EXPLORING THE INTERACTION EFFECTS OF KEY INDICATORS FOR CARBON EMISSION THROUGH PENALIZED REGRESSION MODELS

Zainab Javed[*1], Anam Javaid[2], Hafiz Abdul Sami[3], Dr. Shahbaz Nawaz[4], Dr. Sumbal Javaid[5] Kashan Javaid[6] Arsalan Javed[7]

[*1]*Department of Statistics, The Women University Multan*
[2]*Assistant Professor; Department of Statistics, The Women University Multan*
[3,4]*Punjab Bureau of Statistics, Planning and Development Department Pakistan*
[5]*Nishter Medical University, Multan*
[6]*Bahauddin Zakariya University Multan*
[7]*LLM, Benazir Bhutto Shaheed University, Karachi, Pakistan*

DOI: https://doi.org/10.5281/zenodo.17499089

Keywords
Multicollinearity, Outliers, Boxplot, Ridge, LASSO

Article History
Received: 27 August 2025
Accepted: 20 October 2025
Published: 31 October 2025

Corresponding Author: *
Dr. Anam Javaid

Abstract

Carbon emissions, primarily in the form of carbon dioxide (CO_2), play a critical role in climate change and its associated environmental consequences. This study explores the causes, effects, and potential solutions related to carbon emissions, highlighting their significance in global efforts to mitigate climate change. Pakistan's carbon emissions primarily originate from energy production, industrial activities, transportation, and agricultural practices. The current study focused on the efficient model selection for Carbon emission in Pakistan. The CO_2 (Solid, Liquid, Gaseous) fuel consumption is taken as the response variable while the Population, Total Value, Agriculture, forestry, and fishing, value added (% of GDP) Value, Agricultural land (sq. km) Value, Urban population Value, Gross fixed capital formation (current US$), Industry (including construction), Fertilizer consumption (% of fertilizer production) and GDP taken as the predictors. Five stages are performed for the efficient model selection. In the first stage, multicollinearity diagnosis is applied in term of correlation matrix. For the second stage, outlier diagnosis is performed through the different statistical measures with the graphical analysis in term of Box Plot. The third stage was based on the best model selection by using Ridge and LASSO regression analysis. The fourth stage was consisted on the efficient model selection by using the model selection criteria as Sum Square of Error (SSE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and forecasting efficiency is tested through the Mean Absolute Percentage Error (MAPE). The study result shows that Ridge regression analysis is selected as the best technique on the basis of maximum R^2, minimum MSE, MAPE and LASSO regression.

INTRODUCTION

Long term changes in temperatures and weather patterns are the result of climate change. These changes may be a part of nature, such as a complete reversal of solar cycle by (Haigh et al.,1999). But since the 1800s, human activities, primarily the burning of fossil fuel such as coal, oil and gas, have been the most important cause of climate change by (Woolway et al.,2020). Burning fossil fuel release greenhouse gas that act as a cover around the planet, trapping heat from the sun and increasing temperatures. Carbon dioxide and methane are the two examples of greenhouse gas emissions that affect climate change (Voigt et al.,2017). They are released when cars run on gasoline and buildings are heated with coal. Clearing lands and forests can also release carbon dioxide. Methane emissions are mainly generated from garbage. According to (Younger et al.,2008). Major emission include energy, industry, transportation, buildings, agricultural and land use (Javed et al., 2022).

Climate change is considered as one of the Pakistan's biggest issues. Climate change gives a serious threat to Pakistan. Following the changes in South Asia as a whole, Pakistan's climate has changed over the years (Hussain and Yousaf et al.,2020). These changes have significant on the impact on environment and population of the country. Due to impact of heat, drought and severe weather in the country, Pakistan's climate is as diverse as the country geography, and it is located in a range temperature zone (Khan et al.,2016). More than 150 extreme weather events were reported in Pakistan during 1998 to 2018. In 2022, there have been catastrophic floods at the national level because of one third of the country is underwater (Shehzad et al.,2023). .

The melting of Himalayan glaciers have blocked the major river of Pakistan. Pakistan ranked fifth among the countries that are affected by the extreme weather due to climate change during 199 to 2018. Pakistan has consistently been ranked among the 10 most vulnerable countries for the climate Hazards I index over the past 20 years. With over 10,000 deaths occur from weather related disasters. There is $4 billion economic losses from 173 catastrophic weather events by (Eckstein et al.,2018).

There are many causes responsible for climate change in Pakistan. Due to its frequently warm climate, Pakistan is susceptible to climate change. It is located in an area of the world where temperature risen is predicted to be higher than average by (Ali & Abdullah et al.,2017). According to (Gaur & Squires et al.,2018) the majority of the geographical area is semi-arid and desert; roughly 60% of it gets less than 250 mm of rainfall there annually. Climate change is primarily caused by human activities, including using fossil fuel more consistently and affecting the way of land is used. There are some perspectives of climate change by (Rasul,2012). Pakistan is not immune to global warming. Pakistan precipitation and thermal regimes had a significant impact, particularly in the last 20 years as global atmospheric have risen dramatically. It has a diverse environment ranging from snow caps in the north to south desert by (Laity 2009). .

The world highest mountains surround in the north, acting as a barrier that prevent cold waves from reaching the south in the winter and monsoon rains from moving further north in summer (Whiteman 2000). The summer monsoon brought by the Arabian Sea, which form the southern boundary provides plenty of moisture for domestic, industrial and agricultural water demands. The intensity and frequency of heat waves and rainfall events related to natural disasters (Lodhi et al., 2023). Due to increase in climate temperature, the hydrological cycle has undergone significant changes, including change in weather events, crop correlation and water arability periods. (Banholzer et al.,2014).

Pakistan is not exempt from Global warming and changing climate as neither respect political and geographical borders are respected. Pakistan rainfall and heating systems have undergone changes, particularly in the last 20 years as global atmospheric temperature have increase rapidly (Lieven 2011). Pakistan has a variety of climates, by ice-covered snow caps from the north to scorching deserts in the south. It is surrounded the north by the highest mountain in the world, which acts as a barrier to prevent monsoons from migrating further north in summer and cold waves from reaching the south in winter (Khan 2019).. The Southern border of the country is formed by summer monsoon from Arabian Sea, provide sufficient moisture for industrial use, power generation, or agriculture (Lim et al., 2020). Climate change has altered the hydrological cycle, affecting frequency and intensity of heat wave as well as rainfall, copping patterns, drought and timing of water availability by (Kundzewicz 2008)

METHODOLOGY.

Ordinary Least Square Method

Ordinary Least Squares (OLS) is a method used in regression analysis to estimate the unknown parameters in a linear regression model by (White 1980). Here's how you express it mathematically:

Given a linear regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \epsilon$$

Where:

- y is dependent variable
- $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are the coefficients to be estimated
- $x_1, x_2, \ldots, x_n$ are the independent variables
- ϵ is a error term

The OLS method aims to find the coefficients $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ that minimize the sum of squared residuals (errors) ϵ, expressed mathematically as:

$$min_{\beta_0, \beta_1, \beta_2, \ldots, \beta_n} \sum_{i=1}^{m}(y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in})^2$$

Where:

- m is the number of observations,
- y_i and x_{ij} represent the observed values of the dependent and independent variables, respectively,

The solution to this minimization problem gives the OLS estimates for $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$

In matrix notation, the solution can be written as:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

- $\hat{\beta}$ is the vector of estimated coefficients,
- X is the matrix of independent variables (including a column of ones for the intercept term),
- y is the vector of observed values of the dependent variable.

Ridge Regression

Ridge regression is a regularized version of Ordinary Least Squares (OLS), designed to handle multicollinearity and overfitting by shrinking the coefficients (Marquardt & Snee 1975).

Objectives of ridge regression are as follows:

Given standard linear model:

$$y = X\beta + \in$$

Ridge regression modifies the OLS objective function by adding an L2 penalty term:

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

Where:

- $||y - X\beta||^2 = \sum_{i=1}^{n}(y_i - x_i\ \beta)^2$ residual sum of squares (RSS)
- $||\beta||^2 = \sum_{j=1}^{p}\beta_j^2$ squared Euclidean norm (L2 norm)
- $\lambda \geq 0$ regularization parameter (controls strength of penalty)

Comparison with OLS

- **OLS**

$$\hat{\beta}^{OLS} = (X^T X)^{-1} X^T y$$

- **Ridge regression**

$$\hat{\beta}^{ridge} = (X^T X + \lambda I)^2 X^T y$$

The ridge estimator shrinks the coefficients towards zero (but never exactly zero).

LASSO Regression

LASSO (Least Absolute Shrinkage and Selection Operator) is a linear regression technique that adds an L1 penalty to the loss function by (Tibshirani 1996).. It is primarily used for

- Regularization: Prevents overfitting by penalizing large coefficients.
- Variable selection: Automatically selects important features by shrinking some coefficients to exactly zero.

Lasso use for high-dimensional data (e.g., p≫n), or when many features are irrelevant or correlated, ordinary least squares (OLS) can lead to:

- High variance in estimates
- Poor generalization on new data
- Lasso mitigates this by introducing bias but reducing variance → better prediction. **Summary of Lasso Properties:**

Property	**Lasso regression**
Penalty type	L1 norm

Encourages sparsity	Yes (variable selection)
Shrinks coefficients	Yes
Coefficients exactly zero	Yes
Closed-form solution	No
Solved via	Coordinate descent, LARS, etc.
Needs standardization	Yes

FINDINGS AND DATA ANALYSIS

The dataset of variables used in this study is taken from https://www.kaggle.com/. Total of 56 observations are used for the purpose of analysis. There are 8 variables related to the carbon emission are analyzed in the dataset, CO_2 is taken as dependent variable and all other 7 factors such as Population, Industrialization, GDP, Fertilizer consumption, urban area, Agriculture land and forestry, Gross fixed capital formation was considered as predictors. Interaction effects of the variables are also studied by using regression analysis. The codes are given to all the included variables in the analysis. The list of the codes with their respective names are mentioned in Table 1

Table 1: Variables Codes and Descriptions

Sr No.	Variable Names	Variable Codes
1	CO_2 emissions from solid fuel consumption (kt) Value, CO_2 emissions from liquid fuel consumption (kt) Value, CO_2 emissions from gaseous fuel consumption (kt) Value	**Y**
2	Population, total Value	X_1
3	Agriculture, forestry, and fishing, value added (% of GDP) Value	X_2
4	Agricultural land (sq. km) Value	X_3
5	Urban population Value	X_4
6	Gross fixed capital formation (current US$)	X_5
7	Industry (including construction), value added (constant 2010 US$)	X_6
8	Fertilizer consumption (% of fertilizer production)	X_7
9	GDP	X_8

Table 1 represents that in the climate change due to CO_2 emission the main factors are Population, Industrialization, GDP, Fertilizer consumption, urban area, Agriculture land and forestry, Gross fixed capital formation. CO_2 emission has interacted with all factors.

The unit of the CO_2 emission is taken as in kiloton (KT) while the population total value, Agricultural land, GDP, and urban population is in per 1000 unit. The Agriculture, forestry, Fertilizer consumption, and fishing is in terms of per 100 unit. The Gross fixed capital formation, Industry The main factors of climate change are Population, total Value, Agriculture, forestry, and fishing, value added (% of GDP) Value, per 1 USD. In Table 1, the CO_2 emissions from solid fuel consumption (kt) Value, CO_2 emissions from liquid fuel consumption (kt) Value and CO_2 emissions from gaseous fuel consumption (kt) Value are represents the dependent variable Y and the independent variables are coded by X1, X2, X1, X4, X5, X6, X7 and X8.

Details about Dataset Used in Analysis

For the analysis purpose, the dataset of climate change are selected for this study.

Agricultural land (sq. km) Value, Urban population Value, Gross fixed capital formation (current US$), Industry (including construction), value added (constant 2010 US$), Fertilizer consumption (% of fertilizer production), GDP. The interaction effects of independent variables are used to analyze the significance of regression models.

Multicollinearity Analysis in the Data With Interaction Effect

Correlation matrix is calculated to diagnose the multicollinearity issue in the dataset with interaction effect that is used in the analysis. The result for the matrix are obtained as follows:

For the CO_2 emission

Correlation matrix for the main factors of climate change is calculated by using Excel software with dependent and independent variables with interaction effect, Table 2 shows the information

Table 2 Correlation matrix for the main factors of climate change with interaction effect

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X1X2	X1X3	X1X4	X1X5	X1X6	X1X7	X1X8	X2X3	X2X4	X2X5	X2X6	X2X7	X2X8	X3X4	X3X5	X3X6	X3X7	X3X8	X4X5	X4X6	X4X7	X4X8	X5X6	X5X7	X5X8	X6X7	X6X8	X7X8
Y	1.00																																				
X1	0.99	1.00																																			
X2	-0.71	-0.79	1.00																																		
X3	-0.14	-0.13	0.13	1.00																																	
X4	0.99	1.00	-0.77	-0.13	1.00																																
X5	0.96	0.94	-0.63	-0.15	0.95	1.00																															
X6	1.00	0.99	-0.72	-0.16	0.99	0.98	1.00																														
X7	-0.43	-0.47	0.51	0.15	-0.46	-0.39	-0.44	1.00																													
X8	0.94	0.92	-0.58	-0.11	0.93	0.99	0.96	-0.36	1.00																												
X1X2	0.98	0.98	-0.67	-0.09	0.98	0.95	0.98	-0.42	0.95	1.00																											
X1X3	0.99	1.00	-0.79	-0.09	1.00	0.94	0.99	-0.47	0.92	0.98	1.00																										
X1X4	0.99	0.98	-0.67	-0.12	0.99	0.98	0.99	-0.42	0.98	0.99	0.98	1.00																									
X1X5	0.94	0.91	-0.55	-0.14	0.92	0.99	0.96	-0.35	0.99	0.93	0.91	0.97	1.00																								
X1X6	0.98	0.96	-0.62	-0.14	0.97	0.99	0.99	-0.39	0.99	0.97	0.96	0.99	0.99	1.00																							
X1X7	0.84	0.83	-0.62	-0.09	0.83	0.80	0.83	0.08	0.79	0.83	0.83	0.83	0.78	0.82	1.00																						
X1X8	0.91	0.88	-0.51	-0.10	0.90	0.97	0.93	-0.33	1.00	0.92	0.88	0.95	0.99	0.97	0.76	1.00																					
X2X3	-0.72	-0.79	1.00	0.22	-0.77	-0.63	-0.72	0.52	-0.58	-0.67	-0.79	-0.67	-0.56	-0.63	-0.61	-0.51	1.00																				
X2X4	0.99	0.99	-0.70	-0.10	0.99	0.95	0.99	-0.44	0.95	1.00	0.99	0.99	0.93	0.97	0.83	0.92	-0.70	1.00																			
X2X5	0.96	0.94	-0.62	-0.13	0.95	1.00	0.98	-0.38	0.99	0.96	0.94	0.98	0.99	0.99	0.80	0.98	-0.62	0.96	1.00																		
X2X6	0.99	0.99	-0.70	-0.14	0.99	0.98	1.00	-0.43	0.97	0.99	0.98	1.00	0.96	0.99	0.83	0.94	-0.70	0.99	0.98	1.00																	
X2X7	-0.53	-0.59	0.71	0.17	-0.58	-0.47	-0.54	0.96	-0.44	-0.51	-0.59	-0.51	-0.42	-0.47	-0.10	-0.39	0.72	-0.54	-0.47	-0.53	1.00																
X2X8	0.94	0.92	-0.56	-0.10	0.93	0.98	0.95	-0.36	1.00	0.95	0.91	0.97	0.99	0.98	0.78	0.99	-0.56	0.95	0.99	0.96	-0.43	1.00															
X3X4	0.99	1.00	-0.77	-0.10	1.00	0.95	0.99	-0.46	0.93	0.98	1.00	0.99	0.92	0.97	0.83	0.90	-0.77	0.99	0.95	0.99	-0.58	0.93	1.00														
X3X5	0.96	0.94	-0.63	-0.13	0.95	1.00	0.98	-0.39	0.99	0.95	0.94	0.98	0.99	0.99	0.80	0.97	-0.63	0.95	1.00	0.98	-0.47	0.98	0.95	1.00													
X3X6	1.00	0.99	-0.72	-0.14	0.99	0.98	1.00	-0.44	0.96	0.98	0.99	1.00	0.96	0.99	0.83	0.93	-0.72	0.99	0.98	1.00	-0.54	0.96	0.99	0.98	1.00												
X3X7	-0.42	-0.47	0.51	0.20	-0.46	-0.38	-0.44	1.00	-0.36	-0.42	-0.47	-0.41	-0.35	-0.39	0.08	-0.33	0.52	-0.43	-0.38	-0.43	0.96	-0.35	-0.46	-0.39	-0.44	1.00											
X3X8	0.94	0.92	-0.57	-0.10	0.93	0.99	0.96	-0.36	1.00	0.94	0.92	0.97	0.99	0.99	0.79	1.00	-0.58	0.95	0.99	0.96	-0.44	1.00	0.93	0.99	0.96	-0.36	1.00										
X4X5	0.93	0.90	-0.53	-0.13	0.91	0.99	0.95	-0.34	0.99	0.92	0.90	0.96	1.00	0.98	0.77	0.99	-0.54	0.93	0.99	0.95	-0.41	0.99	0.91	0.99	0.95	-0.34	0.99	1.00									
X4X6	0.97	0.95	-0.60	-0.14	0.96	0.99	0.98	-0.38	0.99	0.97	0.95	0.99	0.99	1.00	0.81	0.98	-0.60	0.97	0.99	0.98	-0.46	0.99	0.96	0.99	0.98	-0.38	0.99	0.99	1.00								
X4X7	0.93	0.93	-0.71	-0.12	0.93	0.90	0.93	-0.14	0.88	0.92	0.93	0.93	0.87	0.91	0.97	0.85	-0.70	0.93	0.90	0.93	-0.30	0.88	0.93	0.90	0.93	-0.14	0.88	0.86	0.90	1.00							
X4X8	0.90	0.87	-0.49	-0.10	0.89	0.97	0.92	-0.32	0.99	0.91	0.87	0.95	0.99	0.97	0.75	1.00	-0.49	0.91	0.97	0.93	-0.38	0.99	0.89	0.97	0.92	-0.32	0.99	0.99	0.98	0.84	1.00						
X5X6	0.91	0.87	-0.50	-0.14	0.89	0.98	0.93	-0.32	0.99	0.90	0.87	0.95	1.00	0.97	0.75	0.99	-0.50	0.90	0.98	0.93	-0.38	0.98	0.89	0.98	0.93	-0.32	0.99	1.00	0.98	0.84	0.99	1.00					
X5X7	0.96	0.94	-0.64	-0.16	0.95	0.99	0.98	-0.33	0.98	0.95	0.94	0.98	0.98	0.99	0.84	0.96	-0.64	0.95	0.99	0.97	-0.44	0.97	0.95	0.99	0.97	-0.33	0.98	0.98	0.99	0.93	0.96	0.97	1.00				
X5X8	0.84	0.80	-0.42	-0.10	0.82	0.94	0.87	-0.28	0.97	0.85	0.80	0.90	0.97	0.93	0.69	0.99	-0.42	0.85	0.94	0.87	-0.33	0.97	0.82	0.94	0.87	-0.28	0.97	0.98	0.94	0.78	0.99	0.98	0.92	1.00			
X6X7	0.99	0.98	-0.72	-0.18	0.98	0.97	0.99	-0.34	0.95	0.97	0.98	0.98	0.94	0.97	0.89	0.91	-0.73	0.98	0.96	0.99	-0.47	0.94	0.98	0.96	0.99	-0.34	0.94	0.93	0.97	0.97	0.90	0.91	0.98	0.85	1.00		
X6X8	0.88	0.85	-0.46	-0.11	0.87	0.96	0.91	-0.30	0.99	0.89	0.85	0.93	0.98	0.96	0.73	1.00	-0.47	0.89	0.96	0.91	-0.36	0.99	0.86	0.96	0.91	-0.30	0.99	0.99	0.97	0.82	1.00	0.99	0.95	0.99	0.89	1.00	
X7X8	0.94	0.92	-0.58	-0.13	0.93	0.98	0.96	-0.32	1.00	0.95	0.92	0.97	0.99	0.99	0.83	0.99	-0.58	0.95	0.99	0.96	-0.41	0.99	0.93	0.99	0.96	-0.32	0.99	0.99	0.99	0.91	0.99	0.98	0.99	0.96	0.96	0.98	1.00

The matrix in Table 2 shows that correlation matrix with interaction variables are predicted mullticollineaity in the data set. According to the factors mentioned in Table 2, there is positive association between two factors (greater than 0.9) and the factors are highly correlated with each other. Correlation matrix for the main factors of climate change is calculated by using Excel software with dependent and independent variables without multicollinearity effect.

Box plots of climate change with interaction

Boxplots of the main factors for climate change is calculated by using the Minitab software with dependent and independent variables is provided with interaction effect in Figures 1 to 13

Journal of Media Horizons
ISSN: 2710-4060 2710-4052 Volume 6, Issue 5, 2025

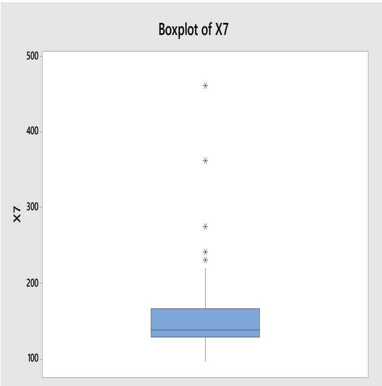

Figure 1

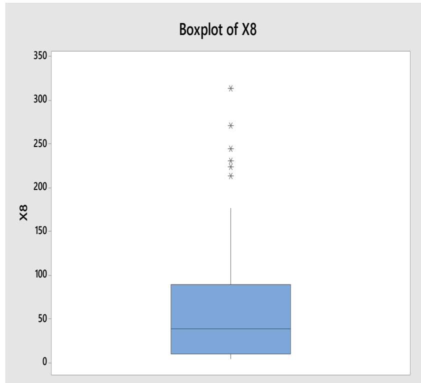

Figure 2

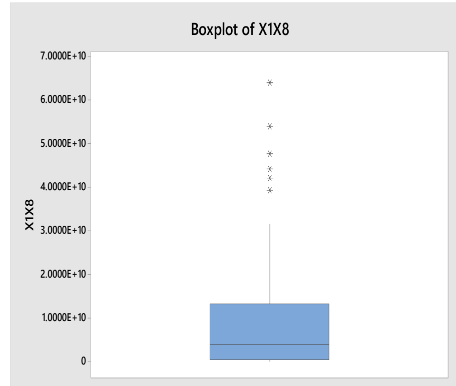

Figure 3

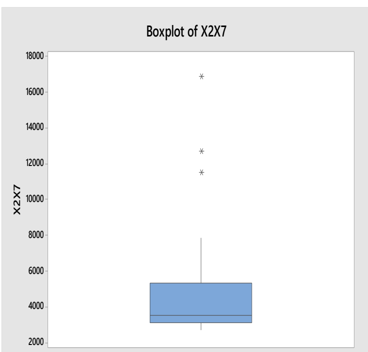

Figure 4

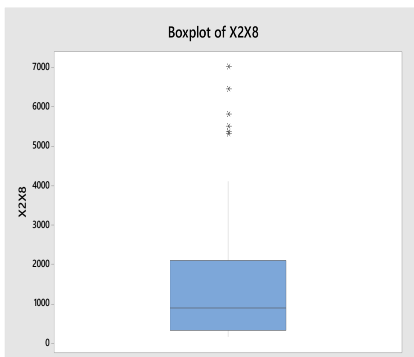

Figure 5

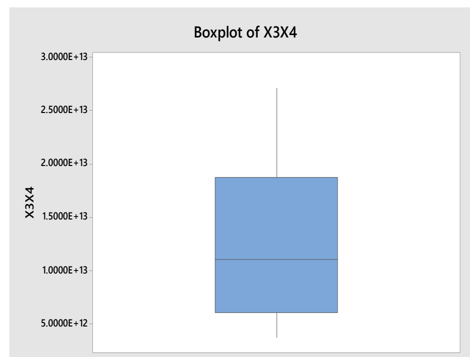

Figure 6

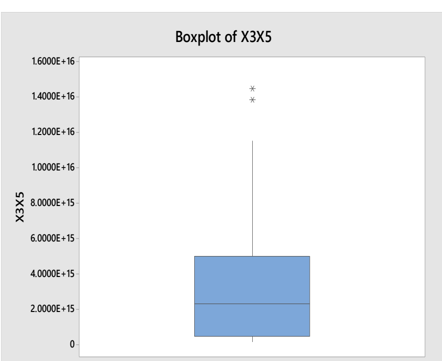

Figure 7

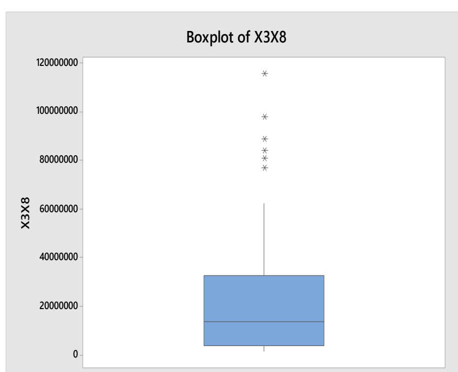

Figure 8

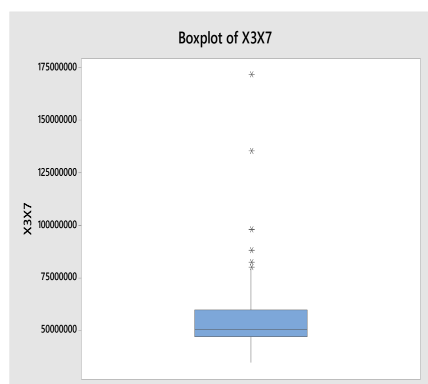

Figure 9

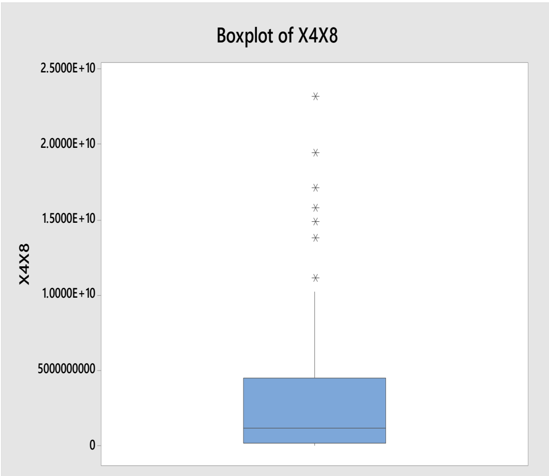

Figure 10

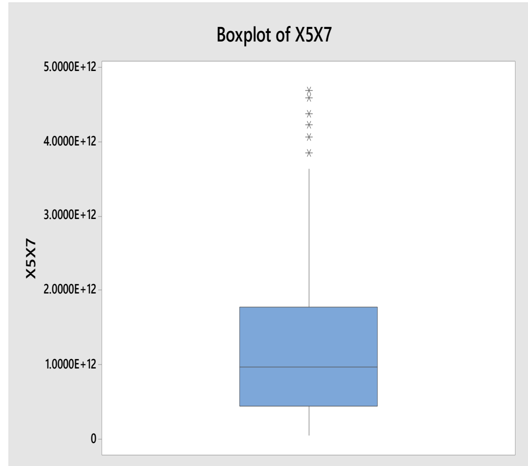

Figure 11

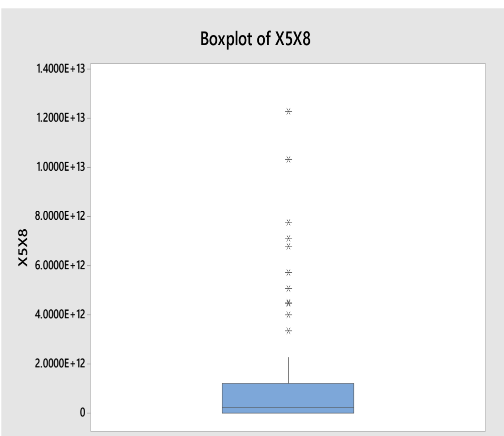

Figure 12

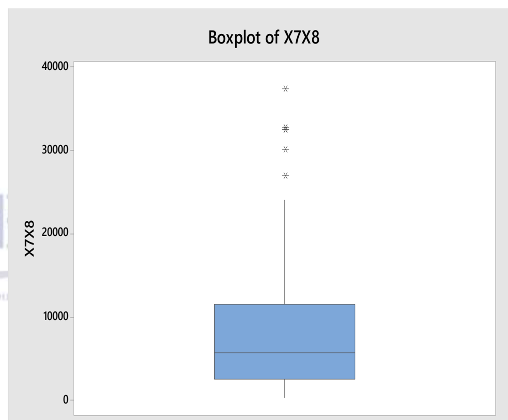

Figure 13

Box Plots representing the interaction variables, namely Population (total value), Agricultural (forestry, and fishing value added % of GDP) value, Agricultural land (sq.km) value, Urban population value, Gross fixed capital formation (current US$), Industry including (construction value added constant 2010 US$), Fertilizer consumption (% of fertilizer production), and GDP. These Box Plots are generated using the Minitab Software and display outliers are observed among the variables.

Best Selection Model

After the analysis of multicollinearity and outliers, the regression analysis is carried out of the best selection model due to presence of multicollinearity and outliers.

Ridge Regression Analysis in the Dataset

To calculate the significant factors in the presence of outliers of climate change Ridge Regression is used.

Ridge Regression Analysis With Interaction Effect

Ridge regression for the main factors for climate change is calculated by using R-studio Software with interaction is provided in Table 3

Table 3: Results of Ridge Regression Analysis with interaction effect

Factors	Coefficients	*P*-value	Significance
X_1	$8.372178e^{-05}$	0.00028	Significant
X_2	$5.945582e^{+01}$	0.00035	Significant
X_3	$-6.426969e^{-02}$	0.00021	Significant
X_4	$2.035955e^{-04}$	0.0004	Significant
X_5	$1.806122e^{-07}$	0.00012	Significant
X_6	$3.194822e^{-07}$	0.00051	Significant
X_7	$-1.736236e^{+01}$	0.0001	Significant
X_8	$4.800941e^{+00}$	0.00023	Significant
X_1X_2	$3.403357e^{-06}$	0.00043	Significant
X_1X_3	$2.179624e^{-10}$	0.00055	Significant
X_1X_4	$7.307350e^{-13}$	0.00041	Significant
X_1X_5	$3.555848e^{-16}$	0.00036	Significant
X_1X_6	$1.084049e^{-15}$	0.00045	Significant
X_1X_7	$3.509629e^{-07}$	0.00035	Significant
X_1X_8	$-2.982374e^{-08}$	0.00074	Significant
X_2X_3	$2.558868e^{-04}$	0.00023	Significant
X_2X4	$7.035499e^{-06}$	0.00054	Significant
X_2X_5	$4.273402e^{-09}$	0.00033	Significant
X_2X_6	$1.111471e^{-08}$	0.00024	Significant
X_2X_7	$1.328889e^{-01}$	0.00013	Significant
X_2X_8	$-1.503651e^{-01}$	0.00011	Significant
X_3X_4	$5.478942e^{-10}$	0.00028	Significant
X_3X_5	$4.676628e^{-13}$	0.00043	Significant
X_3X_6	$8.616280e^{-13}$	0.00062	Significant
X_3X_7	$-4.072266e^{-05}$	0.00011	Significant
X_3X_8	$8.942598e^{-06}$	0.00025	Significant
X_4X_5	$6.339424e^{-16}$	0.00047	Significant
X_4X_6	$2.685392e^{-15}$	0.00051	Significant
X_4X_7	$9.895899e^{-07}$	0.00024	Significant
X_4X_8	$-1.169435e^{-07}$	0.00018	Significant
X_5X_6	$5.967463e^{-19}$	0.00045	Significant
X_5X_7	$8.431804e^{-10}$	0.00030	Significant
X_5X_8	$-6.610492e^{-10}$	0.00015	Significant
X_6X_7	$2.097423e^{-09}$	0.0007	Significant
X_6X_8	$-2.203028e^{-10}$	0.0004	Significant
X_7X_8	$-1.118114e^{-02}$	0.00011	Significant

Table 3 displays the results of the Ridge regression analysis with interaction effects. The variables and their interactions are deemed statistically significant due to their predictors' p-values being below 0.05. The findings indicate the following significant factors and their respective impacts on CO_2 emissions: The total population exhibits a coefficient of ($8.372178e^{-05}$) per 1000 units, suggesting a small positive influence on CO_2 emissions. Thus, for every increase of 1000 units in population, CO_2 emissions are estimated to rise by approximately ($8.372178e^{-05}$) units.

"Agriculture, forestry, fishing" is identified as another significant factor, with a coefficient of ($5.945582e^{+01}$) per 100 units. This implies that for every increase of 100 units in "Agriculture, forestry, fishing," CO_2 emissions are projected to increase by approximately $5.945582e^{+01}$ units in 1000 kilotons. Agricultural land is also a significant factor, with a coefficient of ($-6.426969e^{-02}$) per 1000 square kilometers. This negative coefficient suggests that for every decrease of 1000 square kilometers in agricultural land, CO_2 emissions are anticipated to decrease by approximately ($-6.426969e^{-02}$) units in 1000 kilotons. Urban population demonstrates significance, with a coefficient of ($2.035955e^{-04}$) per 1000 units. This indicates that for every increase of 1000 units in urban population, CO_2 emissions tend to rise by approximately $2.035955e^{-04}$ units in 1000 kilotons. Gross fixed capital formation is found to be significant, with a coefficient of ($1.806122e^{-07}$). It suggests that an increase of 1 unit in USD for gross fixed capital formation corresponds to a rise of $1.806122e^{-07}$ units in 1000 kilotons of CO_2 emissions. The industry is identified as a significant factor, with a coefficient of ($3.194822e^{-07}$). This indicates that the presence of the industry influences CO_2 emissions, as the coefficient is positive. Fertilizer consumption is another significant factor, with a coefficient of ($-1.736236e^{+01}$). This suggests that a decrease of 100 units in fertilizer consumption is associated with a decrease of approximately ($-1.736236e^{+01}$) units in 1000 kilotons of CO_2 emissions. GDP is considered a non-significant factor, with a coefficient of ($4.800941e^{+00}$) units increase in CO_2 emissions. Additionally, several significant interactions between factors are observed: The interaction between population and "Agriculture, forestry, fishing" exhibits a coefficient of ($3.403357e^{-06}$), indicating a significant effect on CO_2 emissions. The interaction between population and agricultural land shows a coefficient of ($2.179624e^{-10}$), signifying a significant impact on CO_2 emissions. The interaction between population and urban population yields a coefficient of ($7.307350e^{-13}$), demonstrating a significant effect on CO_2 emissions. The interaction between population and gross fixed capital formation reveals a coefficient of ($3.555848e^{-16}$), suggesting a significant influence on CO_2 emissions. The interaction between population and industry exhibits a coefficient of ($1.084049e^{-15}$), indicating a significant effect on CO_2 emissions. The interaction between population and fertilizer consumption shows a coefficient of ($3.509629e^{-07}$), suggesting a significant impact on CO_2 emissions. The interaction between population and GDP yields a coefficient of ($-2.982374e^{-08}$), signifying a significant effect on CO_2 emissions. Several other interactions between different factors are also observed, with each showing significant effects on CO_2 emissions. These findings highlight the significant relationships and interactions among various factors and their impacts on CO_2 emissions.

LASSO Regression Analysis With Interaction Effect

LASSO regression for the main factors for climate change is calculated by using R-studio Software with interaction is provided in Table 4

Table 4: Results of LASSO Regression Analysis with interaction effect

Factors	Coefficients	*P*-value	Significance
Intercept	$-3.243644e^{+04}$	0.0001	Significant
X_1	$9.225649e^{-05}$	0.0003	Significant
X_2	$6.432281e^{+02}$	0.0005	Significant
X_4	$4.665809e^{-04}$	0.00025	Significant
X_6	$2.201602e^{-06}$	0.0006	Significant
X_1X_3	$7.789301e^{-11}$	0.0007	Significant
X_1X_7	$1.252515e^{-07}$	0.0007	Significant
X_2X_7	$1.645712e^{-01}$	0.0008	Significant

X_3X_4	$7.059446e^{-10}$	0.0006	Significant
X_3X_6	$2.111283e^{-12}$	0.0009	Significant
X_5X_8	$-1.108011e^{-09}$	0.0004	Significant

Table 4 presents the results of the Lasso regression analysis with interaction effects, indicating the statistical significance of the variables and their interactions, as their p-values are below 0.05. The analysis reveals that CO_2 emissions vary based on different indicators, with an estimated increase of ($-3.243644e^{+04}$) kilotons. One significant factor is the total population, which exhibits a positive impact on CO_2 emissions, with an estimated coefficient of ($9.225649e^{-05}$) per 1000 units. This suggests that for every increase of 1000 units in population, CO_2 emissions tend to rise by approximately ($9.225649e^{-05}$) units. Another significant factor is "Agriculture, forestry, fishing," indicating a positive effect on CO2 emissions. With a coefficient of ($6.432281e^{+02}$), it implies that for every increase of 100 units in "Agriculture, forestry, fishing," CO_2 emissions increase by approximately $6.432281e^{+02}$ units in 1000 kilotons. The industry is also identified as a significant factor that influences CO_2 emissions. The coefficient of ($2.201602e^{-06}$) suggests that the presence of the industry contributes to CO_2 emissions, as the value is positive. Furthermore, several significant interactions between factors are observed: The interaction between population and agricultural land shows a coefficient of ($7.789301e^{-11}$), indicating a significant effect on CO_2 emissions. The interaction between population and fertilizer consumption exhibits a coefficient of ($1.252515e^{-07}$), suggesting a significant impact on CO_2 emissions. The interaction between "Agriculture, forestry, fishing" and fertilizer consumption yields a coefficient of ($1.645712e^{-01}$), signifying a significant effect on CO_2 emissions. The interaction between agricultural land and urban population reveals a coefficient of ($7.059446e^{-10}$), demonstrating a significant impact on CO_2 emissions. The interaction between agricultural land and industry exhibits a coefficient of ($2.111283e^{-12}$), indicating a significant effect on CO_2 emissions. The interaction between gross fixed capital formation and GDP shows a coefficient of ($-1.108011e^{-09}$), suggesting a significant influence on CO_2 emissions. These findings highlight the significant relationships and interactions among various factors and their impacts on CO_2 emissions.

Ordinary Least Square (OLS) with Interaction effect

Ordinary Least Square (OLS) for the main factors for climate change is calculated by using R-studio Software with interaction is provided in Table 5

Table 5: Results of Ordinary Least Square with interaction

Factors	Coefficients	*P*-value	Significance
Intercept	$3.905e^{+06}$	0.10015	Non-Significant
X_1	$-2.408e^{-01}$	0.00866**	Significant
X_2	$-1.011e^{+04}$	0.66408	Non-Significant
X_3	$-8.413e^{+00}$	0.18723	Non-Significant
X_4	$7.095e^{-01}$	0.00625**	Significant
X_5	$1.581e^{-04}$	0.10993	Non-Significant
X_6	$-1.365e^{-04}$	0.01366*	Significant
X_7	$-6.711e^{+02}$	0.32201	Non-Significant
X_8	$-2.732e^{+04}$	0.05835	Non-Significant
X_1X_2	$8.938e^{-04}$	0.00856**	Significant
X_1X_3	$5.444e^{-07}$	0.02434*	Significant
X_1X_4	$1.260e^{-10}$	0.0481**	Significant

X_1X_5	-6.233e^{-12}	0.00482**	Significant
X_1X_6	3.455e^{-12}	0.01256*	Significant
X_1X_7	3.351e^{-05}	0.31216	Non-Significant
X_1X_8	2.691e^{-04}	0.12756	Non-Significant
X_2X_3	-1.585e^{-02}	0.79958	Nom-Significant
X_2X4	-2.699e^{-03}	0.01299*	Significant
X_2X_5	-1.511e^{-06}	0.00887**	Significant
X_2X_6	1.203e^{-06}	0.14874	Non-Significant
X_2X_7	2.157e^{-01}	0.96872	Non-Significant
X_2X_8	1.007e^{+02}	0.02370*	Significant
X_3X_4	-1.639e^{-06}	0.01592*	Significant
X_3X_5	1.383e^{-10}	0.33247	Non-Significant
X_3X_6	9.416e^{-11}	0.49384	Non-Significant
X_3X_7	3.583e^{-04}	0.74509	Non-Significant
X_3X_8	4.149e^{-02}	0.12816	Non-Significant
X_4X_5	1.554e^{-11}	0.00611**	Significant
X_4X_6	-9.582e^{-12}	0.00638**	Significant
X_4X_7	-1.029e^{-04}	0.29508	Non-Significant
X_4X_8	-6.856e^{-04}	0.14751	Non-Significant
X_5X_6	5.125e^{-16}	0.30386	Non-Significant
X_5X_7	-9.864e^{-08}	0.06475	Non-Significant
X_5X_8	-1.970e^{-07}	0.00164**	Significant
X_6X_7	3.990e^{-08}	0.23371	Non-Significant
X_6X_8	2.354e^{-07}	0.04571*	Significant
X_7X_8	1.211e^{+01}	0.04710*	Significant

Table 5 displays the regression results, revealing the statistical significance of the variables X_1, X_4, and X_6 as their p-values are below 0.05. Conversely, the variables X_2, X_3, X_5, X_7, and X_8 are deemed non-significant due to their p-values exceeding 0.05 (Suhaeri et al., 2021). The primary significant factor is the total population, which exhibits a negative impact with an estimated coefficient of (-2.408e^{-01}) per 1000 units. This suggests that for every increase of 1000 units in population, CO_2 emissions tend to decrease by approximately (0.2408) units. The subsequent significant factor is urban population, indicating a positive effect on CO_2 emissions. With a coefficient of (7.095e^{-01}), it implies that for every increase of 1000 units in urban population, CO_2 emissions rise by approximately 0.7095 units. Furthermore, the variable X_6, representing the industry (including construction), is found to be a significant factor. Its coefficient of (-1.365e-04) suggests that an increase of 1 US Dollar unit in the industry leads to a rise of 0.0001365 units in CO_2 emissions, considering other predictors remain constant. Additionally, several significant interactions between factors are observed: The interaction between population and (agriculture, forestry, and fishing) shows a coefficient of (8.938e^{-04}), indicating a significant effect on CO_2 emissions.

The interaction between population and agricultural land exhibits a coefficient of (5.444e^{-07}), suggesting a significant impact on CO_2 emissions (Lim et al., 2020). The interaction between population and urban population displays a coefficient of (1.260e^{-10}), signifying a significant effect on CO_2 emissions. The interaction between population and gross fixed capital formation yields a coefficient of (-6.233e^{-12}), demonstrating a significant influence on CO_2 emissions. The interaction between population and industry reveals a coefficient of ($3.455e^{-12}$), indicating a significant effect on CO_2 emissions. The interaction between (agriculture, forestry, and fishing) and urban population displays a coefficient of ($-2.699e^{-03}$), suggesting a significant impact on CO_2 emissions. The interaction between (agriculture, forestry, and fishing) and gross fixed formation exhibits a coefficient of ($-1.511e^{-06}$), signifying a significant effect on CO_2 emissions. The interaction between (agriculture, forestry, and fishing) and GDP shows a coefficient of ($1.007e^{+02}$), indicating a significant influence on CO_2 emissions. The interaction between agriculture land and urban population yields a coefficient of ($-1.639e^{-06}$), demonstrating a significant effect on CO_2 emissions. The interaction between urban population and gross fixed capital formation reveals a coefficient of ($1.554e^{-11}$), suggesting a significant impact on CO_2 emissions. The interaction between urban population and industry exhibits a coefficient of ($-9.582e^{-12}$), signifying a significant effect on CO_2 emissions. The interaction between gross fixed capital formation and GDP yields a coefficient of ($-1.970e^{-07}$), indicating a significant influence on CO_2 emissions. The interaction between industry and GDP shows a coefficient of ($2.354e^{-07}$), suggesting a significant impact on CO_2 emissions (Javaid and Akbar; 2020). The interaction between fertilizer consumption and GDP exhibits a coefficient of ($1.211e^{+01}$), signifying a significant effect on CO_2 emissions. These findings highlight the significant relationships and interactions among various factors that impact CO_2 emissions.

Results for Without Multicollinearity Variables

Without multicollinearity of the variables by using the R-Studio Software without interaction is provided in Table 6

Table 6: Results of without multicollinearity variables

Factors	Coefficients	*P*-value	Significance
Intercept	$-6.050e^{+04}$	0.172091	Non-Significant
X_3	$3.113e^{-01}$	0.041720*	Significant
X_2X_3	$-3.898e^{-03}$	0.033462*	Significant
X_2X_6	$1.122e^{-07}$	2.24e-05***	Significant
X_2X_7	$1.234e^{+01}$	0.003275**	Significant
X_3X_7	$-1.477e^{-03}$	0.001110**	Significant
X_4X_7	$8.956e^{-06}$	0.000409***	Significant
X_5X_7	$-4.048e^{-10}$	0.885697	Non-Significant

Table 6 presents the regression results, indicating the statistical significance of the variables X_3, X_2X_3, X_2X_6, X_2X_7, X_3X_7, and X_4X_7, as their p-values are below 0.05. On the other hand, the variable X_5X_7 found to be non-significant because its p-value exceeds 0.05 (Javaid et al., 2019).Among the significant factors, the variable X3 (representing Agriculture land) has a positive impact on CO_2 emissions, with an estimated coefficient of ($3.113e^{-01}$) per 1000 units of X3 (in 1000 kiloton CO_2 emission) (Javed et al., 2022). This suggests that for every increase of 1000 units in Agriculture land, CO_2 emissions tend to rise by approximately (0.3113) units. Furthermore, the interaction between the variables (Agricultural, forestry, fishing) and (Agriculture land) as denoted by X2X3 has a significant negative effect on CO_2 emissions, with a coefficient of ($-3.898e^{-03}$). This implies that when both factors increase, CO_2 emissions decrease by approximately (0.003898) units, all other predictors held constant. Similarly, the interactions between (Agricultural, forestry, fishing) and (Industry), (Agricultural, forestry, fishing) and (Fertilizer Consumption), (Agricultural land) and (Fertilizer Consumption), as well as (Urban Population) and (Fertilizer

Consumption) are all found to be significant factors affecting CO_2 emissions. The respective coefficients for these interactions are ($1.122e^{-07}$), ($1.234e^{+01}$), ($-1.477e^{-03}$), and ($8.956e^{-06}$), indicating their individual impacts on the outcome variable when the other predictors are held constant. Overall, these findings demonstrate the significant influence of various factors on CO_2 emissions, particularly in the context of Agriculture land and its interactions with other relevant variables.

Model Selection Criteria

To select a best model in each dataset model selection criteria Loss function is used.

Table 7: Comparison among results of SSE, SST, R^2, MSE, RMSE and MAPE without and with interaction variable.

	SSE	SST	R^2	MSE	RMSE	MAPE
Ridge	1108195999	153398709961	0.9927757	2650814221	51486.05852	50.86453
LASSO	467752489	153398709961	0.9969507	240881825.7	15520.36809	33.04923
OLS				6052381.8	2460.1589	4.648196
Without Multicolinearity				1041181699	32267.34726	67.56007

Table 7 presents the results of SSE, SST, R^2, MSE, RMSE, and MAPE for the model with and without interaction variables in this study (Javaid et al., 2020). The presence of outliers affects the extreme values of these metrics due to the influence of the loss function. The ridge regression model demonstrates the lowest MAPE, which can be attributed to its ability to reduce variance and potentially minimize the mean absolute percentage error. Conversely, the ordinary least squares (OLS) technique exhibits the lowest MSE value. However, caution is required when interpreting OLS results in the presence of outliers since they may violate certain assumptions of OLS as outlined by Gujrati (2008). Therefore, excluding OLS from consideration, a comparison is conducted among the remaining three techniques. LASSO regression with interaction variables shows larger R^2, lower SSE and RMSE values. On the other hand, OLS with interaction variables displays the lowest MSE, RMSE, and MAPE values. Nonetheless, it is important to note that the OLS results can be misleading in the presence of outliers due to the violation of OLS assumptions. Consequently, our preference lies with the ridge regression model without interaction variables on the basis of LASSO regression model with interaction variables on the basis of larger, maximum R^2 , minimum SSE and RMSE.

REFERENCES

Afzal, S. N., Haque, M. I., Ahmedani, M. S., Rauf, A., Munir, M., Firdous, S. S., ... & Ahmad, I. (2008). Impact of stripe rust on kernel weight of wheat varieties sown in rainfed areas of

Pakistan. Pakistan Journal of botany, 40(2), 923-929.

Ali, S., Li, D., Congbin, F., & Khan, F. (2015). Twenty first century climatic and hydrological changes over Upper Indus Basin of Himalayan region of Pakistan. Environmental Research Letters, 10(1), 014007

Eckstein, D., V. Künzel, et al. (2021). The global climate risk index 2021, Bonn: Germanwatch.

Haider, S., & Adnan, S. (2014). Classification and assessment of aridity over Pakistan provinces (1960-2009). International Journal of Environment, 3(4), 24-35.

Parthasarathy, B., N. Sontakke, et al. (1987). "Droughts/floods in the summer monsoon season over different meteorological subdivisions of India for the period 1871–1984." Journal of Climatology **7**(1): 57-70.

Rashid, K., & Rasul, G. (2011). Rainfall variability and maize production over the Potohar Plateau of Pakistan. Pakistan Journal of Meteorology, 8(15), 63-64.

Rasul, G. and B. Ahmad (2012). "Climate change in Pakistan." Pakistan Meteorological Department.

Viechtbauer, W. and M. W. L. Cheung (2010). "Outlier and influence diagnostics for meta-analysis." Research synthesis methods **1**(2): 112-125.

Voigt, C., Lamprecht, R. E., Marushchak, M. E., Lind, S. E., Novakovskiy, A., Aurela, M., ... & Biasi, C. (2017). Warming of subarctic tundra increases emissions of all three important greenhouse gases–carbon dioxide, methane, and nitrous oxide. Global Change Biology, 23(8), 3121-3138.

Woolway, R. I., Kraemer, B. M., Lenters, J. D., Merchant, C. J., O'Reilly, C. M., & Sharma, S. (2020). Global lake responses to climate change. Nature Reviews Earth & Environment, 1(8), 388-403.

Younger, M., H. R. Morrow-Almeida, et al. (2008). "The built environment, climate change, and health: opportunities for co-benefits." American journal of preventive medicine **35**(5): 517-526.

Haigh, J. D. (1999). A GCM study of climate change in response to the 11-year solar cycle. Quarterly Journal of the Royal Meteorological Society, 125(555), 871-892.

Hussain, M., Butt, A. R., Uzma, F., Ahmed, R., Irshad, S., Rehman, A., & Yousaf, B. (2020). A comprehensive review of climate change impacts, adaptation, and mitigation on environmental and natural calamities in Pakistan. Environmental monitoring and assessment, 192, 1-20.

Shehzad, K. (2023). Extreme flood in Pakistan: Is Pakistan paying the cost of climate change? A short communication. Science of The Total Environment, 880, 162973.

Ali, S., Liu, Y., Ishaq, M., Shah, T., Abdullah, Ilyas, A., & Din, I. U. (2017). Climate change and its impact on the yield of major food crops: Evidence from Pakistan. Foods, 6(6), 39.

Gaur, M. K., & Squires, V. R. (2018). Geographic extent and characteristics of the world's arid zones and their peoples. Climate variability impacts on land use and livelihoods in drylands, 3-20.

Javaid, A., Ismail, M. T., & Ali, M. K. M. (2019, December). Model selection for collector efficiency of seaweed drier by using LASSO and multiple regression analysis using 8sc. In *AIP Conference Proceedings* (Vol. 2184, No. 1, p. 050032). AIP Publishing LLC.

Javaid, A., & Akbar, A. (2020). Factors Identification for Human Development Index. *STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH*, *2*(1), 1-8.

Javaid, A., Ismail, M. T., & Ali, M. K. M. (2020). Comparison of sparse and robust regression techniques in efficient model selection for moisture ratio removal of seaweed using solar drier. *Pertanika Journal of Science and Technology*, *28*(2), 609-625.

Javed, Z., Javaid, A., Javaid, S., & Javed, A. (2022). Analysis of Child Mortality on Geographical Basis Over Different Centuries. *STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH*, *4*(2), 1-7.

Laity, J. J. (2009). Deserts and desert environments (Vol. 3). John Wiley & Sons.

Whiteman, C. D. (2000). Mountain meteorology: fundamentals and applications. Oxford University Press.

Lieven, A. (2011). Pakistan: A hard country. Penguin UK.

Lim, H. Y., Fam, P. S., Ali, M., & Khan, M. (2020). Ridge Regression as Efficient Model Selection and Forecasting of Fish Drying Using V-Groove Hybrid Solar Drier. *Pertanika Journal of Science & Technology*, *28*(4).

Lodhi, I., Nawaz, S., Javaid, A., Javaid, S., & Javaid, A. (2023). Geographically Analysis of Wheat Production on Annual Basis. *Statistics, Computing and Interdisciplinary Research*, *5*(1), 29-36.

Khan, S. (2019). Climate Classification of Pakistan: Climate Classification of Pakistan. International Journal of Economic and Environmental Geology, 10(2), 60-71.

Kundzewicz, Z. W. (2008). Climate change impacts on the hydrological cycle. Ecohydrology & Hydrobiology, 8(2-4), 195-203.

White, H. (1980). Using least squares to approximate unknown regression functions. International economic review, 149-170.

Venkatesh, K. A., Mishra, D., & Manimozhi, T. (2023). Model selection and regularization. In Statistical Modeling in Machine Learning (pp. 159-178). Academic Press.

Marquardt, D. W., & Snee, R. D. (1975). Ridge regression in practice. The American Statistician, 29(1), 3-20.

Suhaeri, M. E., Alimudin, Javaid, A., Ismail, M. T., & Ali, M. K. M. (2021, November). Evaluation of clustering approach with euclidean and Manhattan distance for outlier detection. In *AIP Conference Proceedings* (Vol. 2423, No. 1, p. 070025). AIP Publishing LLC.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288.