1. The chosen country is China, and, considering data availability, the period considered will be from 1960 to 2014.
   The first variable, considered dependant (Y) in this subject, will be the annual carbon dioxide emissions of the country. This data is in annual kilo tonne.
   The two other variables, the independents one (Xs), are annual GDP at purchaser's price of the country, and the total population of the country.
   All of these data are from the World Development Indicators database, from the world bank (*databank.worldbank.org*)

2.

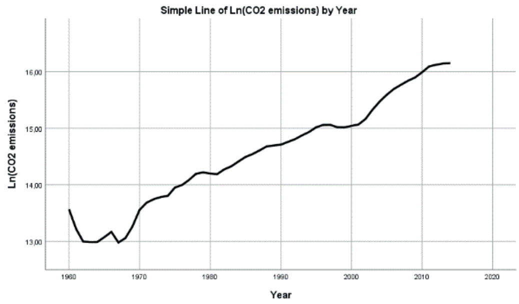

Figure 1 Natural logarithm of annual CO2 emissions in China from 1960 to 2014
Source : *databank.worldbank.org*

This figure shows a general growing trend, despite some crashes. It appears to be non-stationary.

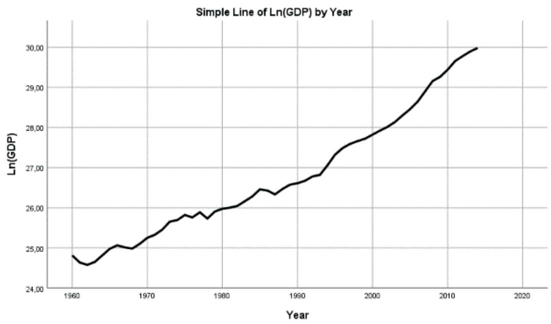

Figure 2 Natural logarithm of annual GDP in China from 1960 to 2014
Source : *databank.worldbank.org*

This figure shows a general growing trend, despite some variations. It appears to be non-stationary.
There are some similarities in the trend between GDP and CO2 emissions.

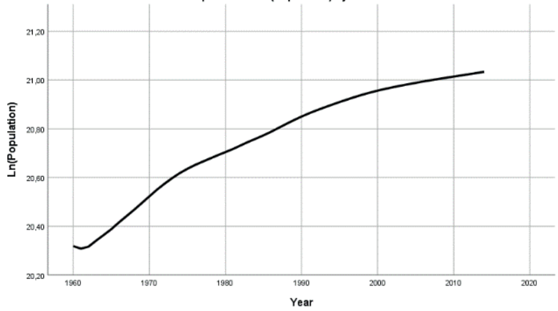

Figure 3 Natural logarithm of population in China from 1960 to 2014
Source : *databank.worldbank.org*

This figure shows an almost linear growing trend. Still, it appears to be non-stationary.

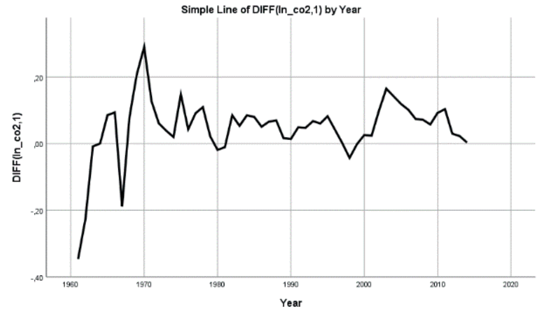

Figure 4 First difference of natural logarithm of annual CO2 emissions in China from 1960 to 2014
Source : computed from data from databank.worldbank.org

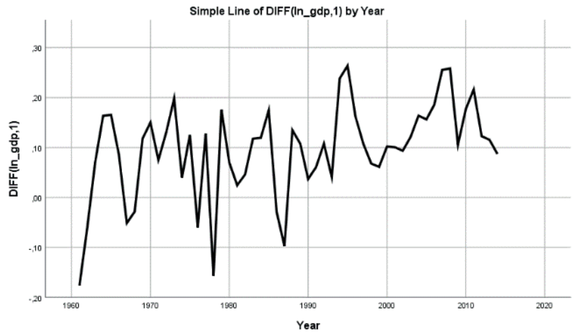

Figure 5 First difference of natural logarithm of annual GDP in China from 1960 to 2014
Source : computed from data from databank.worldbank.org

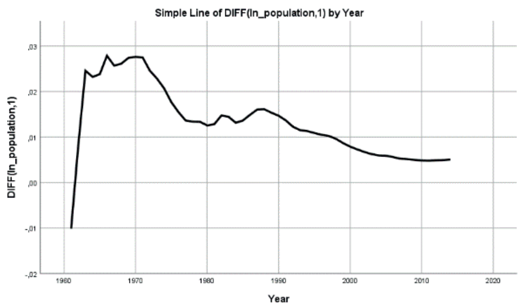

Figure 6 First difference of natural logarithm of population in China from 1960 to 2014
Source : computed from data from databank.worldbank.org

In each of these graphs, we can observe a general growing trend over the time : none of them can be considered as stationary.

3. $\ln Co2_t = \beta 0 + \beta 1 \ln GDP_t + \beta 2 \ln Pop_t + u_t$

4. We should use the Dickey-Fuller test to detect unit root in each of these variables.
The considered variables are fluctuating around a linear trend, not around 0 or an average value : the appropriate Dickey-Fuller test is for random walk with drift and a deterministic trend.
We must estimate the following model for each variable :

$$\Delta \ln(X_t) = \beta 1 + \beta 2 + \beta 3 \ln(X_{t-1}) + u_t$$

Doing this, we find the following results:

	Lag of CO2 emissions	Lag of GDP	Lag of population
t statistic	-4,536	-0,978	0,738

We should now compare these results with those found in the Dickey-Fuller table. We should use the one considering drift and deterministic trend. Since our n value is between 55 and 58 for the three variables, we will consider n=50 for reading the table. We will consider a level of significance of 95%.

From the table, we find that the critical value is -3,5.

The t value for GDP and population is greater than the critical value from the table. It means that we can't reject that both variables are non-stationary. The t value for CO2 emissions is lower than the critical value, meaning that, at 95% of significance, we can reject that it is non-stationary.

5. Since GDP and population are non-stationary variables, it is not recommended to run an OLS regression. However, we need to run it in order to obtain residuals and check for co integration.
After obtaining the residuals, we need to perform a Dickey-Fuller on it.
Since residuals are fluctuating around 0, we should use the Dickey-Fuller test without trend and drift, so we should estimate the following model for the residuals :

$$\Delta\ln(X_t) = \beta 3\ln(X_{t-1}) + u_t$$

Doing this, we obtain the following t value : -4,39.
Considering the Dickey-Fuller table without trend or drift, at a 95% significance level, the critical value is -1,95.
The obtained t value is smaller than the critical value, we can reject the null hypothesis of a unit root in the residuals, meaning that the variables are co integrated : the linear combination of these variables is stationary, we can use regression on these variables: it's not spurious.
The results are :

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-32,308	3,778		-8,553	,000
	Ln(GDP)	,354	,032	,581	11,172	,000
	Ln(Population)	1,800	,219	,428	8,225	,000

a. Dependent Variable: Ln(CO2 emissions)

6. We should now perform Dickey-Fuller test on the variables in first difference. As explained in the question 2), all of the three variables in first difference seems to follow a general growing trend, so the appropriate test is Dickey-Fuller with trend and drift.
Again, we need to estimate the following model for each variable in first difference:

$$\Delta\ln(X_t) = \beta 1 + \beta 2 + \beta 3\ln(X_{t-1}) + u_t$$

We obtain the following values :

	Lag of lag of CO2 emissions	Lag of lag of GDP	Lag of lag of population
t statistic	0,210	0,045	-2,955

The critical value is still -3,5 so even in first difference, none of these variables are stationary.

We estimate the following model :

$$\Delta\ln(CO2_t) = \beta 0 + \beta 1\Delta\ln(GDP_t) + \beta 3\Delta\ln(pop_t) + u_t$$

We obtain the following results :

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-,055	,024		-2,310	,025
	DIFF(ln_gdp,1)	,502	,111	,505	4,502	,000
	DIFF(ln_population,1)	4,096	1,345	,342	3,045	,004

a. Dependent Variable: DIFF(ln_co2,1)

But estimating our model in first differences if they are non-stationary can be spurious. We should first check for co integration.

We run a Dickey-Fuller test with no trend and no drift for the saved residuals.
The t value obtained is -4,958. The value is lower than the critical value from the table, variables in first differences are co integrated.

7. The model obtained through the question 5 (in level) is better. Running a model in first difference is less precise than running one in level and should only be used if the variables in level are non-stationary and no cointegrated. But we showed before that the variables in level are co integrated : that's why we should use the OLS model in level.
The model gives us this estimator :

$$\ln(CO2_t) = -32{,}308 + 0{,}354\ln(GDP_t) + 1{,}8\ln(pop_t)$$

Considering that for each coefficient, the significance level is around 0, all are statistically significant.
This equation is a good estimation of the model, even if a lot of independent variables are probably missing.