2020

DATA PROJECT

THE RELATIONSHIP BETWEEN CO2 EMISSION AND ENVIRONMENT, EDUCATION, AND ECONOMY FACTORS

WU YUTONG

Table of Contents

Key Research Topic

Over the past decades, there is a surge of concerns in the climate change due to the negative effects it placed on the planet as well as the economic development, such as unexpected weather, lack of ecosystem diversity, poor agriculture, health issue, and fall in tourism (ImportantIndia, 2017). According to European Commission, the increasing level of CO2 as a result of human activities is the major cause of climate change.

In light of that, this study aims to examine how the CO2 emissions of a country vary in responds to the changes in factors regarding the domestic environment, education, and economy and empirically tests the hypothesis using cross-sectional data from 137 countries.

Resource and Data Set

Data is extracted from The Quality of Government Institute (QOG) Data Set 2020. The data source is reliable given that the provider - QOG - is an independent, impartial, and professional research institution founded by 2 professors from University of Gothenburg. The data set composed of 2 types of data, cross-sectional and time series. For the purpose of this study, only cross-sectional type of data is adopted; the **unit of analysis** is Country.

Given the data limitation, the **studied year** centers around 2016, with plus or minus 3-year variation. This study assumes that environmental and socioeconomic data of a country will not significantly vary within the 3-year range.

5 variables from 4 categories (Energy and Infrastructure, Environment, Education, and Public Economy) out of the 19 categories are selected for investigation, including CO2 Emission per Capita, Total Ecological Footprint of Consumption per Person, Renewable Electricity Output, Human Capital Index, and GDP per Capita.

Explanation of the Variables

- **Criteria in Variable Selection**

The selection of variable is based on the following considerations:

1) **Number of observations**: variables that contain less than 100 observations will not be considered;
2) **Multicollinearity**: to reduce the likelihood of multicollinearity, the variables chosen for the regression model are considered to be representative and do not significantly correlate with each other. For instance, this study chooses Total Ecological Footprint of Consumption as the environmental variable, rather than Fossil fuel energy consumption or Electricity production from coal sources, since it is believed that the latter ones are part of the Total Ecological Footprint.

- **Dependent Variables**

CO2 emissions (*co2_emission*) is selected as the response variable. The variable measures quantity (metric tons) of CO2 emissions per capita of a country caused from burning of fossil fuel and manufacturing of cement. The data ranges from year 2014 to 2016, with 190 countries available.

- **Independent Variables**

3 types of explanatory variables are prioritized as factors that most importantly correlate with CO2 emission, including Environment, Education, and Economy.

1) Environment

Environmental factors comprise of Total Ecological Footprint of Consumption (*eco_footprint*) and Renewable electricity output (*ele_renewable*). **Total Ecological Footprint of Consumption** measures how much Global Hectares (GHA) of ecological footprint that each individual in a country consumes during a year ranging from 2013 to 2016, with 176 observations in total. **Renewable electricity output** variable measures the proportion of the electricity that is generated by renewable resources in total electricity of a country during year 2015, containing 193 observations.

2) Education

Education factor is represented by **human capital index** (*human_capital*) variable, which is an index based on the years of schooling and assumed returns of 142 countries during year 2016.

3) Economy

Economy factor is represented by **GDP per capita in constant 2010 US dollar**. The variable is calculated by gross domestic product divided by midyear population in year 2016 (2014 or 2016 if data in year 2016 is not available). The variable contains GDP data of 187 countries.

Testable Hypothesis

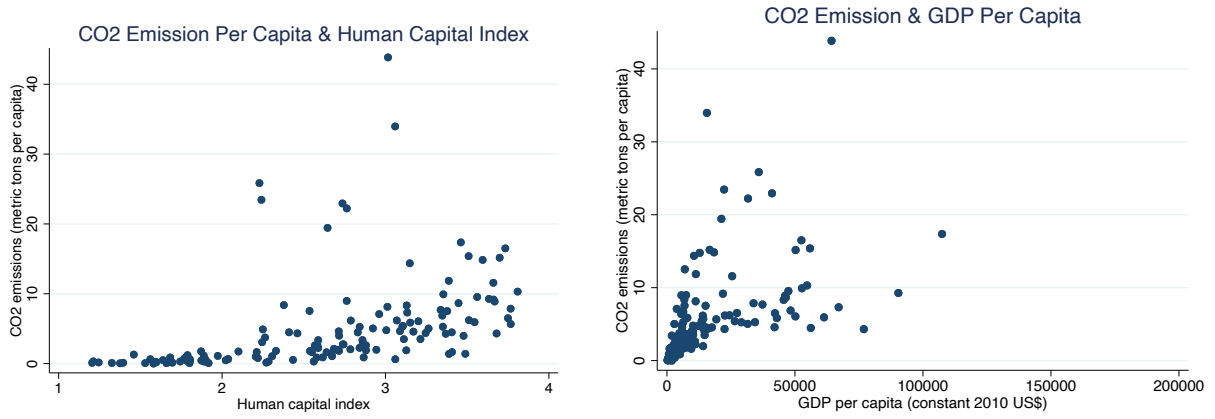
- **Hypothesis 1 and Hypothesis 2 for Environmental Factor:**

Carbon Footprint, which can be expressed in tonnes of carbon dioxide emitted, comprises the largest proportion (60%) of the ecological footprint (Global Footprint Network, 2017). Therefore, the **first hypothesis** is that the CO2 emission will linearly increase with the increase in humanity's ecological footprint, holding all else constant.

Additionally, research point out that the carbon emission from energy can be reduced by 70% with the help of the renewable resources (CimateAction, 2017). Thus, the **second hypothesis** is that everything else being equal, countries with a larger proportion electricity from renewable sources are expected to have a lower level of carbon dioxide emission.

- **Hypothesis 3 and Hypothesis 4 for Socioeconomic Factors:**

Sociodemographic variables are important in explaining emissions (Baiocchi, et al., 2010). However, the relationship might be much more complicated than environmental factors given that they might not be directly correlated with CO2 emission. Therefore, bivariate scatter plots graphs are constructed as follows, trying to observe some patterns.



Based on the data visualization, this study assumes a non-linear correlation between CO2 emission and the studied socioeconomic factors. The **third hypothesis** is that the relationship of CO2 emission and Human Capital Index is negative quadratic, holding other factors constant. Similarly, the **final hypothesis** is that given everything else equal, the relationship of CO2 emission and GDP Per Capita of a country is also negative quadratic. Note that although the quadratic relationships are not apparently demonstrated in the graphs above, it might be the case after controlling other variables in the following analysis.

Regression Model and Interpretation

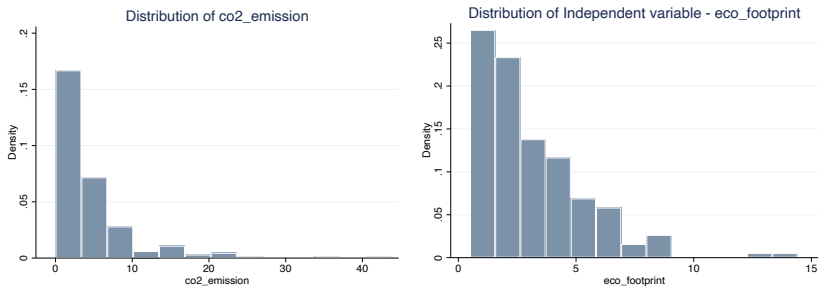
Two statistical models are derived for the project. The regression output of the two specifications are shown below and detailed explanation and interpretation are articulated following the regression output.

VARIABLES	(1) Log (co2_emission)	(2) Log (co2_emission)
Log (eco_footprint)	1.249***	1.006***
	(0.144)	(0.138)
ele_renewable	-0.011***	-0.011***
	(0.002)	(0.002)
human_capital	0.631***	3.794***
	(0.122)	(0.560)
square_human_capital		-0.634***
		(0.109)
gdp_percapita	**-1.12*10^-07**	2.64*10^-5 ***
	(0.000)	(0.000)
square_gdp		-2.13*10^-10 **
		(0.000)
Constant	-1.706***	-5.370***
	(0.264)	(0.678)
Observations	136	136
R-squared	0.840	0.876

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

- **Original Model with Logged Variables (Specification 1)**

At the beginning, histogram distribution of each variables is derived to test the normal distribution. As shown below, the dependent variable and the *eco_footprint* variable are not normally distributed. Thus, *co2_emission* and *eco_footprint* variables are logged to avoid violation of the basic assumption.



Then, the first model is run based on the logged variables. Note that in the first model, all the independent variables are assumed to be linearly correlated with the dependent variable. The equation based on the specification 1 is as follows:

Log (co2_emission)
= 1.249*Log(eco_footprint) - 0.011*ele_renewable – 0.631* human_capital – 1.12*10^-07 * gdp_percapita - 1.706

The adjusted R-squared of the basic model is 0.835, indicating that in overall, 83.5% of the observation can be explained by the basic model. The F value is significantly low, suggesting that the coefficients of the 4 predictor variables on dependent variable are jointly significant.

For each independent variable, excepted for *eco_footprint*, which is positively correlate with *co2_emission*, all other explanatory variables have a negative coefficient on the dependent variable. Notably, based on the basic model, *gdp_percapita* is not significantly correlate with the *co2_emission* given a 95% confidence interval, since the p value is 0.977, which is significantly larger than 0.05.

- **Quadratic Model (Specification 2)**

To test our hypothesis of negative quadratic relationship as stated earlier, the second model includes quadratic terms for *human_capital* and *gdp_percapita* variables based on the first model.

The new equation after adding quadradic terms is as follows.

$$Log(co2\_emission) = 1.01 * \mathrm{Log}(eco\_footprint) - 0.01\, ele\_renewable + 3.79 * human\_capital - 0.63 * human\_capital^2 + 0.0000264 * gdp\_percapita + (-2.13 * 10^{-10}) * gdp\_percapita^2 - 5.37$$

Compared with the basic model, the quadratic model has an adjusted R-square of 87.02%, which is around 3.5% higher than that in basic model and represents a better fit.

For non-quadratic variables, the coefficient of log(*eco_footprint*) is 1.01, which indicates that holding all other variables constant, 1% increase in the ecology footprint per individual of a country is associate with 1.01% rise in annual CO2 emissions per capita. The coefficient of *ele_renewable* is -0.01, suggesting a negative correlation between *ele_renewable* and co2_emission. Specifically, 1 unit increase in the percentage of total electricity that is renewable is expected to associate with a 1% decrease of CO2 emission per capita, holding other variables constant.

With respect to statistical significance, the p-values of *log_eco_footprint* and *ele_renewable* are approximately 0, which means that the possibility of having the coefficients as in the model given that the coefficients are actually 0 is extremely low. Therefore, both *eco_footprint* and *ele_renewable* variables are significantly correlate with *co2_emission*.

As for quadratic variables, the coefficient on the quadratic term of *human_capital* is -0.63. The turning point is 3.01[1]. Therefore, based on the transformed model, we would expect that when human capital index is lower than 3.01, the level of CO2 emission of a country is expected to increase as the human capital index increases; however, when human capital index is higher than 3.01, CO2 emissions are expected to decrease as the human capital increases. For variable *gdp_percapita*, the quadratic coefficient is also negative and based on the model, the CO2 emission

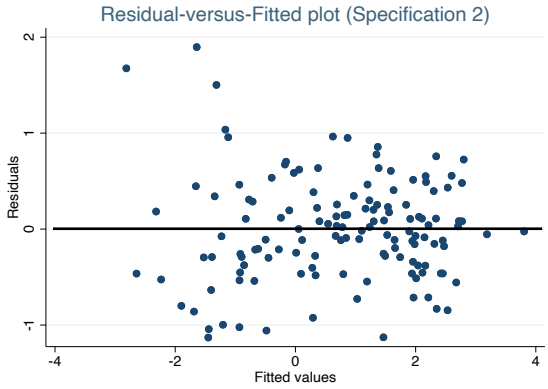
[1] The turning point of human captial variable: 3.79/(2*0.63) = 3.01

of a country reaches the maximum value when the GDP per capita is $61972[2], holding all else constant. This indicate that we would expect a positive correlation between *co2_emission* and *gdp_percapita* if the GDP is lower than $61972 and negative correlation otherwise.

Joint tests are conducted to test the significance of the quadratic terms. The p-value of the F test for joint effect of the *human_capital* and $human\_capital^2$ variables is 0.000 and for joint effect of the $gdp\_percapita$ and $gdp\_percapita^2$ is 0.0048. Thus, the 2 quadratic variables are both significantly correlate with *co2_emission*.

- **Heteroskedastic and Robust Model**

Residual against the fitted value plot is derived below based on the quadratic model. The plot seems to suggest heteroskedastic exist.



A Breusch-Pagan Test was run to validate the visual finding. The p-value of the test result is approximately 0, which is significantly low. Therefore, the null hypothesis is rejected that the residuals have constant variance regardless of the independent variables and there is significant evidence of heteroskedastic.

Following the BP Test, a robust model is run to re-estimates the standard errors in light of heteroskedasticity. However, the p-value of each variable in the model after considering

[2] The turning point of gdp variable: 0.0000264 / (2*2.13*10^(-10))

heteroskedasticity is still significantly low, indicating that the heteroskedasticity level is acceptable in our model.

Results and Conclusion

Since the quadratic models present a more significant result, particularly the coefficient on the economic variable. Besides, the quadratic model has a higher R-squared, representing a better goodness of fit. Therefore, this study concludes that the second specification is a more representative model to describe the correlation between CO2 emission and the four selected explanatory variables.

Based on the regression result, there are sufficient evidences to support the 4 hypothesis that articulated earlier in the report. The following conclusions are drawn based on the findings:

First of all, countries that have a higher ecological per capita tend to have a higher level of CO2 emission.

Secondly, countries with a larger output in renewable electricity are expected to have lower amount of carbon dioxide emission.

Thirdly, for those countries that have a human capital index higher than 3.01, the CO2 emission are expected to decrease as the human capital index increase. However, for countries of which human capital index is lower than 3.01, a unit increase in the human capital index is potentially associated with a percentage increase in CO2 emission. That's to say, countries in which people are in average extremely well-educated and in which citizen are extremely poorly educated tend to have a low level of CO2 emission.

Lastly, the relationship between the GDP per capita and CO2 emission, while less obvious, is similar to that of human capital index and CO2 emission. Therefore, it is expected that countries

with an extremely high GDP per capita and those that are extremely poor will have relatively low-level CO2 emissions, whereas countries with a moderate level in economic development tend to have the highest levels of CO2 emission.

Reference

Baiocchi, G., Minx, J. & Hubacek, K., 2010. The Impact of Social Factors and Consumer Behavior on Carbon Dioxide Emissions in the United Kingdom. *Journal of Industrial Ecology,* 14(1).

CimateAction, 2017. *Renewables can reduce CO2 emissions by 70% by 2050.* [Online] Available at: http://www.climateaction.org/news/renewables_can_reduce_co2_emission_by_70_by_2050

European Commission, n.d.. *Causes of climate change.* [Online] Available at: https://ec.europa.eu/clima/change/causes_en

ImportantIndia, 2017. *What are the major effects of Climate Change?.* [Online] Available at: https://www.importantindia.com/25636/25636/

Global Footprint Network, 2017. *How Ecological Footprint accounting helps us recognize that engaging in meaningful climate action is critical for our own success.* [Online] Available at: https://www.footprintnetwork.org/2017/11/09/ecological-footprint-climate-change/