CHAPTER - 3
METHODOLOGY

The methodology adopted in the study on "Relationship between Greenhouse gas emissions, Trade Openness and Gross Domestic Product- A Comparative Study between India and China", are discussed as under.

3.1. Selection of the problem

3.2. Selection of data

3.2.1 Model Specification

3.2.2 Sources of data

3.2.3 Definition of the variables used in the model

3.3. Tools used

3.3.1Annual Growth rate

3.3.2 Compound Annual Growth rate

3.3.3Decoupling Analysis

3.3.4Unit root test

3.3.5Johansen's Cointegration Test

3.3.6 Vector Error Correction Model (VECM)

3.3.7 Diagnostic tests

3.3.8 Granger causality test

3.3.9 Variance decomposition analysis

3.3.10 Impulse Response Function

3.3.11 Testing the Stability

3.4. Limitations of the study

3.1. Selection of the Problem

India and China are the two major emerging economies of Asia which are in the distinct stages of structural change, technical innovation, trade, energy use and growth (Fan and Hossain, 2018). Sustainable economic growth and development is very much desirable to get better social wellbeing. Every economy has the objective to attain the desired stage of growth and development for a long term without harming environment quality. Both the countries selected for the study have achieved noteworthy economic development in the last few decades. GDP of both collectively account for 25 percent of the global GDP (World Bank databank, 2020). Trade openness is no doubt a primary driving force behind country's development and it is true for India and China too. Faster Economic growth of both the countries has led to the consumption of energy which is projected to be 32 percent in 2035 (Energy Information Administration, 2011). This increase in energy consumption drastically increases the pollution levels. India's and China's contribution to the world's emission increased to 33 percent in 2018(World Bank databank, 2018). In the process of industrialization, these two countries give priority to manufacturing industry which consumes more energy resulting in high carbon dioxide emissions. Further development requires considerably more energy than the countries currently consume. India and China will quicken environmental degradation due to high amount of carbon dioxide emission. "It is inequitable to ask developing countries to sacrifice growth for the sake of reducing emissions" (Truong, 2010). These issues have led to intense debates all over the world that led countries to take steps to reduce greenhouse gas emissions.

Global warming and climate change have paved way for revival of interest to test the nexus between greenhouse gas emissions per capita, trade openness and gross domestic product per capita for India and China during the period 1970 to 2018.

3.2. Selection of Data

3.2.1 Model Specification

The theoretical framework of this thesis is based on Environmental Kuznets Curve (EKC) which means there is an inverted-U shaped relationship between economic growth and the environment, as environmental degradation first increase, then level off and decrease with economic growth. As per the past studies not all environmental problems follow this inverted-U curve model. Since early 1990s, EKC hypothesis has been an important theory that explains the link between growth and environmental degradation. Abundant EKC literature is available that test for linear, as well as quadratic and cubic relationships between per capita income and CO_2 emissions (Soytas et al., 2007). Considering environmental degradation measure(s) as the dependent variable and income as the independent variable, studies provide mixed results (Hill and Magnani, 2002, Dinda, 2004, Copeland and Taylor, 2004 and Stern,2004).

In the present study, the relationship between the greenhouse gas emissions per capita, trade openness and gross domestic product per capita was analysed. The earlier studies (Zhou et al. 2018, Shahbaz et al., 2016, Dogan et al., 2015, Lu, 2017, and Jian et al., 2019) analysed the relationship among carbon emissions, trade openness, energy consumption and financial development. In the above studies, the carbon emission alone was considered as the greenhouse gas emission. But this study has included data which includes all the important gases which causes pollution and harm the environment. Based on the earlier studies, foreign direct investment was included in the study, but due to the non-availability of sufficient data for the study period, the variable was dropped. Finally, the relationship between environmental degradation and economic growth in India and China was analysed based on the greenhouse gas emissions per capita, trade openness and gross domestic product per capita alone. The economic growth was measured in terms of gross domestic product per capita at constant prices in US $ billions, greenhouse gas emissions per capita in kt of CO_2 equivalent and trade openness as percentage of gross domestic product per capita.

In this study, the relationship between the greenhouse gas emissions per capita, trade openness and gross domestic product per capita in the long run was analysed based on the long period data. The ordinary linear regression models could not be specified to estimate the long run relationship between these variables if they are non-stationary. As non-stationary variables included in the ordinary linear regression model leads to spurious regression relationship, it was essential to test the stationarity of greenhouse gas emissions per capita and its determinants, gross domestic product per capita and its determinants, trade openness and its determinants before assessing the relationship among the variables. When the variables are stationary at first difference, the cointegration regression and the error correction models are the models to assess the long run relationship between the greenhouse gas emissions per capita, trade openness and gross domestic product per capita as in the earlier studies (Ozturkand Uddin, 2012, Ang, 2007). Therefore in the current study, the stationarity of the variables were tested by using the unit root test.

When the unit root test was applied to test the stationarity of greenhouse gas emissions per capita, trade openness and gross domestic product per capita, they were non stationary at zero level and became stationary at first order difference. Therefore, the co integration analysis and vector error correction model were specified to assess the relationship between greenhouse gas emissions per capita, trade openness and gross domestic product per capita. To assess the direction of relationship and the causality between greenhouse gas emissions per capita, trade openness and gross domestic product per capita, granger causality test was applied. The decoupling analysis was performed to understand the nature of decoupling over the years in India and China. Variance decomposition analysis showed the percentage of forecast error variance for the chosen variables that point to its own shocks, and fluctuations in other variables. Finally, the stability of the gross domestic product per capita model, trade openness model and greenhouse gas emissions per capita model(Error correction model) was tested based on the CUSUM and CUSUM square test. The specifications of econometric models are as under

In the first model, gross domestic product per capita was the dependent variable influenced by gross fixed capital formation per capita as a proxy for investment, trade openness, inflation, population, energy consumption per capita and greenhouse gas emissions per capita. Though the first few factors have abundant literature (Hasson and Masih, 2017,Cetin et al., 2018, Osiobe, 2019) establishing the relationships, the last variable greenhouse gas emissions per capita's influence on gross domestic product per capita of a country has been studied by very few researchers (Azam, et al., 2015).

Model I

$$\mathbf{LGDPpc_{it} = \beta_{oi} + \beta_{1i}LGFCFpc_{it} + \beta_{2i}TO_{it} + \beta_{3i}INF_{it} + \beta_{4i}POP_{it} + \beta_{5i}LENpc_{it} + \beta_{6i}LGHGpc_{it} + \varepsilon_{it}}$$

Where,

LGDP pc is log of gross domestic product per capita at constant prices in USD (proxy for Economic Growth)

LGFCF pc is log of gross fixed capital formation per capita at constant prices in USD(proxy for Investment)

TO is the trade openness in percentage

INF is the inflation rate in percentage

POP is the population growth of the country in percentage

LEN pc is the log of energy consumption per capita in kg of oil equivalent

LGHG pc is the log of greenhouse gas emissions per capita in kt of CO_2 equivalent

εt is the error term

In the second model, trade openness is the dependent variable influenced by Gross domestic product per capita, gross fixed capital formation per capita**,** exchange rate and population.

Model - II

$$TO_{it} = \beta_{0i} + \beta_{1i}LGDPpc_{it} + \beta_{2i}LGFCFpc_{it} + \beta_{3i}ER_{it} + \beta_{4i}POP_{it} + \varepsilon_{it}$$

Where

TO is the trade openness in percentage

LGDP pc is log of gross domestic product per capita at constant prices in USD (proxy for economic growth)

LGFCF pc is log of gross fixed capital formation per capita in USD (proxy for Investment)

ER is the exchange rate in percentage

POP is the population growth of the country in percentage

εt is the error term

In the third model, green house gas emissions per capita is the dependent variable influenced by gross domestic product per capita, gross domestic product per capita square, energy consumption per capita and trade openness.

Model- III

$$LGHGpc_{it} = \beta_{0it} + \beta_{1i}LGDPpc_{it} + \beta_{2i}LGDPpc^2_{it} + \beta_{3i}LENpc_{it} + \beta_{4i}TO_{it} + \varepsilon t$$

Where

LGHG pc is the log of greenhouse gas emissions per capita expressed in kilo tonnes of CO_2 equivalent

LGDP pc is log of gross domestic product per capita at constant prices in USD (proxy for economic growth)

LGDP pc square is the log of square of gross domestic product per capita at constant prices in USD.

LEN is the log of energy consumption per capita in kg of oil equivalent.

TO is the trade openness in percentage

εt is the error term

3.2.2 Sources of data and period of the study

The data for the study were secondary in nature. For the selected variables viz., GDP per capita, GHG emissions per capita, GFCF per capita, trade openness, inflation, population, energy consumption per capita and exchange rate, data were obtained from the World Bank data bases, Emissions database for global atmospheric research, Our world in data and World resources institute websites for the period 1970 to 2018.The variables such as gross domestic product per capita, greenhouse gas emissions per capita, gross fixed capital formation per capita and energy consumption per capita are taken in the natural logarithm form considering the econometric estimation of the model. The other variables such as trade openness, inflation, population, exchange rate are expressed in percentages. The greenhouse gas emission per capita had started to increase from the year 1970 onwards and it became global issue only from the year 1970 onwards. Therefore, the study period was confined to 1970 -2018.

3.2.3 Variables used in the model

In order to assess the impact of macro variables on the greenhouse gas emissions per capita in India and China, the study used annual data from 1970 to 2018 on selected macroeconomic indicators.

GHG Emissions per capita

Carbon dioxide, methane, nitrous oxide and water vapor (which all occur naturally) and fluorinated gases (which are synthetic) are the gases causing greenhouse effect by trapping heat in the atmosphere and warming the planet.

GDP per capita

GDP per capita is an indicator of economic performance and helps in comparing average living standards and economic wellbeing in different countries.

Gross Fixed Capital Formation per capita

Gross Fixed Capital Formation otherwise known as investment is used as an independent variable determining GDP of a country (Gibescu 2010).Gross fixed capital formation constitutes a substantial share to the total final demand of goods and services in national accounting(Soderstern,2017).

Energy consumption per capita

Energy consumption is important in the economy, both in supply and demand. Gross domestic product and energy consumption can often correlate. Countries with higher GDP consume more energy as they produce more goods because of industrialization.As well, countries that boast a higher GDP are likely to be more populated (the prime examples being India, China, and the US) which also corresponds to more energy use in that country (Donev et al., 2020). Rising demand for energy kindles growth, but it causes greenhouse gas emissions (Lu, 2017).

Trade openness

Trade openness is an important driver of economic activity which affects carbon emissions. A country’s trade openness is measured by the yearly ratio of the sum of its export and imports scaled by a country’s GDP to account for the size of the economy (Gleditsch, 2002 and Frankel, 2003).

Population

Population growth increases the demand for goods and services which means more environmental pressure. At the same time, economic growth, which is regarded as the best way of reducing the rate of population growth, is also associated with severe degradation of natural resources. So far, in developing countries the mechanisms of economic growth have been unable to reduce both the demographic pressure and the economic inequalities, but, at the same time, these mechanisms tend to strengthen the factors of environmental degradation resulting from economic activity. This is called as the paradox of economic growth (Beato and Chiarello, 2000).

Inflation

High level of inflation upsets the smooth functioning of a market economy (Krugman, 1995).Several studies have confirmed the existence of relationship between Inflation and Economic Growth (Barro, 1995; Ahmed and Mortaza, 2005; Sweiden, 2004).

Exchange Rate

A strong exchange rate is often considered to be a sign of economic strength. It can become a symbol of national pride. It is long known that poorly managed exchange rates may be disastrous to Economic growth (Rodrik, 2008).

Log form of the variables

Few variables used in this study for analyzing the data are transformed into their logarithmic form. By doing so the dependent variable is normalised, if the distribution of its residuals is positively skewed (Kleinbaum, Kupper, Muller and Nizam, 1998). This process of transformation has an effect of smoothing a time series to certain extent and removes seasonal trends for the effect of other influences on the data generating process to be observed (Radhika and Devi 2017).

3.3 Tools Used

Statistical techniques are seen as tools that enable researcher to analyze and summarize meaningful observations of the collected data. The current study has postulated the analysis of secondary data using relevant econometric and statistical tools and models with reference to the literature in the area of the research and applicability to suit policy implications. Software EViews was used for most of the analysis performed in the study. EViews is a statistical package for Windows, used largely for time-series oriented econometric study.

3.3.1Annual growth rate

The annual per cent growth rate for the selected macro-economic variables used in the study in India and China was calculated using the following formula

$$AGR = \frac{V_{present} - V_{past}}{V_{past}} \times 100$$

Where AGR is Annual Growth rate

$V_{present}$ is the present value of the variable

V_{past} is the past value of the variable

3.3.2 Compound Annual Growth Rate (CAGR)

Compound Annual Growth Rate is a useful measure of the variables over multiple time periods, especially if the value of selected variables has fluctuated widely during the time period in question. Compound Annual Growth Rate of the variables in the study during the period 1970 to 2018 was calculated using the formula,

$$CAGR = \left(\frac{End\ Value}{Beginning\ Value}\right)^{\left(\frac{1}{n}\right)} - 1$$

Where

n is the number of years

3.3.3 Decoupling method

Decoupling method is an important method to measure the nexus of economic growth and greenhouse gas emissions. Extensive research has been done on decoupling between economic growth and greenhouse gas emissions. Organisation for Economic Cooperation and Development (OECD) pioneered the concept of decoupling in 2002.

Tapio Decoupling Method

Tapio used the elastic coefficient decomposition model to analyse the causes and responses to decoupling between environmental harm and economic growth. Tapio extended Vehmas' decoupling classification and divided it into three major categories: weak, strong and recessive decoupling. Similarly, negative decoupling was divided in three major categories: strong negative, weak negative and expansive negative decoupling. Hao et al. (2019) explored the association between Chinas carbon emissions and economic growth for a panel of 29 provinces by using the Tapio model and EKC framework. The results confirmed

EKC hypothesis showing an inverted U shaped relationship between carbon emissions and per capita GDP.

The decoupling function can be expressed as

$$e = \frac{\Delta C/C}{\Delta Y/Y}$$

Tapio's decoupling status takes the following eight cases shown in Table 3.1.

Table 3.1: Eight cases of relationships of Tapio Decoupling Model

Classification	**Status**	**GHG Emissions per capita Change (∆C)**	**GDP per capita Change(∆Y)**	**Elastic Coefficient**
Decoupling	Weak Decoupling	> 0	> 0	0≤e 0.8
	Strong Decoupling	<0	> 0	e<0
	Recessive Decoupling	<0	<0	e>1.2
Negative Decoupling	Expansive negative decoupling	>0	> 0	e>1.2
	Strong negative decoupling	>0	<0	e<0
	Weak negative decoupling	<0	<0	0≤e< 0.8
Coupling	Expansive coupling	>0	> 0	0.8 ≤e<1.2
	Recessive coupling	<0	<0	0.8≤ e<1.2

Source: Tapio (2005)

3.3.4 Unit root test

The data on all the variables used in the study were time series in nature and hence there was a necessity to test the stationarity of time series data. The stationarity of time series data was examined by unit root test. The study employed Augmented Dickey Fuller Unit root test (Dickey and Fuller, 1979, and

Dickey et.al., 1986) to investigate the same. The ADF test is based on the estimate of the following regression:

$$\Delta X_t = \delta_0 + \delta_1 t + \delta_2 X_{t-1} + \sum_{I=1}^{k} \alpha_i \Delta X_{t-1} + u_t$$

where, Δ is the first-difference operator, X_t is the natural logarithm of the series, δ_0, δ_1, δ_2, and α_i are being estimated and u_t is the error term.

The test is conducted by augmenting each variable with its lags. The ADF test in this study is conducted by including a constant only and a constant and with a time trend. The null hypothesis is that

$H0{:}\delta_2<0$; which means that the time series is non-stationary.

The alternative hypothesis is that

$Ha{:}\delta_2=0$; which means that the variable in the series is stationary

According to Tau statistics, at 5 percent level of significance if the p-value is less than or equals to 0.05, the null hypothesis is rejected or else the null hypothesis is accepted which implies presence of stationarity in the series of the variable.

3.3.5 Johansen's cointegration test

Johansen's (1988) cointegration and VECM were employed to find the nexus between gross domestic product and its determinants, trade openness and its determinants, greenhouse gas emissions per capita and its determinants,. The cointegration test is used to find out whether time series variables have different unit roots(non-cointegrated) or same unit roots(cointegrated). The concept of cointegration is that non-stationary time series are cointegrated if a linear combination of these variables is stationary.

When there is cointegrating relationship among time series variables, It means there is the existence of a standing, long-run relationship, which is an assumption before it is tested empirically. It is necessary to perform Cointegration analysis because if two non-stationary variables are cointegrated, a VAR model in the first difference is mis-specified due to the effects of a common trend. When we have more than two variables in the model, a different approach of testing for co integration is needed (Radhika and Devi, 2017).

To examine the existence of an equilibrium relationship between gross domestic product per capita and its determinants, trade openness and its determinants and greenhouse gas emissions per capita and its determinants Johansen cointegration technique was used in the present study. Johansen developed two test statistics

Trace tests

Trace tests assess the number of linear combinations in a time series data, i.e., K to be equal to the value K_0, and the hypothesis for the value K to be greater than K_0. It is illustrated as follows:

H_0: $K = K_0$

H_0: $K > K_0$

In the above hypothesis, K_0 = zero, to test whether the null hypothesis will be rejected when using the trace test to test for cointegration in a sample. If it is rejected, it can be inferred that there exists a cointegration relationship in the sample. Therefore, the null hypothesis should be rejected to prove the existence of a cointegration relationship in the sample.

Maximum Eigenvalue test

An Eigenvalue is defined as a non-zero vector which transforms by a scalar factor, when a linear transformation is applied to it. The Maximum Eigenvalue test is like the Johansen's trace test. The key difference between the two is the null hypothesis.

H_0: $K = K_0$

H_0: $K = K_0 + 1$

In a case where $K=K_0$ and the null hypothesis is rejected, it means that there is only one likely result of the variable to produce a stationary process. However, in a case where $K_0 = m-1$ and the null hypothesis is rejected, it means that there are likely linear combinations. Such a situation is not possible unless the variables in the time series are stationary.

Order of integration of the variables

Variables used have to be of the same order of integration for constructing a meaningful model of gross domestic product per capita, trade openness and greenhouse gas emissions per capita.

3.3.6. Vector Error Correction Model (VECM)

A VECM is a cointegrated VAR model that can be used for data series which are non stationary. It includes an error-correction mechanism to control for multiple cointegration relationships.

3.3.7 Diagnostic tests

Breusch-Godfrey Test

The Durbin Watson test is limited to detecting first order auto regression, but Breusch Godfrey test can detect auto-correlation up to any pre chosen orderρ.The null hypothesis is that there is no serial correlation of any order up toρ.Since the test is based on the idea of Lagrange Multiplier testing, it is sometimes referred to as an LM test for serial correlation. In Eviews this test is done after a regression.

Breusch-Pagan Test for Heteroscedasticity

Recollecting the OLS assumption that v $(\varepsilon_j) = \sigma^2$ for all j, that is, the variance of the error term is constant, which means there is homoscedasticity. If the error terms do not have constant variance, they are said to be heteroscedastic. If the OLS assumptions are violated it can produce biased and false parameter estimates. The Breusch-Pagan test is designed to find any linear form of heteroscedasticity. Breusch-Pagan tests the null hypothesis that the error variances are all equal versus the alternative that the error variances are a multiplicative function of one or more variables. A large chi-square would indicate the presence of heteroscedasticity.

3.3.8 Granger causality test

Granger developed a widely used definition of causality that is often used by researchers interested in the inter temporal flow of effects between two variables X and Y.Y is said to Granger cause X if information about the past of Y improves one's ability to foresee the behavior of X, above what can be achieved when only information about the history of X is used for this purpose. Thus, if Y does not Granger cause X, Y is firmly exogenous to X. Granger causality tests can be performed within a Box-–Jenkins ARIMA modeling framework or in the context of a traditional OLS regression analysis (Clarke and Granato, 2005).

If the variables are cointegrated then there is causal relationship among them at least in one direction as shown by Granger Causality Test using F statistics. The direction can be unidirectional or bidirectional.

3.3.9. Variance Decomposition Analysis

The variance decomposition analysis shows the percentage of forecast error variance for each of the variables chosen that may point to its own shocks and to fluctuations in other variables. The results of this analysis will provide some further confirmation on the patterns of association amongst variables being considered, as well as contribute to enhancing insights upon how these variables respond to system wide shocks and watch how these responses propagate over time. This forecast error is due to its own innovations and the innovations of other variables in the system. In a statistical sense, if a variable explains most of its own shock, then it does not permit variance of other variables to explain and is therefore said to be relatively exogenous.

3.3.10. Impulse response function

Impulse responses trace out the response of current and future values of each of the variables to a one-unit increase in the current value of one of the VAR errors, assuming that this error returns to zero in subsequent periods and that all other errors are equal to zero. The implied thought experiment of changing one error while holding the others constant makes most sense when the errors are uncorrelated across equations, so impulse responses are typically calculated for recursive and structural VARs.

3.3.11 Testing the stability

The CUSUM test

The CUSUM test (Brown, Durbin, and Evans, 1975) is based on the cumulative sum of the recursive residuals. This option plots the cumulative sum together with the 5 percent critical lines. The test finds parameter instability if the cumulative sum goes outside the area between the two critical lines.

The CUSUM test is based on the statistic:

$$W_t = \sum_{r=k+1}^{t} w_r / s \qquad E(W_t) = 0$$

for $t = k+1, \ldots, T$ where w is the recursive residual defined above, and *s* is the standard deviation of the recursive residuals w_t. If the vector β remains constant from period to period, $E(W_t) = 0$, but if β changes, W_t will tend to diverge from the zero mean value line. The significance of any departure from the zero line is assessed by reference to a pair of 5 percent significance lines, the distance between which increases with t. The 5 percent significance lines are found by connecting the points. Movement of W_t outside the critical lines is suggestive of coefficient instability.

CUSUM of Squares Test

The CUSUM of squares test (Brown, Durbin, and Evans, 1975) is based on the test statistic:

$$S_t = \left(\sum_{r=k+1}^{t} w_r^2 \right) / \left(\sum_{r=k+1}^{T} w_r^2 \right)$$

The expected value of S_t under the hypothesis of parameter constancy is:

$$E(S_t) = (t-k)/(T-k)$$

whichgoes from zero at $t = k$ to unity at $t = T$. . The significance of the departure of S from its expected value is assessed by reference to a pair of parallel straight lines around the expected value. The CUSUM of squares test provides a plot of S_t against t and the pair of 5 percent critical lines. As with the CUSUM test, movement outside the critical lines is suggestive of parameter or variance instability. The cumulative sum of squares is generally within the 5 percent significance lines, suggesting that the residual variance is somewhat stable.

3.4 Limitations of the study

1. Since the study used time series data, all limitations such as the missed values while taking lags is applicable to this study too.
2. The study period is limited to 49 years from 1970 because of the availability of data. According to Borg and Gall (cited in Cohen et al., 2000, p.93, found in a study "The Sampling Issues in Quantitative Research" by Ali Delice, 2010) causal studies require sample size more than 50.

3. Selection of lag length criteria for selecting lag length is a major issue in time series model which is applicable in the present study.
4. In the speedy growth of developing countries such as India and China, foreign direct investment has a role. But it could not be included as an influencing variable in this study due to the non-availability of proper data from 1970 for China.
5. The problem of endogenity in the times series models was not addressed in the current study.