

CSC110 Project Report: The effectiveness of the carbon tax policy on reducing CO2 emission levels

Mojan Majid, Kimiya Raminrad, Dorsa Molaverdikhani, Ayesha Nasir

Monday, December 14, 2020

Problem Description and Research Question

One of the main challenges of our modern world is global warming. According to NASA, this phenomenon is caused by the expansion of greenhouse effects by humans ("The Causes of Climate Change", 2020). The greenhouse effect happens when certain gases like CO2 do not let the heat radiating from the Earth towards space to escape from the atmosphere ("The Causes of Climate Change", 2020). In this era, human activities like burning fossil fuels increase the amount of CO2 in the atmosphere. Global warming has devastating and severe effects like a rise in sea levels, more severe heat waves, defrosting the ices, etc ("The Effects of Climate Change", 2020).

One of the approaches that has been launched by some governments to help reduce CO2 emissions is carbon pricing. Using this approach, governments calculate the external costs of carbon emissions ("Pricing Carbon," n.d.). People are required to pay these costs, which are related to damage to crops due to drought or health care costs related to heatwaves ("Pricing Carbon," n.d.).

Carbon pricing costs will be linked to the sources that cause them and create CO2 pollution, using carbon pricing. In this approach, the emitters and the companies which produce CO2 are responsible ("Pricing Carbon," n.d.). The more CO2 they produce, the more money they have to pay. As a result, the emitters and polluters are encouraged to develop ways that reduce CO2 emission and pollution ("Economics of Climate Change", 2020). In this market-oriented approach, the polluter is responsible for finding a way to reduce its CO2 emission ("Economics of Climate Change", 2020).

Global warming has influenced our planet severely, and if not controlled, the possible damages may be irreparable. Consequently, some approaches should be launched by governments in order to stop industries from emitting excessive CO2. This motivates our group to look at the methods that have been used to mitigate global warming, especially ones implemented by governments, as some industries have significantly contributed to the emission of CO2.

As we explained above, one of the used approaches is carbon pricing. We are curious to determine the effectiveness of carbon tax -one type of carbon pricing- by examining the amount of CO2 for countries in our dataset using this method for several years. Our question is **How effective carbon tax has been in reducing CO2 emissions for countries in our dataset?**

Dataset Description

- Dataset 1: co2-emissions-vs-gdp

Source: Our World in Data - <https://bit.ly/3ngrAIT>

Format: CSV file

Description: This dataset contains the amount of CO2 emissions and Gross domestic product (GDP) per capita in 2011 international dollars of countries throughout the years.

Columns used in the program:

- Entity: changed to 'country' in the program
- Year: changed to 'year'

- Real GDP per capita in 2011US\$, 2011 benchmark (Maddison Project Database (2018)): changed to 'gdp_per_capita'
- Per capita CO2 emissions: changed to 'co2_per_capita'
- Dataset 2: CPI_Data_DashboardExtract
 - Source: The World Bank Carbon Pricing Dashboard - <https://bit.ly/2Kpyc9m>
 - Format: XLS file
 - Description: In the carbon pricing dashboard we make the following selections: STATUS: Implemented, TYPE OF INSTRUMENT: Carbon tax, TYPE OF JURISDICTION: National
 - We then download the relevant dataset which has data on the jurisdiction, type, status, and year of implementation of initiatives implemented in several countries.
 - We then delete the first two rows of the dataset so that the dataset starts with the name of the columns, and convert the Data_Overall tab of the file to a CSV format. The new file is called Data_Overall.csv.
 - Columns used in the program:
 - Year of implementation
 - Jurisdiction covered

Computational Overview

We used the pandas library, which provides us with the `DataFrame` data structure, and we manipulated and filtered our data using the methods it offers. To read the csv files mentioned in the dataset description, we used `pd.read_csv`. In the `cleaning_data.py` program, we created several functions that use pandas and restrict or change aspects of a given dataframe.

The function `rename_the_columns` changes the column names of the given dataframe using `df.columns` from pandas. In the `country_dataframe` function, we create a new dataframe for a given country. We used `df.dropna()` to ensure the new dataframe does not have any missing values. We also used `df.reset_index` to reindex our dataframe after we made some changes to it, and `df.shape` to access the number of rows and columns in our dataframe. We also restrict the data to years before or after the implementation of a carbon tax using pandas in the functions `before_implement` and `after_implement`.

We used scikit-learn library's estimator methods and LinearRegression model. We are interested in finding the changes of association between year and a given y before and after the implementation of a carbon tax and investigate the effectiveness of the initiative. Thus, we are working on a regression problem. We use LinearRegression to fit a linear model to our data.

We also use the estimator method `fit()` to reach our goal. `model.fit()` is useful in our program because it fits the linear model. Also we use `model.coef_[0]` and `model.intercept_` to find the slope and intercept of each line and plot the fitted line. From this data, we can see how the slopes of the two lines compare.

In our codes, we have two functions that find the intercept and slope of the fitted line based on given data. Also, they use our helper functions to find the implementation year and separate the data to years before and after the implementation year. We can use these functions for the given y variable in our data, for example, we can choose y to be CO2 per GDP of the given country or the logarithm of the CO2 per GDP. We use logarithm since we expect the linear association between the years and the logarithm of CO2 per GDP to be stronger.

First, we give our function some necessary variables, such as our data frame, y, and x. After that we define the model using `LinearRegression(fit_intercept=True)`. After using this model, we can find the slope and intercept. Additionally, we use `model.fit(x[:, np.newaxis], y)`, this code will find the fitted line for the given data and, after that, we will take the intercept and slope from these codes. We use `model.coef_[0]` to find the coefficient which is our slope and `model.intercept_` to find the intercept.

We represent our final result using a scatter plot and two regression lines, one for before the tax implementation year and the other for after the implementation year. We create this plot using the `create_plot` function, which takes

four parameters: `country`, `y_value`, `regression_slope` and `regression_constant`

`country` should be one of the country names in the `Data_Overall` and `co2.emissions_vs_gdp` CSV files. `y_values` should be one of the columns in the country dataframe. This function will plot the observations with the year as the value for x and one of the column values as the value for y.

In this function's body, we convert the `co2.emissions_vs_gdp` dataset, which is a CSV file, into a pandas dataframe. Using this and the country that we took as an input, we called our `before_implementation` and `after_implementation` functions from the `cleaning_data` module to obtain new pandas data frames. We also call `look_up_implementation_year` from the `cleaning_data` module in order to obtain the year that `country` start the implementation, and we later use the year that we obtained and `matplotlib.pyplot.axvline` to add a vertical line showing this implementation year.

Then, we called the helper function `plotting_values` two times. First, we called it to obtain a list of x values and y values, y-intercept, and slope of the regression line before the implementation year. Second, we called it to obtain a list of x values and y values, y-intercept, and slope of the regression line after the implementation year. We then create an array for x values and y values using `numpy.array`, and plot the points and the regression lines using `matplotlib.pyplot.plot`. We also use `matplotlib.pyplot.plot` and `regression_slope` and `regression_constant` to create an extension of linear regression line for before the implementation.

The helper function `plotting_values` takes a pandas data frame as `country_df` and two strings as `x_column` and `y_column`, which should be the names of two columns in the `country_df`. In this function, we convert the `x_column` and `y_column` in the given pandas data frame into lists and create arrays for them using `numpy.array`. Then, we obtain the y-intercept and slope of the line of best fit using `numpy.polyfit`. This function returns a tuple containing a list of x values, a list of y values, the slope and y-intercept of the line of best fit.

Mathplotlib is a python library that can be used to generate plots and visualize data. We used mathplotlib in our program to generate a scatter plot for the given country with year as the x-axis and customized y-axis values. In the plot we created a line of best fit that represented the linear regression models before and after the implementation of the carbon tax. This helped us understand the relationship between carbon tax policy and it's effectiveness.

Numpy is a python library that is used to work with arrays. In the `plotting_data.py` file, we used numpy to convert the x and y point values into arrays that can be given to mathplotlib to generate the scatter plot required for data visualization.

Instructions For Obtaining the Datasets

Dataset 1 and 2 can be downloaded from the sources provided in the Dataset Description. Rename dataset 1 to `co2.emissions_vs_gdp`. Remove the first two rows in dataset 2, and then convert the XLS file to CSV. Name the new csv file `Data_Overall`. Alternatively, you can download the two CSV files and the XLS file from markus. Save the two CSV files to the same folder as the folder of the python files.

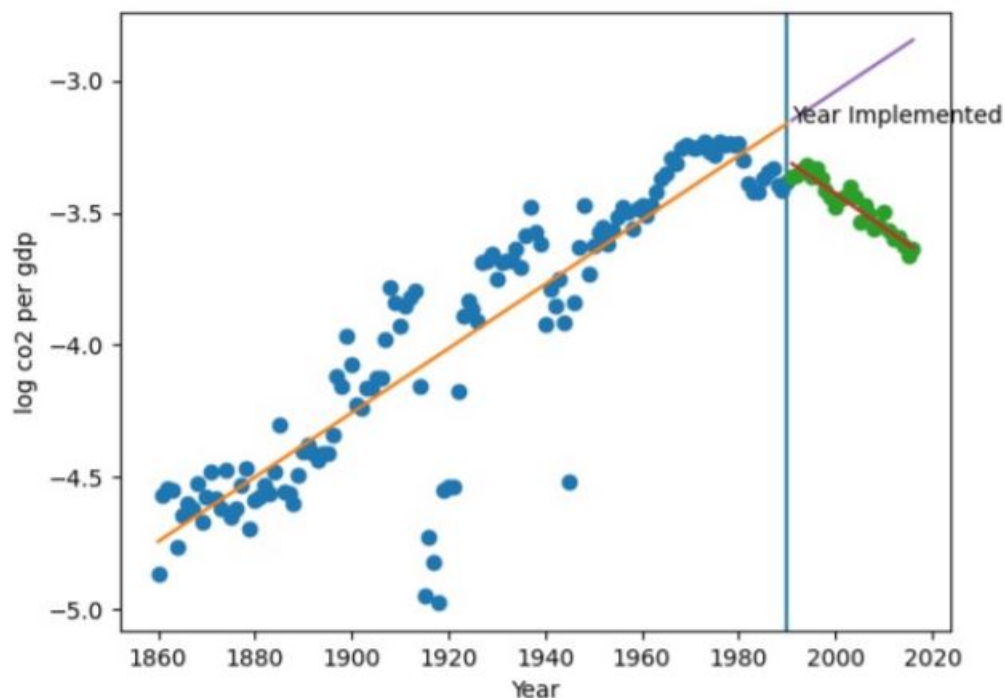
Instructions For Running the Project File:

Once everything has been downloaded properly, open the main file and call the function `country_linear_regression` with the name of the country (first letter capatalized) and the y-axis (column of data) as string parameters. This will create and open a linear regression plot and also automatically open the PDF report in the web browser. When the function is called in the console, a pdf file should be generated that looks like this:

The Effectiveness of Finland's Carbon Tax Policy on Reducing CO2 Emissions.

In this instance of our program, we are testing the effectiveness of Finland's policy of Carbon Tax. Our program generates a scatter plot of the `log_co2_per_gdp` of the country against the years since data was being collected.

We generate the slope and coefficient of the linear regression lines using this data. Before the carbon tax was implemented, the linear regression line for the country of Finland has the slope 5 and the y-intercept as 7. After the carbon tax implementation, the linear regression line was found to have the slope 0 and the y-intercept to be 1. The plotted values of the country's `log_co2_per_gdp` against years is shown in a plot.



From our plotted values and the slopes of linear regression model generated before and after the tax implementation, we can see that there is sizeable difference between the slopes of linear regression model before and after the implementation: the difference is greater than half the size of the slope before implementation. Thus, we may be able to conclude that the implementation of Carbon Tax by Finland has been successful in reducing co2 emissions

If you want to run our main file, you should give some necessary variables to our functions, such as the name of the country you want to find the tax effect.

Also, the y variable is one of the columns' names in our data, but because we want to see the effect on CO2 emission, we recommend giving CO2 per GDP (`co2_per_gdp`). If you want to have a more linear association, give your y variable the logarithm of CO2 per GDP (`log-co2_per_gdp`). In the output which, is the PDF, you can see a plot and also some information.

In the plot, you will see two groups of points. The group of blue points is the data of your dependent variable during the years before the implementation year. The group of green points is the data for the y variable after the tax implementation.

Additionally, there are three lines in our scatter plot. The yellow line is a fitted line for the year before the CO2 tax implementation, and the red line is the suitable line for our points after the implementation. We show the implementation year by the vertical line in our plot. Also, there is a purple line that helps us understand the difference between the levels of CO2 between when the tax is implemented and when we suppose it is not. There is another purple line that serves as an extension of the line of best fit for plot points before the tax was implemented, this line is generated using the slope and intercept values from the linear regression algorithm and serves as a visual depiction of the difference between the before and after of tax implementation on y-axis.

There are two descriptions before and after our plot. The first part at the top of our page will give you general information about the plot and the fitted line's result. For example, it mentions the country and the y variable you used and the intercept and slope of the fitted line for each group.

The second and last part informs you if there is a considerable difference before and after implementing the tax policy. We define a substantial difference based on the cutoff we previously explained.

Changes to the Project Plan

Initially we had planned on investigating the effectiveness of the ETS policy. But we decided to focus on the effects of carbon tax instead, as that is more relevant to the data we have. Moreover, we had planned to look at the emissions of several developed countries, but now we have created functions in our dataset that allow us to see the results for any country in our dataset.

We also realized that it would be better to also consider the GDP of a country as each country wants to reduce its emission levels relative to its productions. As a result, we now use the CO2 per GDP of countries over the years. Additionally, we use the logarithm of that value so that we have a stronger linear association. Because of this change, we now use a new dataset which includes data on per capita CO2 and GDP per capita.

In our original proposal, we planned to use the plotly library to create our plots, but now we use the matplotlib library. Moreover, we have made changes to how we show our results on the plots. We create a vertical line that represents the implementation year of a carbon tax. We also create two fitted lines for the years before and after the implementation. We comment on how these two lines compare to each other and if there is a sizeable difference between them. We define a cutoff such that if the slope of the line for the years before the tax is less than half the slope of the line for the years after it, we claim the tax is effective.

Discussion Section

- Results of computational exploration:

Our computational exploration helps us understand if there are any considerable changes in the slope of the fitted line after the carbon tax gets implemented in a country. So our investigation enables us to learn about the effect of a carbon tax on CO2 emissions and answer our research question.

We regard a change as considerable when the tax makes the fitted line's slope at least twice the amount of what it was before. For example, let's look at the CO2 per GDP and logarithm of CO2 per GDP plots for Finland. We see that the year and the dependant variable are positively associated before the implementation year and negatively associated afterward, so clearly this satisfies our cutoff of the before line having a slope less than half the slope of the after line, indicating a considerable difference.

- Limitations: Firstly, as stated in the changes sections, we changed ETS to the carbon tax in our research question. If we wanted to work with ETS, one of our project's main limitations would be that we could not access each European country since we only had general data about EU countries in our dataset.

Also, our method only considers a significant change in the slope to indicate effectiveness. But if we wanted to make our results more general and consider all possible effects, we would need to explore any change in the

slope. For example, a tax may not pass the cutoff that the after line has a slope of at least twice the before line, and have a smaller difference, but still be an effective policy. However, in our method, we only consider significant effects.

Another limitation in our project is that we do not have data about other factors that may play a part in reducing CO2 emission levels besides the tax. So it is possible that a change we observe is not a direct result of the implemented tax. For example, if many firms in a country close in a specific year because of economic problems, CO2 emission levels will decrease. However, that outcome is not because of the implemented tax policy.

- Further exploration: We create a plot for Finland and the independent variable logarithm of CO2 per GDP and check the change in the fitted line's slope after the tax's implementation. In this plot, we can see a sharp decrease in the y variable for some observations at the end of 1919. At the beginning of 1920, the observations will again have a y value, which is close to the fitted line.

In further exploration, we would be interested in investigating the reason behind this sudden and significant decrease in the y variable for Finland. Similarly, we would be interested in exploring any observations for a given country with a substantial distance from the line.

Moreover, we can study the cause of such reductions separate from the factor of a carbon tax, which may be more effective than the carbon tax. For example we can look at the year of 1919 to see if other countries have had similar reductions for the y variable. After identifying the cause of those observations, we can explore if countries may be able to reduce their emission levels with similar methods.

- Conclusion: In conclusion, in countries in which the fitted line's slope considerably decreases after implementing the tax, assuming that no other factors are affecting our results, we claim that the tax has been effective in reducing CO2 emission levels.

References

- Economics of Climate Change. (2020, July 07). Retrieved November 05, 2020, from <https://bit.ly/3l5vMKf>
- Pricing Carbon. (n.d.). Retrieved November 05, 2020, from <https://www.worldbank.org/en/programs/pricing-carbon>
- The Effects of Climate Change. (2020, August 21). Retrieved November 05, 2020, from <https://climate.nasa.gov/effects/>
- The Causes of Climate Change. (2020, August 18). Retrieved November 05, 2020, from <https://climate.nasa.gov/causes/>
- Pandas Team. (n.d.). *Pandas*.<https://pandas.pydata.org>
- Scikitlearn Team. (n.d.). *Scikit learn*.https://scikit-learn.org/0.20/modules/generated/sklearn.linear_model