Assignment_2

2022-03-28

Project description

I intend to examine to what extent does economic activity cause CO2 emission. I will measure economic activity through GDP, and I will use per capita values for GDP and emissions as well (a confounder)

Downloading and describing data

I have downloaded the following time series data for all countries from 1992-2018 from the World Bank WDI section for all countries:

- GDP per capita data (constant 2015 USD)
- CO2 emission in metric tonnes per capita

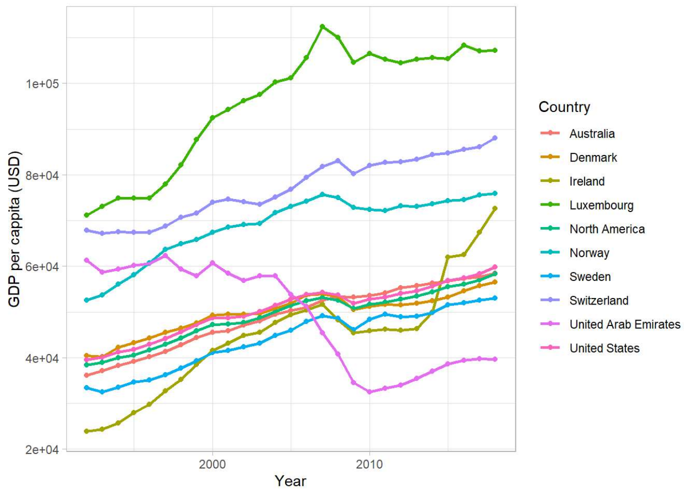
I did the following data cleaning and feature engineering steps:

- I set the NA values to be shown as “NA” instead of “..”.
- I simplified column names
- I excluded grouped observations (e.g. EU, MENA)
- I dropped rows with missing data, and to get a balanced sample, I dropped rows with countries that did not have data for all 27 years
- I set data type to years format
- I created the log of GDP and CO2, and the lag variables (2 and 6)
- Taking the log of GDP and CO2, I also fixed the originally skewed distribution

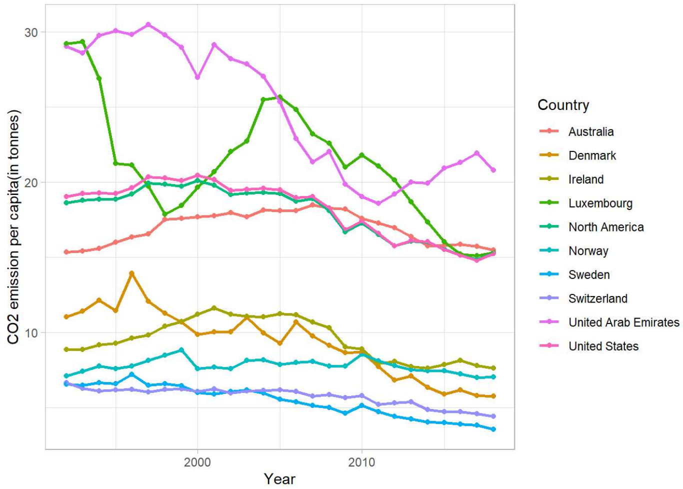
In the end I ended up with 44010 observations for 163 countries, each country having 27 yearly data on GDP and CO2 per capita, from 1992 to 2018. This is a balanced panel data.

Exploratory analysis

We can see that there is a clear trend for both GDP per capita and CO2 emissions per capita increasing over time (for simplicity, I only showed the top10 countries)

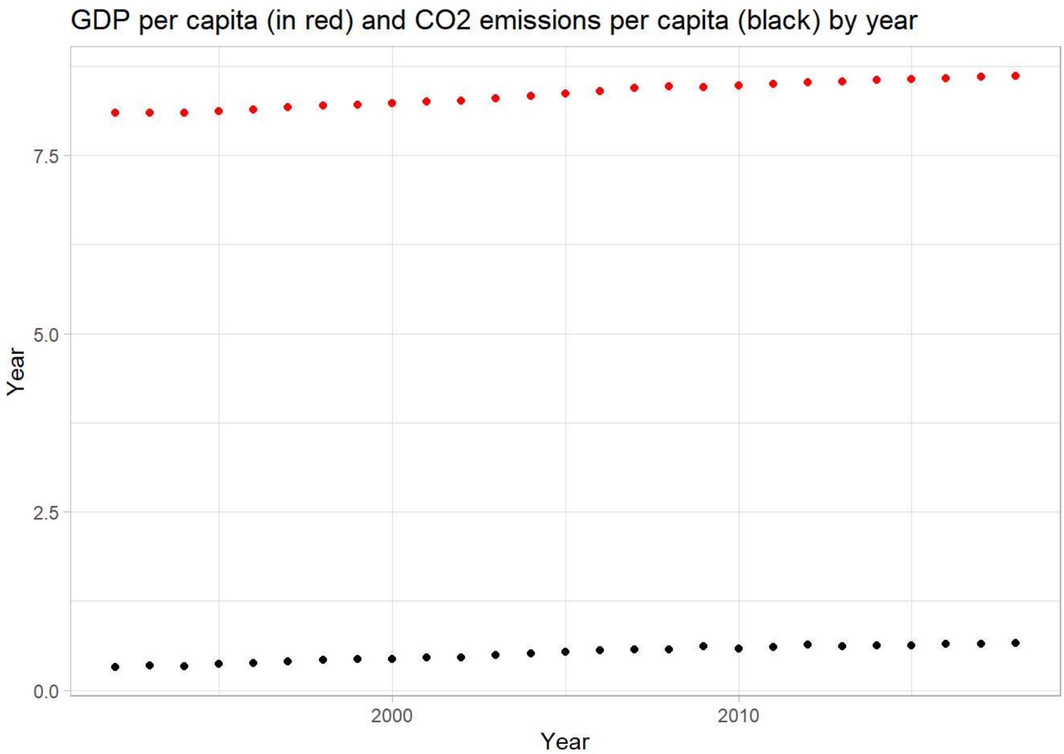


Development of GDP per capita in top10 countries



Development of GDP per capita in top10 countries

Plotting the ln of GDP and CO2 per capita against Year, we can see that they are linear, and are almost paralell.



Regressions

In the following chapter I will interpret the results of 7 models I built.

Cross-sectional OLS regressions for chosen years

First of all I created a basic cross-section OLS models for 1994 and for 2016, with the following formula: ln_CO2PC ~ ln_GDPPC

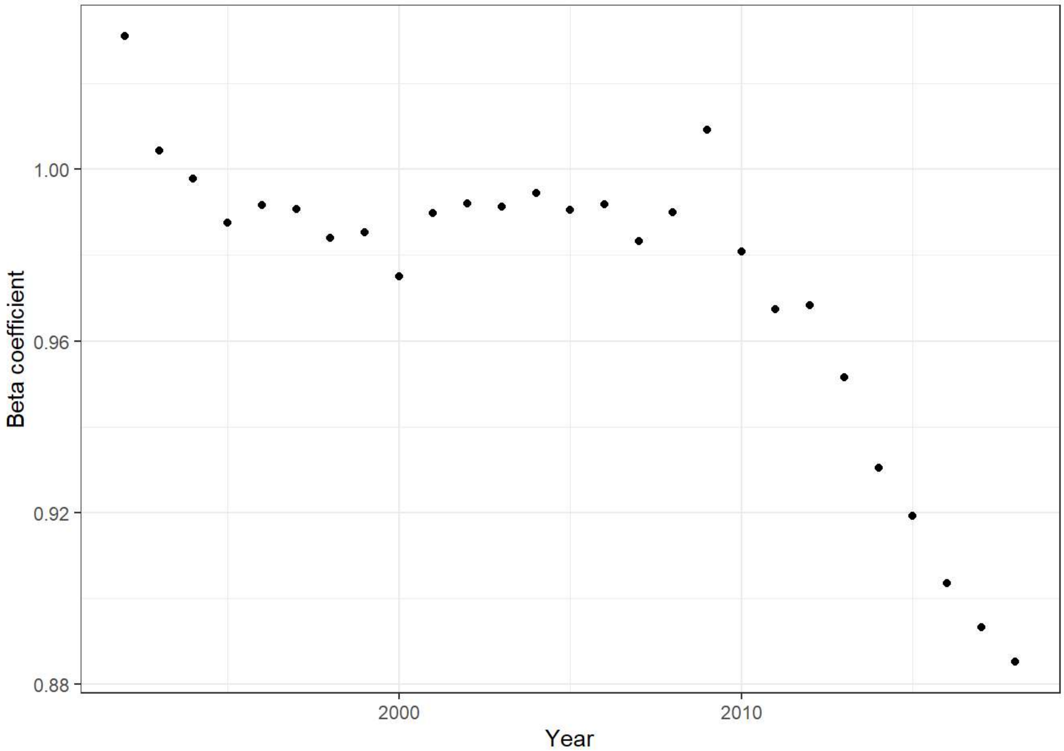
The results shows that, on average, if the CO2 per capita is higher by 1% the GDP per capita is higher by 0.99% in 1994, and 0.90% in 2016.

	Cross-sec 1994	Cross_sec 2016
ln GDP per capita	0.998 ***	0.904 ***
	(0.045)	(0.045)
N	163	163
R2	0.719	0.773

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

I see a declining trend between 1994 and 2016 in the coefficient, so I did further calculations.

By calculating the coefficients for all years and plotting it on a scatter plot, we can see that the beta coefficient is smaller and smaller after 2010. Using domain knowledge makes me form a hypothesis in the 20th century and early 2000s, high industrial production caused both high GDP per capita and high CO2 emissions, however, the trend is that developed countries increase their GDP per capita and decrease their emissions. This can be explained with one the one had a conscious effort since the Paris agreement to lower emissions, and due to the service sector contributing mor and more to GDP growth, while production is outsourced to developping countries.



First difference regression

After the OLS models I built 3 FD models:

- d1ln_CO2PC ~ d1ln_GDPPC + Year
- d1ln_CO2PC ~ d1ln_GDPPC + 2 lags + Year
- d1ln_CO2PC ~ d1ln_GDPPC + 6 lags + Year

Lags take care of delayed effects, while yearly dummies take care of aggregate trends that are non linear.

The results are the following:

- For the 1st model, on average, CO2 emission per capita tends to increase by 0.51% more when GDP per capita increases by 1% more
- For the 2nd model, on average, CO2 emission per capita tends to increase by 0.43% more when GDP per capita increases by 1% more. The cumulative coeff. is 0.121, meaning two years GPD change is 0.121% additional change in CO2 emission per capita
- For the 3rd model, on average, CO2 emission per capita tends to increase by 0.44% more when GDP per capita increases by 1% more. The cumulative coeff. is 0.28, meaning two years GPD change is 0.28%% additional change in CO2 emission per capita
- The coefficient is significant at 99% in all 3 models, and including lags does not change the immediate effect significantly between 2nd and 3rd model, so 2 lags are enough

	FD, no lags	FD, 2 lags	FD, 6 lags
d(ln GDP per capita)	0.512 ***	0.428 ***	0.443 ***
	(0.076)	(0.080)	(0.086)
d1ln_GDPPC lag1		0.135 *	0.227 **
		(0.064)	(0.076)

d1ln_GDPPC lag2		-0.014	-0.050
		(0.041)	(0.053)
d1ln_GDPPC lag3			-0.015
			(0.055)
d1ln_GDPPC lag4			0.076
			(0.072)
d1ln_GDPPC lag5			-0.017
			(0.055)
d1ln_GDPPC lag6			0.062
			(0.042)
Constant	0.683	0.862	0.130
	(0.521)	(0.484)	(0.532)
N	4238	3912	3260
R2	0.038	0.036	0.042

*** p < 0.001; ** p < 0.01; * p < 0.05.

Fixed effects model

For FE model, I used the following formula: ln_CO2PC ~ ln_GDPPC + Year

The coefficient shows that when GDP per capita is 1% higher compared to its average value in a given country, the CO2 emission per capita are, on average, 0.55% higher than its average value within that given country.

	FE
ln GDP per capita	0.551 ***
	(0.085)
N	4401
within_R2	0.979

*** p < 0.001; ** p < 0.01; * p < 0.05.

Long difference model

I made a a long difference model, using log of GPD per capita and log of CO2 emission per capita between 1992 and 2018, and ignoring the rest.

The coefficient in this case means that on a long term we expect CO2 emission per capita to increase by 0.75% more when GDP per capita increases by 1% more.

	LD
LD ln GDP per capita	0.749 ***
	(0.103)
Intercept	-0.053
	(0.071)

N	163
R2	0.249

*** p < 0.001; ** p < 0.01; * p < 0.05.

Summary

I built OLS model, 3 FE models and 1 FD and 1 long difference model, using GDP and CO2 emission per capita data from World Bank. A first estimate is that in case of an additional 1% change in GPD per capita we can expect CO2 emission to increase with 0.5% more. To claim causality this is not enough, we would need to think of this problem in more details.

	FD, no lags	FD, 2 lags	FD, 6 lags	FE	LD 1992-'18
d(ln GDP per capita)	0.512 ***	0.428 ***	0.443 ***		0.749 ***
	(0.076)	(0.080)	(0.086)		(0.103)
ln GDP per capita				0.551 ***	
				(0.085)	
LD ln GDP per capita					
N	4238	3912	3260	4401	163
R2	0.038	0.036	0.042	0.980	0.249

*** p < 0.001; ** p < 0.01; * p < 0.05.

Potential mechanisms

Potential mechanisms include ways for causality to take place: effects through which one variable affects the other. In this case GDP per capita can effect CO2 emission: One example, as discussed above, when more goods are produced, GDP is higher, and emissions are also higher. Another example is that win countries where GDP is higher, people have more money to consume things that directly generate CO2 emissions (not through production but through usage). E.g. using the car, or utrning on the lights, or using AC (given that power is generated from fossils). However, more services on the other hand are can lead to more travelling, more electricity needed, etc. which all can lead to more CO2 emission.

Think of a potential confounder

A potential confounder would be weahter the country joined the Paris agreement to decrease emissions or not. This confounder can further be refined by dividing countries into front runners of green energy such as Germany, or followers such as Hungary.

Another potential confounder is the ratio of industrial output within the GDP, and differentiate mature economies with a large services sector (e.g. Luxembourg with no industry, of California as an IT hub) from production based industries such as China.