CHAPTER THREE
RESEARCH METHODOLOGY

Overview: The present chapter is dedicated to providing a detailed outline of the empirical approach that will be undertaken for the fulfilment of the objectives of the study. Information regarding the study sample, study period, variables considered and empirical models undertaken are elaborated in this chapter for each of the objectives. The chapter also introduces a brief overview of all the hypotheses formulated for the empirical models.

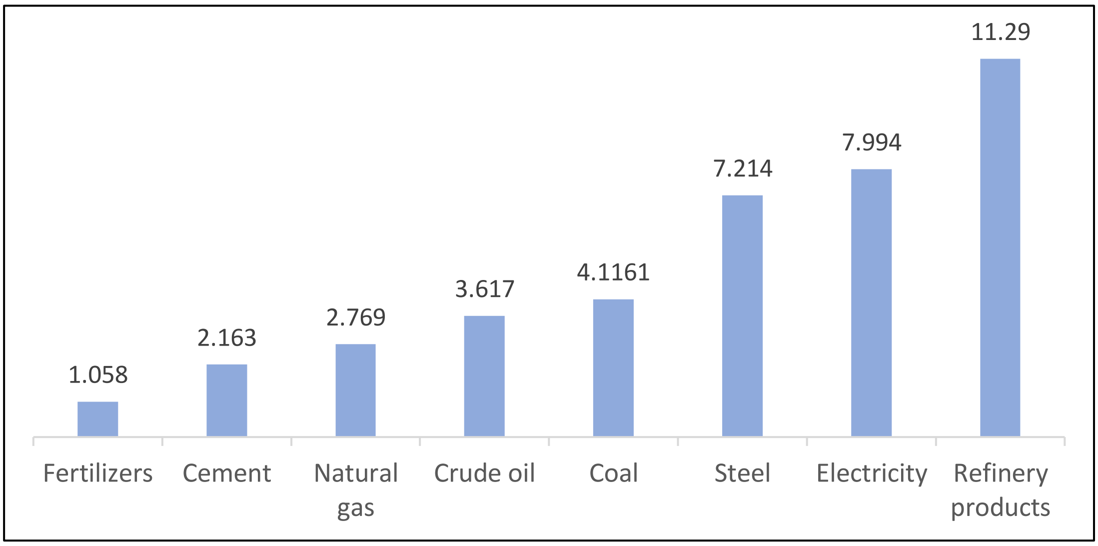
3.1. Introduction:

The research methodology is one of the core components of any research study. The identification of a suitable methodology is necessary to be done based on the nature of the work, its scope, and the final objective. The present chapter discusses the crucial components of the methodology for the present research work and elaborates on the different statistical approaches adopted in the study.

3.2. Sample size and data description:

The present study considers the eight core industries of India as its study sample. The eight core industries are coal, crude oil, natural gas, refinery products, fertilizers, steel, cement, and electricity. The Government of India has recognized these industries as the core industries of the country due to their highest contribution towards economic growth. The core industries accounted for 40.27 percent of India's overall industrial growth in 2021 (Office of Economic Adviser, 2022). Figure 3.1 demonstrates the proportion of the core industries' growth occupied in the Index of Industrial Production (IIP). IIP is an index published by the Office of Economic Adviser, Government of India, which measures the growth and production level of all recognized industries of the nation with reference to the National Industrial Classification (Ministry of Statistics and Programme Implementation, 2011). Even though the core industries' contribution to the economic growth of India is remarkable, it is important to note that these industries are major sources of environmental pollution.

Figure 3.1: Weightage percentage in IIP



(Author's compilation; Source: Office of Economic Adviser, Government of India)

The coal industry has been the primary source of energy since the era of the industrial revolution and to date, it is the dominant energy source in the world. However, its ill effects on the environment and human health cannot be neglected (Bilgen, 2016). The coal quality also determines the pollution level it releases into the atmosphere. A study in Bangladesh has indicated that the hazardous effects of low-quality coal energy are much more than the expected level of environmental degradation (Howladar *et al.*, 2020). Mastalerz and Drobniak (2013) also state that the emission level can be estimated by analyzing the components and type of coal to some extent, if not accurately. Likewise, other studies have also studied the environmental effects of the coal industry in any economic region (Gao *et al.*, 2021; Zhang *et al.*, 2015).

The crude oil industry is marked as one of the most valuable industries, yet the world has witnessed its severe adverse environmental effects. Audu et al. (2016) have stated that despite bringing in much revenue to the economy, the exploitation and exploration of crude oil have introduced various hazardous byproducts into the ecological system of Nigeria. The continuous flaring gas and oil spillover have, without a doubt, polluted the atmosphere of the country and this fact has been established by several studies in Nigeria (Ipingbemi, 2009; Omofonmwan and Odia, 2009; Ugochukwu and Ertel, 2008). The industry affects biodiversity by releasing various toxic elements and chemicals. Hong et al. (2014) have also reported the effects of the Hebei Spirit oil spill in Korea. The study has thoroughly evaluated its negative effects on the environment, living orgasms and the health of the people. Several studies have also empirically presented the significant relationship between the crude oil industry and the environmental conditions of an economy (Ankathi *et al.*, 2022; Sreenu, 2022).

The natural gas industry is primarily involved in causing air pollution by emitting various pollutants into the atmosphere. Carbon dioxide (CO_2) and methane (CH_4) are two dominant components of this industry (Crow *et al.*, 2019; Heydarzadeh *et al.*, 2020). The natural gas industry is related to production and processing, pipeline transport and storage of the outputs. The transportation and storage of natural gas contribute to more than 82 percent of emissions from the industry in the Russism Federation (Uvarova *et al.*, 2014). According to Dinca et al. (Dinca *et al.*, 2007), the natural gas industry can have alarming environmental degradation effects from multiple aspects, such as the impact on public health, agriculture, materials, global warming, etc. So. There is no doubt that the natural gas industry can cause severe damage to the environment if not handled or functioned with preventive measures.

Petroleum refinery products are widely used for various purposes, from industries to ordinary households. However, the adverse environmental effects of the industry can be of large scale on an economic level. This industry is the fourth largest greenhouse gase (GHG) emitter globally (United States Environmental Protection Agency, 2020). Abella and Bergerson (2012) have carefully analyzed how the industry can have severe environmental effects on the environment of North America. The refinery emissions depend on the quality of the crude oil. Irrespective of the crude's quality, unfavourable environmental effects are certain on different levels (Hirshfeld and Kolb, 2012). Alhaddad et al. (2015) have stated that the actual emissions from the refinery industry in Kuwait are more than the international limits, which means hazardous environmental effects. Several studies support the fact that petroleum refinery products significantly impact the emissions of various pollutants (Kwasniewski *et al.*, 2016; Nelson, 2013).

Fertilizers have always played a vital role in increasing agricultural production and quality. However, there is a chance of harming the environment if chemically processed fertilizers are used extensively on agricultural land. It can impact the aquatic and terrestrial habitats, toxicate groundwater sources and human health too (Tilman *et al.*, 2002). A study by Kang et al. (2022) stated that the ammonia emission level from nitrogen phosphorus Potassium oxide fertilizers is very high in South Korea and demands well-structured mechanisms to quantify and minimize its emissions. The nitrogen fertilizers and ammonia production process is even more CO_2 intensive than the global average (Kahrl *et al.*, 2010). Nevertheless, nullifying the damaging effects of fertilizers on nature is challenging. In most cases, it even reduces nutrient efficiency (Chen *et al.*, 2018).

The steel industry is inevitable for any economy because of its extensive usage. However, like the other core industries, this industry has also been causing adverse environmental effects in various economic regions. Especially the wastewater discharged from the steel industry can be one of the most dominant factors in destructing nature (Nguyen *et al.*, 2022; Sun *et al.*, 2019). At the same time, the industry encourages air pollution by emitting various pollutants into the atmosphere (Tang *et al.*, 2020). Moreover, it may also degrade a region's soil quality (Strezov and Chaudhary, 2017). Other past studies have also empirically presented the possible environmental effects of the steel industry on its surroundings (Jozi and Majd, 2014; Serrenho *et al.*, 2016).

The cement industry is one of the biggest emitters in the globe. The energy-intensive process of the cement industry has made it more challenging to minimize the emission effects of the industry (Rehan and Nehdi, 2005). To resolve such issues, advanced technology innovation is required to be implemented in the cement industry. Sanjuán et al. (Sanjuán *et al.*, 2020) have given a detailed analysis of the Spanish cement industry and suggested how the industry's emission levels can be reduced. Studies evaluating the cement industry admit that CO2 is the industry's dominant pollutant, dramatically affecting the environment's quality (Ige *et al.*, 2022; Summerbell *et al.*, 2016).

The electricity industry is no different from the others. Alves and Uturbey (2010) have estimated the environmental cost incurred by the Brazilian electricity sector. Its effect on human health may be severe as they have found that the pollution caused by local plants may be even more dangerous than global warming. Hence, it is essential to determine how electricity is utilized in any industry as an energy source. Another study has concluded that electricity generation is Malaysia's largest source of emissions (Mahlia, 2002). Such findings demand the implementation of new energy sources to protect the environment. Balsalobre-Lorente et al. (2018) have proved that renewable electricity consumption can help improve the environmental quality of a country. Other studies have also empirically demonstrated that the electricity sector significantly affects the environment (Kahouli, 2018; Zhang *et al.*, 2013).

In a nutshell, past studies have recognized the alarming threats that the core industries pose to the environment. These industries have a global reputation for being highly destructive to environmental health. Considering these eight industries' significance in the nation's economic growth and environmental degradation, their environmental assessment has become critical for the country's ecological welfare. Hence, the core industries are considered as the final study sample for the present research work.

The natural gas and fertilizer industries were included in the list of core industries of the nation from 2005 onwards. Therefore, the study period for the work has been chosen from 2005 to 2021, depending on the availability of data. For each objective, secondary panel data has been considered for the respective empirical analyses.

3.3. Methodology for Objective 1 (To compare the levels of carbon emissions across the core industries)

For Objective 1, data from 2005 to 2021 is considered for the core industries. However, data for carbon emissions for the electricity, steel and refinery products industries are not available throughout the period from 2005 to 2021. For the electricity industry, the data is available till 2020; for the steel industry, the data is available till 2018 and for the refinery products industry, the data is available till 2018. Hence, an unbalanced panel dataset is considered for Objective 1.

3.3.1. Variable description:

To measure the industrial growth level of the core industries in India, the Index of Eight Core Industries (ICI) of the respective industries is considered. The index is published by the government to indicate the growth and production levels, particularly for the core industries (Office of Economic Adviser, 2022). The data for ICI is collected from the EPW Research Foundation database.

On the other hand, the CO_2 emissions of each of the core industries have been considered for measuring the environmental pressure imposed by the core industries. Past studies have also used carbon emissions to estimate the level of environmental degradation of a country or industry (Halicioglu, 2009; Hao *et al.*, 2019; Lv *et al.*, 2021), as CO_2 is the most dominant component of all GHGs. International authorities have also recognized the threatening characteristics of CO_2 gas for increasing climate change and global warming issues (IEA, 2023; IPCC, 2014). The Intergovernmental Panel on Climate Change (IPCC) and the United Nations have also considered CO_2 emission levels as a base measure to estimate the environmental effects of other GHGs (IPCC, 2007). Hence, the study considers the yearly CO_2 emission levels (in tonnes) of the respective core industries to measure their environmental degradation level. The mentioned data is extracted from the Global Carbon budget database (Global Carbon Project, 2022), except for the fertilizers, steel, and refinery products industries. Emission levels from the fertilizer industry are available at the FAOSTAT (Food and Agriculture Organization Statistics) database of the United Nations and those of the steel and refinery products industries are published by the CEEW (the Council on Energy, Environment, and Water). The uniformity in the industrial emissions data has been checked and maintained across all industries as they all conform to the guidelines provided by the IPCC.

3.3.2. Empirical approach:

To compare the levels of carbon emissions across industries, the study gives a graphical representation of the industries' carbon emission trends. Further, the study considers the decoupling method suggested by Tapio (2005) for assessing the environmental efficiency of the core industries. This empirical approach will help provide a comparative picture of the core industries' carbon emission levels with the inclusion of their industrial growth factor. In doing so, a true depiction of the industries' environmental efficiency level will be reflected. The elasticity of CO_2 emissions in the Indian core industries will be calculated using Model 1.1.

$$DI = \frac{\frac{CO_{2t} - CO_{2t-1}}{CO_{2t-1}}}{\frac{ICI_t - ICI_{t-1}}{ICI_{t-1}}} = \frac{\Delta CO_2\%}{\Delta \text{ICI}\%} \quad \text{--- (1.1)}$$

In Model 1.1, DI stands for decoupling index, CO_2 for carbon emission levels, and ICI for Index of Core Industries. Here, t and t-1 refer to the current year and the preceding year, respectively. Each industry's decoupling state will indicate the environmental efficiency of the industrial operations.

Further, this study attempts to add a new dimension to the model by assigning decoupling scores to each of the eight elasticity degrees. Table 3.1 provides the scores assigned to each of these elasticity degrees. The scores will range from 1 (the least polluting state) to 8 (the most polluting state). The scores are assigned based on the following criteria:

a. When the $DI < 0$*:*

There are two instances when the DI may have a negative value: strong decoupling and strong negative decoupling. Strong decoupling elasticity is given a score of 1, as it indicates the most desirable decoupling elasticity state for any industry. Here, when industrial growth occurs, it achieves a decrease in carbon emissions. Hence, the rate of change in environmental pressure will have a negative value while the industry continues to attain growth. In contrast, the strong negative decoupling elasticity has been assigned a score of 8, where carbon emissions are seen to be rising even when industrial growth declines. It is the worst state of decoupling state from the environmental perspective, reflecting the industry's inability to balance growth and environmental competence. In this state, the environmental stress level of the industry continues to grow even when the industrial growth rate falls.

b. When the $DI > 0$*:*

The DI value is positive in all the other six decoupling states. In such cases, the scores are assigned following two rules:

- *Environmental Efficiency:* The decoupling elasticity is an indication of an industry's level of environmental efficiency. The current year's environmental efficiency increases when each unit of industrial growth emits less carbon than the previous year. A lower value of decoupling elasticity will mean better environmental efficiency, i.e., the lower the DI, the better for the environment. Hence, the elasticity degree with the range of lower DI values will be given better decoupling scores than the higher ones. According to this rule, weak decoupling will be provided with a better score than expansive decoupling, as the DI value range is smaller in the case of the former.
- *Economic performance*: If the DI values fall under the same range, the elasticity degree with a positive industrial growth rate will be assigned a better score. For instance, weak decoupling and weak negative decoupling elasticities are categorized under the same range of DI values. In that case, the former will be assigned a better score for having a superior and positive industrial growth rate than the latter. A positive industrial growth rate reflects the greater economic performance of the industries, which is always preferred over declining industrial growth shown by a weak negative decoupling state.

Table 3.1: Decoupling scores

Decoupling elasticity	$\Delta CO_2\%$	ΔICI%	Elasticity degree	Score
Strong decoupling	< 0	> 0	$[0, -\infty)$	1
Weak decoupling	> 0	> 0	$[0,0.8)$	2
Weak negative decoupling	< 0	< 0	$[0,0.8)$	3
Expansive coupling	> 0	> 0	$[0.8,1.2)$	4
Recessive coupling	< 0	< 0	$[0.8,1.2)$	5
Expansive negative decoupling	> 0	> 0	$[1.2, +\infty)$	6
Recessive decoupling	< 0	< 0	$[1.2, +\infty)$	7
Strong negative decoupling	> 0	< 0	$[0, -\infty)$	8

(Author's compilation)

Next, an average score will be given to the industries, which will indicate the level of environmental stress imposed by their growth. Each year, each industry's scores will be assigned, reflecting how pollutive or dirty the industry is. Table 3.2 provides the ranges for the average decoupling scores to categorize the industries based on their pollution level. An average decoupling score of 1 to 3.5 indicates a low pollution level for industries. A range of 3.5 to 5.5 indicates a moderate level of pollution. Lastly, a 5.5 or higher score will be provided to the dirty or high-polluting industries. The ranges of scores in each categorization are based on the average decupling elasticity values represented by each assigned score as in Table 3.1. Till the score of 3, the average value of elasticity is less than 0.8, which means industries' relatively healthier environmental performances. For the scores of 4 and 5, the average elasticity score is between 0.8 to 1.2, which is a worse state than the previous score ranges but more preferred over the next set of ranges available. Lastly, in the case of scores 6 and above, the industries show undesirable high-polluting conditions, with average elasticity values higher than 1.2. In order to prepare a continuous class interval, a variation of ±0.50 is allowed for each category and thereafter, the score ranges are arrived at as shown in Table 3.2.

Table 3.2: Average decoupling score range for industries

Average score range	Remarks
1-3.5	Low-polluting/clean industry
3.5-5.5	Moderate-polluting industry
5.5-8	High-polluting/dirty industry

(Author's compilation)

3.4. Methodology for Objective 2 (To identify the driving forces of carbon emissions in the core industries):

For Objective 2, data for carbon emissions for the electricity, steel and refinery products industries are not available from 2005 to 2021. For the electricity industry, the data is available till 2020; for the steel industry, the data is available till 2018 and for the refinery products industry, the data is available till 2018. Therefore, considering an unbalanced panel dataset for the period from 2005 to 2021, the study proceeds with the incorporation of 129 industry-year observations for the fulfilment of Objective 2.

3.4.1. Variable description:

3.4.1.1. Dependent variable:

To measure the environmental degradation levels of the core industries, the carbon emission levels of the respective industries are considered for Objective 2 also. As discussed above, CO2 is known to be the most dominant of GHGs due to its severe detrimental effects on the environment (IEA, 2023; IPCC, 2014). In this objective, the natural log of the carbon dioxide emission levels (in tonnes) is considered for the empirical models. Similar to Objective 1, the carbon emission levels of each core industry are collected from three different sources. However, uniformity in their estimations is ensured as the assessments of all the databases are based on the guidelines provided by the IPCC.

3.4.1.2. Explanatory variables:

At first, the potential drivers are identified that can significantly impact the emission levels of the core industries based on the arguments of the past literature in similar contexts. Most of the earlier studies established that these variables have significant effects on country-level analysis. The present study aims to further investigate their role in terms of industrial emissions with respect to India's core industries. The potential drivers of the core industries' emissions are grouped into four categories: economic, industrial, demographic, and environmental. Measurement and sources of all the explanatory variables are discussed below:

Economic factors:

Economic growth: In this study, economic growth is measured by the annual percentage growth of gross domestic product (GDP) per capita. GDP is the aggregated value of all final goods and services that are produced within a country during a specified period of time. GDP per capita is calculated by dividing the country's total GDP by the total volume of the population.

Foreign investments: In this study, the natural log of foreign direct investments (FDI) inward financial flows is considered. FDI inflow refers to the total value of a foreign country's investments in a host country during a specified period of time. Good quality FDIs are expected to enable the transfer of funds, technologies, skills, production growth, infrastructure development etc. in the host country.

Agricultural activities: It is estimated by the natural log of gross agricultural production value. Gross agricultural production refers to the sum of all agricultural products produced in a nation for a specified period of time. Crops, livestock production, vegetables, fruits, etc. are a few of the most common elements of agricultural production.

Research and development: The indicator for the variable is the annual percentage growth of patent applications by residents of India. A patent application is a formal request to seek legal approval and protection of inventions, allowing for exclusive rights to the applicant to earn revenues through selling and licensing.

Industrial factors:

Industrialization: It is measured by industry (including construction) value added in terms of percentage of GDP. Industrial value added is the net output generated by the industrial sector of the economy within a country during a specified period of time. Here, the values of intermediate inputs are subtracted from the sum of all industrial outputs.

Industries' energy consumption: The variable is measured by the annual percentage growth of electricity energy consumption by Indian industries. It is the total energy consumed by the overall industrial sector of the economy, equivalent to electricity consumption.

Financial credits to industries: It is estimated by the annual percentage growth of deployment of gross bank credit (outstanding amount) in India. It is the total amount of credit facilities offered by financial institutions to the industrial sector. An efficient financial system helps industries easily access finances that are required for their overall growth.

Industrial innovation: This is indicated by the annual percentage growth of industrial design applications by residents of India. Industrial design is primarily concerned with the physical qualities of the industrial products. It is the process of designing and developing sustainable product designs through innovation for mass production. Overall, the products' visual appearances, functionality and composition depend on their quality of industrial design.

Demographic factors:

Population: It is measured by the natural log of population density (people per sq. km of land area). Population density is the number of people living per unit area, which is calculated by the total population divided by the total land area.

Urbanization: It is measured by the annual percentage growth of population volume in urban agglomerations of more than 1 million. It simply refers to the number of people living in urban areas of the country.

Poverty: The variable is indicated by the percentage of people with no access to safely managed sanitation services. Safely managed sanitation refers to the requirement of at least basic sanitation facilities for individual households, which is considered essential for maintaining a minimum standard of living.

Education: It is estimated by the annual percentage growth of the number of pupils in secondary education. Hence, the education level of the country is measured based on the number of total number of students enrolled for secondary education in either public or private schools.

Environmental factors:

Tree cover loss: It implies the scale of tree removals in tree canopy locations during a specified time period. Tree cover loss is the most fundamental cause of deforestation.

Water stress: It is measured by the percentage of water withdrawal from total water sources (in billion cubic metres). Water stress generally refers to a situation where water demand exceeds its level of supply. Thus, creating pressure on the limited local water resources for water extraction.

Sustainable innovation: It is estimated by the annual percentage of the number of patents particularly related to environment-related technologies. Environment-related technologies are innovations that particularly aim at reducing environmental stress levels of industries to promote sustainable growth.

Environmental credits: It is measured by the natural log of the number of CERs issued projects in India. CERs are approved carbon credits based on the results of reducing GHG emissions after formal verification and certification by concerned authorities. These are aimed at improving the environmental productivity of industries and ensuring their sustainable growth.

Table 3.3 summarizes the list of all the variables that are considered for Objective 2. In the table, the supported references of empirical studies that have documented the significant impact of the respective variables on carbon emissions are also included. As mentioned above, these studies have explored the variables in country-level analysis. No study has been found investigating their impact, particularly on the emission levels of the Indian core industries.

Table 3.3: List of variables (Objective 2)

Variables	Indicators, Abbreviations	Data sources	References
Dependent variable:			
Environmental degradation	Carbon dioxide emissions, *lnCO2*	Global Carbon Budget database, FAOSTAT, CEEW	(Afriyie et al., 2023; Sreenu, 2022)
Independent variables:			
Economic factors:			
Economic growth	Gross domestic product per capita, *GDP*	World Bank Open Database	(Aslam et al., 2021; Awan and Azam, 2021)
Foreign investments	FDI inwards financial flows, *lnFDI*	World Bank Open Database	(Sreenu, 2022; Zakaria and Bibi, 2019)
Agricultural activities	Agricultural production, gross value, *lnAGR*	FAOSTAT	(Anwar et al., 2019; Phiri et al., 2021)
R&D	Patent application by residents, *R&D*	World Intellectual Property Organization Database	(Awan and Azam, 2021; Lee et al., 2015)
Industrial factors:			
Industrialization	Industry value added, *IVA*	World Bank Open Database	(Aslam et al., 2021; Mahmood et al., 2020)
Energy consumption	Industry electricity energy consumption, *EC*	Central Electricity Authority of India reports	(Awan and Azam, 2021; Phiri et al., 2021)
Financial credits	Industry Deployment of Gross Bank Credit, *BANK*	Reserve Bank of India Statistics	-
Industrial innovation	Industrial design applications by residents, *INDDEG*	World Bank Open Database	-
Demographic factors:			
Population	Population density, *lnPOP*	World Bank Open Database	(Aslam et al., 2021; Ohlan, 2015)
Urbanization	Population in urban agglomerations of more than 1 million, *URB*	World Bank Open Database	(Afriyie et al., 2023; Mahmood et al., 2020)

Poverty	Population with no access to safely managed sanitation services, *POV*	World Bank Open Database	-
Education	Number of pupils in secondary education, EDU	World Bank Open Database	(Tang et al., 2023; Xin et al., 2023)
Environmental factors:			
Tree cover loss	Tree cover loss, *lnTCL*	Environmental Performance Index Database	(Minlah et al., 2021)
Water stress	Water withdrawal by industries, *WS*	AQUASTAT	(Driscoll et al., 2024; Rajan et al., 2020)
Sustainable innovation	Environment-related technological patents, *ENVTECH*	OECD database	-
Environmental credits	Number of CERs issued projects, *CER*	Ministry of Statistics and Programme Implementation, Govt. of India reports	-

(Author's compilation)

3.4.2. Empirical approach:

3.4.2.1. Descriptive statistics:

Once the variables are identified and categorized, Objective 2 proceeds with the descriptive statistics of the variables. The documentation of the descriptive statistics is essential to provide an overview of significant data characteristics.

Correlation matrix:

The correlation matrix denotes the degree of pairwise correlation between the explanatory variables. The degree of correlation will determine the dataset's suitability for further regression analysis. If the pairwise correlation degree is more than 0.80 between any two explanatory variables, it indicates the possibility of multicollinearity issues in the empirical model (Gujarati and Porter, 2004). In that case, the model may provide spurious results. In the absence of such casualties, the data model can be proceeded for further empirical tests.

3.4.2.2. Regression analysis:

In order to investigate the drivers of carbon emissions, the present research work conducts the Dynamic Ordinary Least Squares **(**DOLS) estimation for Objective 2, considering its strong statistical competence over other techniques. While dealing with panel data analysis, several econometric issues may occur such as serial correlation, heteroscedasticity and endogeneity. In the presence of these issues, the empirical models are expected to provide highly spurious outcomes. Nevertheless, the DOLS approach (Saikkonen, 1991; Stock and Watson, 1993) is capable of yielding higher efficiency in producing consistent and reliable results even for small sample-sized datasets (Numan *et al.*, 2022). By considering leads and lags of the explanatory variables, the DOLS approach overcomes issues of endogeneity and autocorrelation (Narayan and Smyth, 2007). Additionally, the weighted criteria of DOLS can resolve heterogeneity in long-run variance and co-integrated panels (Dogan and Seker, 2016). Therefore, in the presence of such statistical issues, the DOLS approach will be considered.

It is to be noted that to conduct DOLS analysis, the variables can be integrated either at levels or at first-differences, i.e., variables can be stationary either at I(0) or I(1) (Chowdhury *et al.*, 2022; Raihan *et al.*, 2022; Stock and Watson, 1993). Therefore, the ADF-Fisher chi-square and PP-Fisher Chi-square panel unit root tests are performed before applying the DOLS approach to check the suitability of the variables for the estimates. The null hypothesis for both ADF-Fisher Chi-square and PP-Fisher Chi-square panel unit root tests is that each time series in the panel dataset has a unit root. If the test-statistics of the unit root are significant, it would imply rejection of the null hypothesis and stationarity of the variables. If the variables are integrated either at I(0) or I(1), the DOLS approach will be applicable. The following Model 2.1 is applied to get the estimated coefficients of the DOLS approach (Kao and Chiang, 2001):

$$y_{i,t} = \alpha_i + \beta_i x_{i,t} + \sum_{k=-k_i}^{k_i} \gamma_{i,k} \Delta x_{i,t-k} + \varepsilon_{it} \qquad \text{--- (2.1)}$$

Here, y is the dependent variable and x is the independent variable. Then, $-k_i$ and k_i are the lead and lag of the difference, respectively. $\gamma_{i,k}$ is the coefficient of lead and lags that are responsible for possible serial correlation and endogeneity issues in the explanatory variables.

Next, the Random Effects Model (REM) approach will be undertaken to determine the category-wise (i.e., economic, industrial, demographic and environmental) driving factors of carbon emissions. When statistical issues such as serial correlation and endogeneity do not occur, the REM is often applied due to its easy and simplified application. In panel data, the choice of REM and Fixed-Effects Model (FEM) over Pooled Ordinary Least Squares (POOLS) regression model is determined based on the Breusch-Pegan test and Hausman test.

The Breusch-Pegan Test is widely used to determine whether a POOLS model or a REM is appropriate for the observed dataset. The null hypothesis of the test is,

$\boldsymbol{H_0}$: There are no random effects among the error terms.

If the test statistic p-value fails to reject the null hypothesis, the statistical model will be free from heteroscedasticity. In such cases, the pooled OLS will be more appropriate for the study, and vice versa.

Likewise, the Hausman test help conclude if the REM or FEM is appropriate for the study. The null hypothesis of the test is,

$\boldsymbol{H_0}$: There is no correlation between the residuals and the regressors.

If the test statistic fails to reject the null hypothesis, it will imply that REM is a better and more suitable estimation model for the dataset and vice versa. Thus, the appropriate regression model is selected for the empirical analyses based on the outcomes of the Breusch-Pegan test and the Hausman test. Further, to resolve the possible issues of heteroscedasticity, robust standard errors are applied for the panel regression models (White, 1980). It helps improve the robustness and reliability of the empirical results.

For each category of potential driver, one empirical model will be formulated to explore the determinants of carbon emission. Model 2.2 represents the study's generic regression model.

$$DV_{it} = \alpha + \beta . IV_{it} + \varepsilon_{it} \qquad \text{--- (2.2)}$$

In Model 2.2, i and t denote the cross-section item and the time period, respectively. DV stands for the dependent variable and IV represents the independent or explanatory variables considered for the study. Here, α is the constant and ε is the error term of the regression model and β signifies the coefficient values of the independent variables. Based on Model 2.2, the study formulates the following Model 2.3 to investigate the factors of carbon emissions in the core industries. From Table 3.3, a set of eight variables out of the 16 variables are included in this regression model depending upon the results of correlation analysis.

$$lnCO_2 = \alpha + \beta_1 GDP_{it} + \beta_2 R\&D_{it} + \beta_3 EC_{it} + \beta_4 POV_{it} + \beta_5 EDU_{it} + \beta_6 lnTCL_{it} + \beta_7 WS_{it} + \beta_8 lnCER_{it} + \varepsilon_{it} \qquad \text{--- (2.3)}$$

Moreover, the inclusion of all 16 variables in a single regression model may lead to statistical overspecification issues, causing spurious empirical results. At the same time, the exploration of their impact on carbon emissions holds critical significance in providing valuable insights for ensuring sustainable industrial growth in the nation. Therefore, the study develops the following Models 2.4 to 2.7 in order to conduct an in-depth analysis of other potential drivers of industrial emissions. The categorization of the potential drivers into separate regression models will allow the investigation of all potential drivers, which can critically assist better policy formulation.

$$lnCO_2 = \alpha + \beta_1 GDP_{it} + \beta_2 lnFDI_{it} + \beta_3 lnAGR_{it} + \beta_4 R\&D + \varepsilon_{it} \qquad \text{--- (2.4)}$$

$$lnCO_2 = \alpha + \beta_1 IVA_{it} + \beta_2 EC_{it} + \beta_3 BANK_{it} + \beta_4 INDES + \varepsilon_{it} \qquad \text{--- (2.5)}$$

$$lnCO_2 = \alpha + \beta_1 lnPOPD_{it} + \beta_2 URB_{it} + \beta_3 POV_{it} + \beta_4 EDU + \varepsilon_{it} \qquad \text{--- (2.6)}$$

$$lnCO_2 = \alpha + \beta_1 lnTCL_{it} + \beta_2 WS_{it} + \beta_3 ENVTECH_{it} + \beta_4 lnCER + \varepsilon_{it} \qquad \text{--- (2.7)}$$

The abbreviations used in Models 2.3 to 2.7 have already been mentioned in Table 3.3.

3.4.2.3. Diagnostic tests:

Once the results are obtained from the regression analysis, diagnostic tests have to be conducted in order to ensure that the results are not spurious. Firstly, the Breusch-Godfrey LM Test will be conducted to check if the statistical model suffers from any possible issues of autocorrelation. The null hypothesis of the test is that there is no serial correlation. Thus, if the test-statistics are found to be non-significant and the null hypothesis is accepted, it can be concluded that the statistical results are free from autocorrelation or serial correlation. Secondly, the Breusch-Pegan test results will indicate if the model suffers from heteroscedasticity. As explained earlier, in order to confirm that the model does not suffer from heteroscedasticity, the test-statistic of the Breusch-Pegan test must be non-significant. Nevertheless, robust standard errors are applied for the panel regression models in Objective 2, which overcomes the possible issues of heteroscedasticity (White, 1980). Thirdly, the Pesaran CD test will be performed to check the status of cross-sectional dependence, whose null hypothesis is that the residuals are uncorrelated across cross-sections. The non-significant test-statistic and the acceptance of the null hypothesis are favourable for the empirical results, confirming the absence of cross-sectional dependence. Lastly, the test-statistic of the Wald chi^2 test will be observed. The null hypothesis for this test is that the explanatory variables do not have a significant effect on the dependent variable. Accordingly, its significant test-statistics are required to reject the null hypothesis and confirm the overall statistical significance of the study models. If the results of all these diagnostic tests are satisfied, it can be concluded that the empirical outcomes of the models are reliable and are not spurious. The diagnostic tests undertaken in Objective 2 are summarized in Table 3.4.

Table 3.4: Diagnostic tests (Objective 2)

Diagnostic test	Null hypothesis	Inference
Breusch-Godfrey LM test	There is no serial correlation.	A non-significant test-statistic would indicate absence of serial correlation.
Breusch-Pegan test	There are no random effects among the error terms.	A non-significant test-statistic would indicate that the data is not heteroscedastic.
Pesaran CD cross-sectional test	There is no cross-sectional dependence.	A non-significant test-statistic would indicate that the residuals are not cross-sectionally dependent.
Wald chi2 test	All the coefficients are equal to zero.	A significant test-statistic would indicate that the overall model is statistically significant.

(Author's compilation)

3.4.2.4. Robustness check:

As explained earlier, the DOLS approach has many statistical advantages in providing robust and reliable estimates. Therefore, the DOLS is applied as a robustness check for Objective 2 in the present study. The ADF-Fisher Chi-square and PP-Fisher Chi-square panel unit root tests will be conducted to ensure that the variables are integrated either at I(0) or I(1), confirming their suitability for DOLS estimates.

3.5. Methodology for Objective 3 (To analyze the relationship between industrial growth and environmental degradation in the core industries):

For Objective 3, data regarding carbon emission for the electricity, steel and refinery products industries are not uniformly available for the period of 2005 to 2021. For the electricity industry, the data is available till 2020; for the steel industry, the data is available till 2018 and for the refinery products industry, the data is available till 2018. Therefore, unbalanced panel data is considered from 2005 to 2021 for the empirical investigation in Objective 3. The regression models will be tested with the inclusion of 129 industry-year observations.

3.5.1. Variable description:

Objective 3 is focused on investigating the relationship between industrial growth and carbon emissions in the Indian core industries, with reference to the industry-specific environmental Kuznets curve (IEKC) hypothesis. In order to test the hypothesis, the objective considers the dependent and explanatory variables that are presented in Table 3.5.

3.5.1.1. Dependent variable:

The carbon dioxide emission levels of the respective core industries are considered for Objective 3 as the dependent variable for all models in Objective 3, indicating the environmental degradation levels of the industries. It has already been established that CO_2 poses a great deal of threat to global environmental health. The data source of the emission levels of the respective core industries is the same as mentioned in Objective 1.

3.5.1.2. Primary independent variable:

The measure the industrial growth level of the core industries in India, the Index of Eight Core Industries (ICI) of the respective industries is considered as the primary independent or explanatory variable for Objective 3. As discussed earlier, the index is published by the Office of Economic Adviser, Government of India to indicate the growth and production levels, particularly for the core industries (Office of Economic Adviser, 2023). Similar to earlier statements, the data for ICI is collected from the Economic and Political Weekly Research Foundation (EPWRF) database for Objective 3.

3.5.1.3. Other explanatory variables:

Further, the empirical models of Objective 3 considered other explanatory or control variables in the empirical models. The significant drivers of the core industries' emission levels established in Objective 2 will be considered for Objective 3 as control variables. They are the economic growth of the nation (economic factor), energy consumption level by industries (industrial factor), and water stress level imposed on the environment (environmental factor). The empirical models could not accommodate the inclusion of any demographic factors of industrial emissions, as it leads to the presence of multicollinearity issues in the models. The variables considered have similar details to those used in Objective 2. Hence, they are not repeated in this section.

3.5.1.4. Moderating variables:

After analysing the relationship between industrial growth and environmental degradation in the Indian core industries, the present objective further tests the moderating roles of environmental policy stringency and the economy's industrial structure improvement to examine if they have any influence on the aforesaid relationship. A brief detail of these two moderating variables is mentioned below:

Environmental policy stringency: It indicates the level of regulatory pressure applied on industries to initiate applicable efforts for the fulfilment of environmental quality improvement. A stricter policy setting is expected to promote better environmental standards in a nation. To measure this variable, the environmental policy stringency index for the industrial sectors is considered.

Industrial structure improvement: The tertiary and the secondary sectors differ in their fundamental characteristics and operating features. It is generally considered that the tertiary sector imposes less threat on environmental deterioration and hence, its growth is expected to improve the economy's industrial structure. Industrial structure improvement is measured by the proportion of value added by the tertiary sector to that of the secondary sector of the economy.

Table 3.5: List of variables (Objective 3)

Variables	Indicators	Data sources	References
Dependent variable:			
Environmental degradation	Carbon dioxide emissions, *$lnCO_2$*	Global Carbon Project database, FAOSTAT, CEEW	(Afriyie *et al.*, 2023; Sreenu, 2022)
Explanatory variables:			
Industrial growth	Index of Eight Core Industries, *ICI*	EPWRF database	-
Economic growth	Gross domestic products, *GDP*	World Bank Open Database	(Aslam *et al.*, 2021; Awan and Azam, 2021)
Energy consumption	Industrial electricity energy consumption, *EC*	Central Electricity Authority of India reports	(Awan and Azam, 2021; Phiri *et al.*, 2021)
Water stress	Water withdrawal by industries, *WS*	AQUASTAT	(Driscoll *et al.*, 2024; Rajan *et al.*, 2020)
Moderating variables:			
Regulatory pressure	Sectoral Environmental Policy Stringency Index, *SECP*	OECD database	(Çetinkaya *et al.*, 2024; Yoon and Heshmati, 2021)
Industrial structure	The ratio of the value added of the tertiary sector to that of the secondary sector, *INDSTR*	World Bank Open Database	(Zhao *et al.*, 2022; Zhou *et al.*, 2013)

(Author's compilation)

3.5.2. Empirical approach:

3.5.2.1. Descriptive statistics:

At first, the descriptive statistics of all considered variables in Objective 3 will be presented. As mentioned earlier, it will help provide some indication of the basic characteristics of all the variables.

Correlation matrix:

The empirical models will be suitable for regression analysis only when the pairwise correlation degree among the explanatory variables is less than 0.80 (Gujarati and Porter, 2004). To ensure such suitability, the correlation matrix has to be tabulated before proceeding for further analysis, depicting the pairwise correlation degree between all explanatory variables considered in Objective 3.

3.5.2.2. Regression analysis:

As mentioned earlier, Objective 3 is aimed at assessing the relationship between the core industries' growth and their carbon emission levels, with reference to the IEKC hypothesis. To test the same, the following regression model- Model 3.1 is considered.

$$DV_{it} = \alpha + \beta . IV_{it} + \gamma . CV_{it} + \varepsilon_{it} \quad \text{--- (3.1)}$$

In this model also, i and t represent the cross-section item and the time period, respectively. DV is the dependent variable, IV represents the independent or explanatory variables, and CV denotes the control variables. In the case of the moderating roles, IV also includes the interaction term of the independent variable and the moderating variable. Then, α is the constant and ε is the error term. β signifies the coefficient values of the independent variables., whereas γ captures the coefficient values for the control variables. With reference to Model 3.1, the study develops Models 3.2 to 3.5 for Objective 3.

The base model: To test the IEKC hypothesis, a quadratic model (Model 3.2) is adopted to account for the non-linear relationship between industrial growth and carbon emission. Here, ICI and ICI^2 are the two primary independent variables that will determine the validity of the IEKC hypothesis. The empirical results will confirm the presence of an inverted U-shaped IEKC if $\beta_1 > 0, \beta_2 < 0$. Any deviation from this condition will result in the hypothesis being rejected with respect to the Indian core industries.

$$lnCO_2 = \alpha + \beta_1 ICI_{it} + \beta_2 ICI_{it}^2 + \beta_3 GDP_{it} + \beta_4 EC_{it} + \beta_5 WS_{it} + \beta_6 SECP_{it} + \beta_7 INDSTR_{it} + \varepsilon_{it} \quad \text{---(3.2)}$$

Moderating effect of environmental policy stringency: The moderating role of SECP in the relationship between industrial growth and environmental degradation is also examined in Objective 3 by using Model 3.3. To do so, the interaction term *SECP_ICI* is introduced in Model 3.3, whose coefficients will determine the moderating roles of SECP. If the coefficient of the interaction term is found to be positively significant, it will imply that the implementation of stricter environmental policies will magnify the effects of industrial growth on environmental degradation. In contrast, its negatively significant coefficient will indicate that stricter environmental regulations in practice will help reduce the harmful environmental consequences from the core industries' growth. The stringency of environmental regulation will help attain sustainable industrialization. Based on the level of regulatory pressure or the stringency of the regulation, the strength of the relationship between industrial growth and environmental degradation can differ.

$$lnCO_2 = \alpha + \beta_1 ICI_{it} + \beta_2 ICI_{it}^2 + \beta_3 SECP\_ICI_{it} + \beta_4 GDP_{it} + \beta_5 EC_{it} + \beta_6 WS_{it} + \beta_7 INDSTR_{it} + \varepsilon_{it} \qquad \text{---(3.3)}$$

Moderating effect of industrial structure improvement: The interaction term *INDSTR_ICI* is included in Model 3.4 to capture the moderating effect of INDSTR in the relationship between the core industries' growth and environmental degradation levels. Depending on the proportion of tertiary and secondary sectors in the economy, the environmental outcomes of the core industries' growth may vary. The investigation of the moderating role of the industrial structure improvement in the relationship between core industries' growth and environmental degradation will assert whether the emphasis should be placed on the growth of the secondary sector or the tertiary sector in India for environmental benefits. A positive and significant coefficient of the interaction term will establish that an improved industrial structure will increase the adverse impact of the core industries' growth on the environment. On the other hand, the confirmation of a negative and significant moderating role will indicate that industrial structure improvement will help reduce the magnitude of the degrading environmental effects of the core industries' growth.

$$lnCO_2 = \alpha + \beta_1 ICI_{it} + \beta_2 ICI_{it}^2 + \beta_3 INDSTR\_ICI_{it} + \beta_4 GDP_{it} + \beta_5 EC_{it} + \beta_6 WS_{it} + \beta_7 SECP_{it} + \varepsilon_{it} \qquad \text{---(3.4)}$$

The abbreviations for variables considered in Models 3.2 to 3.4 have been stated in Table 3.5.

While dealing with panel data, there is always a chance that the model suffers from various econometric issues such as autocorrelation, heteroscedasticity and endogeneity. However, the Fully Modified Ordinary Least Squares (FMOLS) approach (Kao and Chiang, 2001; Pedroni, 2000; Ramirez, 2007) is capable of providing robust estimations while overcoming the possible issues of serial correlation and endogeneity in the empirical model (Chowdhury *et al.*, 2022; Farhani and Balsalobre-Lorente, 2020; Zafar *et al.*, 2020). Moreover, the FMOLS approach is suitable for offering asymptotically unbiased, normally distributed coefficient estimates even when a small sample size is considered (Kao and Chiang, 2001; Pedroni, 2000; Ramirez, 2007). Therefore, the FMOLS approach will be applied for the empirical analysis of Objective 3 to determine the long-run relationships between industrial growth and carbon emissions. Model 3.5 gives the generic form of the FMOLS (Pedroni, 2000):

$$\hat{\beta}_{FMOLS} = \frac{1}{N}\sum_{i=1}^{N}[(\sum_{t=1}^{T}(X_{it} - \bar{X}_{it})^2)^{-1}(\sum_{i=1}^{T}(X_{it} - \bar{X}_i)Y_{it}^* - T\hat{\gamma}_i)] \qquad \text{--- (3.5)}$$

Here, $Y_{it}^* = Y_{it} - \bar{Y}_i - (\hat{\Omega}_{2,1,i}/\hat{\Omega}_{2,2,i})\Delta X_{it}$ and $\hat{\gamma}_i = \hat{\Gamma}_{2,1,i} + \hat{\Omega}_{2,1,i}^0 - (\hat{\Omega}_{2,1,i}/\hat{\Omega}_{2,2,i})(\hat{\Gamma}_{2,2,i} + \hat{\Omega}_{2,2,i})$. Ω_i^0 is the contemporaneous covariance and Γ_i is a weighted sum of autocovariance.

To perform the FMOLS estimation, the variables are required to be integrated at first order, i.e. I(1) (Kao and Chiang, 2001; Pedroni, 2000). To ensure the level of integration meets this criterion, the stationarity of the variables is determined. The study conducts both ADF-Fisher chi-square and PP-Fisher chi-square panel unit root tests, as undertaken in Objective 2 (Guan et al., 2023; Xu and Lin, 2017).

3.5.2.3. Diagnostic tests:

Similar to the empirical approach of Objective 2, diagnostic tests are conducted with respect to Models 3.2 to 3.4. The Breusch-Godfrey LM test will be conducted to check the possible autocorrelation issue in the empirical models. The non-significant test-statistic is desirable because it indicates the absence of serial correlation. Similarly, the non-significant test-statistic of White's heteroscedasticity test indicates that the model is free form heteroscedasticity. Next, the results of the Pesaran CD cross-sectional test will be reported. Its non-significant test-statistic will validate the non-existence of cross-sectional dependence in the empirical models. Lastly, the possibility of endogeneity and the simultaneity bias will be tested by the Durbin test and the Wu-Hausman test. Their non-significant test-statistics will confirm the absence of such issues in the regression models. Once the test results for all these diagnostic tests are satisfied, the empirical results will be declared as non-spurious. In Table 3.6, the diagnostic tests undertaken in Objective 3 are summarized.

Table 3.6: Diagnostic tests (Objective 3)

Diagnostic test	Null hypothesis	Inference
Breusch-Godfrey LM test	There is no serial correlation.	A non-significant test-statistic would indicate absence of serial correlation.
White's heteroscedasticity test	The variance of the residuals is constant.	A non-significant test-statistic would indicate that the data is not heteroscedastic.
Pesaran CD cross-sectional test	There is no cross-sectional dependence.	A non-significant test-statistic would indicate that the residuals are not cross-sectionally dependent.

(Author's compilation)

3.5.2.4. Robustness check:

In Objective 3 also, the DOLS approach is adopted as a robustness test for the estimation results obtained from FMOLS. The advantages of the DOLS estimates are already explained with respect to Objective 2. Hence, the same is not repeated in this section of the study. The results of this approach will help determine whether the reported results are consistent and reliable. Similar to Model 2.6, Model 3.6 is applied with respect to Models 3.2 to 3.4 (Kao and Chiang, 2001):

$$y_{i,t} = \alpha_i + \beta_i x_{i,t} + \sum_{k=-k_i}^{k_i} \gamma_{i,k} \Delta x_{i,t-k} + \varepsilon_{it} \qquad \text{--- (3.6)}$$

3.5.2.5. Turning point analysis of the IEKC hypotheses:

The empirical results will determine whether an inverted U-shaped IEKC or a U-shaped IEKC exists in India's core industries. Either way, the estimation of their turning points will add further significance to the findings of the objective of building a greener industrial setup. Therefore, the present research work further conducts an additional analysis to estimate the turning points of the core industries' growth levels, beyond which they are expected to lead to a healthier nation, as per the traditional inverted U-shaped IEKC arguments. In order to estimate the turning or threshold point, Models 3.7 and 3.8 are used in the study (Gill *et al.*, 2019). Here, Y* represents the turning point of the inverted U-shaped IEKC hypothesis, without considering any moderating effects of the studied variables (i.e., SECP and INDSTR). Next, Y** calculates the turning point of the core industries' growth by incorporating the moderating effects of both variables. This empirical approach will also provide empirical evidence for the moderating effects of the considered variables.

$$Y^{*} = -\frac{\beta_1}{2\beta_2} \qquad \text{--- (3.7)}$$

$$Y^{**} = \frac{-(\beta_1 - \beta_3.MV)}{2\beta_2} \qquad \text{--- (3.8)}$$

Here, β_1 is the coefficient of ICI and β_2 is the coefficient of ICI^2 in Models 3.2. to 3.4. Next, β_3 is the coefficient of the interaction terms *SECP_ICI* in Model 3.3 and *INDSTR_ICI* in Model 3.4. Further, *MV* is the average value of the moderating variables (SECP and INDSTR, in this case).

3.5.2.6. The N-shaped IECK hypothesis:

The exploration of an N-shaped IEKC hypothesis in the Indian core industries is critical in predicting and understanding the long-term environmental influences of its industrialization. Such exploration will help take a proactive role in enabling sustainability instead of playing a reactive role. The existing literature has asserted that even after the attainment of the desirable inverted U-shaped IEKC, the industries can potentially cause further environmental damage at the attainment of a second turning point. However, no prior study has attempted such an investigation of the hypothesis in the context of the core industries. Considering the gap and its significance, Objective 3 of the present research work formulates Model 3.7. To assess the N-shaped IEKC, a cubic equation is required. Therefore, the cubic term ICI^3 is additionally included in Model 3.2. The results will confirm the N-shaped IEKC hypothesis in the core industries only if $\beta_1 > 0, \beta_2 < 0, \beta_3 > 0$. Any variation in the β-values will lead to the rejection of the N-shaped EKC hypothesis.

$$lnCO_2 = \alpha + \beta_1 ICI_{it} + \beta_2 ICI_{it}^2 + \beta_3 ICI_{it}^3 + \beta_4 GDP_{it} + \beta_5 EC_{it} + \beta_6 WS_{it} + \beta_7 SECP_{it} + \beta_8 INDSTR_{it} + \varepsilon_{it} \quad \text{---(3.9)}$$

3.6. Chapter summary:

The present chapter describes a detailed empirical approach that will be conducted for fulfilling the research objectives of the present study. It considers an unbalanced secondary panel dataset from 2005 to 2021 consisting of the eight core industries in India. In a nutshell, Objective 1 presents a comparative picture of the absolute carbon emission levels of the core industries and their environmental efficiency levels by applying the decoupling approach. Next, Objective 2 is focused on exploring the significant driving factors of the core industries' carbon emission levels by applying the REM. Here, the driving factors are categorized into economic, industrial, demographic, and environmental factors. Lastly, the relationship between industrial growth and carbon emission levels of the core industries is empirically investigated in Objective 3, with reference to the IEKC hypothesis, by applying the FMOLS estimates. Further, the estimations of the turning points of the IEKC hypothesis are also depicted and the moderating roles of two variables (i.e., environmental policy stringency and industrial structure improvement) are assessed in the relationship between the core industries' industrial growth and carbon emission levels in Objective 3. These additional analyses are expected to contribute and add more value to the findings of the overall research work. In the cases of Objectives 2 and 3, the DOLS approach is also applied as a robustness check of the reported results. The empirical results for all three objectives are discussed chapter-wise in detail in the following chapters.