Python Project

Anjali Agrawal and Eileen Zhang

May 2022

1 Notes

-This is a very, very, rough draft of our writeup. We jotted down quick explanations of how we went about the code and analysis, and the intent of exploring the extensions that we did.
-Please offer any suggestions on how we could best format our ideas.
Additionally, we are trying to create this kind of plot `https://dataplotplus.com/add-text-labels-scatter-plot-in-python/` where -we can assign specific colours to our dummy variables (low income countries, middle income countries, high income countries) but are struggling. Would you be able to offer some advice on how to go about this?

2 Topic and Explanation

- We wanted to explore the impact of CO2 growth on GDP growth. We found this relationship to be particularly interesting given the imperative nature of climate action.
- Our data is from the world bank and looks at the relationship between GDP per capita and CO2 emissions per capita (controlling for time and fixed effects).

3 Files we included

- Programmatic download: In this file we used an API called wbgapi that programmatically accesses the world bank database. Using this we pulled the variables of our interest into a dataframe, renamed them, and exported it as a csv. Users can run this code file to programmatically download our dataset.
- Data analysis: the data analysis file houses all the work we did on cleaning our dataset, reshaping it. In this file users can run the analysis that we did i.e. all our regressions and also programmatically download our visualisations.

4 Analysis and Methods

4.1 Impact of carbon emission on GDP

$$log\_GDP = \alpha + \beta log\_CO2 \tag{1}$$

For our initial regression we ran an OLS linear regression on log GDP and log CO2. To begin with we pulled the GDP per capita and CO2 per capita. We logged these values and then dropped NaN values. Furthermore we filtered the time period to be between 2000-2019. With the log GDP per capita as our dependent variable, our model achieved an R^2 value of 0.743. The regression results are as follows, and a scatter plot as well as a residual plot that explains the fitness of our model is below:

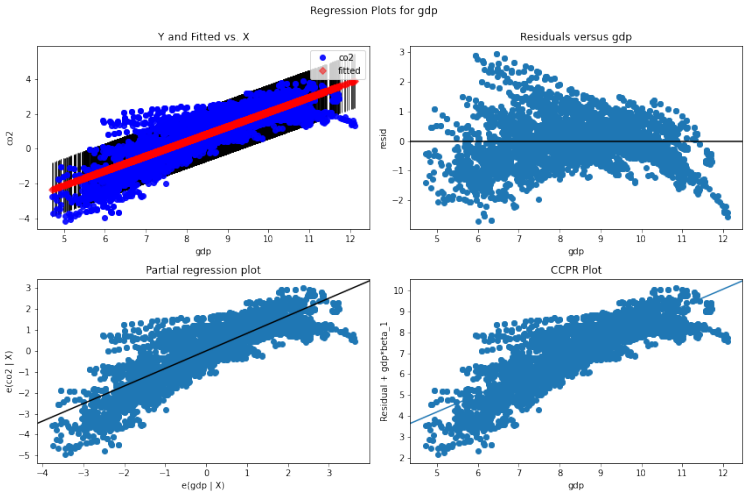

Figure 1: Fitted explanations of Regression 1

	log_gdp
const	7.76***
	(0.01)
log_co2	0.86***
	(0.01)
R-squared	0.72
R-squared Adj.	0.72

4.2 Panel Regression

- We also created a panel dataset that introduced log methane emission, and log nitrous oxide emissions to the data. We also controlled for time effects and fixed effects from 2000-2019. In doing so we found that our initial fit for the data is significantly reduced.Table 2 summarizes the results

5 Extension

5.1 Methods

- Dummy: Now flipping the dependent and independent variables. We created helper functions that would run through the list of economies and return a 1 value for any country that was in the income level and 0 if not. This improved our R^2 up to 0.775.
- Interaction terms: we added interaction terms for our dummy and log gdp to accommodate that each dummy likely has a unique slope compared to the other dummy group. By doing this we were able to improve the fit of our data even more to an R^2 of 0.792.

5.2 Analysis

- We suspected that our model suffered from simultaneous endogeneity and so we decided to flip the regression and see what results we would get.

(a) Given that we were now analysing the relationship between growth of GDP on growth of CO2 we decided to analyse the environmental Kuznets curve with our data. The environmental Kuznets curve (EKC) is a hypothesized relationship between various indicators of environmental degradation and per capita income. In the early stages of economic growth, pollution emissions increase and environmental

Dep. Variable:	log_gdp	**R-squared:**	0.2108
Estimator:	PanelOLS	**R-squared (Between):**	0.2334
No. Observations:	5321	**R-squared (Within):**	0.2198
Date:	Wed, May 04 2022	**R-squared (Overall):**	0.2536
Time:	20:37:41	**Log-likelihood**	629.25
Cov. Estimator:	Unadjusted		
		F-statistic:	455.41
Entities:	185	**P-value**	0.0000
Avg Obs:	28.762	**Distribution:**	F(3,5115)
Min Obs:	3.0000		
Max Obs:	38.000	**F-statistic (robust):**	455.41
		P-value	0.0000
Time periods:	19	**Distribution:**	F(3,5115)
Avg Obs:	280.05		
Min Obs:	269.00		
Max Obs:	289.00		

	Parameter	**Std. Err.**	**T-stat**	**P-value**	**Lower CI**	**Upper CI**
const	4.7730	0.2082	22.921	0.0000	4.3648	5.1813
log_co2	0.4261	0.0165	25.863	0.0000	0.3938	0.4584
log_me	0.2725	0.0250	10.890	0.0000	0.2235	0.3216
log_nox	0.0998	0.0264	3.7748	0.0002	0.0480	0.1517

Table 1: PanelOLS Estimation Summary

quality declines, but beyond some level of per capita income (which will vary for different indicators) the trend reverses, so that at high income levels, economic growth leads to environmental improvement. This implies that environmental impacts or emissions per capita are an inverted U-shaped function of per capita income.

$$log\_C02 = \alpha + \beta_1 log\_GDP + \beta_2 Low\_Income + \beta_3 High\_Income + \tag{2}$$

$$\beta_4 Low\_Income * log_G DP + \beta_5 Low_I ncome * log_G DP \tag{3}$$

(b) Looking at our data we can see that with being in a high income country the co2 growth increases compared to other income level countries, however, with each increase in 1 % of the growth rate we see a 0.46 % decrease in carbon dioxide emissions per capita compared to middle income countries. Which suggests that there is some weight to the Kuznets argument.

Table 2

	log_co2
const	-5.74***
	(0.11)
log_gdp	0.80***
	(0.01)
Low_dummy	-2.99***
	(0.26)
High_dummy	4.53***
	(0.25)
Interaction_low	0.30***
	(0.04)
Interaction_high	-0.47***
	(0.03)
R-squared	0.79
R-squared Adj.	0.79