Data Report on Project "Evaluation of the effectiveness of carbon pricing in greenhouse gas emission reduction"

Main Question

In my project, I want to investigate the effectiveness of carbon pricing in reducing greenhouse gas emissions. Specifically, I want to answer the following questions: How effective are carbon pricing policies in reducing greenhouse gas emissions?

Data Sources

Therefor, I will use the following two data sources: World Bank's Carbon Pricing Dashboard and Emissions Database for Global Atmospheric Research: Greenhouse Gas emissions from all countries in the world.

World Bank's Carbon Pricing Dashboard

The World Bank's Carbon Pricing Dashboard Data can be downloaded here. It provides information on carbon pricing initiatives around the world. It's based on the World Bank's annual State and Trends of Carbon Pricing report series. The data includes, but isn't limited to, the following information. The name of the carbon pricing initiative and the instrument used, the country implementing the policy, some statistical information about the country, prices per tonne in US$ from 1990 to 2024, the sector covered by the policy, the coverage of the policy.

Data Structure and Quality

The data is available in XLSX format and is updated regularly by the World Bank. The data appears to be quite accurate as it is based on the World Bank's annual State and Trends of Carbon Pricing report series of the world bank. It appears to be fairly complete as it includes data for 142 pricing programmes around the world. The data is well structured and appears to be consistent. It contains data up to 2024, which is perfect. The relevance is given as the data is from an official source and fits the purpose of my analysis.

License

The data is available under the Creative Commons Attribution 4.0 International License. As this is a standard open-data license, I am allowed to use the data as long as I give appropriate credit to the World Bank, indidacte changes made and provide a link to the licence. I plan to fulfil this obligation by citing the source of the data in my final report and by attributing the sources where possible.

Emissions Database for Global Atmospheric Research: Greenhouse Gas emissions of all world countries

The emissions list can be downloaded here. It provides information on greenhouse gas emissions from all countries in the world. The data are based on the EDGAR (Emissions Database for Global Atmospheric Research) Community GHG Database. The data include the GHG emissions of CO_2 (fossil only), CH_4, N_2O and F-gases. They are aggregated using Global Warming Potential values from IPCC AR5 (GWP-100 AR5). There

are different sheets that contain the data for different levels of interest. The data are available for the years 1970-2022.

Data Structure and Quality

The data is available in XLSX format. The data seems to be quite accurate as it is based on the EDGAR Community GHG Database. It seems to be quite complete as it contains data for 210 countries around the world. The data is well structured and appears to be consistent. It contains data up to the year 2022, which should be sufficient for my analysis. The relevance is given as the data is from an official source and fits the purpose of my analysis.

License

The data is available under the Creative Commons Attribution 4.0 International License. As this is a standard open data licence, I am allowed to use the data as long as I give appropriate credit to the World Bank, indidacte changes made, and provide a link to the license. I plan to fulfil this obligation by citing the source of the data in my final report and by attributing the sources where possible. The data is ordinaly form die EDGAR Community GHG Database(EDGAR (Emissions Database for Global Atmospheric Research) Community GHG Database (a collaboration between the European Commission, Joint Research Centre (JRC), the International Energy Agency (IEA), and comprising IEA-EDGAR CO2, EDGAR CH4, EDGAR N2O, EDGAR F-GASES version 8.0, (2023) European Commission. IEA-EDGAR CO2 (v2), a component of the EDGAR (Emissions Database for Global Atmospheric Research) Community GHG database version 8.0 (2023) including or based on data from IEA (2022) Greenhouse Gas Emissions from Energy, www.iea.org/statistics, as modified by the Joint Research Centre). Users of the data are obliged to acknowledge the source of the data also with reference to the EDGAR report webpage (https://edgar.jrc.ec.europa.eu/report_2023) and EDGARv8.0 website (https://edgar.jrc.ec.europa.eu/dataset_ghg80) and/or relevant reports).

Data Pipeline

First I tried to do my pipeline in jayvee, which didn't work because I could delete the first row in an excel sheet, I switched to python. I used the pandas library to read the data from the excel sheet and to extract the sheets of interest. The automatic datatype detection worked quite well. But I still had to do some manual work on the carbon price sheet.Because if there is no price for a year, the cell is "-" and the datatype is recognised as text. I wanted to change it to be a float value for better calculation later in my project, so every "-" had to be changed to NaN. I also had to delete the first row of the carbon price sheet because it was a timestamp. The emissions sheet was already in a good format. If a data source is not available or is not in the expected format, the pipeline will throw an error and stop. The user has to fix the problem and restart the pipeline. Only if everything works as expected will the pipeline proceed to the data type conversion and data cleansing steps.

Result and Limitations

The resulting data is stored in a sqlite database. The file is called "data.sqlite" and is located in the data directory. This makes it easy to query the data in future analyses. The data is stored in two tables. One table for the carbon price data (carbonPrice) and one for the greenhouse gas emissions data (emissions). As the data fits really quite well, the data stored in the database is quite similar to the original data. The only differences are the small changes for better calculation in the carbonPrice table. The quality of the data seems to be very good as it comes from official sources. The data is up to date and seems to contain all the

information I need for my analysis. The only problem I see at the moment is that the emissions data go up to 2022 and the carbon price data up to 2024. So I have to think about how to deal with this in my analysis.