Causal Effects of Renewable and Fossil Fuel Energy Sectors

Project proposal

Luca Dibattista, George Maratos

1 INTRODUCTION

The United Nations Sustainable Development Goals[1] are a set of objectives that every nation should take as a reference to improve the quality of life on the whole planet. Some examples of the 17 goals are: "no poverty," "zero hunger," "good health and well-being," "quality education," etc. Starting from the UN SDG, we focused on goal 7 "affordable and clean energy." We will search for causalities between the energy source and other factors.

2 DATA

We plan to use multiple UN data sets[2], collected by the United Nations Statistics Division, and combine them. Those are the National Accounts and Main Aggregates and the following databases: Energy Statistics, Demographic Statistics, Environment Statistics, Commodity Trade Statistics, and Industrial Commodity Statistics. Also the Greenhouse Gas Inventory Data from the United Nations Framework Convention on Climate Change.

Using these, we can collect data in the following topics from numerous countries sampled at one-year intervals, over multiple years[3]: the production of electricity from various sources like coal and solar, economic indicators like GDP and GNI, greenhouse gas emissions, water and waste, demographic information like the number of people employed in various sectors and education levels, electrical consumption and demand, and finally import and export quantities of various products that could be relevant to the renewable energy and fossil fuel industries either directly or through their supply chains.

Since the data is for various countries and over many years, we choose to use those as units of measurement for the relevant variables in our model. So we will merge all of our data on the number of years or country. These data sets are very large, there are typically around 100k or more entries, where each entry covers a single country in a single year for some indicator of interest.

3 PROBLEM DESCRIPTION

In the last few years, governments have made huge investments in the renewable energy sector. Our goal is to better explore the effects of the transition from fossil fuel to renewable energy sources. Figure 1 shows the causal graph of the main causes and consequences of the two sectors. The graph contains not only variables of the energy sector itself, but also variables concerning the quality of life. All variables are obtained from datasets discussed in Section 2.

The most important root node of the graph is the electrical demand: this is the amount of energy demanded by the population. Commonly, an increase in demand is followed by an increase in energy production (and, vice versa, a lower demand causes a lower production). We can take as an example a specific company that offers energy to consumers and businesses. If the energy demanded by that company from the market increases, it is likely that the company will hire new people: this is why we indicated that the demand causes employment in the energy industry.

Energy production (both from renewable and fossil fuel sources) causes the production of CO_2. For instance, while the coal emanates, on average, 888 tonnes of CO_2 per GWh, the wind only emits 26 tonnes of CO_2/GWh [3]. The production of energy also causes its consumption by the population. The energy production from renewable sources usually depends on the weather: for example solar panels, wind turbines, or ocean power plants.

Greenhouse gases may worsen human development (e.g. through pollution).

We already discussed that when the energy demand increases, more people are employed. We also added an edge between education and employment, since in the hiring process a person's education is taken into account.

Employment and the production of raw materials both affect GDP. We will use GDP per capita as an indicator in the constant version (i.e., prices are adjusted for the effects of inflation).

In our work, we will answer questions regarding the causes that renewable and fossil fuel energy sources have on the economy (with the GDP) and on society (with the employment and human development variables). The human development variable is given by the Human Development Index (HDI)[4], an index that considers life expectancy, education, and the GNI (gross national income) index.

4 RESEARCH PLAN

Our data has a temporal aspect to it, so we will use this information to construct temporal models. This temporal nature is useful because it allows us to quantify how indicators influence each other by observing how they change over time. For example, we could observe the change in electrical consumption in one year and see how that influences the change in energy production in the next and that would allow us to measure how much electrical consumption influences energy production. We assume Granger causality [4] and say that some variable Y in our graph is better described by some function that includes previous times steps of other variables, for example X, along with previous Y values. This would make sense in our case. One might be able to predict the current year's energy production just by knowing last year because on the large scale of countries they probably won't make dramatic changes in production, but you might better predict this year's output by also knowing what was last year's energy demand.

The above scenario is a naive view of the causality of energy demand over time. It might be the case that meeting the demand requires more than a year, maybe because it could be the case that

[1] https://www.un.org/sustainabledevelopment

[2] You can find them at http://data.un.org/

[3] The number of years varies per dataset, but we estimate that we will have at least 20 years of workable data

[4] http://hdr.undp.org/en/content/human-development-index-hdi

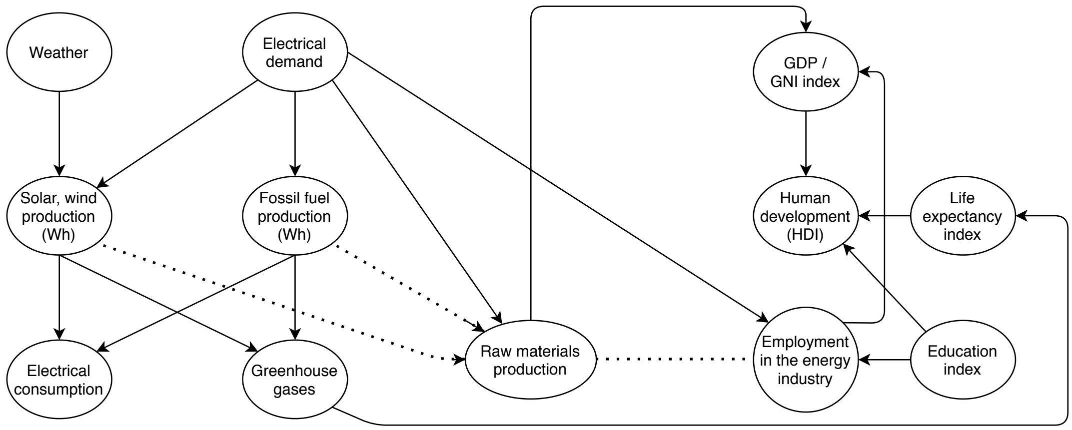

Figure 1: The causal graph. It includes the most relevant effects and causes of the energy sector. A dotted arrow means that we are still not sure whether to add it or not (i.e., we are not sure about the causal effect of one variable on another one).

a specific power plant took longer than usual to build in some country. Therefore, it is more realistic to assume that the lag might vary based on a number of factors in our real-world data. We can consider two competing frameworks for discovering Granger Causality, there is a collection of Granger based methods that rely on a fixed lag, some of which are described in [2], and the other being VL-Granger [1] which allows us to consider causes with variable time delays. We can try both to discover the role that time plays in the model described in Figure 1.

There are some aspects that might make us change our approach later. As discussed earlier, we believe that the graph needs more careful analysis to determine what is the true structure. Second, it could be the case that when we merge the data, our sequences are not long enough, and we might have trouble with the sequence-based causality models. Our plan going forward is to merge as much data as we can from the sources described in Section 2, in an attempt to develop long and descriptive sequences of the variables in our model. Then we will test out the fixed length and the variable length Granger methods described in the previous paragraph.

REFERENCES

[1] Chainarong Amornbunchornvej, Elena Zheleva, and Tanya Y Berger-Wolf. "Variable-lag Granger Causality for Time Series Analysis". In: *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. 2019, pp. 21–30.

[2] Andrew Arnold, Yan Liu, and Naoki Abe. "Temporal causal modeling with graphical granger methods". In: *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2007, pp. 66–75.

[3] Greenhouse Gas Emissions. "Comparison of Lifecycle Greenhouse Gas Emissions of Various Electricity Generation Sources". In: ().

[4] Clive WJ Granger. "Investigating causal relations by econometric models and cross-spectral methods". In: *Econometrica: journal of the Econometric Society* (1969), pp. 424–438.