

DA4 - Assignment 2: CO2 emission and GDP

A. BOGNAR

4/8/2022

A technical note: the code and data can also be found in this Github repo.

1. Data

To investigate the relationship and assess possible causal relationship between economic activity and CO2 emission I downloaded different World Bank Indicators and used different regression-based methods. Below I describe my findings.

The indicators have data for total 217 countries. The specific indicators I considered were:

- **GDP per capita** (constant 2015 US\$) for economic activity as causal variable. My data is between 1992 and 2018 and this the most recent indicator available for most countries.
- For **CO2 emissions** I had two options: either take the full CO2 emissions in kilo tonnes or metric tons per capita and divide it by population. Since the two numbers would be exactly the same, I chose by checking which one has more values in the dataset. As you can see below the difference was only one country, so I chose total emissions.

columns	number_countries_missing
metric tons per capita	27
metric kt or population	28

The next decision to make was how to deal with missing values, I took the following approach:

- First, drop all countries where either indicator was missing for all years.
- Second, drop countries with high number (over 10) of missing values in either variable.
- For countries with low missing values I decided to impute the next year's value (e.g. Kuwait had no emission data before 1995 so this year's value was added to 1992-94). I also checked the results had I dropped all missing values, the results turn out to be not statistically different, you can see them in the Annex.

The last feature engineering step is to log-transform both indicators, this means that our results can be interpreted in percentages, commonly called as elasticity. This is likely easier to interpret than flat values. 2 countries in the data had 0 emissions recorded. These are likely not really 0 but very small emissions. However, I still decided to drop them as they had several missing values and seem to be too small countries (Montenegro, Lebanon) to generalize based on their results.

2. Regressions

After the data cleaning I am left with results for 181 countries. First, I checked the general association between the two variables at the beginning and end of the dataset.

Cross-sectional OLS 1992 and 2018

In 1992 1% higher GDP was associated with just over 1% higher CO₂ emission per capita rate among the 181 countries in the data. In 2018 this association is weaker, only 0.89%

Table 1: Simple Cross Sectional Regression

	(1)	(2)
Intercept	-7.90*** (0.391)	-6.99*** (0.379)
Log GDP per capita 1992	1.02*** (0.044)	
Log GDP per capita 2018		0.892*** (0.043)
Observations	181	181
R2	0.715	0.761

First difference with and without lags

Moving onto more sophisticated methods to investigate possible causal link, the first difference models account for within-country and yearly trends. The first FD model shows that countries where GDP per capita increased by 1% in the same year a nearly 0.5% emission increase was expected on average. If we include effects over 2 years after the increase, we see that a total of 0.55% increase is expected most of which takes place in the same year as the GDP increase. If we look at even longer after-effect, for 6 years we don't see too much change, it seems that most of the effect takes place within the year of economic growth. In fact, the effects starting from the 2nd year are not statistically different from 0.

	(1)	(2)	(3)
Intercept	0.002 (0.002)	0.001 (0.002)	-0.003 (0.002)
First Diff of GDP per capita	0.495 *** (0.061)		
Cont Diff GDP per capita		0.438 *** (0.055)	0.453 *** (0.056)
1st Lag of Diff GDP per capita		0.119 * (0.049)	0.140 ** (0.050)
2nd Lag of Diff GDP per capita		0.014 (0.033)	-0.003 (0.041)
3rd Lag of Diff GDP per capita			-0.029 (0.034)
4th Lag of Diff GDP per capita			0.072 (0.043)
5th Lag of Diff GDP per capita			0.021 (0.037)
6th Lag of Diff GDP per capita			0.048 (0.029)
N	4706	4344	3620
R2	0.043	0.044	0.047

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

Fixed Effect and Long Difference

If we don't want to look at year-over-year effects but overall outcomes, we have too look at fixed effect or long difference models. The below fixed effect model shows that in years where a country's gdp per capita is 1% higher than its avg. throughout the 27 year period, a total of 0.537% higher emission was experienced than the country's avg. emission in the same period. The long difference model affirms a similar result and found that from 1992 to 2018 a 1% increase in gdp per capita resulted in 0.639% increase in emissions on avg in 2018.

	(1)	(2)
Log GDP per capita	0.537 *** (0.077)	
year	0.001 (0.002)	
Intercept		-0.014 (0.070)
Lag of GDP per capita		0.639 *** (0.141)
N	4887	181
R2	0.980	0.197

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

3. Summary, limitations

In summary, the models seem to show a positive effect between economic growth and emission. The magnitude seems to be around half percent higher emission for one percent gdp per capita increase in most cases. However, its important to note that these models only control for country specific trends and overall time trends. They do not control for any changes that affect all countries over the full period. They also don't control for changes in either indicator whose affect can change from year to year. For instance, we saw the association decreased somewhat between the beginning and end of the period. This might just be coincidence in the data. But we have seen a trend for some of the most developed countries moving to nuclear power such as France or developing countries such as China becoming green super powers. The energy composition of a country can affect both its economic growth and its emission and this affect can change rapidly. To be more confident in establishing a causal relationship we should control for such variables.

Still, while some argue that economic growth can be achieved without higher emissions, for now most people would consider that scientific and sociological changes are needed to achieve that so until new evidence is presented we can at least take away from these findings that growth did typically lead to higher emissions in this period of human history. With industrial production still literally fueling the growth of emerging countries such as China it is hard to say if this can change before humanity will have to face dire consequences of growth.

Annex - Comparing imputation to dropping missing values

Results for the same regression when I dropped all countries with any missing data instead of imputation. The same size is smaller by 20 countries. As you will see most coefficients' confidence intervals overlap with the imputation results. Keep in mind that the countries with missing data are generally really small, so if our question included weights, the difference would be even less.

Cross-section

Table 2: Simple Cross Sectional Regression with all missing dropped

	(1)	(2)
Intercept	-8.02*** (0.396)	-6.97*** (0.406)
Log GDP per capita 1992	1.03*** (0.045)	
Log GDP per capita 2018		0.885*** (0.046)
Observations	161	161
R2	0.725	0.758

First difference and lags

	(1)	(2)	(3)
Intercept	0.003 (0.002)	0.002 (0.002)	-0.004 (0.002)
First Diff of GDP per capita	0.511 *** (0.076)		
Cont Diff GDP per capita		0.432 *** (0.080)	0.443 *** (0.086)
1st Lag of Diff GDP per capita		0.135 * (0.064)	0.227 ** (0.076)
2nd Lag of Diff GDP per capita		-0.017 (0.041)	-0.051 (0.053)
3rd Lag of Diff GDP per capita			-0.015 (0.055)
4th Lag of Diff GDP per capita			0.076 (0.072)
5th Lag of Diff GDP per capita			-0.017 (0.055)
6th Lag of Diff GDP per capita			0.061 (0.041)
N	4186	3864	3220
R2	0.037	0.035	0.042

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

FE and Long Diff

	(1)	(2)
Log GDP per capita	0.550 *** (0.085)	
year	0.002 (0.002)	
Intercept		-0.051 (0.074)
Lag of GDP per capita		0.748 *** (0.145)
N	4347	161
R2	0.980	0.249

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