DA4 - Assignment 2

Oszkar Egervari

2022-04-08

Introduction

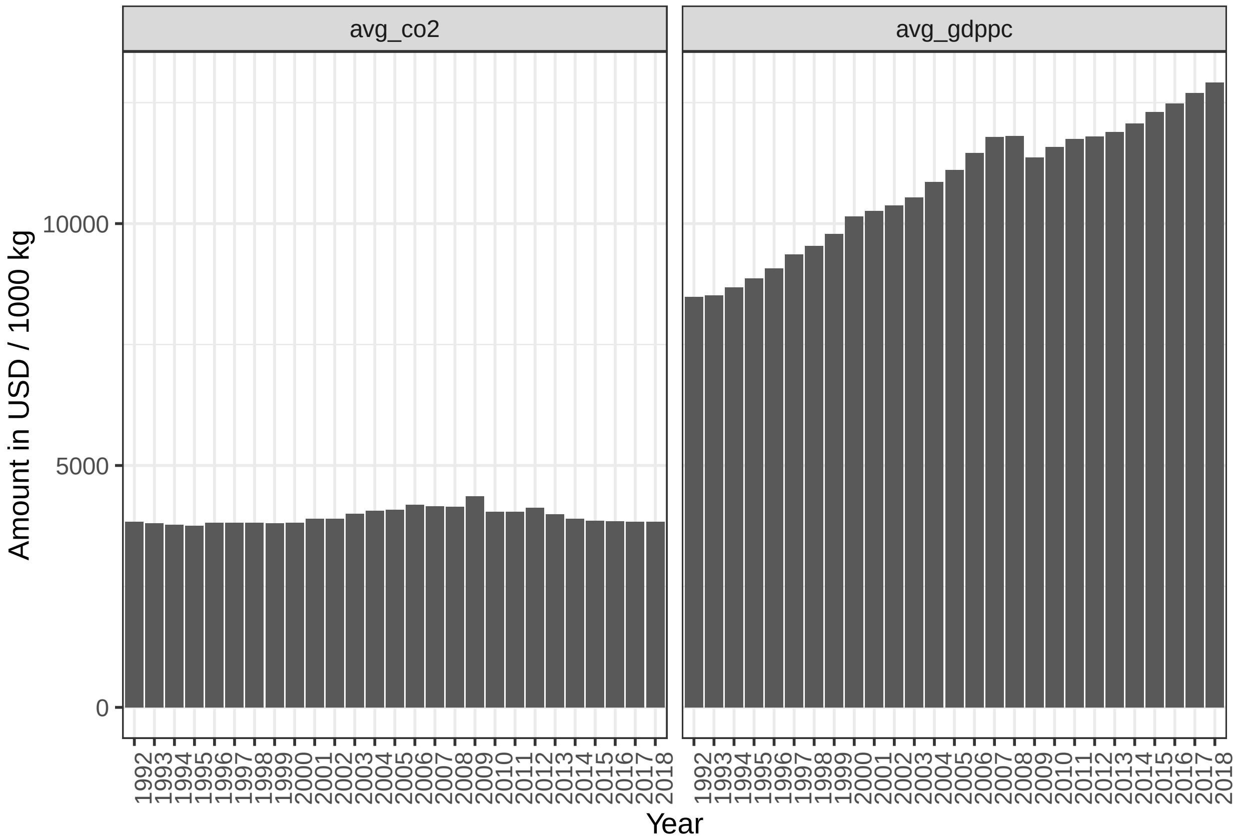
This assignment I will try to find out, to what extent does economic activity cause CO2 emission. I am going to use GDP per capita to measure the economic activity of a country. I'm going to explore this relationship with the help of various models ranging from cross-sectional OLS through fixed effects models to multiple first difference models.

The Data

The data is acquired from World Bank through the WDI R library, both the GDP per capita and the CO2 per capita emissions. The gdppc is in USD, while the co2 is in metric ton, which I converted to kgs. The time period is between 1992 and 2018 and for all countries having a full set of data for both indicators for all years. I decided to keep only countries without missing values, because I couldn't think of a good solution make up for the missing data. If I use 0 instead of NAs, the means will be lowered quite significantly, however if I take the average values, it increase the mean, because mostly developing countries with lower than average values of gdppc and co2 emissions are missing these values. After dropping them the data still has 161 countries, so I figured this is the best option.

Data description

When we take a look at the data before diving into the models, an interesting discovery we may make, is that the average CO2 emission hasn't really changed much in the last 27 years. On the other hand, the average GDP per capita has raised from 8488.68 USD to 12912.8 USD. I didn't include the same plot for the sum, but it looks very similar to plot of the averages seen above.



Models

Model 1

The first model is a cross-section OLS for year 2018.

```
## OLS estimation, Dep. Var.: EN.ATM.CO2E.PC
## Observations: 161
## Standard-errors: IID
##                   Estimate Std. Error t value   Pr(>|t|)
## (Intercept)    2048.502580 316.882205 6.46456 1.1876e-09 ***
## NY.GDP.PCAP.KD    0.138682   0.014097 9.83786  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 3,270.7   Adj. R2: 0.374471
```

The intercept is 2048.5. That means that if a country has a gdp per capita of 0, its expected CO2 emission per year is 2048.5 kg. The beta is 0.14, which means that if the GDP per capita increases by 1 USD, the country's CO2 emission is expected to increase with 0.14 kg.

Model 2

The second model is a first difference model with time trend and without lags. The intercept means that if gdppc doesn't change, co2 emission is expected to change by 0.05 kg. Meanwhile beta shows us the average change in co2 emission when gdppc changes during the same period. So in case gdppc increases by 1 USD, based on the model we expect co2 emission to increase by 0.16 kg.

```
## OLS estimation, Dep. Var.: d_co2
## Observations: 4,346
## Standard-errors: Clustered (country)
##             Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 0.052633  14.696225 0.003581 9.9715e-01
## d_gdp       0.161285   0.027914 5.777947 3.8410e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1,194.7   Adj. R2: 0.246536
```

Model 3

```
## OLS estimation, Dep. Var.: d_co2
## Observations: 4,024
## Standard-errors: Clustered (country)
##              Estimate Std. Error  t value   Pr(>|t|)
## (Intercept) 14.574157  16.414089 0.887905 3.7592e-01
## d_gdp        0.161740   0.027894 5.798486 3.4719e-08 ***
## l(d_gdp, 1)  0.060375   0.045741 1.319939 1.8874e-01
## l(d_gdp, 2)  0.046305   0.032111 1.442053 1.5124e-01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1,234.4   Adj. R2: 0.249199
```

The third model is also a first difference model with time trend, but this time with 2 years of lag. These values show the lagged associations, the effect of the change of gdppc on co2 emission in the following years. So 0.05 is the coefficient of the year 2 lag, meaning that in case of 1 USD increase in gdppc, co2 emission is expected to raise by 0.05 kg on average 2 years later.

Model 4

```
## OLS estimation, Dep. Var.: d_co2
## Observations: 3,380
## Standard-errors: Clustered (country)
##              Estimate Std. Error   t value   Pr(>|t|)
## (Intercept) 10.383955  19.615560  0.529373 5.9728e-01
## d_gdp        0.163635   0.027906  5.863842 2.5136e-08 ***
## l(d_gdp, 1)  0.056820   0.044451  1.278258 2.0301e-01
## l(d_gdp, 2)  0.066093   0.034421  1.920120 5.6622e-02 .
## l(d_gdp, 3) -0.037109   0.064967 -0.571188 5.6867e-01
## l(d_gdp, 4)  0.102395   0.075062  1.364142 1.7444e-01
## l(d_gdp, 5)  0.008689   0.038856  0.223627 8.2333e-01
## l(d_gdp, 6) -0.085371   0.027952 -3.054239 2.6436e-03 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1,332.1   Adj. R2: 0.254649
```

The fourth model is the last of the first difference models, with 6 years of lag this time. This is pretty much the same as the previous model, but with more lags. We can see, that year 6 lag has a negative coefficient interestingly, resulting in a decrease in expected co2 emission by 0.09 kg in case of 1 USD increase in gdppc within 6 years.

Model 5

```
## OLS estimation, Dep. Var.: EN.ATM.CO2E.PC
## Observations: 4,347
## Fixed-effects: country: 161
## Standard-errors: Clustered (country)
##                  Estimate Std. Error   t value Pr(>|t|)
## NY.GDP.PCAP.KD   0.020415   0.060607  0.336838  0.73668
## year1993       -39.185549  36.842821 -1.063587  0.28912
## year1994       -70.508843  55.762611 -1.264447  0.20791
## year1995       -92.399257  74.937361 -1.233020  0.21938
## year1996       -42.197606  89.808442 -0.469862  0.63909
## year1997       -46.185894 102.487561 -0.450649  0.65285
## year1998       -51.392412 108.388590 -0.474150  0.63604
## year1999       -63.997160 120.249616 -0.532203  0.59532
## ... 19 coefficients remaining (display them with summary() or use argument n)
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 1,020.8     Adj. R2: 0.951486
##                 Within R2: 0.024498
```

The fifth model is a fixed effects model with time and country fixed effects The coefficient estimate means that when we compare years with the same countries, in years when the gpdpc is higher by 1 USD, co2 emission tends to be 0.02 kg higher than its average within the country.

Model 6

```
## OLS estimation, Dep. Var.: d_co2
## Observations: 161
## Standard-errors: Clustered (country)
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 332.38057 269.081685  1.23524  0.21855
## d_gdp        -0.07586   0.054288 -1.39735  0.16424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 2,231.7   Adj. R2: 0.05522
```

The last of the models is a long difference model, which takes the difference of the variables from two years, which are far apart. In this case I chose the first and last years, 1992 and 2018 respectively. The intercept means that if the gdppc doesn't change, the expected average change in co2 emission is 332.38 kgs. The beta coefficient shows us, that if the gdppc increases by 1 USD (from 1992 to 2018), the expected average change in co2 emission is -0.08, so it decreases.

Summary

The main finding of the assignment is that there is a positive correlation between gdppc and co2 emission. However it's not that simple, because there is a clear trend - most likely in first world countries - to decrease or at least not increase their co2 emissions in the last couple years, meanwhile their gdppc is increasing still. This is however only my assumption, further analysis would be required to say more on this topic, perhaps aggregating the indicators to at least developing and developed countries, or even something more sophisticated.

I believe that interestingly the same mechanism is behind both outcomes, increasing co2 emission for developing countries and decreasing co2 emission for developed countries, which is increasing gdp or gdppc. Developing countries doesn't yet have the resources to abandon fossil fuel sources, meanwhile developed countries - since they are already developed - can focus on research and can afford to lose (at least initially) resources in order to switch to more sustainable energy sources.