CAN NATURAL FACTORS EXPLAIN ANY CROSS-COUNTRY DIFFERENCES IN CARBON DIOXIDE EMISSIONS?

ANSON MA (#20916612) and TAHMID BARI (#20864394)
GROUP 9

Department of Management Sciences
University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

(First Draft 27 November 2020; Final Draft 7 December 2020)

Keywords: GDP; Temperature; Transportation; Renewable and Fossil Resources

1 INTRODUCTION

The objective of this paper is to determine whether natural factors in additional to gross domestic product (GDP) can explain any cross-country differences in carbon dioxide (CO_2). The analysis conducted is mainly based on and compared against four countries of different income groups as categorized by the World Bank, namely Ethiopia representing low-income), Vietnam representing lower-middle-income, Thailand representing middle-income, and Japan representing high-income. CO_2 is an important heat-trapping gas, which is released through human activities such as deforestation and burning fossil fuels, as well as natural processes such as respiration and volcanic eruptions (NASA[1], 2008).

Over the past decade, human activities have raised atmospheric concentrations of CO_2 by 47% above pre-industrial levels found in 1850 (NASA, 2020). This has presented one of the world's most pressing challenges to combat climate change. Nevertheless, far from stabilizing concentrations, the global CO_2 emissions are in fact still rising and accumulating. Figure 1 shows the annual total CO_2 emissions, by world regions. It is not difficult to see that nations around the world has yet to peak in emissions.

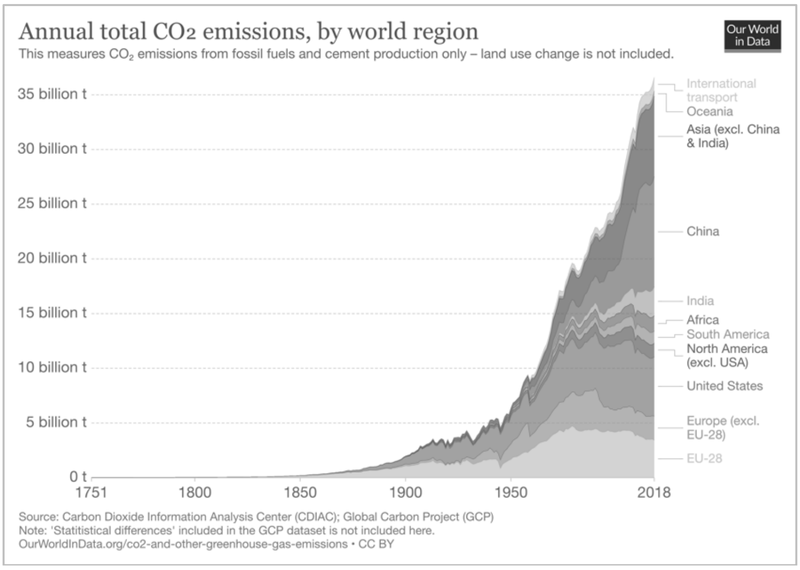
A changing climate has a range of potential ecological, physical, and health impacts, including extreme weathers, sea-level rise; altered crop growth; disrupted water systems. Between 2030 and 2050, climate change is expected to cause approximately 250 000 additional deaths per year, from malnutrition, malaria, diarrhoea and heat stress. The direct damage costs to health issue are estimated to be between USD 2-4 billion/year by 2030 (WHO[2], 2018). This is a critical to each and every one of us as we have all been experiencing these negative impacts already, regardless which part of the world we situate in.

1 The National Aeronautics and Space Administration
2 The World Health Organization

If countries achieved their current pledges and targets, there would be an improvement on climate change in the coming years. But if our aim is to limit global warming to well below 2 Degrees Celsius as laid out in the Paris Agreement, we are clearly still far from the rates of progress we would need to achieve international targets. To begin with, countries will have to identify factors that affect their CO_2 emissions (Ritchie, H et al.,2017).

The purpose of this project is therefore to examine and reproduce the relationship between CO_2 emissions, GDP and natural factors through statistical methods as established by Neumayer (2002). Natural factors include the differences in average climate temperatures, the proportion of urban land areas, and the availability of renewable and fossil fuel resources While many studies have shown the empirical relationship between CO_2 emissions and income, the question to what extent natural factors determine cross-country differences in CO_2 emissions has been somehow neglected, which will be addressed in this paper.

FIGURE 1 Annual total CO_2 emissions, by world regions (Global Carbon Project, 2019)



2 LITERATURE REVIEW

We started off our project by researching for some literatures done in the past years. This section reviews recent literature that seeks to determine the factors. Tucker (1995) examined the relationship between CO_2 emissions and global GDP. A positive relationship between CO_2 emissions, the most important greenhouse gas implicated in global warming, and GDP is shown in his paper, examining per capita income and CO_2 emissions of 137 countries across 21 years.

Tucker (1995) also shows that as per capita incomes accelerate across countries emissions increases, for the most part, tend to decelerate. It could be that higher income levels lead to increased demand for environmental protection. Only emissions reduction proposals that assure incomes will not be adversely affected, particularly those of less developed countries, will have any possibility of successful implementation. It is obvious that there are some other factors that leads to CO_2 emissions among countries.

Li et al. (2019) investigated the impacts of modernization on CO_2 emissions based on the situation of China. Modernization refers to the general trend of developmental progress that occurs within human societies. As the world's largest developing economy and carbon emitter, China faces the dual challenge of peaking carbon emissions by 2030 while realizing basic modernization by 2035. The results demonstrate that industrialization, agricultural modernization, informatization, and urbanization exerted positive effects on CO_2 emissions during the study period, suggesting these aspects of modernization led to increased CO_2 emissions. A negative correlation between ecological modernization and CO_2 emission was identified, indicating that ecological modernization helped to abate CO_2 emissions.

Zhang et al. (2017) used a panel data of 141 countries over the period of 1961–2011 and analyzed the impact of urbanization on CO_2 emissions empirically. As a crucial indicator of modernization, urbanization has significant effects on CO_2 emissions. The results show that there is an inverted U-shaped relationship between urbanization and carbon emissions and the turn point is around 73.80%. But excessive urban concentration can claim the benefits of high-level urbanization.

Kerkhof et al. (2009) looked into the variation in household CO_2 emissions between and within countries to identify some determinants of national household CO_2 emissions and their distribution across income groups. For that purpose, the study quantifies the CO_2 emissions of households in the Netherlands, UK, Sweden and Norway around the year 2000 by combining a hybrid approach of process analysis and input–output analysis with data on household expenditures. The results show that average households in the Netherlands and the UK give rise to higher amounts of CO_2 emissions than households in Sweden and Norway. Moreover, CO_2 emission intensities of household consumption decrease with increasing income in the Netherlands and the UK, whereas they increase in Sweden and Norway. A comparison of the national results at the product level points out that country characteristics, like energy supply, population density and the availability of district heating, influence variation in household CO_2 emissions between and within countries.

Moutinho et al. (2017) examined the method to estimate the efficiency of 26 different European Countries over 2001 and 2012 comparing their performance using Data Envelopment Analysis and quantile regression technique. The results indicate that share of renewables and non-renewable energy sources are important to explain differences in emissions. They suggest a significant change in the trend of economic and environmental efficiency in European countries and put forward the high disparities existing among them.

Bhattacharya et al. (2017) conducted a comprehensive and robust analysis of the role of renewable energy consumption and institutions on economic growth and in combating CO_2 emissions across the regions and income groups through annual data from 85 developed and developing economies across the world over the period from 1991 to 2012. The results from the system and fully modified ordinary least square analysis indicate that the growth of renewable energy consumption has a significant positive and negative impact on economic output and CO_2 emissions, respectively. Both renewable energy deployment and institutions are significant in promoting economic growth and reducing CO_2 emissions.

3 METHODOLOGY

We utilized RStudio and Microsoft Excel to execute the model. RStudio is an integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. RStudio is available in open source and commercial editions and runs on the desktop on different operating platforms. An assessment of the normality of data is a prerequisite for many statistical tests because normal data is an underlying assumption in parametric testing. There are two ways to assess normality: graphically and numerically.

Dependent Variable test of normality – Kolmogorov-Smirnov test (K-S test): An attractive feature of this test is that the distribution of the K-S test statistic itself does not depend on the underlying cumulative distribution function being tested. Another advantage is that it is an exact test (the chi-square goodness-of-fit test depends on an adequate sample size for the approximations to be valid). Despite these advantages, the K-S test has several important limitations: It only applies to continuous distributions. It tends to be more sensitive near the centre of the distribution than at the tails. Perhaps the most serious limitation is that the distribution must be fully specified. That is, if location, scale, and shape parameters are estimated from the data, the critical region of the K-S test is no longer valid. It typically must be determined by simulation.

Dependent Variable test of normality – Shapiro-Wilk (S-W test): The S-W test is a way to tell if a random sample comes from a normal distribution. The test gives us a W value; small values indicate our sample is not normally distributed we can reject the null hypothesis that our population is normally distributed if our values are under a certain threshold).

The null-hypothesis of this test is that the population is normally distributed. Thus, if the p-value is less than the chosen alpha level, then the null hypothesis is rejected and there is evidence that the data tested are not normally distributed. On the other hand, if the p-value is greater than the chosen alpha level, then the null hypothesis (that the data came from a normally distributed population) cannot be rejected (e.g., for an alpha level of .05, a data set with a p-value of less than .05 rejects the null hypothesis that the data are from a normally distributed population).

Like most statistical significance tests, if the sample size is sufficiently large this test may detect even trivial departures from the null hypothesis (i.e., although there may be some statistically significant effect, it may be too small to be of any practical significance); thus, additional investigation of the effect size is typically advisable, e.g., a Q–Q plot in this case. Applying the sample t-test to compare the mean of DV on 4 countries: The dependent sample t-test is a member of the t-test family. All tests from the t-test family compare one or more mean scores with each other. The t-test family is based on the t-distribution, sometimes also called Student's t. Within the t-test family the dependent sample t-test compares the mean scores of one groups in different measurements. It is also called the paired t-test, because measurements from one group must be paired with measurements from the other group. The dependent sample t-test is used when the observations or cases in one sample are linked with the cases in the other sample. This is typically the case when repeated measures are taken, or when analysing similar units or comparable specimen.

Construct 95% confidence intervals around the largest and the smallest mean: There are two types of estimates for each population parameter: the point estimate and confidence interval (CI) estimate. For both continuous variables (e.g., population mean) and dichotomous variables (e.g., population proportion) one first computes the point estimate from a sample. Recall that sample means and sample proportions are unbiased estimates of the corresponding population parameters.

For both continuous and dichotomous variables, the confidence interval estimate (CI) is a range of likely values for the population parameter based on the point estimate, e.g., the sample mean and the investigator's desired level of confidence (most commonly 95%, but any level between 0-100% can be selected) and the sampling variability or the standard error of the point estimate.

Multivariate Analysis - Ordinary Least Squares (OLS) Regression Model: The OLS regression is more commonly named linear regression (simple or multiple depending on the number of explanatory variables). OLS chooses the parameters of a linear function of a set of explanatory variables by the principle of least squares: minimizing the sum of the squares of the differences between the observed dependent variable (values of the variable being observed) in the given dataset and those predicted by the linear function.

There are several different frameworks in which the linear regression model can be cast in order to make the OLS technique applicable. Each of these settings produces the same formulas and same results. The only difference is the interpretation and the assumptions which have to be imposed in order for the method to give meaningful results. The choice of the applicable framework depends mostly on the nature of data in hand, and on the inference task which has to be performed.

Residual analysis to check for normality: The standard assumption in linear regression is that the theoretical residuals are independent and normally distributed. The observed residuals are an estimate of the theoretical residuals, but are not independent (there are transforms on the residuals that remove some of the dependence, but still give only an approximation of the true residuals). So, a test on the observed residuals does not guarantee that the theoretical residuals match.

To take note that the tests of normality are rule out tests, they can us you that the data is unlikely to have come from a normal distribution. But if the test is not significant that does not mean that the data came from a normal distribution, it could also mean that we just do not have enough power to see the difference. Larger sample sizes give more power to detect the non-normality, but larger samples and the central limit theorem mean that the non-normality is least important. So, for small sample sizes the assumption of normality is important but the tests are meaningless, for large sample sizes the tests may be more accurate, but the question of exact normality becomes meaningless.

Breusch-Pagan test to check for Heteroskedasticity: Heteroskedasticity occurs when the variance for all observations in a data set are not the same. In this demonstration, we examine the consequences of heteroskedasticity, find ways to detect it. Breusch Pagan Test is used to test for heteroskedasticity in a linear regression model and assumes that the error terms are normally distributed. It tests whether the variance of the errors from a regression is dependent on the values of the independent variables. It is a x^2 test.

In the presence of heteroskedasticity, there are two main consequences on the least square's estimators: The least squares estimator is still a linear and unbiased estimator, but it is no longer best. That is, there is another estimator with a smaller variance. The standard errors computed for the least square's estimators are incorrect. This can affect confidence intervals and hypothesis testing that use those standard errors, which could lead to misleading conclusions.

Most real-world data will probably be heteroskedastic. However, one can still use OLS without correcting for heteroskedasticity because if the sample size is large enough, the variance of the least square's estimator may still be sufficiently small to obtain precise estimates. The following shows the proposed model of this paper.

$$Y_i = a_i + b_{i1}x_{i1} + b_{i2}{x_{i1}}^2 + b_{i3}x_{i3} + b_{i4}x_{i4} + b_{i5}x_{i5} + b_{i6}x_{i6} + b_{i7}x_{i7} + e_i$$

(Neumayer E, 2002)

$Y = CO_2\ emission\ per\ capita$
$x_1 = GDP\ per\ capita$
${x_1}^2 = squared\ GDP\ per\ capita$
$x_3 = lowest\ monthly\ average\ temperature$
$x_4 = highest\ monthly\ average\ temperature$
$x_5 = percentage\ of\ urban\ land\ in\ total\ land\ area$
$x_6 = percentage\ of\ renewable\ energy\ use\ in\ total\ energy\ use$
$x_7 = percentage\ of\ fossil\ fuel\ consumption\ \ in\ total\ energy\ use$
$\varepsilon\ = error\ term$

$i = 1{:}\ Ethiopia;\ 2{:}\ Vietnam, 3{:}\ Thailand, 4{:}\ Japan$

4 DATA

The raw data is consisted of a panel covering 25 years from 1991-2015 on 4 countries collected form the World Bank to maintain consistency, expecting a better significant mode. Table 1 depicts the descriptive statistics for each variable. The appendix contains a description of each variable, its type, how it was constructed and source.

CO_2 emissions per capita, forming the dependent variable, is based on the data from CO_2 Information Analysis Center (CDIAC). CO_2 emissions are those stemming from the burning of fossil fuels and the manufacture of cement. These emissions also include CO_2 produced during consumption of solid, liquid, and gas fuels and gas flaring.

Income, as one of the independent variables, is based on the data from World Bank National Accounts and OECD National Accounts. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. GDP per capita is obtained by diving GDP by mid-year population. Higher income countries would have greater emissions to sustain economic development and activities, and vice versa.

The lowest and highest monthly average temperatures are based on the data from World Bank Climate Change Knowledge Portal (CCKP). The temperatures are the average of temperatures of a month, which are then compared against each other within the year to identify the lowest and the highest of the year. It is expected that cold countries would have greater heating demands while hot countries would have greater cooling demands. Cold countries and hot countries would therefore have higher CO_2 emissions.

The percentage of urban areas is based on the data from CIESIN Urban-Rural Population and Land Area Estimates and the Food and Agriculture Organization. The urban area is computed on a combination of population counts, settlement points, and the presence of nighttime lights. The numbers are then divided by the country total land area to obtain the percentage. Countries with less urban areas are sparsely inhabited and have higher transportation demands to move goods and people over long distances. Higher transportation demands would have higher emissions, and vice versa.

The percentage of renewable energy in total energy use is based on the data from World Bank Sustainable Energy for All (SE4ALL). Renewable resources encompass hydroelectric, geothermal, solar and wind resources as well as “fuel and waste”, which comprise biomass and animal products, gas/liquids from biomass, industrial waste, and municipal waste. It is expected that countries that have access to domestic renewable energy resources would have lower emissions than countries that lack such resources.

The percentage of fossil fuel consumed is based on the data from IEA Statistics. Fossil fuel comprises coal, oil, petroleum, and natural gas products. The higher the consumption, the fewer the reserve, vice versa. Countries that have fewer fossil fuel reserves should have lower CO_2 emissions than countries that are rich in such reserves. This is for two reasons: First, because of the emissions generated in the extraction and possibly the transport and processing of such resources. Second, because of countries that lacked major domestic fossil fuel reserves have had strong incentives to develop in a less fossil fuel intensive way to cut down on energy import.

TABLE 1 Descriptive Statistics

Variable	Variable Name	Description	Type	Mean	SD	Minimum	Maximum
Y	co2_emission	carbon dioxide emission per capita	continuous	3.440	3.740	0.041	9.881
x1	gdp_capita	US$ gross domestic product per capita	continuous	10,764	12,139	347	40,396
x2	gdp_capita2	squared US$ gross domestic product per capita	continuous	261,744,033	430,488,940	120,120	1,631,856,465
x3	lowest_temp	lowest monthly average temperature	continuous	15.754	9.755	-1.908	24..544
x4	highest_temp	highest monthly average temperature	continuous	25.807	2.497	20.688	30.336
x5	urban_area	percentage of urban land in total land area	continuous	10.247	11.907	0.517	29.816
x6	renewable_energy	percentage of renewable energy use in total energy use	continuous	43.679	35.481	3.568	97.740
x7	fossil_fuel	percentage of fossil fuel consumption in total energy use	continuous	54.392	32.842	2.245	94.633

5 RESULTS

Dependent Variable Test of Normality K-S test: The Kolmogorov-Smirnov test (K-S test) compares the data with a known distribution and tells us if they have the same distribution. It is commonly used as a test for normality to see if our data is normally distributed.

The hypotheses: H0: The data is normally distributed and H1: The data is not normally distributed. The K-S test p-value is less than α = 0.05, therefore did not fail to reject H0 concluding the data are not normally distributed. Again, this is expected given the dv_normality object was created via the rnorm() and co2 emission as runif() function:

```
        Two-sample Kolmogorov-Smirnov test

data:  dv_normality and co2_emission
D = 0.48, p-value = 0.0002033
alternative hypothesis: two-sided
```

Shapiro-Wilk test: The shapiro.test() function in R (program) employs the Shapiro-Wilk test (S-W test) on data to test whether the data are normally distributed. Use of the Shapiro-Wilk test is (less-sensitive) and contingent on univariate and continuous data.

The hypotheses: H0: The data is normally distributed and H1: The data is not normally distributed. The S-W test p-value is greater than α = 0.05, therefore failing to reject H0 concluding the data are normally distributed. Again, this is expected given the dv_normality object was created via the rnorm() function.

```
        Shapiro-Wilk normality test

data:  dv_normality
W = 0.99876, p-value = 0.7254
```

TABLE 2 Results of the dependent variable (DV) after performing the S-W test

Using the CO2_emission (dependent variable), we have performed the S-W Test using R					
Country	**W**	**p-value**	**alpha**	**Results**	**Comments**
Japan	0.9359	0.1072	0.05	p-value > alpha	Normally Distributed
Thailand	0.95715	0.3386	0.05	p-value > alpha	Normally Distributed
Vietnam	0.92054	0.04629	0.05	p-value < alpha	Not Normally Distributed
Ethiopia	0.86383	0.002671	0.05	p-value < alpha	Normally Distributed

TABLE 3 Results of the (square) dependent variable (DV) after performing the S-W test

By squaring the CO2_emission (dependent variable), we have performed the S-W Test using R					
Country	**W**	**p-value**	**alpha**	**Results**	**Comments**
Japan	0.94169	0.1474	0.05	p-value > alpha	Normally Distributed
Thailand	0.96402	0.4768	0.05	p-value > alpha	Normally Distributed
Vietnam	0.87207	0.003933	0.05	p-value < alpha	Not Normally Distributed
Ethiopia	0.74232	2.13E-05	0.05	p-value > alpha	Normally Distributed

While performing the S-W test on the dependent variable (DV = CO2_emission) we have observed that the p-value of Japan and Thailand is greater than the value of alpha (0.05). On the other hand, the p-value of Vietnam and Ethiopia is smaller than the value of alpha (0.05)

Again, by squaring the dependent variable (DV = CO2_emission) we have found after applying the S-W test; the p-value of Japan, Thailand and Ethiopia is greater than the value of alpha (0.05). However, the p-value of Vietnam is smaller than the value of alpha (0.05)

Univariate Analysis: Sample t-test: Paired two sample for Means among the 4 countries; Japan, Thailand, Vietnam, Ethiopia in our dataset.

TABLE 4 Japan: mean (9.391375457) > Thailand: mean (3.103617)

t-Test: Paired Two Sample for Means

	Japan	*Thailand*
Mean	9.391375457	3.103617
Variance	0.108738244	0.568407
Observations	26	26
Pearson Correlation	0.447874747	
Hypothesized Mean Difference	0	
df	25	
t Stat	47.55965533	
P(T<=t) one-tail	2.41779E-26	
t Critical one-tail	1.708140761	
P(T<=t) two-tail	4.83558E-26	
t Critical two-tail	2.059538553	

TABLE 5 Japan: mean (9.391375457) > Vietnam: mean (0.979067422)

t-Test: Paired Two Sample for Means

	Japan	*Vietnam*
Mean	9.391375457	0.979067422
Variance	0.108738244	0.299928883
Observations	26	26
Pearson Correlation	0.216847863	
Hypothesized Mean Difference	0	
df	25	
t Stat	74.63079834	
P(T<=t) one-tail	3.35339E-31	
t Critical one-tail	1.708140761	
P(T<=t) two-tail	6.70678E-31	
t Critical two-tail	2.059538553	

TABLE 6 Japan: mean (9.391375457) > Ethiopia: mean (0.070484)

t-Test: Paired Two Sample for Means		
	Japan	*Ethiopia*
Mean	9.391375457	0.070484
Variance	0.108738244	0.000553
Observations	26	26
Pearson Correlation	0.06671267	
Hypothesized Mean Difference	0	
df	25	
t Stat	144.4497966	
P(T<=t) one-tail	2.35866E-38	
t Critical one-tail	1.708140761	
P(T<=t) two-tail	4.71732E-38	
t Critical two-tail	2.059538553	

TABLE 7 Thailand: mean (3.103616614) > Vietnam: mean (0.979067422)

t-Test: Paired Two Sample for Means		
	Thailand	*Vietnam*
Mean	3.103616614	0.979067422
Variance	0.568406843	0.299928883
Observations	26	26
Pearson Correlation	0.937013138	
Hypothesized Mean Difference	0	
df	25	
t Stat	35.22874312	
P(T<=t) one-tail	3.93694E-23	
t Critical one-tail	1.708140761	
P(T<=t) two-tail	7.87387E-23	
t Critical two-tail	2.059538553	

TABLE 8 Thailand: mean (3.103616614) > Ethiopia: mean (0.070484)

t-Test: Paired Two Sample for Means		
	Thailand	*Ethiopia*
Mean	3.103616614	0.070484
Variance	0.568406843	0.000553
Observations	26	26
Pearson Correlation	0.73200966	
Hypothesized Mean Difference	0	
df	25	
t Stat	20.98822351	
P(T<=t) one-tail	1.0811E-17	
t Critical one-tail	1.708140761	
P(T<=t) two-tail	2.1622E-17	
t Critical two-tail	2.059538553	

TABLE 9 Vietnam: mean (0.979067422) > Ethiopia: mean (0.070484)

t-Test: Paired Two Sample for Means		
	Vietnam	*Ethiopia*
Mean	0.979067422	0.070484373
Variance	0.299928883	0.000552898
Observations	26	26
Pearson Correlation	0.858473926	
Hypothesized Mean Difference	0	
df	25	
t Stat	8.780893293	
P(T<=t) one-tail	2.06402E-09	
t Critical one-tail	1.708140761	
P(T<=t) two-tail	4.12804E-09	
t Critical two-tail	2.059538553	

Japan: mean (H_0 = 9.391375457) > Thailand: mean (H_1 = 3.103617)
Japan: mean (H_0 = 9.391375457) > Vietnam: mean (H_1 = 0.979067422)
Japan: mean (H_0 = 9.391375457) > Ethiopia: mean (H_1 = 0.070484)
Thailand: mean (H_0 = 3.103616614) > Vietnam: mean (H_1 = 0.979067422)
Thailand: mean (H_0 = 3.103616614) > Ethiopia: mean (H_1 = 0.070484)
Vietnam: mean (H_0 = 0. 979067422) > Ethiopia: mean (H_1 = 0.979067422)

TABLE 10 Comparison of mean of the dependent variable (DV) among 4 countries

Country (mean)	**H_0**	**H_1**
Japan > Thailand	9.391375457	3.103617
Japan > Vietnam	9.391375457	0.979067422
Japan > Ethiopia	9.391375457	0.070484
Thailand > Vietnam	3.103616614	0.979067422
Thailand > Ethiopia	3.103616614	0.070484
Vietnam > Ethiopia	0. 979067422	0.979067422

What we have observed from the aforementioned table data is that the mean of the high-income countries is higher than the mean of the middle-income/ lower-middle-income/ lower-income countries.

Confidence Interval: We have constructed the 95% confidence interval between the largest mean and smallest mean data. At first, we have taken the largest mean data from the high-income country (Japan). Then we have taken the smallest mean data from the low-income country (Ethiopia). We had to find the differences between the largest and the smallest mean of the data. Here, confidence coefficient = 0.95 and alpha = 0.05. Then we have found the standard deviation of the sample mean and thus using the data, found the margin of error.

TABLE 11 Constructing the 95% confidence intervals around the largest mean and the smallest mean

(Largest) Mean - Japan	**(Smallest) Mean - Ethiopia**	**Difference**					
9.391375457	3.103616614	6.287759		(Largest) Mean	9.320891	Sample SD	4.206154
9.391375457	0.979067422	8.412308		(Smallest) Min	0.908583	Margin of Error	5.829324914
9.391375457	0.070484373	9.320891		(Largest) Mean LB (Mean - E)	3.491566	confidence coefficient	0.95
3.103616614	0.979067422	2.124549		(Largest) Mean UB (Mean + E)	15.15022	alpha	0.05
3.103616614	0.070484373	3.033132		(Smallest) Min LB (Mean - E)	-4.92074		
0.979067422	0.070484373	0.908583		(Smallest) Min UB (Mean + E)	6.737908		
	(Largest) Mean - Japan	**9.32089**		**(Smallest) Mean - Ethiopia**	**0.90858**		
	Sample SD	3.48848		Sample SD	3.48848		
	Margin of Error	3.660936		Margin of Error	3.660936		

OLS Regression Model: We have used the RegressIt – an excel macro tool to execute the regression model keeping CO2_emission as our dependent variable and fossil_fuel, gdp_capita, gdp_capita2, highest_temp, lowest_temp, renewable_energy, and urban_area as the independent variable. After executing the regression, we have found that the value of R-Squared and Adj. R-Sqr. was = 0.998. The p-value for gdp_capita, gdp_capita2, lowest_temp, and urban_area was 0. At the same time, we have found the coefficient value for gdp_capita2, highest_temp, lowest_temp was a negative value.

TABLE 12 Using RegressIt to generate the Model_1 based on the dependent variable (DV) = CO_2_emission

Model:	Regression							
Dependent Variable:		co2_emission						
	R-Squared	**Adj.R-Sqr.**	**Std.Err.Reg.**	**Std.Dep.Var.**	**# Fitted**	**# Missing**	**Critical t**	**Confidence**
	0.998	0.998	0.181	3.740	100	4	1.986	95.0%
Variable	**Coefficient**	**Std.Err.**	**t-Statistic**	**P-value**	**Lower95%**	**Upper95%**	**VIF**	**Std. Coeff.**
Constant	0.009380	1.095	0.009	0.993	-2.166	2.184	0.000	0.000
fossil_fuel	0.015	0.009280	1.660	0.100	-0.003029	0.034	280.701	0.135
gdp_capita	0.000295	0.000015	19.538	0.000	0.000265	0.000325	101.212	0.956
gdp_capita2	-4.766E-09	2.981E-10	-15.986	0.000	-5.358E-09	-4.174E-09	49.776	-0.549
highest_temp	-0.034	0.026	-1.337	0.185	-0.085	0.017	12.302	-0.023
lowest_temp	-0.035	0.011	-3.037	0.003	-0.057	-0.012	37.162	-0.090
renewable_energy	0.013	0.010	1.289	0.201	-0.007050	0.033	389.579	0.124
urban_area	0.149	0.016	9.195	0.000	0.117	0.181	112.320	0.474
	Mean Error	**RMSE**	**MAE**	**Minimum**	**Maximum**	**MAPE**	**A-D* stat**	
Fitted (n=100)	0.000	0.174	0.130	-0.919	0.366	32.1%	1.00 (P=0.012)	

Residual Analysis:

FIGURE 2 Residual -vs- Predicted of co2_emission (dependent variable)

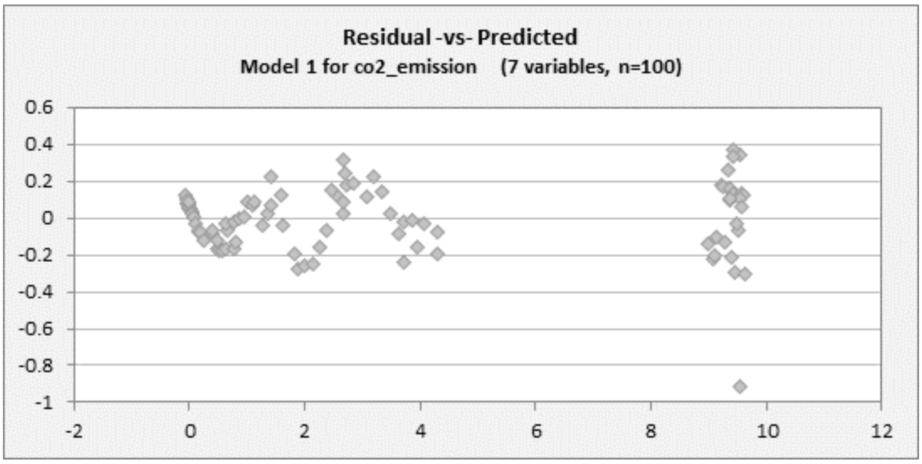


FIGURE 3 Histogram of Residuals for co2_emision (dependent variable)

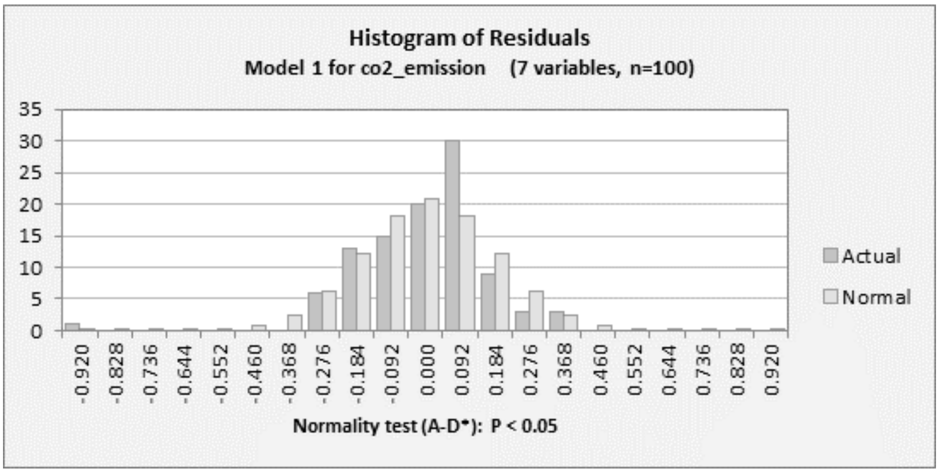


FIGURE 3 Normal Quantile Plot for co2_emission (dependent variable)

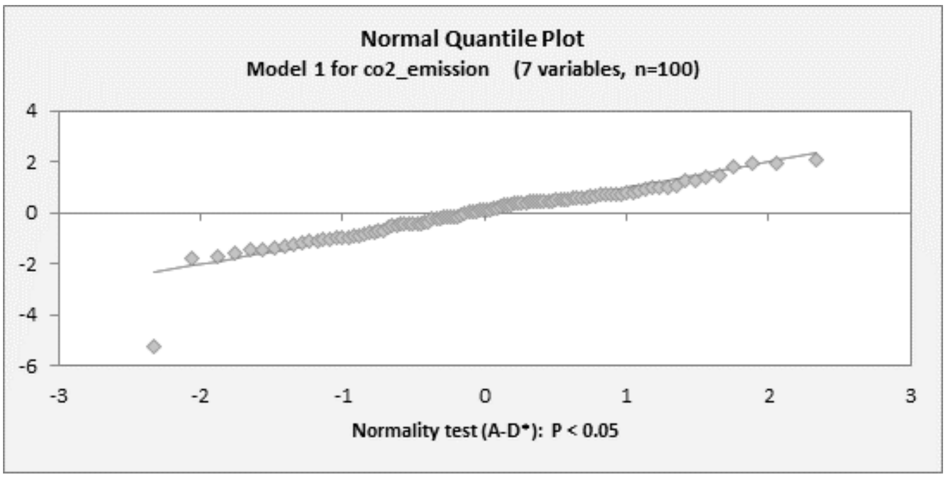


FIGURE 4 Residual -vs- fossil fuel for co2_emission (dependent variable)

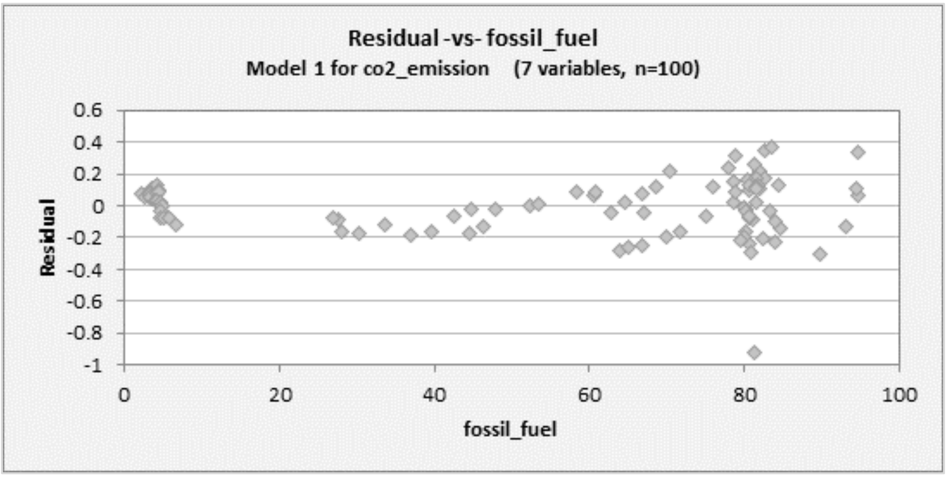


FIGURE 5 Residual -vs- gdp_capita for co2_emission (dependent variable)

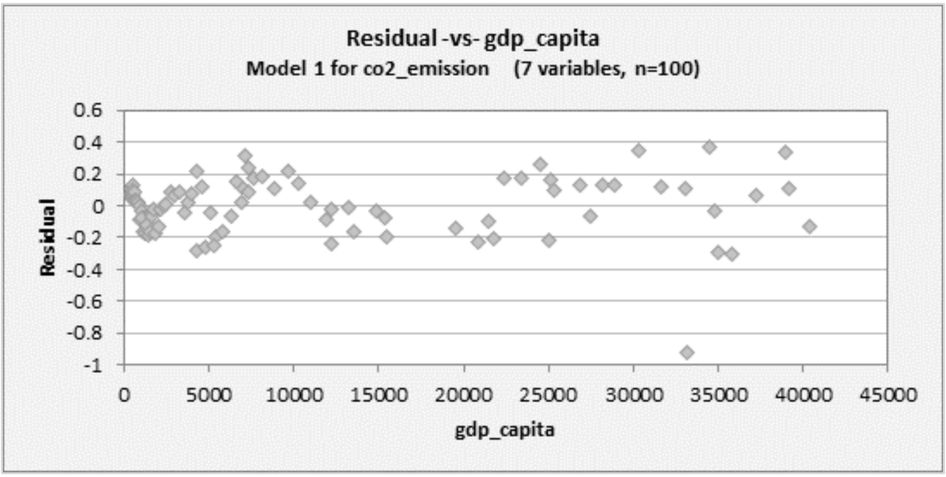
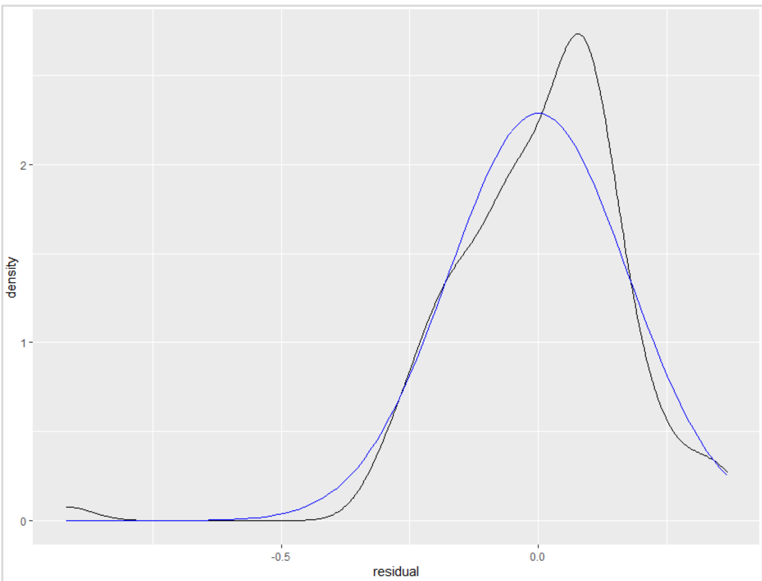
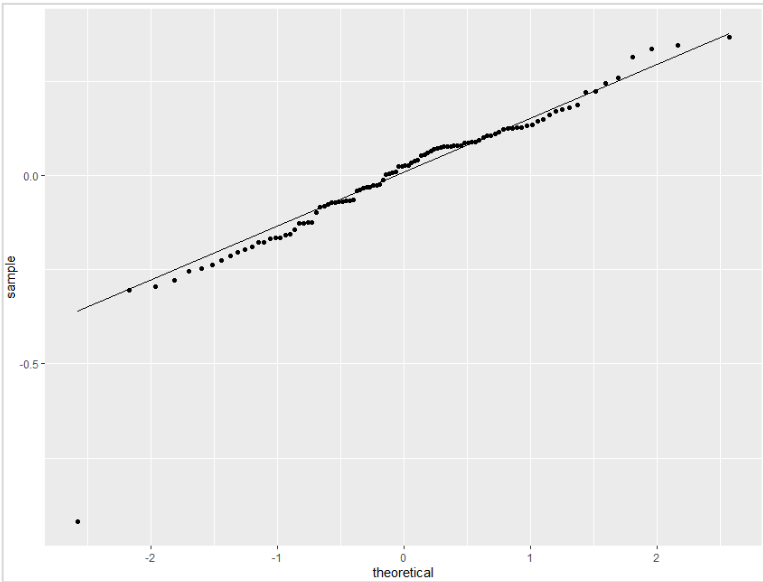


FIGURE 6 Density plot of the residuals



The density plot in FIGURE 6 provides a visual judgment about whether the distribution is bell shaped. The black line represents the residual data, and the blue line represents the normal distribution given the μ and σ values of residual. Note: The residual data appears as bell shaped, and does fit the normal distribution model given the parameters calculated via the residual data.

FIGURE 7 G-G plot of the residuals



Most of the points in the QQ plot fall inside the region defined by the dashed lines, further suggesting the residual data is likely normally distributed. Lastly, a Shapiro-Wilk test can confirm whether the residual data is normally distributed:

Shapiro-Wilk normality test
data: d3_residual$residual
W = 0.91241, p-value = 5.696e-06

```
Shapiro-Wilk normality test
data:  d3_residual$residual
W = 0.91241, p-value = 5.696e-06
```

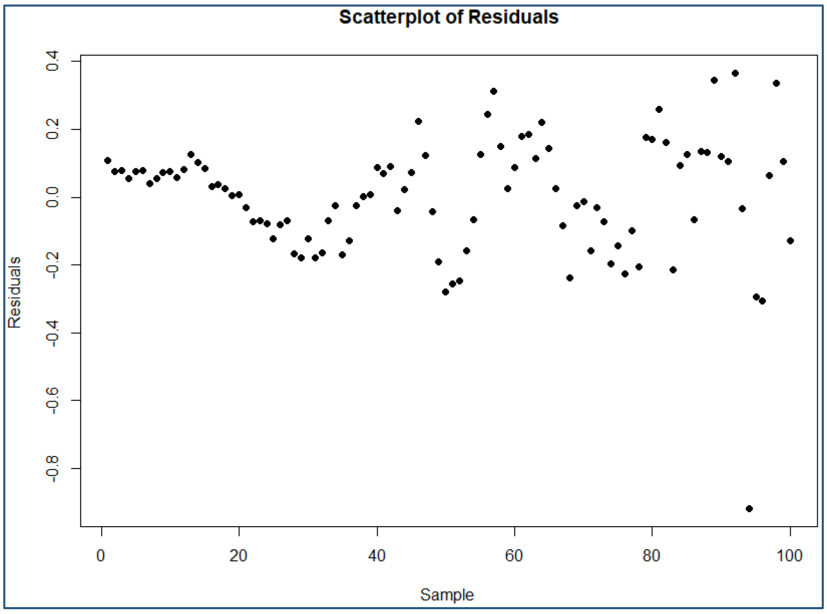
describe(d3_residual$residual)

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	100	0	0.17	0.02	0.01	0.14	-0.92	0.37	1.29	-1.35	6.04	0.02

The Shapiro-Wilk test p-value is less than α = 0.05, leading to reject H0: data are normally distributed. In conclusion, the residual data is not normally distributed. **Note:** Though the visual plots are likely enough to confirm the residual data are normally distributed however, based on the p-value of residual data we can conclude that the data is not normal.

Check for heteroskedasticity: There seems to be no evident pattern. However, it does seem to look as if there's more variation in residuals in this sample data from the linear regression.

FIGURE 8 Scatterplot of Residuals



6 IMPLICATIONS AND CONCLUSION

From the data and discussions, we can say that in natural factors can explain cross-country differences in CO_2 emissions only to some limited extent only. Also, Countries having colder climate, or a lower availability of renewable resources have higher fossil fuel consumption than countries with warmer climates or a higher availability of renewable resources. We have seen that there is a relationship between CO_2 emissions with GDP growth and fossil fuels. As the value of GDP, urban area, renewable energy increases; CO_2 emissions decreases. Besides, countries with having low natural conditions (temperature) demanded higher carbon emissions than countries at roughly the same income levels.

This does not mean that for individual countries natural factors cannot play an important role in determining their CO_2 emissions. The cold, big, fossil fuel rich, but renewable resource poor Soviet Union had of course higher emission requirements than warm, comparatively small, fossil fuel poor, but renewable resource rich Ethiopia, for example.

References

Bhattacharya M, Churchill S A, Paramati S R (2017). 'The dynamic impact of renewable energy and institutions on economic output and CO_2 emissions across regions'. Renewable Energy 111, Pp 157-167

Hu Z G, Yuan J H, Hu Z (2011). 'Study on China's low carbon development in an Economy Energy-Electricity-Environment framework.' Energy Policy 39 (5), Pp 2596–2605

Kerkhof C A, Benders M J R, Moll C H (2009), 'Determinants of variation in household CO_2 emissions between and within countries' Energy Policy 37 (4), Pp 1509-1517

Li S, Zhou C, Wang S (2019). 'Does modernization affect CO_2 emissions? A panel data analysis' Science of the Total Environment 663, Pp 426-435

Moutinho V, Madaleno M, Margarita R (2017), 'The economic and environmental efficiency assessment in EU cross-country: Evidence from DEA and quantile regression approach' Ecological Indicators 78, Pp 85-97

Nathanael A. Heckert, James J. Filliben, C M. Croarkin, B Hembree, William F. Guthrie, P Tobias, J Prinz (2002). 'Handbook 151: NIST/SEMATECH e-Handbook of Statistical Methods' [Publication] Published November 1, 2002.

National Aeronautics and Space Administration (2008/2020). '"Global Climate Change: Vital Signs of the Planet' [Online Resource]. https://climate.nasa.gov/carbon-dioxide/

Neumayer, E (2002). 'Can Natural Factors Explain Any Cross-Country Differences in CO_2 Emissions?' Energy policy, 30 (1). Pp. 7-12.

Ritchie, H and Roser, M (2017). 'CO_2 and Greenhouse Gas Emissions' [Online Resource]. Our World in Data. https://ourworldindata.org/CO2-and-other-greenhouse-gas-emissions

United Nations Framework Convention on Climate Change (1998). 'Kyoto Protocol to the United Nations Framework Convention on Climate Change' [Official Document] http://unfccc.int/resource/docs/convkp/kpeng.pdf

United Nations (2015). 'Paris Agreement' [Online Resource]. https://unfccc.int/process-and-meetings/the-paris-agreement/theparisagreement

Tucker M (1995). 'Carbon dioxide emissions and global GDP' [Online Resource]. Ecological Economics 15 (3). Pp. 215-223.

World Health Organization (2018). 'Climate Change and Health' [Online Resource]. https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health

World Bank, World Development Indicators. (2016). 'CO_2 emissions (metric tons per capita)' [Data file]. https://data.worldbank.org/indicator/EN.ATM.CO2E.PC

World Bank, World Development Indicators. (2019). 'GDP per capita (current US$)' [Data file]. https://data.worldbank.org/indicator/NY.GDP.PCAP.CD

World Bank, World Development Indicators. (2016). 'Climate Change Knowledge Portal' [Data file]. https://climateknowledgeportal.worldbank.org/download-data

World Bank, World Development Indicators. (2013). ‘Urban land area (sq.km)’ [Data file]. https://data.worldbank.org/indicator/AG.LND.TOTL.UR.K2

World Bank, World Development Indicators. (2018). ‘Land area (sq.km)’ [Data file]. https://data.worldbank.org/indicator/AG.LND.TOTL.K2

World Bank, World Development Indicators. (2015). ‘Renewable energy consumption (% of total final energy consumption)’ [Data file] https://worldbank.org/EG.FEC.RNEW.ZS

World Bank, World Development Indicators. (2015). ‘Fossil fuel energy consumption (% of total final consumption)’ [Data file]. https://worldbank.org//EG.USE.COMM.FO.ZS

Zhang N, Yu K, Chen Z (2017). ‘How does urbanization affect CO_2 emissions? A cross-country panel data analysis’ Energy Policy 107. Pp 678-687

APPENDIX A: DESCTIPTION OF VARIABLES

Table A1 contains the name of each variable, the description, the type, and its source.

TABLE A1 Description of Variables

Variable Name	Description	Type	Source	Website
country	countries under analysis	categorical	–	–
country_code	country code	categorical	–	–
year	calendar year	integer	–	–
co2_emission	carbon dioxide emission per capita	continuous	The World Bank	https://data.worldbank.org/indicator/EN.ATM.CO2E.PC
gdp_capita	US$ gross domestic product per capita	continuous	The World Bank	https://data.worldbank.org/indicator/NY.GDP.PCAP.CD
gdp_capita2	squared US$ gross domestic product per capita	continuous	The World Bank	https://data.worldbank.org/indicator/NY.GDP.PCAP.CD
lowest_temp	lowest monthly average temperature	continuous	The World Bank	https://climateknowledgeportal.worldbank.org/download-data
highest_temp	highest monthly average temperature	continuous	The World Bank	https://climateknowledgeportal.worldbank.org/download-data
urban_area	percentage of urban land in total land area	continuous	The World Bank	https://data.worldbank.org/indicator/AG.LND.TOTL.UR.K2
renewable_energy	percentage of renewable energy use in total energy use	continuous	The World Bank	https://data.worldbank.org/indicator/EG.FEC.RNEW.ZS
fossil_fuel	percentage of fossil fuel consumption in total energy use	continuous	The World Bank	https://data.worldbank.org/indicator/EG.USE.COMM.FO.ZS