Distribution Pattern of Carbon Footprint Using Machine Learning Algorithms

Tinni Chaudhuri[1*] , Banhi Guha[2], Pankaj Kumar[3], Pulak Konar[4]

[1]Dept of Statistics, Amity University, Major Arterial Road (South-East), Rajarhat, Newtown,West Bengal 700135, India.

[2] Xavier Business School, St. Xavier's University, Kolkata, West Bengal – 700 160, India.

[3]Dept of Statistics, Amity University, Major Arterial Road (South-East), Rajarhat, Newtown, West Bengal 700135, India.

[4]Dept. of Mathematics, The ICFAI University, The ICFAI University, Tripura, Kamalghat, Mohanpur,West Tripura – 799210, India.

[1]*tinnistat1991@gmail.com** [2]*banhi.guha@gmail.com*, [3]*pankajkumar@kol.amity.edu* [4]*pulak konar@gmail.com*

Article Info

Page Number: 4602 - 4619

Publication Issue:

Vol 71 No. 4 (2022)

Article History

***Article Received**: 25 March 2022*

***Revised**: 30 April 2022*

***Accepted**: 15 June 2022*

***Publication**: 19 August 2022*

Abstract

The article deals with global carbon emission distribution through comparisons with different types ofeconomic determinants among countries. The association amongCO2 emission, Export, Import, FDI net inflows, GDP and Population has been performed through construction of different clusters based on their CO2 emission values followed by the application of fuzzy regression model along with LASSO and ridge methods. The methodology and results thus obtained are analyzed with respect to the current global situation. Several investigations have been established to depict the emission scenario throughout the diversified set of countries.

Key words :carbon emission, fuzzy regression, LASSO, ridge regression, FDI, GDP, population.

I. INTRODUCTION

Human activities are causing enormous changes in the global ecosystem. The increased concentration of greenhouse and ozone damaging gases in the atmosphere is indisputably linked to human activities (Spangenberg, 2007). Economic growth and environmental protection are the two fundamental issues that humanity is currently facing. Preserving the environment has become one of the top contemporary issues for both developed and developing countries as concerns about global warming and climate change are heightened by the decline in environmental quality. The primary causes of climate change and increased global warming are greenhouse gas emissions. (Frankel and Romer, 1999) financial liberalization and development were demonstrated to can draw FDI and higher levels of R&D investment, which can speed up economic growth and, in turn, have an effect on the dynamic of environmental performance.Financial development, according to both (Frankel and Rose, 2002) and (Birdsall and Wheeler, 1993) presented developing nations with the motivation and opportunity to apply modern technology to aid in clean and environmentally friendly production, so strengthening both the sustainability of regional development and the global environment.

In the study (Kumar, Shukla, Muhuri, and Lohani,2019) proposed a novel method for estimating a country's GDP from statistics on its carbon emissions.They have proposed different models and techniques such as transfer learning, neural network and machine learning. They have also modelled the dataset with type-I fuzzy and type-II fuzzy sets and predicted GDP of a developing nation.

Economic variables primarily influence carbon emissions as a result of economic development levels and advancement (Wang and Wang, 2014). Economic growth tends to raise consumption and production demand, which are the key variables influencing carbon emissions (Yang, Wang and Pan, 2014). An inverted U-shape is commonly found between development and emissions that is connected to levels of local economic development (Ahmad, Zhao, Shahbaz, Bano, Zhang, and Wang S. et al, 2016). As a result, the economic development model also describes regional emission reductions. Long-term investment, for example, has been utilized to foster economic development in several Chinese cities, allowing technical advances and energy structure improvements to offset some of the disadvantages of rising emissions (Zheng, Zhang, Davis, Ciais, Hong, Li, et al, 2016)

There are quite number of studies that have studied the relation between CO2 and economic or financial development. However, in this study we have incorporated other developmental indices like the effect of energy consumption, FDI, import-export and most importantly population.The models we have built have tried to explain and account for the major significant factors responsible for worldwide increment of carbon footprint. Since the range of CO2emitted is highly variable depending on if the level of other covariates in a given nation, we have created clusters diving the countries into three groups namely Low, Mid and High emission groups. This segregation has led to better and more precise results within each group by all the three methodologies of Fuzzy regression, LASSO and Ridge regression models.

II. LITERATURE REVIEW

The majority of the research discovered a connection between energy use and economic expansion., particularly in OECD nations (Lee et al., 2008), G7 nations (Narayan and Smyth, 2008), African nations (Akinlo, 2008; Wolde-Rufael, 2009), Central American nations (Apergis and Payne, 2009), South American nations (Yoo and Kwak, 2010), Middle Eastern nations (Al-Iriani, 2006), and Asian nations (Sharma, 2010).They demonstrate that, in the long run, economic growth has a Granger causal effect on energy consumption, and that, in the near run, energy use predicts output growth.

(Murthy, Panda and Parikh, 2007) examined the potential impact of the various CO2 emission reduction strategies on economic expansion. They worked with over a 35-year time data for India and came to the conclusion that for quite some time India does not have any obligation to reduce its carbon emissions as it will result in lower GDP and higher poverty. Also, to reduce its carbon emission by 30% for the next 35 years, India will need $87 billion to $280 billion of capital inflow which may come from private firms by setting up carbon reduction projects but to maintain the welfare at the same level if any CDM project gives India its share less than $14 per tC, India should not accept it. Moreover, India can remain a net seller only for around 25 years and if this opportunity is missed, it would be difficult to persuade them to join in a global effort to reduce carbon emissions. Using Granger's approach and Dolado and Lütkepohl's approach, the static and dynamic causal relationship between primary energy consumption, gross domestic product, and CO2 emissions for India between 1970 and 2007 was studied (Tiwari, 2011). They found no evidence of cointegration relationship among the test variable in the presence of structural breaks. He suggested Indian policy makers to cut primary energy consumption as it is not a source of economic growth.

On the continental level also, studies suggested (Sharma, Kautish et.al, 2019) how financial crisis has contributed to the pollution level in the using generalized method of moments. This study was conducted on five south Asian countries where the authorsfound that in the forthcoming three decades primary energy across globe will have an increase in demand by 30 percent while in developing south Asian countries it is expected to increase by 90 percent resulting in significant energy driven pollution across the globe by 45 percent.China, the Middle East, and South Asian nations will account for 75% of the growth in energy-related pollution. The main reason for this is because of the inflows to India from different countries in the manufacturing and constructions sectors which are the largest emitters of pollutants, and it produces a positive impact on CO_2 emissions in the said region.

(Chotia and Pankaj, 2019) studied data of 9 countries on CO_2emission, population and economic growth for over 30 years data using stochastic model and concluded that there is a positive relationship between population and carbon emissions and an increase of 1% population will lead to more than 1.7% increase in carbon emissions and this figure will be higher for developed countries. Their suggestion was that nations should primarily focus on population control and developed countries should focus more on using clean energy while for developing countries it will take much more time to achieve the same feat of using clean energy comparatively.

Several country specific panel data andtime series models were built by researchers on energy consumption, GDP and related indicator variables where an inverted U-shaped relationship between was observed between FDI inflows and CO_2 emissions.(Bharadwaj,2020) studied with GDP net inflows, energy use and CO_2 emissions data in between 1980-2018 stated that FDI inflows, GDP have a positive impact on CO_2 emissions. He suggested to utilize renewable energy technologies and it can be achieved by a policy mix of mandatory and non-mandatory environmental regulation with economic incentives.

Similar variables were used in the study conducted by (Rana and Sharma, 2020)in which data comprised of FDI, CO_2 emissions, trade openness, energy consumption and technology gap for the period 1980-2014 of Indian subcontinent. They observed uni-directional causality among CO_2 and energy consumption and no noticeable causality from energy consumption to CO_2values. According to the outcomes of the studythey found insights on FDI in Indian economy suggesting the Indian government to take strict measures and keep an eye on the foreign investment inflows since our country is expecting cleaner technologies however it is not getting materialized which in turn is evoking environmental hazards in the name of economic growth.

Though we see many studies at country level or at various specific geographical zones of the world on this subject, however, there is a research gap that exists in identification of contributing factors on a global level encompassing cross sectional data of all countries in a fuzzy environment that is responsible for the increased level of carbon di oxide, the main and most dangerous greenhouse gas. Our aim here is to understand which development growth indices the prime causes are leading to this detrimental effect worldwide.

III. METHODOLOGY

The principal objective of this research is to include as many countries as we can, as well as considerable financial and developmental indices as covariates in order to study their effect on carbon deposit. Our master dataset as obtained from the worldbank websiteconsisted of information on CO_2 Emission, Export, Import, FDI Net Inflows, GDP and Population of a total of 265from 1960 till 2020. However, keeping in mind our target of study, we have sorted and filtered the data for the year 2018, keeping only those countries with minimal missing values of covariates that we are keeping in this work. This reduces and condenses the data to 179 countries with six study variables, CO2 emission being the response variable.

The variables we have for the 179 countries are as follows:

Table I: Data description

Sl. No.	Indicator name	Unit	Note
1	CO_2 emissions	kt (kiloton)	The carbon dioxide emissions are generated during the production of cement and during the burning of fossil fuels; among these are CO2 emissions from gas flaring and the consumption of solid, liquid, and liquid fuels.
2	Export	value index (2000 = 100)	These are current value of exports represented as a percentage of the average w.r.t basein U.S. dollar.
3	Import	value index (2000 = 100)	Current import value (c.i.f.) is converted to US dollars and expressed as a percentage of the base period's (2000) average to create import value indexes.
4	Foreign Direct Investment (Net inflows)	BoP, US$	The total of equity capital, reinvested earnings, and similar other capital makes up this direct foreign inflow. Ownership of 10% or more of the voting stock's common shares is required to demonstrate the existence of a direct investment link.
5	GDP	current US$per million	GDP at purchaser's pricing is calculated as the total of the gross value added by all resident producers, plus any relevant product taxes, excluding any subsidies that aren't represented in the cost of the items. Using statutory exchange rates for a year, the GDP statistics are translated to dollars.
6	Population	Total	The de facto definition of population is used to calculate the total population, which includes all residents regardless of citizenship or legal status. The numbers presented are midyear projections.

We want to create some statistical models that will enable us to definitively determine which elements are directly impacting the discharge of greenhouse gases. Here, it's crucial to realize that our goal is to identify those independent variables that collectively influence the dependent variable, or the CO2 emission levels across the whole cross-sectional data, among the three emission groups, i.e., CO2emission values. Hence first we have created clusters segregating the countries into three level of emission using Hierarchical clustering method so that the three groups are statistically distinguishable (*Table 6,* Annexure 1). Fuzzy regression, Ridge and LASSO techniques were implemented and fitted separately in each of the groups to find out the significant factors uniformly over all the countries.

In the following section we have elucidated the three methodologies used for our analysis viz. Fuzzy Regression, LASSO and Ridge regression. A detailed description of hierarchical clustering has also been presented as we have extensively used this algorithm to create the three emission groups.

- **Fuzzy Regression:**

One of the most common and useful techniques in statistical analysis is regression analysis, which focuses on determining how closely a variable depends on one or more other variables. Anarithmetic in one or more parameterized types often describes the dependence. The primary goal of regression analysis is to estimate the model parameters based on actual data.The linear form:

$$y = \gamma_0 + \gamma_1 x_1 + \cdots + \gamma_n x_n$$

Where y is an output variable,$x_1, x_2, \dots, x_n$ are input variables, and $\gamma_1, \gamma_2, \dots, \gamma_n$are parameters.

When a set of observed data $\{a_1, b_1\}$, $\{a_2, b_2\}$, …, $\{a_m, b_m\}$ are given for the pair of variables {x, y}, we seek to compute the values of γ_0 and $\gamma_1$at which the overall error of the predicted points along the straight line becomes minimum with respect to the corresponding observed points. Using least squares error, the overall error expressed as:

$$\sum_{i=1}^{m} [b_i - \{\gamma_0 + \gamma_1 a_i\}]^2$$

The optimal values of γ_0 and $\gamma_1$can be easily determined by solving the optimization problem as:

$$\gamma_1 = \frac{m \sum_{i=1}^{m} a_i b_i - \sum_{i=1}^{m} a_i \sum_{i=1}^{m} b_i}{m \sum_{i=1}^{m} a_i{}^2 - (\sum_{i=1}^{m} a_i)^2}$$

$$\gamma_0 = \frac{\sum_{i=1}^{m} b_i - \gamma_1 \sum_{i=1}^{m} a_i}{m}$$

The reliance of an output variable on an input variable is indicated by the form in linear model with fuzzy parameters.

$$Y = C_1 x_1 + C_2 x_2 + \cdots + C_n x_n$$

Where $C_1, C_2, \dots, C_n$ are fuzzy numbers, and $x_1, x_2, \dots, x_n$ are real valued input variables; for each n-tuple of values of the input variables, the value of the output variable defined by the above equation is a fuzzy number Y.

The triangular fuzzy numbers assuming that the parameters are symmetric are given as:

$$C_i(c) = \begin{cases} 1 - \dfrac{|c - c_i|}{s_i} & when\ c_i - s_i \le c \le c_i + s_i \\ 0 & otherwise \end{cases}$$

Where c_i is the point for which $C_i(c_i) = 1$ and $s_i > 0$ is the spread of C_i. Let, C_i, which expresses the linguistic terms approximately c_i or around c_i, be denoted by C_i = {c_i, s_i} for all i∈ N_n. Then it becomes easy to prove by the extension principle that Y is also a symmetric triangular fuzzy number which is given by

$$Y(y) = \begin{cases} 1 - \dfrac{|y - x^T c|}{s^T |x|} & when\ x \neq 0 \\ 1 & when\ x = 0, y \neq 0 \\ 0 & when\ x = 0, y = 0 \end{cases}$$

For all y ∈ℝ, where

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ c_2 \\ \dots \\ c_n \end{bmatrix}, \quad s = \begin{bmatrix} s_1 \\ s_2 \\ \dots \\ s_n \end{bmatrix}, \quad |x| = \begin{bmatrix} |x_1| \\ |x_2| \\ \dots \\ |x_n| \end{bmatrix}$$

And T denotes the operation of transposition.

Two criteria of goodness are usually employed here

i. For every pair of data point(a_j, b_j), where a_j a vector containing the values of the input variables, b_j should have a grade that is greater or comparable to some specified valueand correspond to the appropriate fuzzy number h ∈ [0,1].
ii. The total non-specificity of the fuzzy parameters must be minimized.

Finally, a fuzzy regression problem can be formulated as

$$minimize \sum_{i=1}^{n} S_i$$

$$s.t.\ \ (1-h)s^T\left|a_j\right| - \left|b_j - a_j^T e\right| \ge 0, \qquad j \in \mathbb{N}_m$$

$$s_i \ge 0, \qquad i \in \mathbb{N}_m$$

- **Ridge Regression:**

This method of estimation is used when the dataset suffers from multicollinearity resulting in model coefficients getting penalty. This method shrinks the parameters and reeduces the model complexity by coefficient shrinkage.Ridge gives almost same weight to the correlated variables.

The minimized loss function:

$$L_{Ridge}(\hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{p} \hat{\beta}_j^2$$

$$= \| y - X\hat{\beta} \|^2 + \lambda \| \hat{\beta} \|^2$$

The estimates, bias and its variance of these models are:

$$\hat{\beta}_{Ridge} = (X'X + \lambda I)^{-1} (X'Y)$$

$$Bias(\hat{\beta}_{Ridge}) = -\lambda (X'X + \lambda I)^{-1}\hat{\beta}$$

$$Var(\hat{\beta}_{Ridge}) = \sigma^2 (X'X + \lambda I)^{-1} X'X (X'X + \lambda I)^{-1}$$

where, I = identity matrix,

λ = regularization parameter.

Choice of λ is considered in two ways:

Traditional approach: with help of AIC, BIC or information criteria or by Machine learning or modern approach: Cross validation with aim to reduce sum of squares with respect to some measure, say RMSE or MSE etc.

- **LASSO Regression:**

For a more precise prediction using shrinkage, LASSO regression is a regularization approach used over regression algorithms. The LASSO technique supports the use of straightforward models retaining only variates of significant importance (i.e., models with fewer parameters). For models with high levels of multicollinearity or when some elements of model selection, such as selecting features and parameter removal, need to be handled, this particular type of regression is acceptable.The loss function to be minimised:

$$L_{Lasso}(\hat{\beta}) = \sum_{i=1}^{n} (y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j=1}^{p} \| \hat{\beta}_j \|$$
$$= \| y - X\hat{\beta} \|^2 + \sum_{j=1}^{p} \lambda \| \hat{\beta}_j \|$$

Main difference between LASSO and Ridge is their loss function or specifically how they penalize the coefficients.

- **Hierarchical Clustering:**

When using cluster analysis techniques, data sets are examined to see whether they can be summarised meaningfully in terms of a small number of groups or clusters of objects or individuals that resemble one another but are distinct from one another in some ways.

The categorization entails a number of partitions as opposed to categorising the data into a specific number of classes or clusters in a single step. It might go from every person in a group to each of the n clusters that each contain one single observation. Agglomerative approaches are arguably the most used among hierarchical methods. Here, a number of data partitions are created. The first is made up of n clusters with a single member. At each level, the individuals or groups that are most similar to one another or nearest to them merge, and this process continues until a single group made up of all n individuals is generated. Divisive method is just the opposite in algorithm to this agglomerative method.

A technique developed by Ward in 1963 that bases the fusion of two clusters on the magnitude of an erroneous sum-of-squares criterion. Minimizing the increase in the overall within-cluster error sum of squares defined by E, which is provided by, is the goal at each stage as follows-

$$E = \sum_{m=1}^{g} E_m$$

$$E_m = \sum_{l=1}^{n_m} \sum_{k=1}^{P_k} (x_{ml,k} - \bar{x}_{m,k})^2$$

In which $\bar{x}_{m,k} = \frac{1}{n_m} \sum_{l=1}^{n_m} x_{ml,k}$ (the mean of the m^{th} cluster of the k^{th} variable),

$x_{ml,k}$ being the score on the k^{th} variable (k = 1, …, p) for the l^{th} object (l = 1, …, n_m) in the m^{th} cluster (m = 1, …, g).

One of the most well-known standard agglomerative hierarchical clustering method is Ward's method which is mostly referred to as residual sum of squares,and it is usually used with the distance which requires raw data. In this case, following fusion, the distance between groupings is seen as an increase in the total of squares within each cluster then summed over all variables. In order to interpret geometry, it is assumed that points have a Euclidean space representation. It frequently identifies spherical groups of the same size, which are typically sensitive to outliers.

IV. RESULTS AND DISCUSSIONS

Here, we have a total of 179 countries in our final data which are scattered all over the world. Their region is not same, their income groups are different, they have varying social, economic and geometrical backgrounds. Thus, it was not a wise decision to include all the countries in a single model and the need to divide these countries into different groups arises.

As our primary variable of interest is CO_2 emission,this being our dependent variable, we decided to cluster the countries based on their CO_2 emission (kt) value and was able to cluster them in 3 separate groups with the following levels of $CO_{2.}$

- The first group (A) has a total of 90 countries which is the lowest CO_2emittinggroup with a minimum value of 70 kt and maximum value of 12,270 kt.
- The second group (B) withmid-range of CO_2 emission between 14,050 kt to 82,760 kt has a total of 48 countries.
- The third group (C)has highest 41 CO_2 emitting countries with range of 86,620 kt to 1,03,13,460 kt.
  The grouping of the countries are given below in annexure 1.

Fitting the *PLRLS method of fuzzy regression model* separately for all *three groups* cases we get the following coefficients:

TableII: Fuzzy regression coefficient values

		Centre	Left Spread	Right Spread
Group A CO_2 emission vs***Export***	(Intercept)	4446.1233754	4406.978	7953.901
	export	-0.2395264	0.000	0.000
Group A CO_2 emission vs ***Import***	(Intercept)	2937.577132	3045.700760	8.796823e+03
	import	2.833431	2.533398	-2.384186e-07
Group A CO_2 emission vs ***FDI***	(Intercept)	4.290191e+03	4262.819	34101152
	fdi	1.904458e-08	0.000	0
Group A CO_2 emission vs ***GDP***	(Intercept)	1840.381766	1793.977303	7935.539
	gdp	0.1575889	0.1131169	0.000
Group A CO_2 emission vs ***Population***	(Intercept)	3.523344e+03	3.488499e+03	7891.611
	population	9.065070e-05	6.597058e-05	0.000
Group B CO_2 emission vs ***Export***	(Intercept)	43907.78808	23265.207884	32344.63682
	export	-10.34055	3.233792	21.67608
Group B CO_2 emission vs ***Import***	(Intercept)	44610.12851	2.554839e+04	29044.14033
	import	-12.64513	-2.384186e-07	26.70941
Group B CO_2 emission vs ***FDI***	(Intercept)	3.881079e+04	25054.9	44323.87
	fdi	2.831806e-08	0.0	0.00
Group B CO_2 emission vs ***GDP***	(Intercept)	3.176329e+04	1.9771183e+04	3.842308e+04
	gdp	4.397661e-02	9.712080e-03	3.348320e-03
Group B CO_2 emission vs ***Population***	(Intercept)	3.619539e+04	2.241267e+04	36868.15
	population	1.397969e-04	1.201870e-04	0.00
Group C CO_2 emission vs ***Export***	(Intercept)	107801.966	25579.353	2009304.20
	export	1537.391	1542.233	12137.47
Group C CO_2 emission vs ***Import***	(Intercept)	65532.909	0.000	2442378.84
	import	1515.787	1505.565	10622.68
Group C CO_2 emission vs ***FDI***	(Intercept)	4.820369e+05	4.911368e+05	7.953812e+05
	fdi	1.170266e-05	8.691809e-06	2.705296e-05
Group C CO_2 emission vs ***GDP***	(Intercept)	6.367878e+04	2.275189e+04	3.485075e+05
	gdp	3.788967e-01	2.854045e-01	3.751583e-01
Group C CO_2 emission	(Intercept)	9.095046e+04	9.078545e+03	0.0000000
	population	4.735607e-03	4.532264e-03	0.0103303

vs ***Population***				

In case of group A, it can be observed that, Import has the highest impact on CO_2 emission as given by the model equation-

$$CO_2 emission = 2937.577 + 2.8334 * import$$

GDP is the second most impactful explanatory factor in this case even thoughFDI and population are also positively affectingthe emission of CO_2 while export has a negative influence for these low emitting group countries.

In case of group B, GDP affects the CO_2 emission most while both import and export has a negative impact. But surprisingly, in case of the highest CO_2 emitting countries i.e., Group C, import and export has the highest influence on CO_2 emission unlike the other two groups. The most important impression is that GDP is the indicator common in all the three clusters having major impact of carbon emission. Also, noteworthy is the fact that for high emission group population is a highly significant independent variable which is plausible as this group consists of nations like China, India and others most populous countries of the world.

The rest of the equations can be derived in similar manner for each of the clusters, thus indicating the seriousness of each factor on CO_2 emission. We also need to keep in mind here that not only do we want to measure the gravity of a covariate on CO_2 emission, but also the direction of relationship is also of paramount significance in this study. The left and right spreads are parts of triangular fuzzy numbers (TFNs) coefficients that give the range of CO2 emission as the output with its lower and upper boundaries

The total error of fit values for all the models are given as below:

Table III: Total error of fit values

	Group 1	***Group 2***	***Group 3***
CO_2 emission vs ***Export***	5.56239e+11	1.62895e+12	1.604564e+14
CO_2 emission vs ***Import***	5.87568e+11	1.603893e+12	1.637849e+14
CO_2 emission vs ***FDI***	1.534935e+15	1.6651e+12	5.837171e+13
CO_2 emission vs ***GDP***	5.17144e+11	1.446944e+12	3.244346e+13
CO_2 emission vs ***Population***	5.37407e+11	1.476841e+12	4.314351e+13

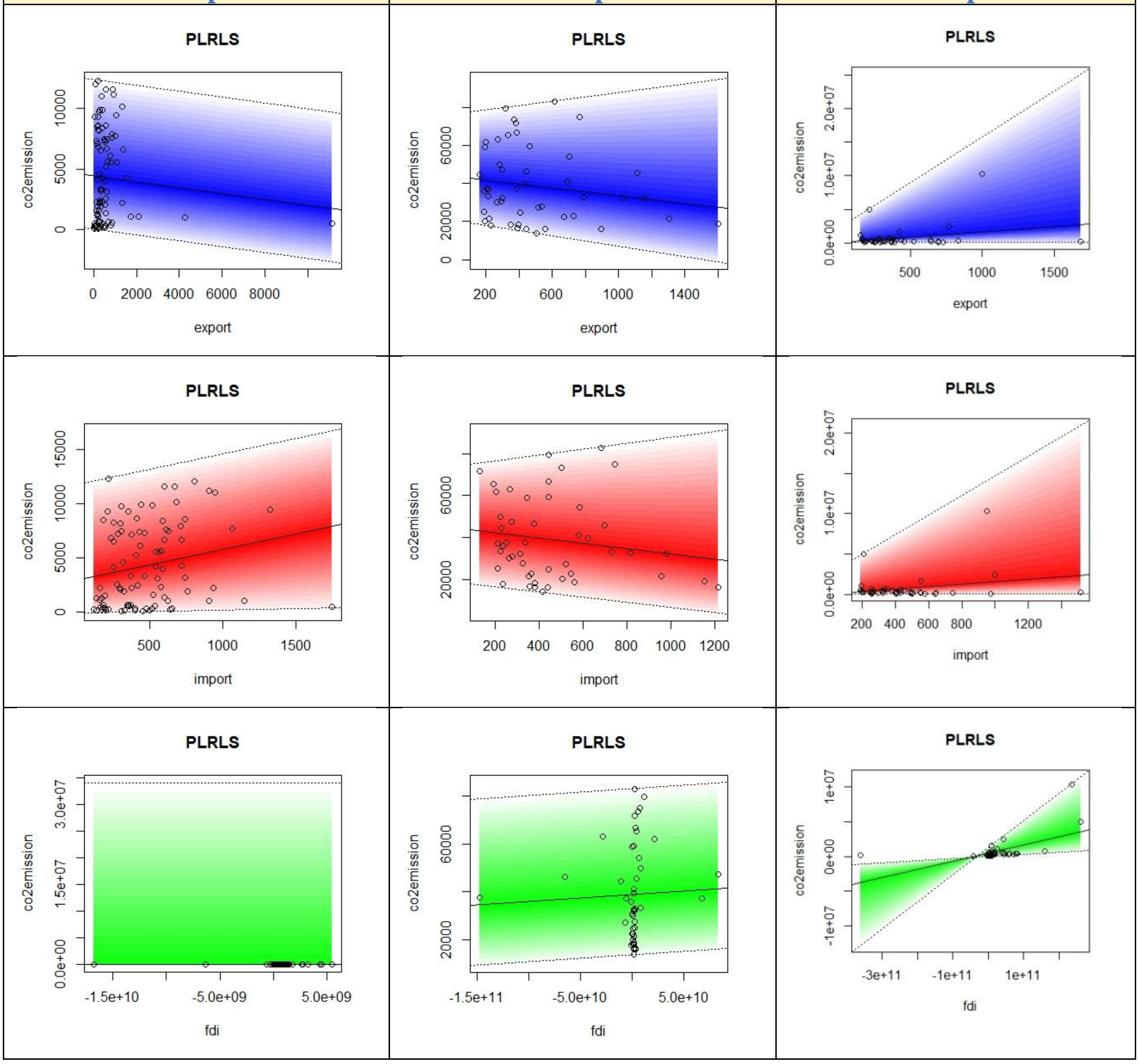
It can be observed that the lowest value of total error of fit is for the group A,CO2emission vs GDP. So, we can say that this particular model gave us the best fit out of all the 15 models. Overall, we

ISSN: 2094-0343
2326-9865

have got significantly better predictive equations for the independent variable GDP in all the 3 groups.

The 15 linear models fitted using the PLRLS method for the three groups (5 equations in each group, one for every independent variable) along with upper and lower boundary of coefficients is given below in Table IV. We can get the lower boundary of the model support interval by subtracting the variable intercept from the centre intercept values while in case of the upper boundary we get the same by adding these two as obtained in Table II. Different colour code has been used for every factor in each cluster.

Group A	Group B	Group C



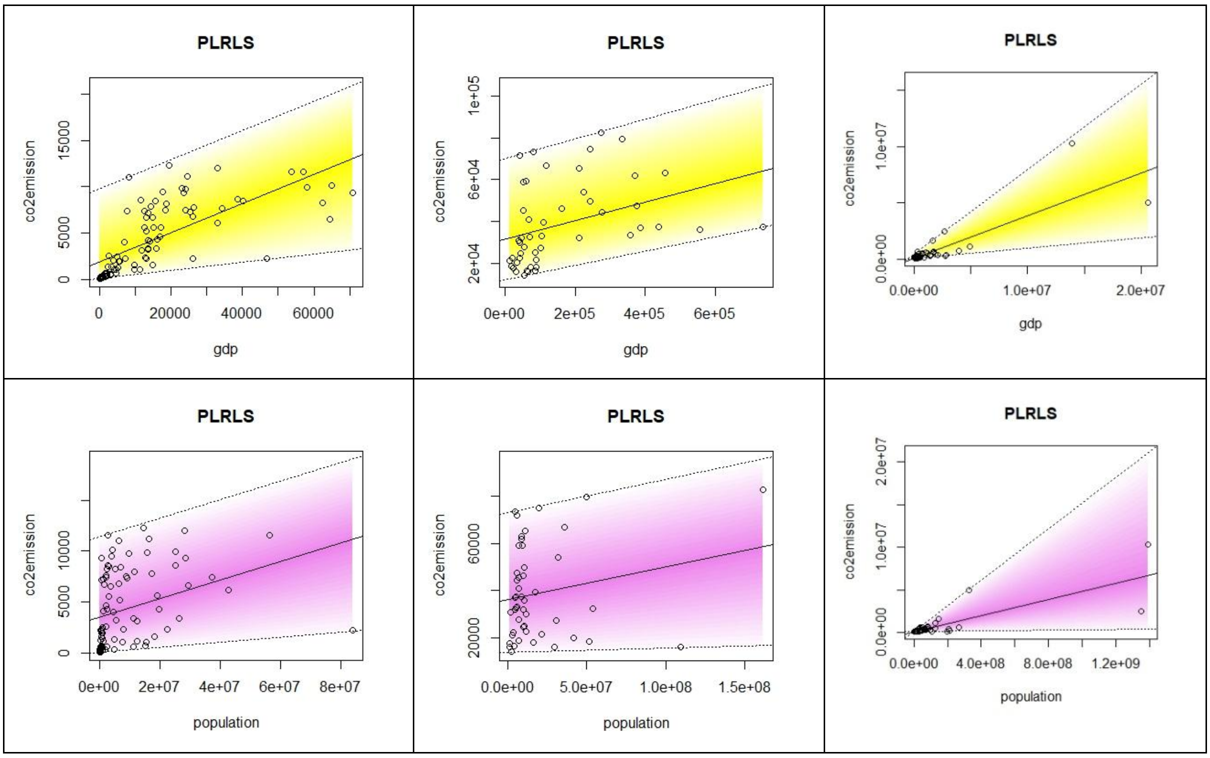


Figure I: Fitted model plots using Fuzzy regression for three groups using 5 variables

Summating the output,we can comparethe fuzzy regression models fitted with the PLRLS method for the 3 different groups in all the 5 cases as follows:

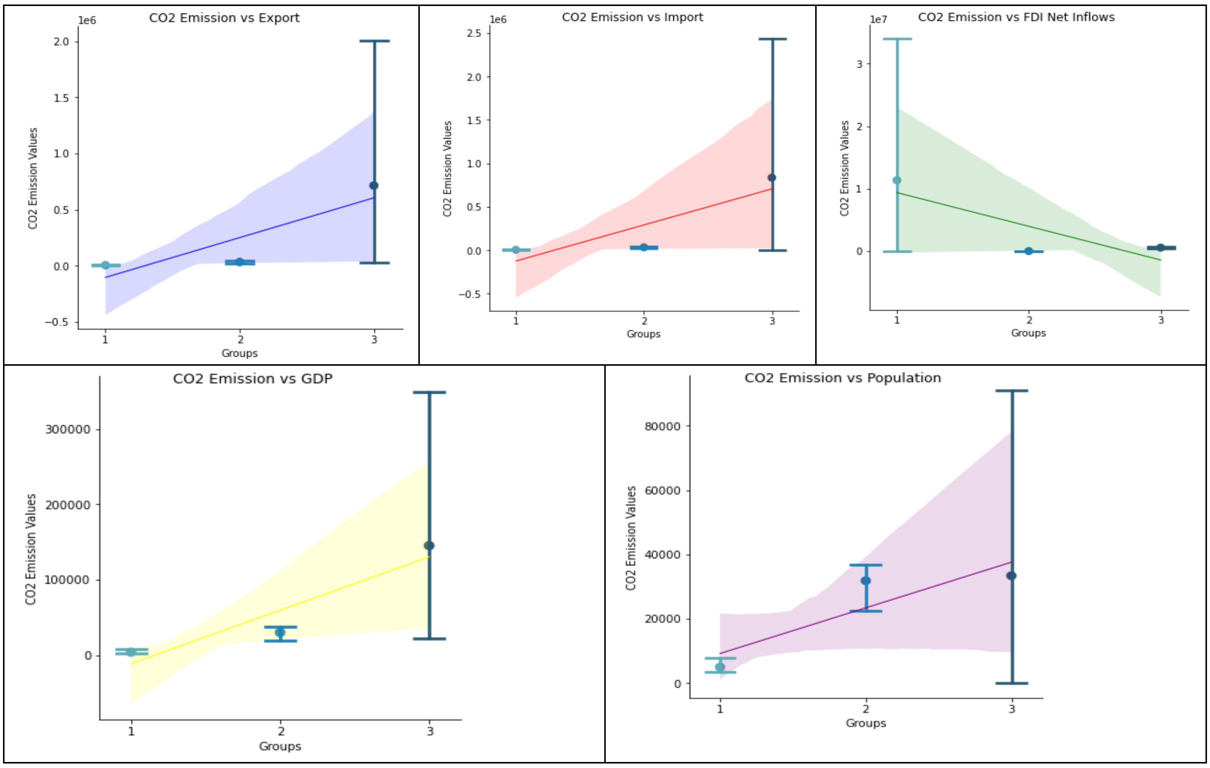


Figure II: Points representing centre values and for the fitted Fuzzy regression models for each independent variable in study

We have also performed LASSO and Ridge regression for the three groups keeping in mind the correlated explanatory variables and the need to identify those indicators which are of fundamental importance in contributing to increased carbon release in the present set up. Results obtained from Ridge regression is given in the following table. The results corroborate the fact the GDP is having positive and high impact on CO2 emission throughout all three groups, followed by population.

Table IV: Coefficient values for all the groups in case of Ridge Regression

VARIABLE	Group A	Group B	Group C
(Intercept)	2.488672e+03	3.653395e+04	4.350490e+05
Export	-2.326117e-01	-8.879185e-36	-1.595400e+01
Import	1.567048e+00	-1.309789e-35	2.156960e+00
FDI	1.313124e-08	-1.331462e-45	8.526721e-07
GDP	7.551801e-02	3.220335e-38	4.232009e-02
population	1.534866e-05	1.973776e-40	3.781296e-04

Clearly both these models perform well for the given dataset though LASSO outperforms Ridge in terms of the RMSE value.

Table V: RMSE values for both Ridge and Lasso Regression

	Group A	Group B	Group C
RIDGE	**2474.432**	**27355.21**	**2871268**
LASSO	**2519.358**	**26903.5**	**2069709**

The following graphFig. IIIgives the significance of covariates in the study both by LASSO and Ridge methods. One important point to note here is the fact that since LASSO is a selection operation of L1 regularization and works best in presence of large number of covariates so here it drops several variables from model in all three groups. The most interesting characteristic is that GDP is retained in LASSO models for all the three clusters thus solidifying and verifying our previous models of fuzzy and ridge regression.The blue and orange bars represent coefficient values for Ridge and LASSO respectively. We can hence conclude that GDP and population are two most important and significant variables in emission of CO2 .

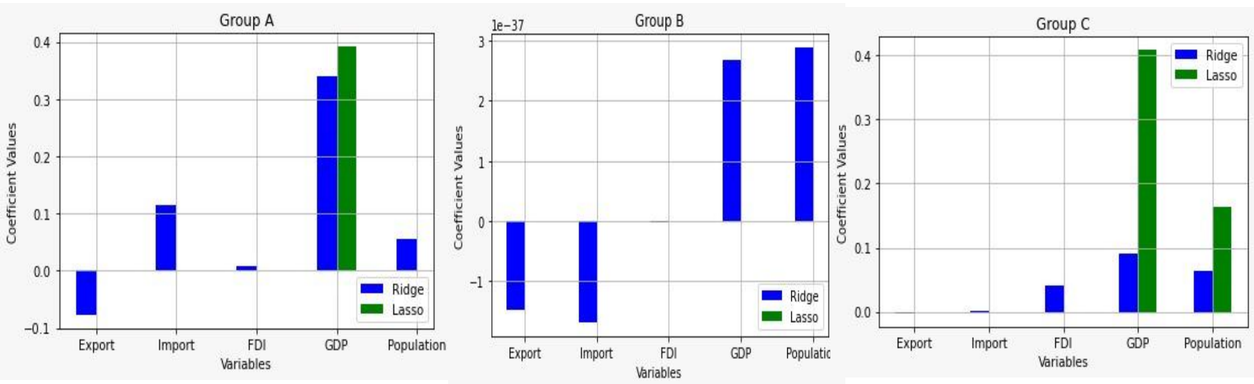


Figure III: Coefficient importance for LASSO and Ridge

V. CONCLUSION

Carbon emission is nowadays a problem for every country. Due to industrialization and other development works, carbon emission is inevitable for the world. In our article, we analyse how the emission is affecting developing and other first world country and their pattern of emission based on statistical and machine learning algorithms. The critical variables such as export-import, GDP, population of different countries are considered for regressive proportion in fuzzy environment. With the help of these three methods of Fuzzy, Ridge and LASSO regression models, it can be exclusively concluded with significance that “GDP” and “population” two are the key factors in the emission episode of carbon and this is true for all countries irrespective of its emission level. The hierarchical Clustering method provides us with a clear picture of carbon emission in the three varying groups. In group C. which is the highest emission group, import-export businesses are also among most responsible factors. In group B, however, there is a negative impact on these types of businesses upon carbon emission. Future research can be extended in the post covid period to find out more conclusive remarks on the distribution pattern of the carbon emission and how it is impacting environment in the long-term scenario. Also, material engineering and novel technologies is highly applicable in these predicaments to search for aninvariable solution for non-renewable energy sources. We can also attemptto include more covariates in the study though it may seem challenging to gather the data for all the countries. One can also build some sustainability models to relate our problem.

APPENDIX I: After clustering the 179 countries into 3 groups the countries are in the following groups as follows:

Table VI: Clustered grouping of the 179 countries

Serial No	**Low** (A)	**Mid** (B)	**High** (C)
1	Afghanistan	Angola	United Arab Emirates
2	Albania	Austria	Argentina
3	Armenia	Azerbaijan	Australia
4	Antigua and Barbuda	Bangladesh	Belgium
5	Burundi	Bulgaria	Brazil
6	Benin	Bahrain	Canada
7	Burkina Faso	Bosnia and Herzegovina	Chile
8	Bahamas	Belarus	China
9	Belize	Bolivia	Czech Republic
10	Barbados	Switzerland	Germany
11	Brunei Darussalam	Colombia	Algeria
12	Bhutan	Denmark	Egypt, Arab Rep.
13	Botswana	Dominican Republic	Spain
14	Central African Republic	Ecuador	France
15	Cote d'Ivoire	Estonia	United Kingdom
16	Cameroon	Ethiopia	Indonesia
17	Congo, Dem. Rep.	Finland	India

18	Congo, Rep.	Ghana	Iran, Islamic Rep.
19	Comoros	Greece	Iraq
20	Cabo Verde	Guatemala	Italy
21	Costa Rica	Croatia	Japan
22	Cyprus	Hungary	Kazakhstan
23	Djibouti	Ireland	Korea, Rep.
24	Dominica	Israel	Kuwait
25	Fiji	Jordan	Mexico
26	Gabon	Kenya	Malaysia
27	Georgia	Lao PDR	Nigeria
28	Guinea	Lebanon	Netherlands
29	Gambia	Libya	Pakistan
30	Guinea-Bissau	Sri Lanka	Philippines
31	Equatorial Guinea	Morocco	Poland
32	Grenada	Myanmar	Qatar
33	Guyana	Mongolia	Russian Federation
34	Honduras	Norway	Saudi Arabia
35	Haiti	New Zealand	Thailand
36	Iceland	Oman	Turkey
37	Jamaica	Peru	Ukraine
38	Kyrgyz Republic	Portugal	United States
39	Cambodia	Romania	Uzbekistan
40	Kiribati	Sudan	Vietnam
41	St. Kitts and Nevis	Singapore	South Africa
42	Liberia	Serbia	
43	St. Lucia	Slovak Republic	
44	Lesotho	Slovenia	
45	Lithuania	Sweden	
46	Luxembourg	Turkmenistan	
47	Latvia	Trinidad and Tobago	
48	Moldova	Tunisia	
49	Madagascar		
50	Maldives		
51	Marshall Islands		
52	North Macedonia		
53	Mali		
54	Malta		
55	Mozambique		
56	Mauritania		
57	Mauritius		
58	Malawi		
59	Namibia		
60	Niger		
61	Nicaragua		
62	Nepal		

63	Nauru		
64	Panama		
65	Palau		
66	Papua New Guinea		
67	Paraguay		
68	Rwanda		
69	Senegal		
70	Solomon Islands		
71	Sierra Leone		
72	El Salvador		
73	Somalia		
74	Sao Tome and Principe		
75	Suriname		
76	Eswatini		
77	Seychelles		
78	Chad		
79	Togo		
80	Tajikistan		
81	Tonga		
82	Tanzania		
83	Uganda		
84	Uruguay		
85	St. Vincent and the Grenadines		
86	Vanuatu		
87	Samoa		
88	Yemen, Rep.		
89	Zambia		
90	Zimbabwe		

References:

1. Murthy N. S., Panda M., Parikh K., 2007. CO2 Emission Reduction Strategies and Economic Development of India.Margin. The journal of Applied Economic Research 1:1; Sage Publications Los Angeles/London/New Delhi/Singapore.
2. TiwariA. K., July 2011.Primary Energy Consumption, CO2 Emissions and Economic Growth: Evidence from India. Applied Econometrics and International Development 11(2):165-189.
3. KumarS., ShuklaA. K., MuhuriP. K., LohaniQ. M. D., June 2019. Transfer Learning based GDP Prediction from Uncertain Carbon Emission Data. 2019 IEEE International Conference on Fuzzy Systems.
4. Sharma R., KautishP., UddinG. S., January 2020. Do the international economic endeavors affect CO2 emissions in open economies of South Asia? An empirical examination under nonlinearity. Management of Environmental Quality an International Journal 31(1):89-110.

5. Chotia V., Pankaj P., 2019. Impact of Population and Economic Growth on Carbon Emissions of Developed and Developing Countries. Int. J Green Economics, Vol. 13, Nos. ¾; 276-287.
6. Shahani R., Raghuvansi K., Carbon Dioxide Emissions and Economic Growth: A Bivariate Co-integration Analysis for Two Emerging Markets of India and China. 9-22.
7. SenguptaP. P., SinhaM., DuttaU. P., June 2019.Economic and environmental performances in manufacturing industries: A comparative study. Periodicals of Engineering and Natural Sciences; Vol. 7, No. 1, ISSN 2303-4521; 99-108.
8. Muthusamy A., Rani P. J., September 2019. FDI, GDP, and CO2 Emission: ARDL Bound Cointegration Relationship Examination. International Journal of Recent Technology and Engineering (IJRTE); ISSN: 2277-3878, Volume-8 Issue-2S10; 132-138.
9. Mathur K.S.,Sahu S. 2016. Does Domestic Energy Consumption Affect GDP of a Country? A Panel Data Study. Global Economy Journal. 16(2)
10. Nain Md Z., Ahmad W., Kamaiah B., Economic growth, energy consumption and CO2 emissions in India: a disaggregated causal analysis, International Journal of Sustainable Energy.
11. RaiS. K., BembeyA. M., SarfareD., 2019. Empirical verification of causality between CO2 emissions, energy consumption, foreign direct investment, gross domestic product, and openness of the economy: evidence from India. Int. J. Sustainable Economy, Vol. 11, No. 3; 237- 257.
12. BharadwajL. N., May 2020. The Nexus among Foreign Direct Investment, Economic Growth and Carbon Emissions: Evidence from India. International Journal of Scientific & Research Volume 9, Issue 05; ISSN 2277-8616; 124-130.
13. Rana R., SharmaM., July-September 2020. Dynamic Causality among FDI, Economic Growth and CO2 Emissions in India with Open Markets and Technology Gap. International Journal of Asian Business and Information Management; Volume 11, Issue 3; 15-31.
14. Wang Q. and Wang S. Decoupling economic growth from carbon emissions growth in the United States: the role of research and development. Journal of Cleaner Production. 2019;(234):702–713.
15. Ahmad A., Zhao Y., Shahbaz M., Bano S., Zhang Z. and Wang S. et al. Carbon emissions, energy consumption and economic growth: An aggregate and disaggregate analysis of the Indian economy. Energy Policy. 2016;(96):131–143.
16. Yang L., Wang J. M., and Pan H. Relationship between energy consumption, economic development and carbon emissions in China. Environmental Engineering and Management Journal. 2014;13(5):1173–1180.
17. Zheng B., Zhang Q., Davis S.J., Ciais P., Hong C., Li M., et al. Infrastructure shapes differences in the carbon intensities of Chinese cities. Environmental science & technology.2018;52(10):6032–6041.
18. Lee, C.-C., Chang, C.-P. (2008). Energy consumption and economic growth in Asian economies: a more comprehensive analysis using panel data.Resources and Energy Economics 30, 50–65.
19. Lee, C.-C., Chang, C.-P., Chen, P.-F. (2008). Energy-income causality in OECD countries revisited: the key role of capital stock. EnergyEconomics 30, 2359–2373.

20. Narayan, P.K., Smyth, R. (2008). Energy consumption and GDP in G7 countries: new evidence from panel Co-integration with structural breaks.Energy Economics 30, 2331–2341
21. Akinlo, A.E. (2008). Energy consumption and economic growth: evidence from 11 African countries. Energy Economics 30, 2391–2400.
22. Wolde-Rufael. (2009). Energy consumption and economic growth: the experience of African countries revisited. Energy Economics 31, 217–224.
23. Apergis, N., Payne, J.E. (2009). Energy consumption and economic growth in Central America: evidence from a panel Co-integration and errorcorrection model. Energy Economics 31, 211–216.
24. Al-Iriani, M. (2006). Energy–GDP relationship revisited an example from GCC countries using panel causality. Energy Policy 34, 3342–3350.
25. Yoo, S.H., Kwak, S.-Y. (2010). Electricity consumption and economic growth in seven South American countries. Energy Policy 38, 181–188.
26. Sharma, S.S. (2010). The relationship between energy and economic growth: empirical evidence from 66 countries. Applied Energy 87, 3565–3574.