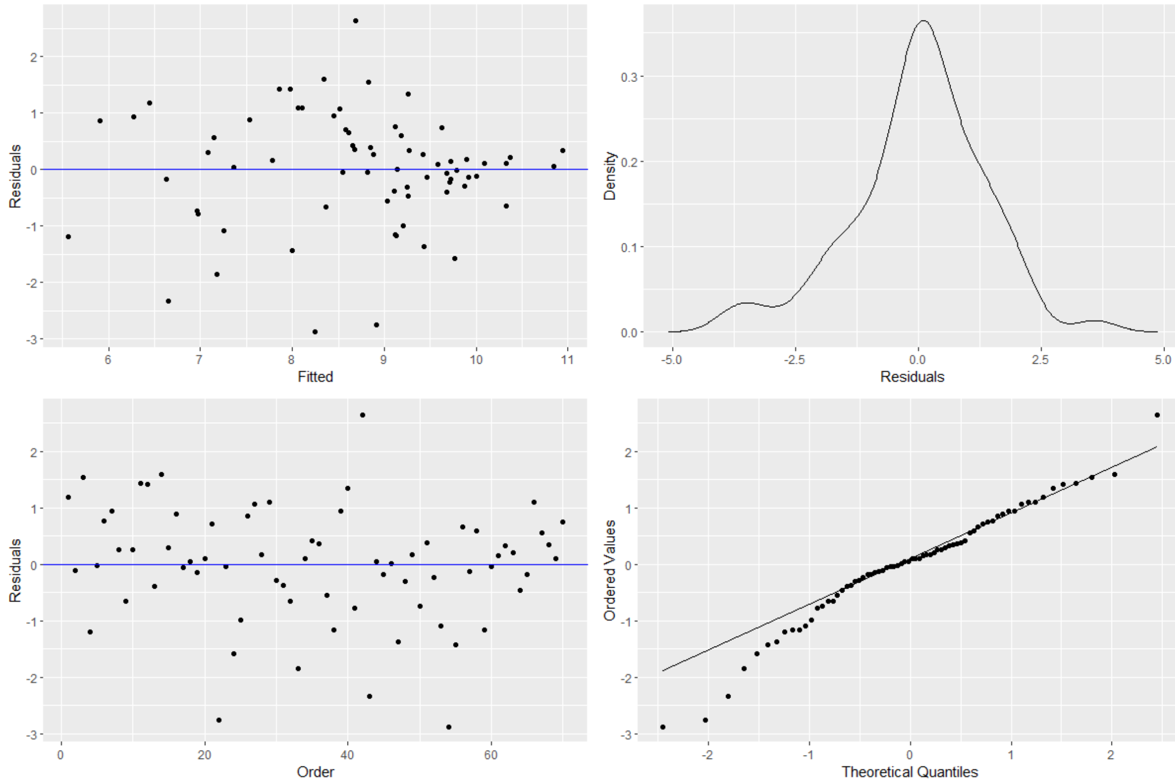
	Est.	S.E.	t val.	p
(Intercept)	13.5943	1.7487	7.7740	0.0000
erp	-3.6858	0.7823	-4.7117	0.0000
l_area	0.2428	0.1052	2.3078	0.0243
sum_oil_gas	-0.1598	0.0434	-3.6821	0.0005
l_coalcap	0.0546	0.0453	1.2051	0.2326
l_nuccap	0.0269	0.0421	0.6396	0.5247

Linear model 2 shows promising support of the stated hypothesis. The F-Statistic for the overall regression implies that the overall model is statistically significant. Additionally, the adjusted R^2 of this model implies that around 39% of the variation in the logged renewable energy production can be explained by the predictors. The standard error of the regression is 1.408, which indicates that the dependent variable could roughly be predicted within $\pm\, 2\hat{\sigma} = \; 2.816$ ln(renewable energy production per capita) using linear model 2. The coefficients can be interpreted as follows (each of these interpretations assume holding all else fixed and the dependent variable refers to renewable energy production per capita):

- (See linear model 1 for intercept interpretation)
- A one percent increase in ERP is associated with a -3.69% change in the dependent variable. This relationship is statistically significant.
- A one percent increase in area is associated with a 0.24% change in the dependent variable. This relationship is statistically significant.
- A one percent increase in oil and gas production per capita is associated with a -0.16% change in the dependent variable. This relationship is statistically significant.
- A one percent increase in oil production per capita is associated with a 0.05% change in the dependent variable. This relationship is **not** statistically significant.
- A one percent increase in oil production per capita is associated with a 0.03% change in the dependent variable. This relationship is **not** statistically significant.

An interesting development in the generalized version of linear model 1 is the statistically significant effect oil and gas production per capita have on the production of renewable energy production per capita. From this result, one could infer that the forces outlined in Section 2.1 potentially exist in some capacity. The relationships between the dependent variable and ERP and Area remained statistically significant.

Graphs 7a – 7d – <u>Standardized</u> Residual Plots for Linear Model 2



Although linear model 2 looks more promising than linear model 1, we still observe strong left skewedness in our standardized residuals.

Section 4 – Model Choice

Although linear model 2 looks promising, it is important to evaluate if the two models are significantly different from one another.

Output 4 – Partial F-Test

```
Model 1: l_rencap ~ erp + l_area + l_oilcap + l_gascap + l_coalcap + l_nuccap

Model 2: l_rencap ~ erp + l_area + sum_oil_gas + l_coalcap + l_nuccap

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     63 126.54
2     64 126.94 -1  -0.40504 0.2017 0.6549
```

The p-value of 0.6549 suggests that we cannot reject the null hypothesis that the models are significantly different from one another, in which case we choose to go with the more generalized model: linear model 2. Going with the second model also makes more sense from a qualitative perspective.

Section 5 - Conclusion and Limitations

The initial hypothesis for this paper was: *the larger a country's equity risk premium and the more endowed the country is with fossil fuels, the higher returns required from a given clean infrastructure investment, and therefore, countries with higher ERPs and fossil fuel endowments will produce lower amounts of renewable energy per capita*.

Linear model 2 provided some support for this hypothesis, suggesting that larger ERPs and larger gas and oil endowments were negatively associated with renewable energy production per capita in our sample of countries that had ERP estimations that were informed by the given country's CDS spread. The effect of ERP was largest, suggesting that a countries perceived investment grade plays a large role in the decision to invest in domestic renewable energy projects. If one were to draw policy conclusions from this preliminary analysis, it could be suggested that domestic subsidies for renewable energy projects could crowd out investment into renewable energy projects in developing countries with higher ERPs.

It should be noted that the analysis performed in this paper has severe limitations in its predictive value. Some of these limitations are:

1. The log-plus-one transformation performed on some of the predictors limits our ability to accurately interpret the coefficients for these variables, and therefore limits the overall predictive value of both linear models 1 and 2.
2. The data used to build these models is from 2019, including the ERP. Although ERP does not generally vary a lot from year to year for more developed countries, it does tend to vary a lot for developing countries. Given that the investment in a renewable energy production started at time t, will not yield renewable energy until some time t+n, there is risk that some of the renewable energy projects yielding production in 2019 were started in some year t in which the ERP for a given country was lower. It would therefore be interesting to do this analysis across time and potentially explore lagged effects of ERP on new renewable energy production per year, and run the same analysis with other energy sources as the dependent variable. This was out of scope for this paper.

The R-code for this paper can be found here: https://github.com/AlexNever/multi-reg/blob/main/assignment2_f/analysis_assign3.R