Chapter 3
RESEARCH METHODOLOGY

3.1 Introduction

The research attempts to investigate the most discussed hypothesis of economic growth and environemntal quality with reference to India. The plethora of literature on the EKC (Environemntal Kuznets Curve) hypothesis have tried to study this relationship using various combination of environmental indicators and economic growth determinants. Mostly the studies are cross-sectional. Very rarely we find time series analyses. The individual country analysis is not that popular as these studies do not comrephensively investigate the underlying casuses of the growth-environment hypothesis, if it exists.

This research attempts to answer a few questions based on the following research objectives:

- To analyze and evaluate the indicators responsible for an inverted-U shape curve of Environmental Kuznets Curve in Indian context (Decomposition analysis).
- To study the behavior of environmental pollution i.e. shape of the Environemntal Kuznets Curve, over time in India during pre-liberalization and post-liberation periods.
- To locate the turning point of the Environmental Kuznets Curve for India in terms of GDP per capita.
- To do a Structural analysis of the Envrionmetnal Kuznets Curve for India.

3.2 Research Inquiries

Based on the above objectives the study attempts to explore answers to the following research questions:

- Is there an Environmental Kuznets curve for India? If yes, which portion of the curve does India lie on?
- Is the environemntal quality related to economic growth in India?
- Is the environmental quality related to demographic factors of the economy?
- Is the environemtnal quality related to governance factors of the economy?
- Is the environemtnal quality related to environmental factors of the economy?

3.3 Research gap

1. Much of EKC literature supports the "inverted-U-shape" effect and the presence of the cubic term influences the results. Whereas the structural transition models do not support income being the only decisive factor in the economic growth and environmental quality hypothesis. The need and the impact of other factors is inevitable if appropriate policy implications are to be concluded from the hypothesis.

2. The underlying cause of the inverted –U shape is considered as autonomous to income growth. As income is inappropriate proxy of a number of other variables. The use of structural model provides a better understanding of the EKC relationship.

3. Use of panel data for various countries at a point in time.

4. The starting point of all countries is different in a pool.

5. The use of various pollutants for analysis. Some have a very low turning point and some have high due to the priority of abatement and cost of abatement.,

6. There is variation in pollutant measurements. Some studies use concentration levels, some use ambient concentrations and some use emissions.

The literature on EKC is broadly based on using reduced form models and panel data analysis for a group of countries. The standpoint of these studies is to analyse the impact on various envrionmental indicators with change in the economic growth of the countries. The selection of environmental indicator depends primarily upon the availability of data. But there are some studies that have justified the selected pollution indicator along with its characteristics. The chararcteristics of the pollutants analyzed in the literature on EKC can be categorized in the following ways:

- The environemntal indicators belong to different category of environmental degradation as per the source of origin. To mention a few: water pollution, air pollution, land pollution and land pollution.
- The zone of influence of the pollutant considering the horizontal as well as vertical dimension, is defined as the damage caused by the pollutant to a particular extent/space. The damage caused by a local pollutant is experienced near the source

of emissions, whereas the regional pollutant has a larger reach. The case of global pollutant is related to the damage that affects the entire planet.

- Concentrations and emissions – The vertical zone of influence of the pollutants labels the damage as the total amount of a pollutant present in the atmosphere. Whereas an emission is the amount of pollutant released in the atmosphere from a specific pollutant source at specific and regualr inteval of time.
- The cost of abatement – The risk that the pollutants cause becomes the main criteria on which the sequence of action is decided. Some pollutants that cause severe health risks and are cheaper to abate draw attention of the authorities. The existence of EKC relationship and the concentration of studies on certain pollutants like Sulphur dioxide (SO_2) and Suspended particulate matter (SPM)demontrates the fact that certain pollutants that are cheaper to abate are taken care of at very low income levels.

- The nature of the pollutant: stock or flow pollutant – The mechanism by which waste is emitted determines the damage done to the environment. This creates a pressure on the capacity of environment to assimilate the waste generated which is termed as absorptive capacity, of the environment. A stock pollutant is a pollutant for which the environment has little or no absorptive capacity (Tietenberg, Tom 2006). They get accumulated and are persistent over time. For such enduring pollutants, the targets should be in terms of emissions path over time, and efficient level of emisisons should be thought of in a steady state of equilibrium (Perman., Roger 2018).

  On the other hand a flow pollutant, also known as fund pollutant, are pollutants for which the environment has some absorptive capacity (Tietenberg, Tom 2006). In this case as long as the emissions are less than the absorptive capacity the accumulation does not take place. Many pollutants that are emitted get absorbed and transformed into less-harmful matter for e.g. carbon dioxide gets aborbed by the plants.

- Time span of a pollutant – Although all pollution indicators under the EKC scanner are stock pollutants, they all have very short life span in the atmoshpere and can therefore can be considered as flow pollutants over the long-run. The lifespan of Sulphur dioxide (SO_2) is 1-4 days and that of Nitrogen oxide (NOx) is 4-5 days. The lifespan of Carbon Monoxide (CO) is 1-3 months while hydrocarbons have a lifespan of a few hours. The reasons for EKC studies on Carbon dioxde is mainly because of its lifeapn in the atmoshpere, that is, about 125 years, as it is a global stock pollutant.

Reason Why CO_2 emissions are important to the world

1. It accounts for more than half of the effect of greenhouse warming (Stern, Young and Drackman, 1992)

2. As less action has been taken on curbing carbon emissions as it was considered less harmful. For countries expecting an inverted U curve for CO_2 emissions would mean development with efficiency.

3. According to World Resource Institute 1991, 1992 and 1993 carbon emissions per capita has varied across the world by over two thousand fold.

4. The data is available for largest span for most of the countries of the world. This makes the analysis more relevant.

3.4 Variables and Data collection

The data on environmental indicators have been the biggest challenge for almost all researchers. The availability and reliability of data is the major issue. The reason there are varying results on EKC hypothesis is because of lack of proper data since in different countries the data is not measured using common methods. In India the collection of environmental data started in the late 80s and for many countries even later. So the availability of longer time series data on environmental indicators is missing.

The data for most of the variables used are retrieved from the World Development Indicators and International Country Risk guide. The specification of each variable is given below:

The World Development Indicators is a compilation of relevant, high-quality, and internationally comparable statistics about global development and the fight against poverty.

CO_2 emissions per capita

Variable	Unit	Time	Source	Expected Sign
CO_2 emissions per capita	Metric tons per capita	1960-2017	WDI	+ (when regressed with income per capita)

Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.

CO_2 emissions are largely a by-product of energy production and use, accounting for largest share of greenhouse gases, which is strongly associated with global warming.

The burning of carbon based fuels has increased since the industrial revolution taking the concentrations of CO_2 emissions to a high level: from an annual average of 280 ppm in the late 1700s to 401 ppm as measured at Mauna Loa in 2015—a 43 percent increase (as reported by UN- EPA, 2015). This addition is totally credited to man made activities. Through various aggrements and conventions, countries around the globe are taking measures to control carbon dioxide emissions.

Economic Indicators

Gross Domestic Product per Capita

Variable	Unit	Time	Source	Expected Sign
GDP per capita (current local currency unit)	per capita	1960-2017	WDI	Negative

GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current local currency.

Electricity production from coal sources (% of total)

Variable	Unit	Time	Source	Expected Sign
Electricity production from coal sources	(%) Percentage of total electricity production from all sources	1960-2017	WDI	Negative

Sources of electricity refer to the inputs used to generate electricity. Coal refers to all coal and brown coal, both primary (including hard coal and lignite-brown coal) and derived fuels (including patent fuel, coke oven coke, gas coke, coke oven gas, and blast furnace gas).

Electricity production is total number of kWh generated by power plants separated into electricity plants and CHP plants. The International Energy Agency (IEA) compiles data on energy inputs used to generate electricity. IEA data for countries that are not members of the Organisation for Economic Co-operation and Development (OECD) are based on national energy data adjusted to conform to annual questionnaires completed by OECD member governments. Electricity production from coal sources in India is approx. 74% of the total electricity production.

The use of energy is important for improving standard of living. But it can also damage the environment. Burning of coal releases twice as much carbon dioxide as does burning of natural gas.

Trade Openness (% of GDP)

Variable	Unit	Time	Source	Expected Sign
Trade openness	(%) Percentage of GDP	1960-2017	Our World Data	Negative

It is measured as ratio of sum of export and import to that of GDP of a country. It measures the extent to which a country is engaged in global trade. It is used to measure the relative importance of international transactions relative to domestic transactions. A lower ratio does not necessarily mean high trade barriers but may due to the size of the economy or geographical position vis-a-viz its trading partners. The aggregate value of international trade in goods and services reflects countries' integration into the world trade. Developing countries are generally more integrated as compared to the developed countries. Higher exports and imports with its developed trading partner fulfills the domestic demand.

Higher trade integration necessarily implies liberal environmental policy standards. The "pollution haven" hypothesis implies transfer of dirty industries from stricter environmental regime to a more relaxed environmental regime. Consequently the developing nations have highly polluted industries.

Industry (including construction), value added (% of GDP)

Variable	Unit	Time	Source	Expected Sign
Industry (including construction)	Value added (%) Percentage of GDP	1960-2017	WDI	Negative

Industry corresponds to ISIC divisions 10-45 and includes manufacturing (ISIC divisions 15-37). It comprises value added in mining, manufacturing (also reported as a separate subgroup), construction, electricity, water, and gas. Value added is the net

output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3 or 4.

Industrial production benefits the economy to fulfill the growth targets. But it damages the environment through air pollution, water pollution, land/soil pollution and noise pollution. It is also considered as a key factor in degrading the environment which in turn is a key factor in climate change and global warming.

Foreign direct investment, net outflows (% of GDP):

Variable	Unit	Time	Source	Expected Sign
Foreign Direct investment, net outflows	(%) Percentage of GDP	1991-2017	WDI	Negative

Foreign direct investment refers to direct investment equity flows in an economy. It is the sum of equity capital, reinvestment of earnings, and other capital. Direct investment is a category of cross-border investment associated with a resident in one economy having control or a significant degree of influence on the management of an enterprise that is resident in another economy. Ownership of 10 percent or more of the ordinary shares of voting stock is the criterion for determining the existence of a direct investment relationship. This series shows ratio of net outflows of investment from the reporting economy to the rest of the world to GDP.

The impact of FDI on the environment is a source of the EKC debate. The movement of investments from higher income economies to lower income economies is the idea behind relocating dirty industries to economies with sloppy environmental regulations. There is plethora of research on how trade and foreign investments are impacting the environmental quality.

Foreign direct investment, net inflows (% of GDP):

Variable	Unit	Time	Source	Expected Sign
Foreign Direct investment, net inflows	(%) Percentage of GDP	1991-2017	WDI	Negative

Foreign direct investment are the net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor. It is the sum of equity capital, reinvestment of earnings, other long-term capital, and short-term capital as shown in the balance of payments. This series shows ratio of net inflows (new investment inflows less disinvestment) in the reporting economy from foreign investors to GDP.

Fossil fuel energy consumption (% of total)

Variable	Unit	Time	Source	Expected Sign
Fossil fuel energy consumption	(%) Percentage of total	1960-2017	WDI	Negative

Fossil fuel comprises coal, oil, petroleum, and natural gas products.

Fossil fuels are non-renewable resources because they take millions of years to form, and reserves are being depleted much faster than new ones being regenerated. In developing economies growth in energy use is closely related to growth in the modern sectors - industry, motorized transport, and urban areas - but energy use also reflects climatic, geographic, and economic factors (such as the relative price of energy). Energy use has been growing rapidly in low- and middle-income economies, but high-income economies still use almost five times as much energy on a per capita basis.

Total energy use refers to the use of primary energy before transformation to other end-use fuels (such as electricity and refined petroleum products). It includes energy from

combustible renewables and waste - solid biomass and animal products, gas and liquid from biomass, and industrial and municipal waste. Biomass is any plant matter used directly as fuel or converted into fuel, heat, or electricity. Fossil fuel energy consumption of the total energy consumption accounts for approximately 74% in India. As it is closely knitted to the environmental quality higher energy use depletes the natural resources and creates an irreversible damage.

Demographic Indicators

Population density (people per sq. km of land area)

Variable	Unit	Time	Source	Expected Sign
Population density	Per sq. km	1960-2017	WDI	Negative

Population density is midyear population divided by land area in square kilometers. Population is based on the de facto definition of population which counts all residents regardless of legal status or citizenship--except for refugees not permanently settled in the country of asylum, who are generally considered part of the population of their country of origin. Higher population density exerts enormous pressure on the existing infrastructure and civic amenities which in turn degrades the environment around.

Several of the most densely populated territories in the world are city-states, microstates, or dependencies. These territories share a relatively small area and a high urbanization level, with an economically specialized city population drawing also on rural resources outside the area, illustrating the difference between high population density and overpopulation.

Population ages 15-64 (% of total)

Variable	Unit	Time	Source	Expected Sign
Population ages 15-64	% percentage of total population	1960-2017	WDI	Negative

Population between the ages 15 to 64 as a percentage of the total population. Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship.

Patterns of development in a country are partly determined by the age composition of its population. Different age groups have different impacts on both the environment and on infrastructure needs. Therefore the age structure of a population is useful for analyzing resource use and formulating future policy and planning goals with regards infrastructure and development.

Urban population (% of total)

Variable	Unit	Time	Source	Expected Sign
Urban population	% percentage of total	1960-2017	WDI	Negative

Urban population refers to people living in urban areas as defined by national statistical offices. The data are collected and smoothed by United Nations Population Division. For the Census of India 2011, the definition of urban area is a place having a minimum population of 5,000 of density 400 persons per square kilometer (1,000/sq. mi) or higher, and 75% plus of the male working population employed in non-agricultural activities. Places administered by a municipal corporation, cantonment board or notified town area committee are automatically considered urban areas.

In principle, urban areas offer a more favorable setting for the resolution of social and environmental problems than rural areas. Urban areas generate jobs and income and deliver education, health care and other services. Urban areas also present opportunities for social mobilization and women's empowerment which help in reducing the pressure on environment.

Literacy rate

Variable	Unit	Time	Source	Expected Sign
Literacy rate	% (percentage of total population)	1960-2017	Unesco Institute for Statistics.	Negative

Percentage of population age 15 and above who can, with understanding, read and write a short, simple statement on their everyday life. Generally, 'literacy' also encompasses 'numeracy', the ability to make simple arithmetic calculations. This indicator is calculated by dividing the number of literates aged 15 years and over by the corresponding age group population and multiplying the result by 100.

The current level of overall literacy is 64.8%, the male literacy rate is 75.3% and female is 53.7%. The gap is wider in rural areas as compared to urban areas.

Poverty rate

Variable	Unit	Time	Source	Expected Sign
Poverty rate (% of population in rural area)	% Percentage	1960-2017	RBI – various economic surveys	Negative

Poverty is a state or condition in which a person or community lacks the financial resources and essentials for a minimum standard of living. According to World Bank, poverty is pronounced deprivation in well-being and comprises many

dimensions. It includes low incomes and the inability to acquire basic goods and services necessary for survival with dignity. Poverty also encompasses low levels of health and education, poor access to clean water and sanitation, inadequate physical security, lack of voice, and insufficient capacity and opportunity to better one's life.

The Suresh Tendulkar committee provided a new method of calculating the poverty line based on per capita consumption expenditure per month or day. For rural areas, it was Rs 816 per month or Rs 27 per day. For urban areas, it was Rs 1000 per month or Rs 33 per day.

Poverty rate

Variable	Unit	Time	Source	Expected Sign
Poverty rate (% of population in urban area)	% percentage	1960-2017	WDI	Negative

Poverty is a state or condition in which a person or community lacks the financial resources and essentials for a minimum standard of living. According to World Bank, Poverty is pronounced deprivation in well-being, and comprises many dimensions. It includes low incomes and the inability to acquire the basic goods and services necessary for survival with dignity. Poverty also encompasses low levels of health and education, poor access to clean water and sanitation, inadequate physical security, lack of voice, and insufficient capacity and opportunity to better one's life.

The Suresh Tendulkar committee provided a new method of calculating the poverty line based on per capita consumption expenditure per month or day. For rural areas, it was Rs 816 per month or Rs 27 per day. For urban areas, it was Rs 1000 per month or Rs 33 per day.

Environmental Indicators

Alternative and nuclear energy

Variable	Unit	Time	Source	Expected Sign
Alternative and nuclear energy (% of total energy use)	% percentage	1960-2017	WDI	Positive

Clean energy is non carbohydrate energy that does not produce carbon dioxide when generated. It includes hydropower and nuclear, geothermal, and solar power, among others.

Alternative energy is produced without the undesirable consequences of the burning of fossil fuels, such as high carbon dioxide emissions, which is considered to be the major contributing factor of global warming.

Access to clean fuels and technologies for cooking

Variable	Unit	Time	Source	Expected Sign
Access to clean fuels and technologies for cooking (% of population)	% percentage	1960-2017	WDI	Positive

It is the proportion of total population primarily using clean cooking fuels and technologies for cooking. The source of the data is World Bank, Sustainable Energy for All (SE4ALL) database from WHO Global Household Energy database.

Data for access to clean fuels and technologies for cooking are based on the World Health Organization's (WHO) Global Household Energy Database. The data is

collected among different sources: only data from nationally representative household surveys (including national censuses) are used by WHO.

If higher proportion of the population uses cleaner cooking fuel, emissions from traditional cooking fuels will reduce. Firewood is the main cooking fuel in rural India. Relationship with

Access to electricity

Variable	Unit	Time	Source	Expected Sign
Access to electricity (% of population)	% percentage	1960-2017	WDI	Positive

It is the percentage of population with access to electricity. Electrification data are collected from industry, national surveys and international sources. It is collected on annual basis.

Keeping energy supplies reliable and stable when attempting to rapidly decarbonize power systems is a major challenge for countries all over the world. To promote economic development and community stability, a growing number of countries are relying on safe and stable electricity supplies. As more powerful and less carbon intensive forms of power become available, this dependence is expected to increase.

People practicing open defecation

Variable	Unit	Time	Source	Expected Sign
People practicing open defecation (% of population)	% percentage	1960-2017	WDI	Negative

It refers to the percentage of the population defecating in the open, such as in fields, forest, bushes, open bodies of water, on beaches, in other open spaces or disposed of with solid waste.

Governance Indicators

The Worldwide Governance Indicators (WGI) are a research dataset summarizing the views on the quality of governance provided by a large number of enterprise, citizen and expert survey respondents in industrial and developing countries. These data are gathered from a number of survey institutes, think tanks, non-governmental organizations, international organizations, and private sector firms. The WGI do not reflect the official views of the World Bank, its Executive Directors, or the countries they represent. The WGI are not used by the World Bank Group to allocate resources.

Political Stability and Absence of Violence/Terrorism

Variable	Unit	Time	Source	Expected Sign
Political Stability and Absence of Violence/Terrorism	ranges from approximately -2.5 (weak) to 2.5 (strong) governance performance)	1996-2018	WGI	Negative

Political Stability and Absence of Violence/Terrorism measures perceptions of the likelihood of political instability and/or politically-motivated violence, including terrorism. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5.

Government Effectiveness

Variable	Unit	Time	Source	Expected Sign
Government Effectiveness	ranges from approximately -2.5 (weak) to 2.5 (strong) governance performance)	1996-2018	WGI	Negative

It reflects the perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5.

Regulatory Quality

Variable	Unit	Time	Source	Expected Sign
Regulatory quality	ranges from approximately -2.5 (weak) to 2.5 (strong) governance performance)	1996-2018	WGI	Negative

It reflects the perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5.

Rule of Law

Variable	Unit	Time	Source	Expected Sign
Rule of Law	ranges from approximately -2.5 (weak) to 2.5 (strong) governance performance)	1996-2018	WGI	Negative

It reflects the perceptions of the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, property rights, the police, and the courts, as well as the likelihood of crime and violence. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5.

Control of Corruption

Variable	Unit	Time	Source	Expected Sign
Control of Corruption	ranges from approximately -2.5 (weak) to 2.5 (strong) governance performance)	1996-2018	WGI	Negative

It reflects the perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests. Estimate gives the country's score on the aggregate indicator, in units of a standard normal distribution, i.e. ranging from approximately -2.5 to 2.5.

3.5 Research Design

The research framework here is divided into three parts based on the kind of methodology adopted to investigate the research questions:

3.6 Part A: Traditional methodology of EKC

In this part the traditional EKC model is used to analyze and evaluate the economic growth-environment relationship in India. The following equation is used

$$P_{it} = \alpha + \beta_1 Y_{it} + \beta_2 Y_{it}^2 + \beta_3 Y_{it}^3 + \varepsilon_{it} \quad \textbf{(1)}$$

Where p = CO_2 emissions per capita, y = GDP per capita measured, t is an index for time and ε is the normally distributed error term. Allowing to test the various forms of environmental-economic relationship, a) $\beta_1 > 0$ and $\beta_2 = \beta_3 = 0$ reveals a monotonically increasing linear curve, signifying that rising incomes are associated with rising levels of emissions, b) $\beta_1 < 0$ and $\beta_2 = \beta_3 = 0$ reveals a monotonically decreasing linear curve, signifying that rising incomes are associated inversely, c) $\beta_1 > 0$, $\beta_2 < 0$ and $\beta_3 = 0$ reveals a quadratic relationship depicting an EKC and d) $\beta_1 > 0$, $\beta_2 < 0$ and $\beta_3 > 0$, reveals a cubic polynomial representing the N-shaped figure (de Bruyn, S. M., et.al, 1998).

Broadly, the time series analysis in earlier studies can be categorized as:

- Studies based on the assumption of stationarity of the times series adopted curve estimation technique to forecast the income turning points and
- Studies based on the assumption that time series is non stationary and using unit root test and cointegration analysis.

Friedl & Getzner (2003) examined the existence of EKC for an individual country and found an N-shaped curve. Galeotti et. al. (2006) examined the existence of EKC for OECD and Non-OECD countries and found that EKC relationship existed for OECD countries only. The use of moving average, differencing and smoothing can be seen in the testing EKC hypothesis.

Hypothesis developement

The earlier literautre on EKC has mostly used panel data and has conducted a cross section analysis mainly due to unavailability of proper time series data. The empirical studies show the turning points and even the shapes of the EKC which vary widely across countries and pollutants. The availability of data determines the pollutant used for the study. The data on many environmental indiciators is either missing or not measured at all. This makes the studies focus either on air pollution or water pollution. This study focuses on the air pollution indicator -carbon dioxide emissions per capita in India. The reasons for using the air pollution indicator and the advantages and disadvantages of emissions per capita and concentration data has been mentioned earlier.

The use of gross domestic product measured in various dimensions has been widely used to measure the income of countires. The use of net domestic product nd green gross domestic product has been proposed by earlier researchers. No strong evidence was found in use of other concepts of gross domestic product. The results more or less remained the same.

Here the study mainly focuses, on the evidence of decoupling between economic growth represented by per capita gross domestic product and environemntal quality represented by carbon dioxide emissions in India. To test this the following hypotheses have been formulated for time series analysis. The nationwide data for India on CO2 emissions per capita and GDP per capita (at market prices) from the period of 1960 to 2019 is collected from the World Development Indicators database.

Theoretical hypothesis

Null hypothesis - H0: Economic growth and Environmental quality (CO_2 emissions) have no significant relationship.

Alternate hypothesis - H1: Economic growth and Environmental quality (CO_2 emissions) have significant relationship.

Statistical hypothesis

Null hypothesis - H0: $\beta_1 = \beta_2 = \beta_3 = 0$

Alternate hypothesis - H1: $\beta_1 \neq \beta_2 \neq \beta_3 \neq 0$

3.7 Part B: Decomposition analysis

Structural Equation Modelling (SEM) is a family of statistical models to depict the relationship between the observed variables with the aim to test the theoretical model developed by the researcher. SEM can be used to answer research questions involving direct and indirect reltionship of one or more independent variables or one or more dependent variables. However, the primary aim of SEM is to determine and validate the hypothesised causal model. Hence, SEM is a confirmatory technique. Therefore, a model is to be specified a priori along with the number of parameters, covariances, path coefficients and variances (Schumacker, R. E., & Lomax, R. G., 2004)

The advantages of using SEM over regression model are:

1. SEM requires multivariate normality whereas some methods do not require normality. In a real world situtation there are differentials. Such an assumption of normality is expendable.
2. Minimum loss of information as factor reduction technique (such as Principal Component anslysis) is a based on linear transformation of the regressors.
3. With the dispersed set of outcomes, factor reduction technique is more helpful as it maximises the variance rather than minimizing the distance.
4. SEM includes more flexible assumptions to allow interpretations in case of multicollinearity.
5. Use of confirmatory technique reduces the measurement error as it allows for multiple indicators per latent (unobserved) variable.
6. The desirablity of the SEM models to be tested overall rather than testing coefficients individually.

7. The ability of SEM to test mediating models rather than testing an addititve model makes it more robust.
8. It provides structure to the basic econometric models which are otherwise simply reduced form equations.
9. The path coefficients become unbiased by error terms, as the strucutral model allows for estimation of erro term on each observed variable.
10. SEM allows for assimilation of many statistical strucutral relationships into one model including causal models.

This study uses non-traditional SEM models for analyzing the relationship between economic growth and environmental quality in India. For this purpose, 21 variables are divided into four categories as follows:

- Economic factors
- Demographic factors
- Environmental factors
- Governance factors

The basis of selection of these variables is the skilled theoretical knowledge and the research study by Bollen, K., & Lennox, R. (1991) on the applicability of the guidelines of criteria to select good measures for the study. The study uses the variables to create a composite index using Principal component analysis technique. The oldest and the best known technique for investigating relationship between variables and cosntructs is factor analysis. It is a data reduction technique. Confirmatory factor analysis (CFA) picks an established set of observed and latent variables based on theoretical and empirical knowledge. It tries to focus on how and to what extent the observed variables are related to latent variables. Within the framework of SEM, the CFA model represents what is called as measurement model.

Principal component analysis linearly transforms an original set of variables into uncorrelated smaller set of factors. Consider random variables x_1, x_2, x_3 ….

$$\mathbf{y1 = a11x1 + a12x2 + a13x3 + \cdots \ldots a1nxn = \sum_{i=1}^{n} aijxi} \qquad \mathbf{(2)}$$

where y1 is the component created, which is defined as variance of y_1 is maximised subject to the constraint that sum of squared weights is equal to 1. That is,

$$\sum_{i=1}^{n} aijxi = 1 \qquad (3)$$

If the sum of sqaured is maximised then the varance of y_1 is also maximised.

According to earlier studies a shift in the observed variables leads to an expected shift in the indicator. These measures are termed as effect indicators (Bollen & Ting 2000). They are not appropriate for all situations. Some variables are better treated as determinants rather than effect indicators. Variables when treated as determinants are called causal indicators. Incorrect specification of variables as effect or cause indicators can produce misspecified models.

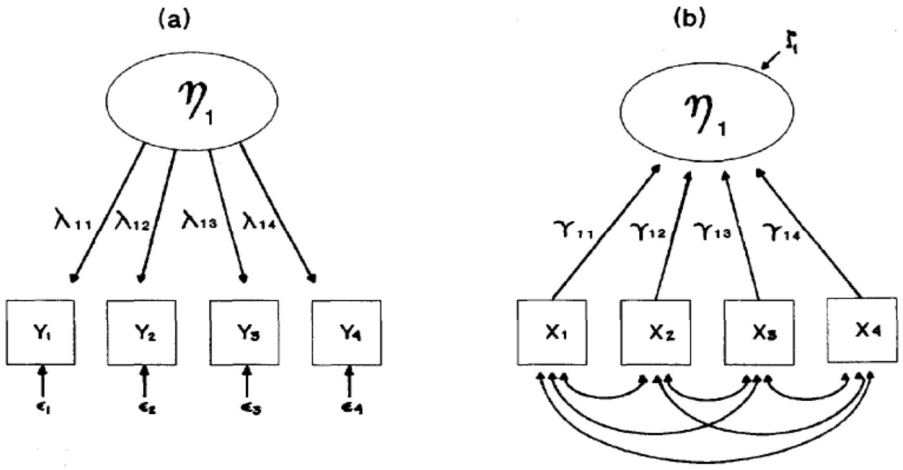

Figure 3.1: Path diagram a) effect indicator b) causal indicator
Source: Bollen, K., & Lennox, R. (1991)

The study models the variables as causal indiciators as these variables are the determinants of the factors.

3.8 Research question with Hypothesis

The factors that are extracted using factor analysis can also be termed as the causes of environemtnal degradation in the economy.

Economic factors - The most prominent of these factors is the economic factors that influence the environmental quality of the economy. In developing countries

environmental quality is often considered a luxury good as compared to developed countries where it is considered a neccessity. Therefore, society pays more attention to production and trade activities. Environemtnal regulations are neglected and emphasis is on economic growth rates. The influence of foreign investment incorporated in the pollution-haven hypothesis have been observed in the case of developed countries. On the basis of the foundational argument, that trade and foreign investment influence environmental quality the study tests the following alternate hypotheses:

H1a: Degree of globalization in the economy influences the environmental quality depicting an EKC relationship.

The EKC hypothesis suggests that in the initial stage of the growth process the consumption of energy is very high with pollution intensive technologies. As growth accelerates, higher income leads to improved enviromental quality leading to a shift towards less-pollution intensive technologies. Consequently, developing countries have higher energy consumption from natural resources which is pollution intensive. As income increases more environmental friendly techniques are adopted which improves the environmental quality in the economy.

H1b: A higher level of fossil fuel consumption influences the enivronmental quality depicting an EKC relationship.

The strucutre of the economy has a great influence on the environment. An agrarian economy in the initial stage extracts natural resources. With higher level of natural resource exploitation the environmetnal quality degrades. On the other hand, an industrial economy with higher share of manufacturing in the GDP exerts negative externatilities on environment. A service based economy is less pollution intensive.

H1c: A higher share of manufacturing in GDP degrades the environmental quality depicting an EKC relationship in the long run.

Developing countries are characterised by the use of higher electricity consumption as it reflects higher manufacturing activities in the economy. However, electricity consumption in the country is also an indication of higher GDP from industrial and manufacturing activities. More importantly the source of electricity production matters.

Use of renewable resources to produce electricity requires highly efficient technologies. Developing countries need more capital resources to resort to these technologies. Thus higher electricity generation from non-renewable resource leads to depletion of natural resources degrading environmental quality in the economy.

H1d: A higher share of electricity generation from non renewable resources reflects the EKC relationship.

Demographic factors – The development of the economy plays a vital role in improving environmental quality as it is one of the important development indicator. The literature on EKC has tested demographic variables as independent to find empirical evidence for EKC. Demographic variables might have mixed influence on environmental quality of the country. The percentage of the population living in urban areas (urbanization) impacts environmental quality both positively and negatively. Increased urbanization can cause better civic amenities and waste disposal techniques emphasizing the need for cleanliness and better environmental quality. On the other hand, increased urbanization can also lead to increased industrial and construction activities and transportation degrading the environmental quality. The net impact of urbanization depends on the level of income and development in the economy. In developing economies this net effect is positively correlated with urbanization activities, implying that higher levels of urbanization exert pressure on environmental quality.

H1e: A higher share of urban population in total population negatively influences the environmental quality depicting an EKC relationship.

Higher urbanization levels also imply higher population density putting pressure on civic amenities and natural resources around the urban area. The connectivity with nearby rural areas enhances road and rail connectivity with increased transport facilities being built causing exploitation of natural resources. The impact of higher population density on air quality also causes severe respiratory diseases. The age composition of the population indirectly exerts pressure on environmental quality. Resource mobility to provide better and suitable employment opportunities to the youth increases the

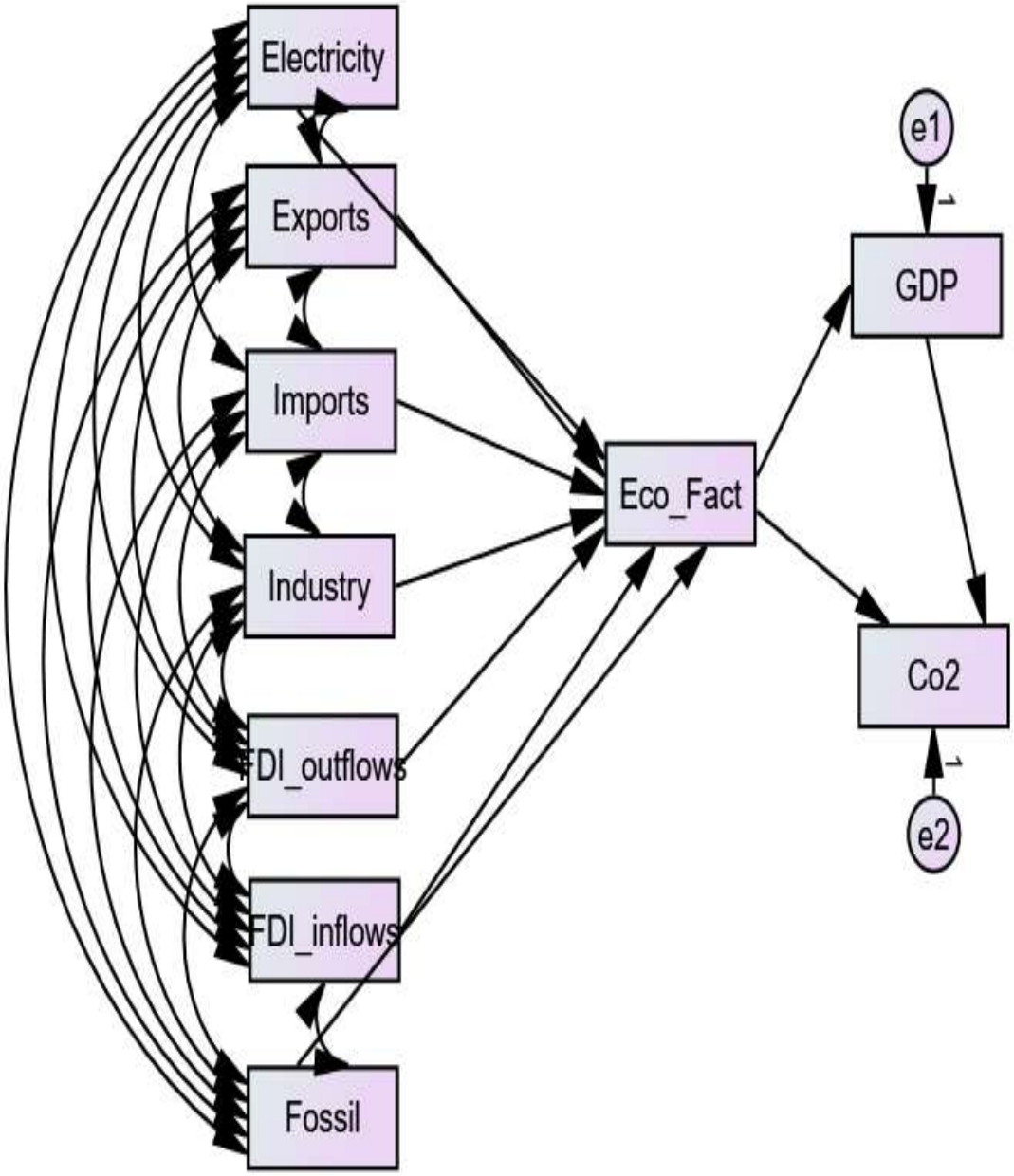

Figure 3.2: Economic effect (factor)
Source: Created by researcher in AMOS

industrial and manufacturing activities in the economy. The older population group feel vulnerable to health problems caused by pollution forcing the authorities to impose stricter environmental regimes.

H1f: Higher population density negatively influences the environmental quality causing an EKC relationship.

H1g: Age composition of the population indirectly exerts pressure on the environmental quality in the economy.

The other demographic development indicators are literacy and poverty. Higher literacy rates directly influence income of the individuals. The opening up of employment opportunities fulfills basic requirements. Consequently, the standard of living improves reducing the poverty numbers in the economy.

H1h: Higher literacy indirectly influences the environmental quality causing an EKC relationship.

H1i: Poverty rates negatively influence environmental quality degrading the environment.

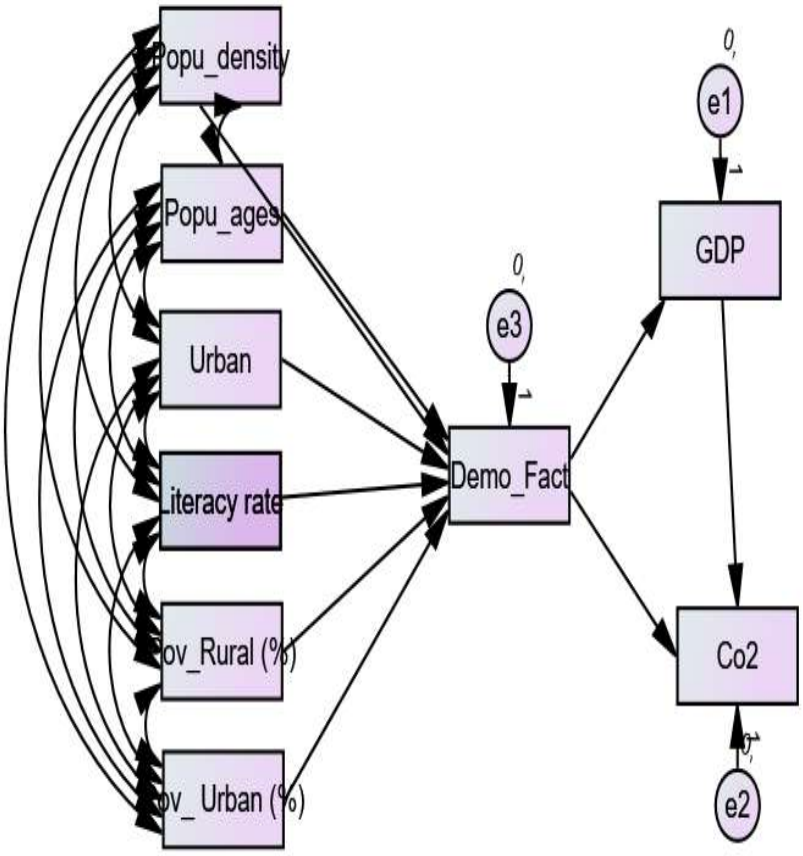

Figure 3.3: Demographic effect (factor)
Source: Created by researcher in AMOS

H1j: Higher share of alternative fuel in total fuel consumption reduces environmental damage.

H1k: Higher access to clean fuel technologies improves environmental quality.

H1l: Access to electricity indirectly impacts environmental quality depicting an EKC relationship.

H1m: Open defecation directly damages the environment depicting an EKC relationship.

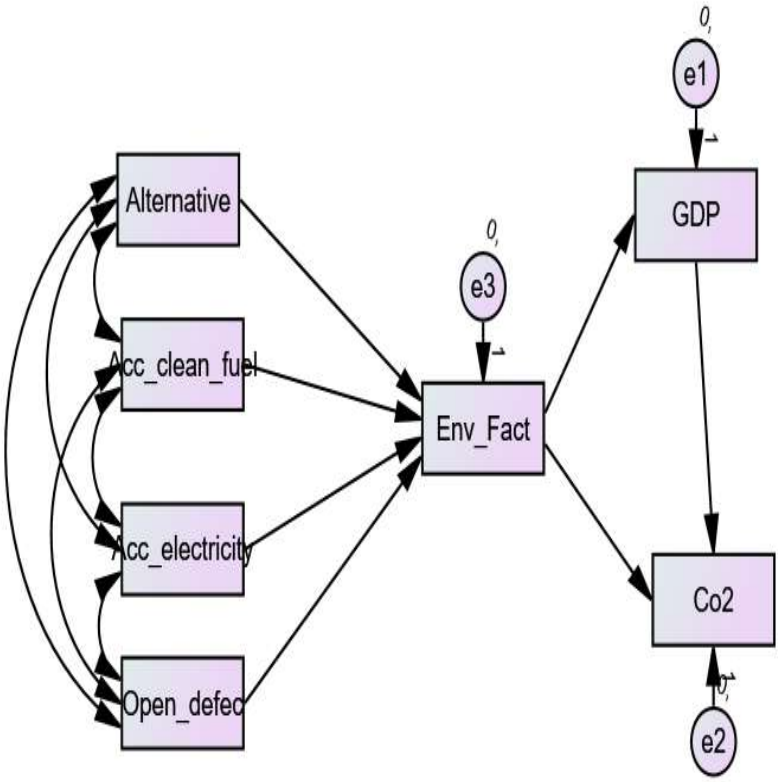

Figure 3.4: Environment effect (factor)
Source: Created by researcher in AMOS

Governance factors - In developing countries where economic growth and earning higher incomes, is the only objective for venturing into economic activities, many encounters with the law hinders the growth process. However, these general rules and regulations pertaining to economic activities are designed to maintain efficiency and equity in the system. These variables measures the quality of public institutions and

governance. Good governance is measured by level of corruption, stability of government, property rights and democratic accountability. The World Governance Indicators measures an aggregate indicator using six dimensions of governance. These dimensions are: voice and accountability, control of corruption, government effectiveness, rule of law, regulatory quality and political stability. These dimensions are of vital importance in developing countries. Amidst weak judicial system, widespread corruption and lack of democracy reduces the effectiveness of the efforts initiated to improve the stability and environmental standards in the country.

H1n: Corruption reduces the enforcement of rules and regulations indirectly influencing environmental quality.

H1o: Political stability and absence of violence indirectly enhances environmental quality.

H1p: The effectiveness of government directly impacts environmental quality.

H1q: Better quality of public institutions leads to less environmental pressure.

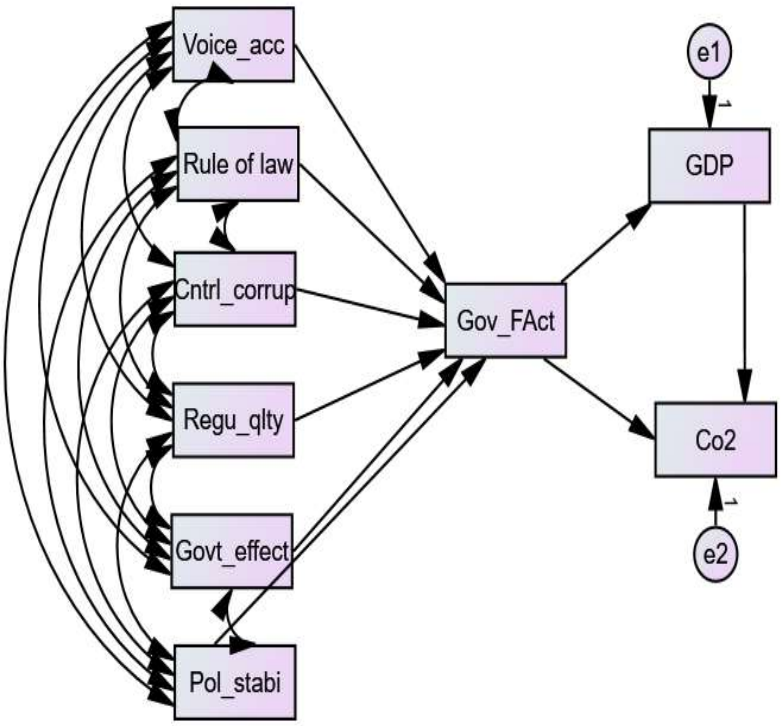

Figure 3.5: Governance effect (factor)

Source: Created by researcher in AMOS

3.9 Part C: Structural Equation Model

Structural equation modeling (SEM) serves purposes similar to multiple regression, but in a more influential way which takes into account the modeling of interactions, nonlinearities, correlated independents, measurement error, correlated error terms, multiple latent independents each measured by multiple indicators, and one or more latent dependents each with multiple indicators (James Gaskin – statwiki.kolobkreations.com).

Structural equation modelling is a family of statistical techniques which incorporates the mesaurement model and the path analysis or the structural model. The general structural model can be decomposed into two submodels: a measurement model and a structural model. A measurement model is a part of the entire structural equation model diagram that is analogous to factor analysis. It includes all the variables that "load" onto the contruct, their realtionships, variance and erorrs. A structual model is a part of structural equation model that includes all the variables to be measured in the model. The measurement model defines the relationship between the observed dependent and the observed independent. The strucutal equation model is defined by two equations:

Structural equation: $\eta_2 = \beta\eta_1 + \Gamma\xi + \zeta$ **(4)**

Measurement model equations: $y = \Lambda y + \varepsilon$ and $x = \Lambda x + \delta$ **(5)**

where **x** and **y** are vectors of observed indicators of exogenous (predictor) and endogenous (response) composite variables, $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are vectors containing the individual exogenous and endogenous composite variables, $\boldsymbol{\Lambda}\mathbf{x}$ and $\boldsymbol{\Lambda}\mathbf{y}$ are vectors of coefficients relating indicators to composites, **B** and $\boldsymbol{\Gamma}$ are coefficient matrices for effects of endogenous and exogenous observed variables on endogenous observed variables, $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are vectors of measurement errors for **x** and **y**, and $\boldsymbol{\zeta}$ is a vector of errors for the $\boldsymbol{\eta}$ variables. For exogenous observed variables, $\boldsymbol{\xi}$, their variances are represented by the diagonal elements of the matrix $\boldsymbol{\Phi}$ while the off-diagonal elements of the matrix are the covariances. The variances of the error terms for the endogenous observed variables ($\boldsymbol{\zeta}$) are contained within the diagonal elements of the ψ matrix while

the off diagonal elements of that matrix represent any covariances among errors. Errors are expected to be uncorrelated.

The composites are specified by $\boldsymbol{\Gamma}$ that corresponds to variables that create the composite. Any of these composites, can have zero error variance. This is only possible when multidimensional variables are used to create a composite. Blalock (1964), points out that the classic representation of latent variables is not always appropriate, because in particular, the indicator variables might be determinants of the exogenous variable rather than effects. Many researchers (Bollen 1984; Bollen and Lennox 1991) have questioned the traditional approach of SEM in representing the theoretical constructs. A related issue is the use of composite variable in representing the theoretical constructs in social sciences based solely on causal (determinants) indicators. The use of composite variables lays great importance to the theoretical expertise in defining the observed theoretical concepts, also known as constructs. The construct serves as technique in understanding the underlying cause between the two observed variables and theory. The composites have the capability of representing heterogeneous causes, which is very useful in social sciences. The observed construct (exogenous) is a collective influence of its determinants. As the error variance can be specified as zero (0), the values of a construct are determined by its causes.

The composites can be categorized into: a) fixed weight composites and b) unfixed weight composites. This study uses principal component analysis in deriving the composites pertaining to an observed construct.

The model is strictly confirmatory as it proposes a single model on the basis of which the hypothesis is either accpeted or rejected. A full strucutral model that specifies causality from a single direction is termed as recursive model. This study focuses on recursive model as the one represented in figure 3.6. This path diagram represents a visual framework of the relations between observed independent and observed dependent that the study assumes for research. A path analysis functions the mediation which assumes that a variable can influence the outcome directly as well as indirectly

through another variable (Yi Fan, et. al., 2016). Mediation refers to a situation where the variables included have a causal relation with each other. The study applies mediation to evaluate the direct and indirect causes of environmental degradation in India. Mediation describes a much complex relationship. There is a direct effect between an independent variable and a dependent variable. There is also an indirect effect between an independent variable and a mediator variable and between a mediator variable and a dependent variable. When we include a mediator, the direct effect changes as a result of decomposing the causal process into direct effects. The degree to which the direct effect changes as a result of decomposition of the causal process is called as the mediating effect The causes are economic, demographic, environmental and governance indicators. The test is to be conducted using AMOS 22.0, a software for formulating, fitting and testing structural equation models. The data file used in AMOS was processed in SPSS.

Assessing the model fit includes: Incremental fit, Absolute Fit and Parsimony Fit measure i.e. checking the statistical validity, convergence of the procedure, model identification statistics, parameter estimation and comparing covariance matrix to the goodness to fit measure.

The Absolute fit measures compare the intercorrelation matrix to the identity matrix. That is, the number of variables is compared to number of parameters. If both are equal, no point in running the model. If the intercorrelation matrix is not equal to the identity matrix, there is difference between the observed correlation matrix to the estimated correlation matrix. This measure thus uses model chi-square as the common fit test. In AMOS chi-square value is called CMIN. If the chi-square is not significant the model is regarded as acceptable. The value of CMIN should be less than 3. Relative chi-square is the chi-square index divided by degrees of freedom (CMIN/*df*). The fit criteria vary across researchers. RMSEA (Root mean square residual) represents square root of the average of the covariance residuals. The value of Zero is considered as perfect fit, but the maximum can be infinite.

The Incremental fit measures compare a hypothesized model with an independent or null model. The comparative fit index (CFI) compares the hypothesized model to

independent model on the one extreme and saturated model to the other. An independent model is the one with number of variables equal to number of parameters, that is, each variable represents a factor. It is also called null model because there is zero correlation among the variables. On the other hand, a saturated model is that in which the number the estimated parameters is equal to number of data points (Bryne., B. 2010). In incremental fit index (IFI) first difference between chi-square of the independent model and chi-square of the hypothesized model is estimated. Though values that exceed 0.9 are acceptable, IFI can exceed 1. Therefore, the comparative fit statistics CFI, IFI, NFI and TLI should have values closer to 1. But it may vary as per researcher.

The Parsimony fit, also known as information theory goodness fit measure, is applicable when the model is estimated using maximum likelihood estimation. The parsimonious fit indices are AIC (Akaike Information Criterion), BCC (Browne-Cudeck Criterion) and CAIC (Consistent AIC). These indices are used to compare different models. The model with the high fit is considered meaningful. The values closer to zero are ideal.

3.10 Research question with Hypothesis

Based on the arguments discussed in the previon sections and earlier chapters, the study formulates the following hypotheses with respect to the Environmental Kuznets curve (EKC).

- Economic factors influence environmental quality as the mediating factor.
- Demographic factors influence environmental quality as the mediating factor.
- Environmental factors influence environmental quality as the mediating factor.
- Governance factors influence environmental quality as the mediating factor.

The mediation hypothesis is supported if the interaction path between the various constructs, GDP and CO_2 emissions is significant. Therefore, through mediation this study attempts to search for the underlying factors that influence the economic growth - environmental quality relationship. The decomposed mediated models are shown in figure 3.2, 3.3, 3.4and 3.5. The overall structural equation model is presented in figure 3.6.

All the hypothesized models explore the relationship between (i) economic factors and environmental quality (ii) demographic factors and environmental quality (iii) environmental factors and environmental quality (iv) governance factors and environmental quality in India. Here the study captures the impact of individual factors and compiles these results to estimate the overall model.

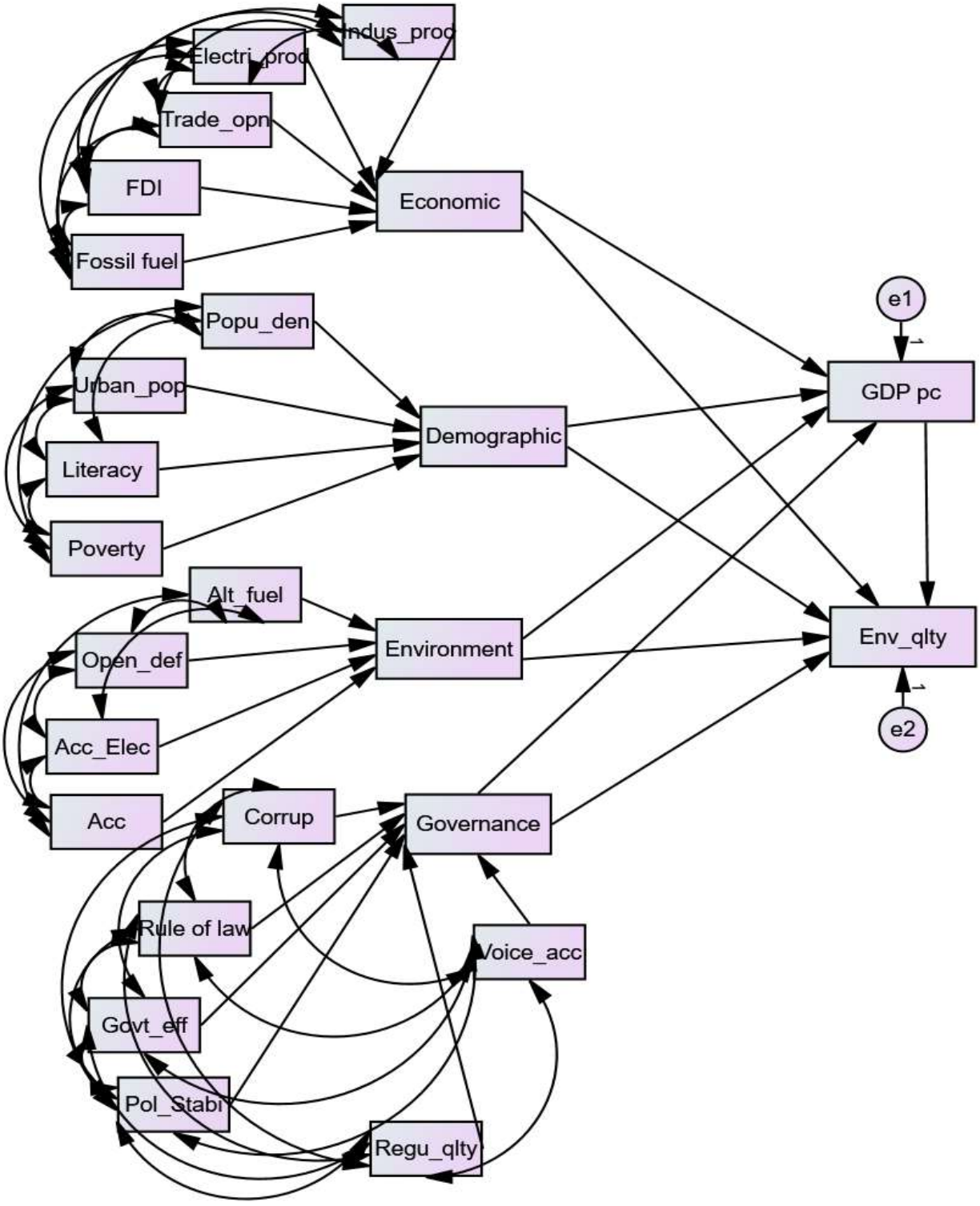

Figure 3.6: Hypothesized Path model.

Source: Created by researcher in AMOS

References

Blalock, H. M. JR. (1964).Causal Inferences in Nonexperimental Research. Chapel Hill: University of North Carolina Press.

Bollen, K. A., & Ting, K. F. (2000). A tetrad test for causal indicators. Psychological methods, 5(1), 3.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. Psychological bulletin, 110(2), 305.

Byrne, B. M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. International journal of testing, 1(1), 55-86.

De Bruyn, S. M., van den Bergh, J. C., & Opschoor, J. B. (1998). Economic growth and emissions: reconsidering the empirical basis of environmental Kuznets curves. Ecological Economics, 25(2), 161-175.

Folmer, H., & Tietenberg, T. H. (Eds.). (2005). The international yearbook of environmental and resource economics 2005/2006: A survey of current issues. Edward Elgar Publishing.

Friedl, B., & Getzner, M. (2003). Determinants of CO2 emissions in a small open economy. Ecological economics, 45(1), 133-148.

Galeotti, M., Lanza, A., & Pauli, F. (2006). Reassessing the environmental Kuznets curve for CO2 emissions: A robustness exercise. Ecological economics, 57(1), 152-163.

Gaskin, J. [James Gaskin]. (2014, May 8). SEM BootCamp 2014 Series [Video Files]. Retrieved from https://www.youtube.com/watch?v=C_Jf4l0PFl8

Schumacker, R. E., & Lomax, R. G. (2004). A beginner's guide to structural equation modeling. psychology press.

Stern, P. C., Young, O. R., & Druckman, D. E. (1992). Global environmental change: Understanding the human dimensions. National Academy Press.

World development indicators 2019. Washington, D.C.: The World Bank.