DA4 Assignment 2

Jana Hochel

2023-10-04

CO2 and GDP

To what extent does economic activity cause CO 2 emission?

Carbon dioxide (CO2) emissions are a critical factor in understanding the human impact on climate change. CO2 is the most abundant greenhouse gas trapping heat in the atmosphere and contributing to global warming. CO2 is primarily produced by human activity, such as the burning of fossil fuels for energy, transportation, and industrial processes. Global CO2 emissions have been increasing rapidly since 1970, exacerbating the impacts of climate change, such as rising sea levels and more frequent and severe heat waves. To mitigate the worst effects of climate change, it is essential to reduce global CO2 emissions through strategies such as transitioning to renewable energy sources and improving energy efficiency.

To make this happen, the world must first understand what are the determinants of CO2 emissions. Earlier studies suggests that developed coutries emit majority of the CO2 emissions (Dietz and Rosa, 1997). Later studies suggest it is more complex and focus on the reverse effect, such as the impact of CO2 emissions and policies on GDP (Llanos et al., 2022). This study reviews the causal link between GDP per capita from 266 countries and the CO2 emissions per capita.

Dietz, T., & Rosa, E. A. (1997). Effects of population and affluence on CO2 emissions. Proceedings of the National Academy of Sciences, 94(1), 175-179.

Llanos, C., Kristjanpoller, W., Michell, K., & Minutolo, M. C. (2022). Causal treatment effects in time series: CO2 emissions and energy consumption effect on GDP. Energy, 249, 123625.

Download data and describe it (1p)

The data used in the model are from World Bank repository for 266 countries from 1992 till 2021. Apart from GDP and CO2 per capita, I have added an urbanisation rate (% of population), access to electricity (% of population), and available fresh water per capita as providing potable water may be energy-consuming.

Data Quality

Here we display only countries with missing data. Some countries have missing values. Usually these are small, developing nations, city states, and islands. Thus, I have decided to drop everything with more than 10 missing observations (years). This way, we retain Venezuela and Afghanistan (n_missing = 10 years). This is mostly true for very early data (1992-2000) but also the most recent (2021).

There is a moderate level of multicollinearity between the GDP and Electricity variables. However, the Water and Urbanisation variables seem to have low levels of multicollinearity with the other variables.

Table 1: Final number of countries

count
225

Table 2: Key Variables

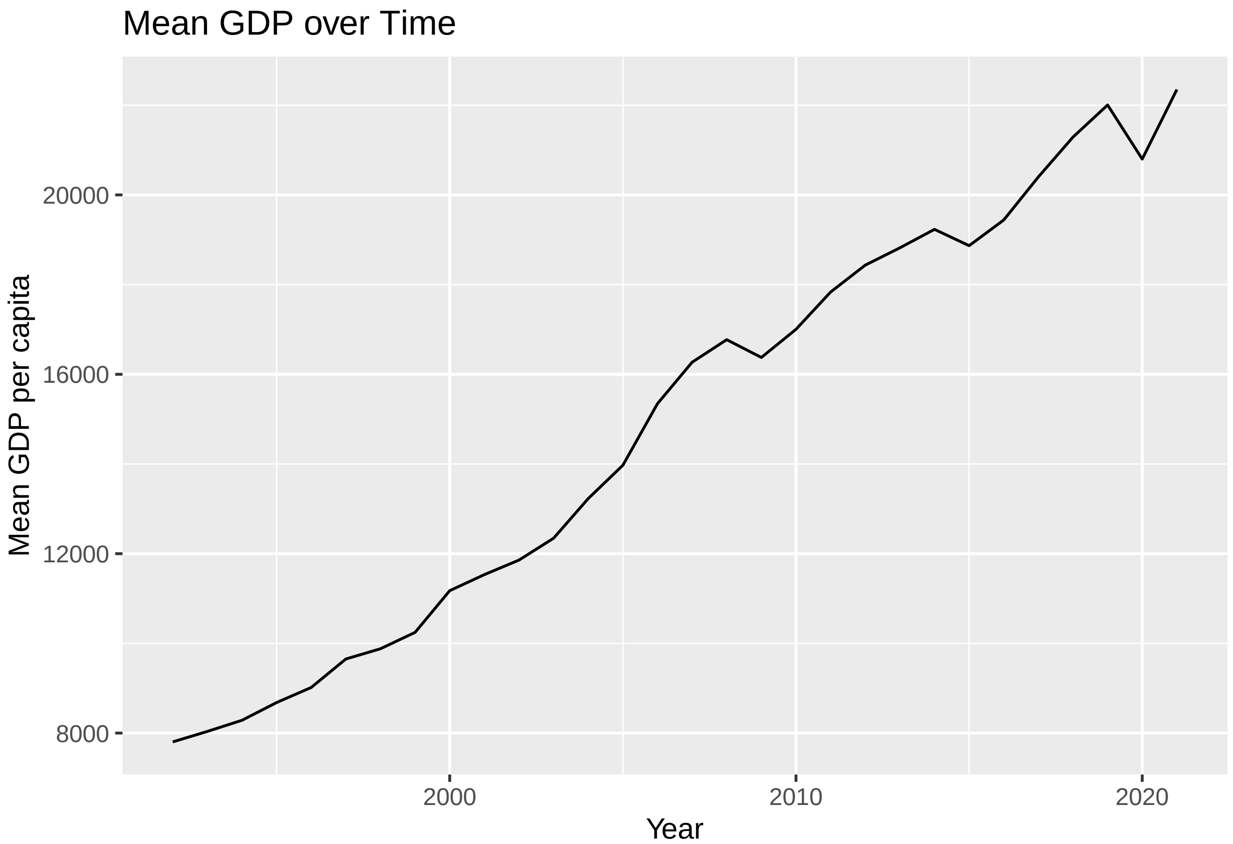
	GDP	CO2
	Min. : 285.1	Min. : 0.0000
	1st Qu.: 3050.4	1st Qu.: 0.6832
	Median : 8176.7	Median : 2.4849
	Mean : 15047.7	Mean : 4.2033
	3rd Qu.: 20129.4	3rd Qu.: 6.1631
	Max. :163219.5	Max. :47.6513
	NA's :912	NA's :1300

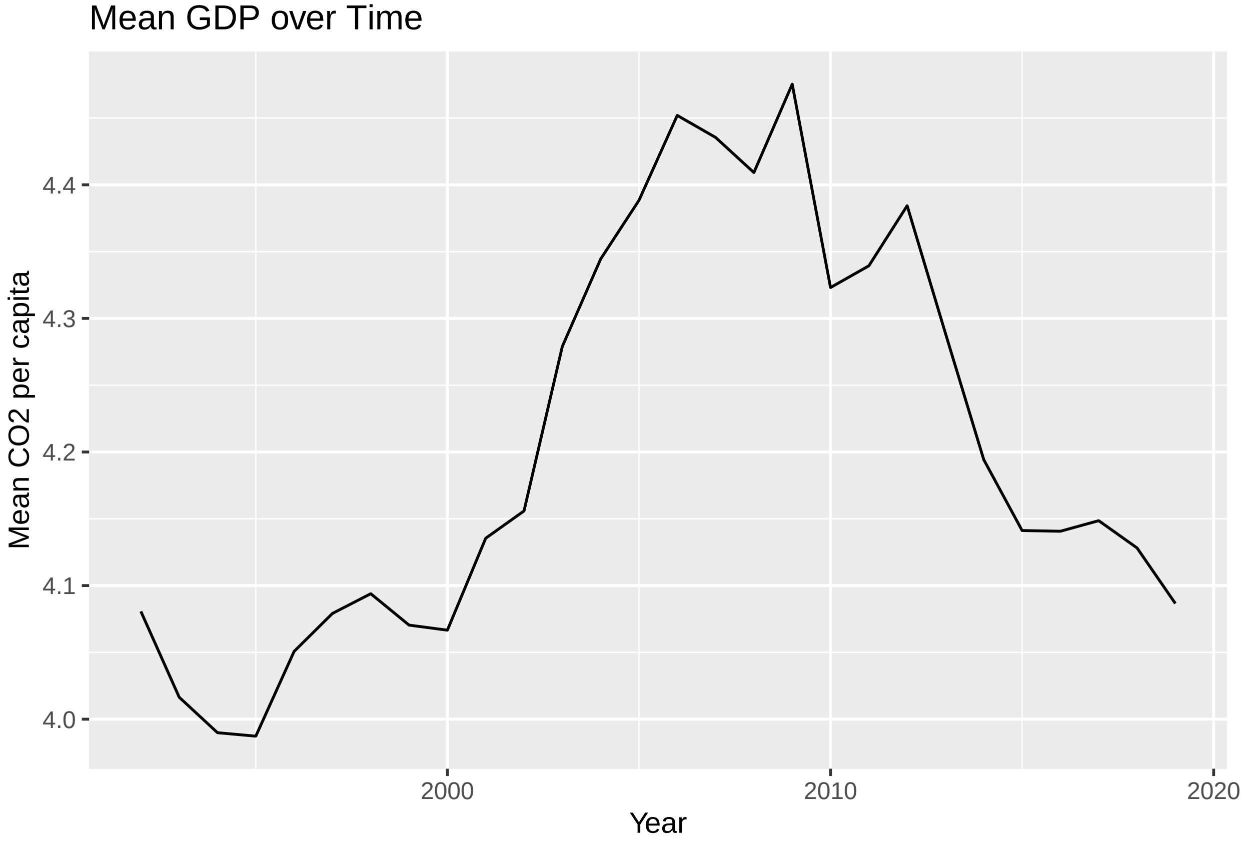
Time Trends and missing values

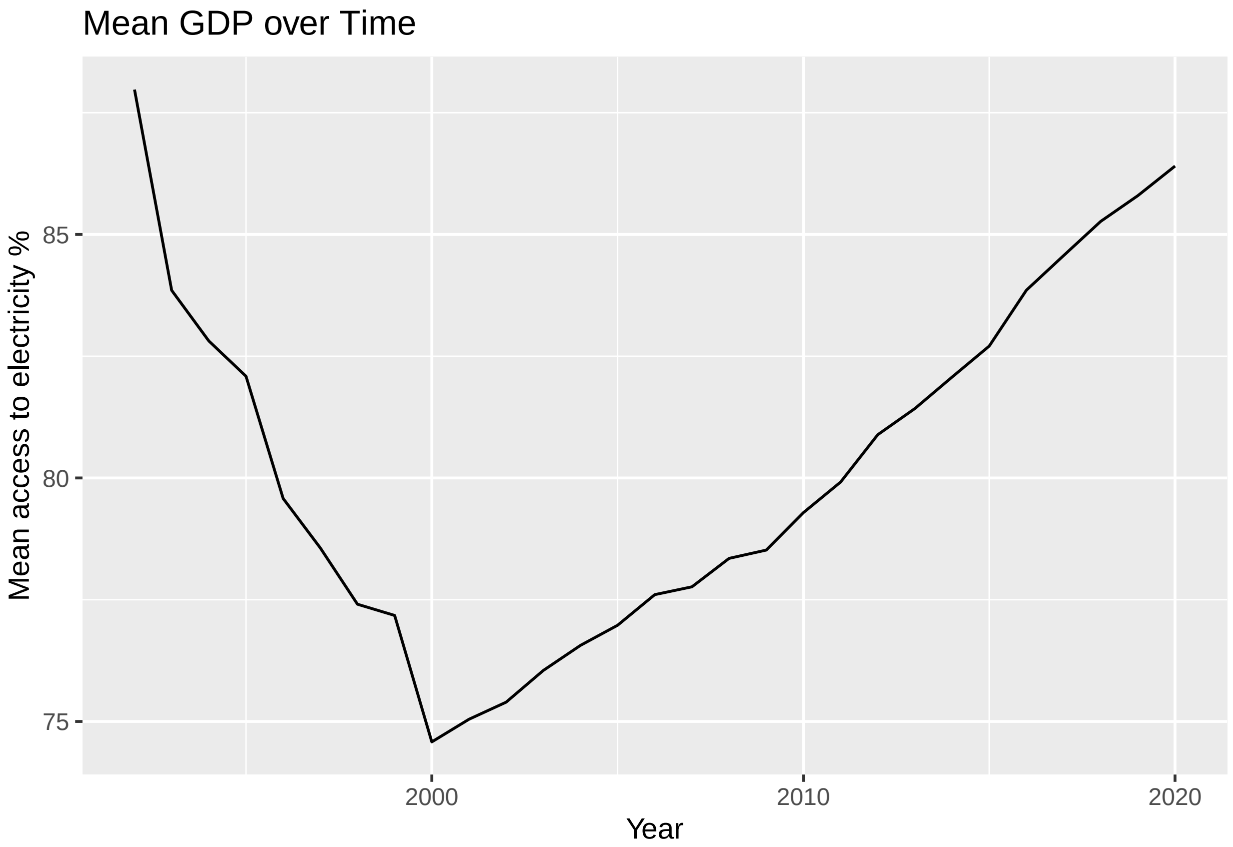
Till 2000 there were mostly data from developed countries. This would explain why the access to electricity dramatically decreased as well ass the gdp between 1992 and 2000.

Table 3: Multicollinearity

Variables	Tolerance	VIF
GDP	0.578	1.73
Water	0.992	1.01
Electricity	0.505	1.98
Urbanisation	0.395	2.53


Mean GDP over Time
Mean GDP per capita
20000
16000
12000
8000
2000
2010
2020
Year


Mean GDP over Time
Mean CO2 per capita
4.4
4.3
4.2
4.1
4.0
2000
2010
2020
Year



The models

1. Cross section OLS for the year 2005

The explanatory power - the ability to capture variance is strong R^2 > 0.90. F statistic is low which means the results are statistically significant.

Reg Equation

$logCO2 = -8.5878 + 0.8952 * logGDP + 0.0169 * Electricity + 0.0001 * Urbanisation - 0.0000005 * Water$

Interpreting Coefficients

In log-log models, the coefficients represent elasticities - percentage change of dependent variable and subsequent 1% change in the independent variable.

Only, logGDP and Access to Electricity (%) seem to display a statistically significant pattern. This coefficient is positive. It means that a 1% increase in GDP is associated with a 0.8952% increase in CO2 emissions when holding other variables constant. The increase for Electricity is lower.

Urbanisation is not only insignificant but also very small, so technically, it could be dropped.

```
Call:
```

```
lm(formula = logCO2 ~ logGDP + Electricity + Urbanisation + Water,
    data = subset(data2, Time == 2005))

Residuals:
    Min      1Q  Median      3Q     Max
-1.4401 -0.3237 -0.0358  0.2954  1.4993

Coefficients:
                 Estimate   Std. Error t value             Pr(>|t|)
(Intercept)  -8.587769491  0.396240727  -21.67 < 0.0000000000000002 ***
logGDP        0.895225875  0.061161211   14.64 < 0.0000000000000002 ***
Electricity   0.016882746  0.001894789    8.91  0.00000000000000023 ***
Urbanisation  0.000113946  0.002660792    0.04                 0.97
Water        -0.000000512  0.000000697   -0.74                 0.46
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.502 on 214 degrees of freedom
  (47 observations deleted due to missingness)
Multiple R-squared:  0.902, Adjusted R-squared:   0.9
F-statistic:  490 on 4 and 214 DF,  p-value: <0.0000000000000002
```

2. Cross section OLS for a year of your choice

I have selected 2019 as the most recent year with good data coverage.

The explanatory power is lower but the ability to capture variance remains strong R^2 > 0.87. F statistic is low which means the results are statistically significant.

Reg Equation

$$logCO2 = -8.8897 + 0.9000 * logGDP + 0.0135 * Electricity + 0.0005 * Urbanisation - 0.0000011 * Water$$

Interpreting Coefficients

Water and Urbanisation still remain insignificant. The coefficients for electricity and logGDP remain almost the same. Thus, for the interpretation please refer to the exercise #1.

```
Call:
lm(formula = logCO2 ~ logGDP + Electricity + Urbanisation + Water,
    data = subset(data2, Time == 2019))

Residuals:
    Min      1Q  Median      3Q     Max
-1.3110 -0.3520 -0.0128  0.3992  1.2929

Coefficients:
                 Estimate   Std. Error t value             Pr(>|t|)
(Intercept)  -8.889651656  0.393840700  -22.57 < 0.0000000000000002 ***
logGDP        0.900021783  0.063053180   14.27 < 0.0000000000000002 ***
```

```
Electricity   0.013526911  0.002339819    5.78          0.000000026 ***
Urbanisation  0.000486456  0.002315089    0.21                 0.83
Water        -0.000001053  0.000000813   -1.30                 0.20
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.496 on 216 degrees of freedom
  (45 observations deleted due to missingness)
Multiple R-squared:  0.872, Adjusted R-squared:  0.87
F-statistic:  369 on 4 and 216 DF,  p-value: <0.0000000000000002
```

3. First difference model, with time trend, no lags

I used different methods for time trends.

he results are almost identical for Basic FD and Time Trend Sequence OLS, for 1% increase in GDP, there is a 1.16% increase in CO2 emissions with the a robust R^2>0.79.The results suggest that GDP is more important determinant than time. Time trend is insignificant due to various ups and downs cause by diversity of countries and macroeconomics fundamentals over the past 30 years and the growth is not linear over time.

This is where assigning Time as a factor was more helpful as some later years there was a statistically significant negative coefficient which may be a sign that the climate change awareness and sustainability is coming to a fore. Significant coefficient was also for 2008 when during an economic crisis carbon-heavy factories and firms might have slowed down their operations. Nonetheless, it may be also a noise as this is only 1* significance. R-square is even more robust and F-score is low for all models pointing to statistical significance (mostly the statistically significant effect of GDP on CO2 emissions).

Basic FD

```
OLS estimation, Dep. Var.: logCO2
Observations: 6,331
Standard-errors: Clustered (Country)
            Estimate Std. Error t value  Pr(>|t|)
(Intercept)    -9.73     0.3040   -32.0 < 2.2e-16 ***
logGDP          1.16     0.0322    36.1 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.692464   Adj. R2: 0.795826
```

Time as a trend sequence (year)

```
OLS estimation, Dep. Var.: logCO2
Observations: 6,331
Standard-errors: Clustered (Country)
             Estimate Std. Error t value  Pr(>|t|)
(Intercept) -9.795180   0.313900 -31.205 < 2.2e-16 ***
logGDP       1.162565   0.032125  36.189 < 2.2e-16 ***
trend        0.000017   0.000017   0.995   0.32063
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.691371   Adj. R2: 0.796438
```

Time as a factor

```
OLS estimation, Dep. Var.: logCO2
Observations: 6,331
Fixed-effects: Country: 234
Standard-errors: Clustered (Country)
                Estimate Std. Error t value           Pr(>|t|)
logGDP           0.47403    0.06432   7.370 0.0000000000029571 ***
factor(Time)1993 -0.02017    0.00794  -2.541 0.0116915154823341 *
factor(Time)1994 -0.04387    0.01478  -2.967 0.0033187488152019 **
factor(Time)1995 -0.02324    0.01632  -1.424 0.1558419305480987
factor(Time)1996 -0.00372    0.02807  -0.133 0.8946699792388724
factor(Time)1997 -0.02638    0.02132  -1.237 0.2172267684940407
factor(Time)1998 -0.00733    0.02976  -0.246 0.8057299515262850
factor(Time)1999 -0.01954    0.03169  -0.617 0.5381279135487320
factor(Time)2000 -0.04149    0.03561  -1.165 0.2450796533555851
factor(Time)2001 -0.03782    0.03790  -0.998 0.3193513842037244
factor(Time)2002 -0.05511    0.03959  -1.392 0.1651869263922041
factor(Time)2003 -0.04192    0.04104  -1.022 0.3080115952365325
factor(Time)2004 -0.04688    0.04434  -1.057 0.2915140467903610
factor(Time)2005 -0.05745    0.04706  -1.221 0.2234230544511681
factor(Time)2006 -0.07533    0.05103  -1.476 0.1412379097607787
factor(Time)2007 -0.09608    0.05427  -1.770 0.0779544201241328 .
factor(Time)2008 -0.11956    0.05643  -2.119 0.0351601273434271 *
factor(Time)2009 -0.08696    0.05886  -1.477 0.1409258500508455
factor(Time)2010 -0.11220    0.05911  -1.898 0.0589071054992973 .
factor(Time)2011 -0.11217    0.06118  -1.833 0.0680089462363907 .
factor(Time)2012 -0.10119    0.06354  -1.593 0.1126037837029419
factor(Time)2013 -0.12731    0.06517  -1.954 0.0519548154438213 .
factor(Time)2014 -0.13898    0.06730  -2.065 0.0400305845731624 *
factor(Time)2015 -0.13104    0.06718  -1.951 0.0523026644148080 .
factor(Time)2016 -0.12756    0.06896  -1.850 0.0656068065142109 .
factor(Time)2017 -0.13689    0.07140  -1.917 0.0564457631766775 .
factor(Time)2018 -0.14943    0.07342  -2.035 0.0429610150073420 *
factor(Time)2019 -0.16613    0.07596  -2.187 0.0297372156337839 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
RMSE: 0.246482     Adj. R2: 0.973023
                 Within R2: 0.255672
```

4. First difference model, with time trend, 2 year lags

In this case, the R^2 > 0.08 which indicates it explains only a small percentage of variation. It may not be a good model and 2 year lags may not explain the workings of CO2 emissions as they are emitted immediatelly after burning or producing a product. F statistic is low which means the results are statistically significant but it still remains a poor model.

Reg Equation

$$logCO2_it - logCO2_i(t-1) = -0.01023 + 0.37580 * logGDP_it - 0.20383 * logCO2_i(t-1) - 0.04569 * logCO2_i(t-2) + 0.17742 * logGDP_i(t-1) + 0.06600 * logGDP_i(t-2) + \epsilon_it$$

Interpreting Coefficients

I have used a first-difference model with one-way (individual) effects. In other words, logCO2 and logGDP, and lags of both variables.

A 1% increase in GDP there is 0.3758% increase in CO2 emissions. There remains some residual effect. According to the coefficients, if there were high emissions this year people are less likely to pollute the next year. The opposite goes for GDP as if people were richer they are more likely to emit CO2 the next two years. The coefficient 0.17742 means that a one percent increase in the lagged GDP variable is associated with a 0.17742 percent increase in the current CO2 variable.

Nonetheless, as mentioned earlier, the model has a low R-squared value, indicating that it explains only a small portion of the variability and may not be a good fit.

```
Oneway (individual) effect First-Difference Model

Call:
plm(formula = logCO2 ~ logGDP + trend + lag(logCO2, 1:2) + lag(logGDP,
    1:2), data = pdata, model = "fd", index = c("Country", "Time"))

Unbalanced Panel: n = 225, T = 2-26, N = 4990
Observations used in estimation: 4765

Residuals:
    Min.  1st Qu.   Median  3rd Qu.     Max.
-0.98868 -0.03752 -0.00333  0.03202  1.26534

Coefficients:
                   Estimate Std. Error t-value             Pr(>|t|)
(Intercept)        -0.01023    0.00224   -4.57          0.000004988 ***
logGDP              0.37580    0.03098   12.13 < 0.0000000000000002 ***
lag(logCO2, 1:2)1 -0.20383    0.01438  -14.18 < 0.0000000000000002 ***
lag(logCO2, 1:2)2 -0.04569    0.01423   -3.21               0.0013 **
lag(logGDP, 1:2)1  0.17742    0.03142    5.65          0.000000017 ***
lag(logGDP, 1:2)2  0.06600    0.03012    2.19               0.0285 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    52.3
Residual Sum of Squares: 48.2
R-Squared:      0.0779
Adj. R-Squared: 0.077
F-statistic: 80.4342 on 5 and 4759 DF, p-value: <0.0000000000000002
```

5. First difference model, with time trend, 6 year lags

In this case, the R^2 > 0.10 which indicates it explains only a small percentage of variation. It may not be a good model and 6 year lags explain only a little better the variation in CO2 emisssions. F statistic is low which means the results are statistically significant but it still remains a poor model.

Reg Equation

$$logCO2(i,t) - logCO2(i,t-1) = \alpha i + \beta(logGDP(i,t) - logGDP(i,t-1)) + \Sigma\delta j * lag(logCO2(i,t), j) + \Sigma\gamma k * lag(logGDP(i,t), k) + \epsilon_i t$$

Interpreting Coefficients

This model suggest that the trend is significant only the first few years. CO2 emissions may cause lower CO2 emissions next year because people already consumed the emissions and do not need to do it anymore. Log GDP has a positive effect the following year on the CO2 emissions.

```
Oneway (individual) effect First-Difference Model

Call:
plm(formula = logCO2 ~ logGDP + trend + lag(logCO2, 1:6) + lag(logGDP,
    1:6), data = pdata, model = "fd", index = c("Country", "Time"))

Unbalanced Panel: n = 224, T = 1-22, N = 4092
Observations used in estimation: 3868

Residuals:
    Min.  1st Qu.   Median  3rd Qu.     Max.
-0.96620 -0.03712 -0.00258  0.03016  1.26489

Coefficients:
                  Estimate Std. Error t-value             Pr(>|t|)
(Intercept)       -0.01361    0.00307   -4.43            0.0000095 ***
logGDP             0.35773    0.03551   10.07 < 0.0000000000000002 ***
lag(logCO2, 1:6)1 -0.23382    0.01601  -14.61 < 0.0000000000000002 ***
lag(logCO2, 1:6)2 -0.03817    0.01629   -2.34                0.019 *
lag(logCO2, 1:6)3  0.14761    0.01632    9.05 < 0.0000000000000002 ***
lag(logCO2, 1:6)4  0.02751    0.01633    1.68                0.092 .
lag(logCO2, 1:6)5  0.00777    0.01647    0.47                0.637
lag(logCO2, 1:6)6 -0.00310    0.01590   -0.19                0.846
lag(logGDP, 1:6)1  0.14455    0.03659    3.95            0.0000795 ***
lag(logGDP, 1:6)2  0.00690    0.03626    0.19                0.849
lag(logGDP, 1:6)3 -0.05202    0.03612   -1.44                0.150
lag(logGDP, 1:6)4  0.07706    0.03544    2.17                0.030 *
lag(logGDP, 1:6)5  0.02045    0.03699    0.55                0.580
lag(logGDP, 1:6)6  0.03844    0.03583    1.07                0.283
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    45.3
Residual Sum of Squares: 40.5
R-Squared:      0.105
Adj. R-Squared: 0.102
F-statistic: 34.9551 on 13 and 3854 DF, p-value: <0.0000000000000002
```

6. Fixed effects model with time and country fixed effects

The R^2 is tiny which indicates it does not explain much variation. This may be due to overfitting. F statistic is low - still statistically significant.

Reg Equation

$CO2_it - mean(CO2_i) = -0.00001011 * (GDP_it - mean(GDP_i)) + u_it$

Interpreting Coefficients

When controlling for so many variables we may see the effect of GDP is signifciantly negative. This can mean that in fact GDP does not have such an large effect but rather other underlying factors that are country-specific.

The coefficient estimate for GDP is -0.00001011, which means that a 1% increase in GDP is associated with a 0.001011% decrease in carbon dioxide emissions. The standard error for the coefficient estimate is 0.00000218, and the t-value is -4.64, indicating that the coefficient is statistically significant at the 1% level. The R-squared value is 0.00411, indicating that GDP explains only a small portion of the variation in carbon dioxide emissions.

```
Oneway (individual) effect Within Model

Call:
plm(formula = CO2 ~ GDP, data = pdata, model = "within", index = c("Country",
    "Time"))

Unbalanced Panel: n = 225, T = 4-28, N = 5443

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-10.50290  -0.23642  -0.00241   0.24995  11.51097

Coefficients:
       Estimate  Std. Error t-value  Pr(>|t|)
GDP -0.00001011  0.00000218   -4.64 0.0000035 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    5480
Residual Sum of Squares: 5460
R-Squared:      0.00411
Adj. R-Squared: -0.0388
F-statistic: 21.551 on 1 and 5217 DF, p-value: 0.00000353
```

To what extent does economic activity cause CO 2 emission?

In the first few models, GDP seems to play an important role in CO2 emissions. These models suggest a high robustness but perhaps GDP is only a proxy for some other udnerlying variables or there is a parallel trend. This may include production, consumption, industry-mix incl. heavy manufacturing and energy mix.

Another cofounders may include the size of population and the economies of scale altough this was partically captured by urbanisation earlier in the models.

Thus, the workings of the CO2 emissions may be more complicated but in general we can agree that richer countries per capita generate more CO2. There may be a causal link but perhaps due to other variables not included in model that are correlated with GDP per capita.

The mechanism could be that net-exporters are more likely to be rich in terms of GDP per capita but also product more CO2 emissions.