Evaluation and comparison of multiple linear regression and contraction algorithms (Ridge, Lasso, and Elastic net) using artificial neural networks

H. M. Almosawi , M. H. Behzad

hanan_almosawi@yahoo.com , behzadi.stat@gmail.com

Department of Statistics, Science and Research Branch Islamic Azad University, Tehran, Iran

Abstract

The Goal of this paper is the statistical analysis of the data of the global warming phenomenon caused by the increase in the amount of carbon dioxide, methane, nitrous oxide and some other gases in the atmosphere that affect climate and environment pollution. by using artificial neural networks. The least squares method of estimating coefficients of a linear model is usually relied upon to find the fit line with the smallest prediction error among all observations. Usually, the method of least squares is used to estimate the coefficients of the linear model and to find the fit line for all observations with the least prediction error. Usually, the method of least squares is used to estimate the coefficients of the linear model and to find the fit line for all observations with the least prediction error. After statistical analysis, we found that the ordinary least squares (OLS) regression method is better than the rest of the regression of Ridge, Lasso and Elastic Network. Since its estimates are unbiased and have minimum variance, the least squares regression is BLUE. The conclusion of this

paper is to obtain a linear model estimated by the least squares method with the best measurements: MSE, RMSE, MAE, Adjusted R-Squared, R-Squared.

Keywords: multiple linear regression, ridge regression, lasso regression, elastic net regression, artificial neural networks.

Abbreviations

MLR	Multiple linear regression	λ	Lamda
OLS	Ordinary least squares	*RR*	Ridge Regression
$\hat{\beta}$	parameter estimation	*CV*	Cross- validation
MAE	mean absolut error	X'	X Transpose
MSE	mean squares error	*GHG*	greenhouse gases

1. Introduction

Multiple linear regression (MLR), is a statistical technique used on many explanatory variables in order to predict the outcome of a dependent variable one of the Goal of multiple linear regression is to model the linear relationship between the explanatory (independent) variables and the dependent variables (Yang,2017)[1]. in its content, multiple linear regression is similar to ordinary least squares (OLS) regression because it contains more than one explanatory variable. OLS regression has Characterized in easily checking model assumptions such as linearity, constant, variance, and influence of outliers by using simple graphics (Hutcheson, G & Sofroniou. N, ,1999)[2]. (J. T. Kilmer & R. L. Rodriguez,2016)[3] applied the OLS method in in studies of allometry. The meaning of linearity refers to the nature of the linear relationship between the explanatory variables, as the explanatory variables should not exist on the same line, if it appears that one of the linear variables depends on another linear variable, that variable can be isolated from the model. The main problem with previous studies that used the multiple linear regression method was the misconceptions about the multiple regression assumptions, the absence of a collinearity between between predictors, and the sufficient sample size. In the research (Williams et, 2013)[4] Multiple Regression Assumptions: Correcting Two Misconceptions, it was shown that multiple regression models estimated using ordinary least squares require the assumption of normally distributed errors for reliable conclusions. In research (Osborne & Waters, 2013)[5] four assumptions of multiple regression that researchers should always test. Researchers have intensified their efforts to raise awareness of the need to check assumptions when using regression.

2. Materials and Methods

This data, consisting of 150 observations, was collected over a period of time from 1990 to 2019 in the country of Iraq, and this data represents the amount of greenhouse gas emissions (CO2, N2O, CH4, FGAS) pulled from the official global website (Climate Watch) . The Research Ethics Committee on the site agreed to obtain the data for the purpose of the study after writing down my e-mail. All statistical procedures were performed using R and R Studio version 4.2.2 for Windows 10 , 64-bit. A scatter plot was created to display the relationship between the pair of variables.

Pearson correlations were calculated for the relationship between the principal factors. Regression analysis contributes to estimating the relationship between variables and the effect of independent variables on outcome variables within the population. Hence, regression analysis methods were used for the variables of interest. In order to determine the methods of this study, the decision support systems used to predict global warming were examined. Using MLR and ANN methods prediction metrics

3. multiple linear regression

Multiple Linear Regression (MLR), is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. It is a generalization of simple linear regression

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon, \qquad k = 1,2,\dots,n$$

Y: dependent variable observations.

X: independent variables.

β_o: intercept, which is the value of $E(Y/X = x)$ when x equals zero

β_k: slope coefficients for each explanatory variable

ε: the model's error term (also known as the residuals)

To write the forms for each of the observation n

$$Y = X\beta + \varepsilon \qquad (1)$$

$$Y_{n\times 1} = \underbrace{X_{n\times(k+1)}\beta_{(k+1)\times 1}}_{n\times 1} + \varepsilon_{n\times 1}$$

To write these equations in matrix form

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

In order to reduce the sum of squared differences between the observed and expected values, we use the OLS method (Douglas ,2012)[6].

Where $Y|_{n\times 1}$, $X|_{n\times k+1}$, $\beta|_{(k+1)\times 1}$, $\varepsilon|_{n\times 1}$. To find the smallest least squares vector $\beta^{\wedge}$ and With the usual assumptions such as

$$E(\varepsilon) = 0 \, , \begin{cases} var(\varepsilon_i / X) = \sigma^2 I_n, & i = 1,2,\dots\dots,n \\ cov(\varepsilon_i \varepsilon_j / X) = 0, & \text{if } \; i \neq j \end{cases}$$

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \acute{\varepsilon}\varepsilon = [Y - X\beta]'[\,Y - X\beta]$$

$$S(\beta) = Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta$$

$$S(\beta) = Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

Where $\beta'X'Y$ is a 1×1 matrix and transpose is the same scalar

$$(\beta'X'Y)' = Y'X\beta$$

$$\left.\frac{\partial S}{\partial \beta}\right|_{\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0, \qquad X'X\hat{\beta} = X'Y$$

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad (2)$$

The presence of the inverse matrix $(X'X)^{-1}$ indicates that the regression factors are linearly independent, so there is no column consisting of a sum of the other columns in the matrix X. When estimating the OLS model, it aims to estimate the coefficients $\hat{\beta}$ to keep the error term to a minimum.

Several assumptions must be met in order for these assumptions to give information about whether the OLS model is capable of this or not.

- Linearity of residuals (Jeffrey ,2009)[7].
- Normality of residuals (Wilson& Morgan,2007)[8].
- Residuals have equal variance (Berry and Feldman, 1985)[9], (Tabaknik and Fidel, 1996)[10].
- Independence of Residuals (Pedhazu, 1997)[11].

4. Ridge regression

Ridge regression is a method for estimating the coefficients of multiple regression models in scenarios where the independent variables are highly dependent on each other. This method provides improved performance in parameter estimation problems against an acceptable amount of bias. As a potential solution to the inaccuracy of least squares estimators when linear regression models contain some multicollinear independent variables, by creating a ridge regression estimator. This provides a more accurate estimate of the parameters, as the variance and mean square estimator are often smaller than previously derived least squares estimators. To show the derivation of the edge regression estimator equation, the sum of squared error $MSE(b) = E||b - \beta||^2$, we divide the least squares estimators into two parts, bias and variance.

$$E||b - \beta||^2 = \sum E(b_j - \beta_j)^2$$

$$= \sum \left[E\left(b_j\right) - \beta_j\right]^2 + var(b_j)$$

According to the Gauss-Markov theorem, least squares have the lowest variance compared to other unbiased estimators. But this does not mean that it has the minimum sum of squared error (MSE). Since $E(\hat{\beta}_{OLS}) = \beta$ and the variance matrix $var(\hat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}$the error boxes can be written as (Jibril ,2014)[12]

$$MSE(\hat{\beta}_{OLS}) = E\|\hat{\beta}_{OLS}\|^2 - \|\beta\|^2$$

$$= tr[\sigma^2(X'X)^{-1}] = \sigma^2 tr[(X'X)^{-1}]$$

$$E\|\hat{\beta}_{OLS}\|^2 = \|\beta\|^2 + \sigma^2 tr[(X'X)^{-1}]$$

Note that if the $X'X$ matrix is weak, the least squares estimator will have large dimensions $\|\beta\|^2$, which is associated with the reduction of variance and instability of the estimators. Here the role of ridge regression to solve the problem of multicollinearity by finding biased estimators is presented, but it has the advantage of having a more stable variance (Jafari et,2016)[13].

$$\hat{\beta}_{RR} = [X'X + \lambda I]^{-1}X'Y \quad (3)$$

In other words, we add a penalty equal to the square of the coefficient. The term L2 is equal to the square of the magnitude of the coefficients. We also add a coefficient λ to control that penalty term. In this case, if $\lambda = 0$, the OLS equation is basic, and if $\lambda > 0$, it adds a restriction to the coefficient. As the value of λ increases, this limitation causes the value of the coefficient to tend to zero. This results in a trade-off of higher bias (dependence on specific coefficients of 0 and specific coefficients tending to be very high, which makes the model less flexible) for lower variance.

$$\sum_{i=1}^{n}(Y_i - X_i\beta)^2 + \lambda_1 \sum_{j=1}^{p}\left|\beta_j\right|_2^2 \qquad (4)$$

5. Lasso regression:

A type of linear regression that uses shrinkage. Contraction is where the data values have shifted towards a central point. This type of regression is suitable for models that exhibit high levels of multicollinearity, or when we want to select specific parts of the model, such as selecting a variable/dropout. A parameter. In this model, a function (penalty) is added as a penalty equivalent to the absolute value of the transaction volume, and this type of organization can remove some coefficients from the model. Larger penalty functions bring coefficient values closer to zero, which is ideal for generating simpler models. Lasso regression model algorithm with the objective of minimizing the value (Wessel ,2021)[14]

$$L(\beta,\lambda) = \|Y - X\beta\|_2^2 + \lambda_1\|\beta\|_1$$

$$\sum_{i=1}^{n}(Y_i - X_i\beta)^2 + \lambda_1 \sum_{j=1}^{p}\left|\beta_j\right| \qquad (5)$$

Some values reduce exactly to zero, resulting in an easily interpretable regression model. The tuning parameter λ controls the strength of the penalty function, since λ is essentially a shrinkage value, no parameter is removed when λ=0, and more coefficients are set to zero as the value of λ increases (theoretically when λ=∞. All coefficients are dropped), so the bias increases, but the variance increases as the value of λ decreases (Al-Azzawi ,2020)[15].

6. Elastic Net Regression

Elastic network regression is useful when there is multicollinearity, and this method, like the previous two methods, does not need to assume the residual value (error) is normal. In RR and LASSO, this soft term is called L1,L2. The measurement term is added to the cost function in the regression. This expression penalizes the regression cost function. in such a way that the sum of the absolute values of the coefficients increases. Sometimes, lasso regression can introduce a small bias in the model where the overprediction depends on a particular variable. In these cases, elastic mesh has proven to combine better lasso and ridge adjustment. The advantage is that it does not remove the high collinearity coefficient easily.

Note: In lasso regression and Ridge regression, we assume the normal residual, but in elastic regression we do not assume this, the multiple linear regression model can be considered as follows (Hojjatollah ,2020)[16].

$$\sum_{i=1}^{n}(Y_i - X_i\beta)^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j|_2^2 + \lambda_2 \sum_{j=1}^{p}|\beta_j| \quad (6)$$

7. Data for the study

This data for Iraq, consisting of 150 observations, was collected over a period of time from 1990 to 2019, and this data represents the amount of greenhouse gas emissions (CO2, N2O, CH4, FGAS,) from a web World site (WMO). And as shown in table (1).

Table (1) emission of greenhouse gases in Iraq

Year	GHG	CO2	CH4	N2O	FGas	predict
2019	320.82	174.56	134.69	5.67	5.9	320.82453
2018	303.53	163.15	130.22	4.78	5.38	303.53329
2017	292.61	155.08	127.95	4.71	4.86	292.60342

2016	276.85	142.23	126.14	4.13	4.34	276.84295
2015	253.93	133.17	113.17	3.77	3.82	253.93179
2014	241.98	134.04	99.79	4.7	3.44	241.97228
2013	241.8	139.1	94.42	5.22	3.06	241.80260
2012	228.19	129	91.3	5.2	2.68	228.18279
2011	202.03	113.04	81.75	4.95	2.3	202.04229
2010	188.67	108.55	73.57	4.63	1.92	188.67137
2009	172.69	94.2	72.04	4.71	1.74	172.69194
2008	161.61	88.14	67.88	4.04	1.55	161.61080
2007	142.99	76.65	61.31	3.67	1.37	143.00013
2006	146.44	82.96	58.51	3.8	1.18	146.44994
2005	145.79	85.13	55.8	3.86	1	145.78979
2004	150.62	88.11	58.26	3.3	0.95	150.61900
2003	129.41	81.28	44.26	2.97	0.9	129.40759
2002	155.42	91.16	57.25	6.15	0.85	155.41342
2001	169.07	97.55	64.81	5.9	0.81	169.07341
2000	160.81	87.63	68.18	4.24	0.76	160.81139
1999	143.56	72.71	66.02	4.17	0.65	143.55166
1998	153.73	91.78	57.07	4.33	0.55	153.73054
1997	162.64	118.91	39.05	4.23	0.45	162.63797
1996	129.07	98.88	25.83	4.02	0.34	129.06731
1995	130.52	100.07	26.1	4.12	0.24	130.52748
1994	132.29	102.12	25.55	4.41	0.22	132.29783
1993	116.23	89.56	21.96	4.5	0.2	116.21814
1992	88.87	67.34	17.86	3.5	0.17	88.86700
1991	62.92	45.55	14.39	2.83	0.15	62.91642
1990	120.28	64.21	51.78	4.16	0.13	120.28094

8. artificial neural networks

Artificial Neural Network (ANN) is a flexible and powerful machine learning technology. It has the same function as the human nervous system. There are input and output signals that travel from input nodes to output nodes, and they acquire their knowledge through training and store this knowledge using communication forces within neurons called synaptic weights. After feeding the network with data and by using the backpropagation algorithm which is a supervised learning algorithm used to train neural networks. 6 neural nodes were distributed on two hidden layers (3,2), which were determined by training and several computer experiments to obtain the best neural network with the least mean square error. (MSE=16.12642), (network error = 0.0009788). Modifying and testing the neural network again and again is the best way to find the parameters that best fit the model. When this neural network is trained, the black lines representing the connections between each layer can be weighted to find coefficients that fit the data with optimal weights, and the blue lines represent the incremental bias at each step.

Figure (1) Forming a neural network

After creating the neural network, we evaluated the accuracy and power of the model. We calculated the MSE and performed a cross-validation (CV) analysis. In cross validation, we have verified that the accuracy of the model changes with the length of the training set. We provided training sets with lengths ranging from 10 to 65. In this network, for each length, 100 samples are randomly selected and the MSE is averaged. We showed that when the training set is large, the accuracy of the model decreases. Before using the model for prediction, it is important to check the performance ability through cross-validation.

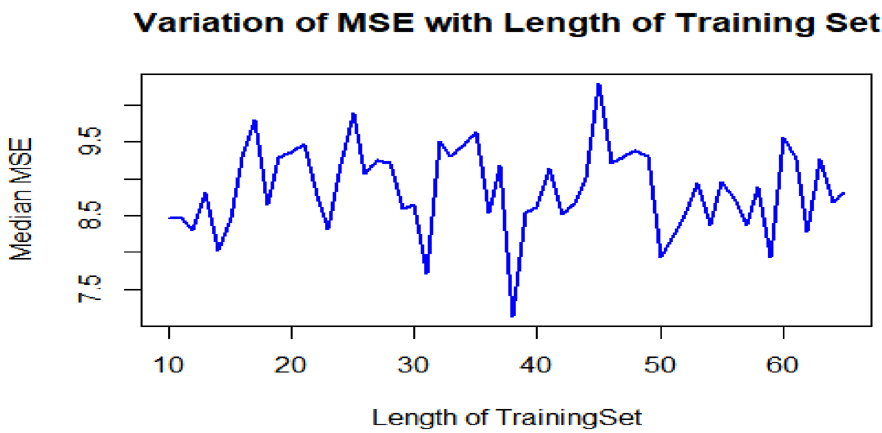

Figure (2) MSE changes

9. Regression model data diagnosis

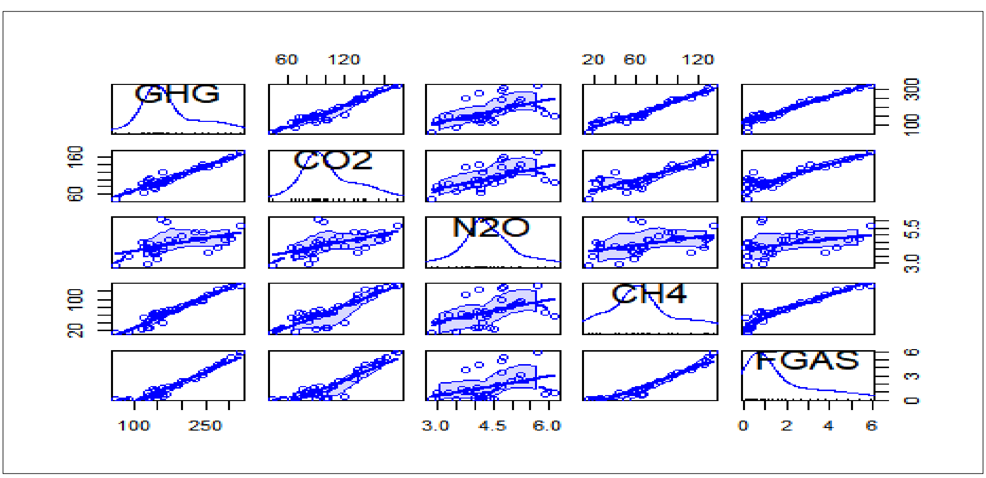

Figure (3) shows the emission of greenhouse gases for the period 1990-2019

Table (2) testing the model data

Heteroscedasticity			**p.value**	**df**
White		7.42	0.492	>8
Brush-Pagan	BP	1.7319	0.7849	4
	ncv	0.0148194	0.90311	1
Autocorrelation				
Durbin-Watson		1.9966	0.2792	
Breusch-Godfrey		0.22967	0.9727	3
Normality				
Shapiro-Wilk		0.9765	0.7270	
Kolmogorov-Smirnov		0.0984	0.9062	
Cramer-von Mises		9.881	0.0000	
Anderson-Darling		0.2808	0.6171	
Box-Ljung		0.23274	0.9937	4
Normal distribution				
Global Stat		2.25581	0.6888	
Skewness		0.10161	0.7499	
Kurtosis		0.32149	0.5707	
Random residual				
runs=13,n1=15,n2=15,n= 30		-1.1148	0.2649	
outlier	max	min	p < 0.05	
	2.166644	0.040413		
Multicollinearty				
Variable	CO2	N2O	CH4	FGAS
Tolerance	0.14677376	0.57113158	0.0874853	0.0530878
VIF	6.813207	1.750910	11.430491	18.836737

Using statistical software (R), we check the results of the tests (non-collinearity, autocorrelation, normality, normal distribution, random residual) in which the p-value is large (p-value > 0.05). We do not reject the null hypothesis. Therefore, we assume that the residual values (collinearity, non-correlation, normality, normal distribution, random residual) are satisfied.

10. Variance Inflation Factor (VIF)

The variance inflation factor (VIF) is used as a measure to detect multicollinearity and identify the independent variable responsible for it. It is defined by the following equation (David .1991)[17]

$$VIF = \frac{1}{1 - R_j^2} = \frac{1}{Tolerance}, \qquad j = 1, \dots, k$$

Where, tolerance is the opposite of VIF. The lower the Tolerance, the higher the probability of a multicollinearity relationship between the variables. A VIF value >10 indicates that the regression coefficients are poorly estimated with a multicollinearity.

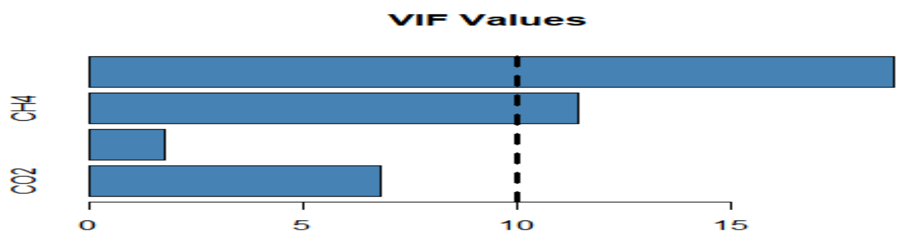

Figure (4) VIF

11. Pearson's Correlation Coefficients

Pearson's correlation coefficient helps to verify the existence of collinearity of the independent variables. Figure (2) shows the correlation analysis between the explanatory variables resulting from the positive correlation, some of which are moderate and significant, and some of which are not moderate, as in the association of the FGAS variable with the rest of the variables at a high level. The correlation coefficient is calculated using the formula:

$$r = \frac{n(\sum xy) - \sum x \sum y}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}}$$

Where r = correlation coefficient , n = number of observations , X = first variable and Y = second variable (Belinda ,2014)[18].

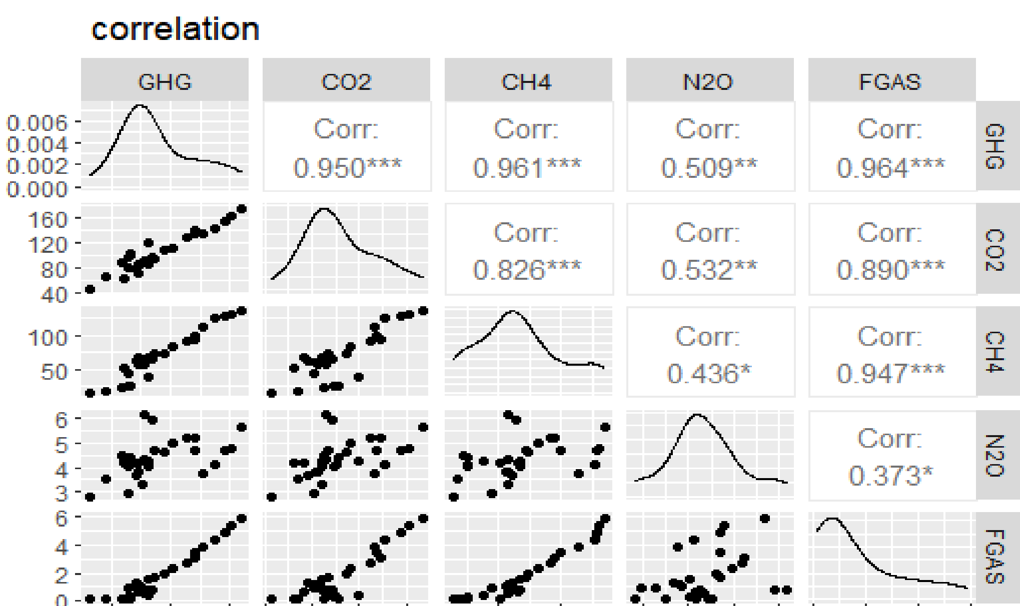

Table (3)

Model-1-Coefficients				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.0076287	0.0084284	-0.905	0.374038221540634
CO2	0.9999643	0.0001013	9869.387	6.5406773685639e-84
N2O	1.0016094	0.0020202	495.805	1.9479087177434e-51
CH4	1.0000812	0.0001161	8615.369	1.9544724656244e-82
FGAS	0.9997167	0.0030586	326.851	6.5000107143096e-47

p-value < 0.05 for each t-statistic it indicates the significanc of each explanatory variable and the existence of a relationship between the dependent variable and the explanatory variables. The null hypothesis was rejected. Therefore the estimated effect of CO2 on global warming is 0.9999643, while the estimated effect of N2O is 1.0016094, the estimated effect of CH4 is 1.0000812, and the estimated effect of FGAS = 0.9997167.

Table (4)

RSE	df	R	R^2	Adj R^2	F	p-value
0.006419	25	-0.00892200	1	1	7.331e+08	< 2.2e-16

We note that $R^2 = 1$, which means that 100% of the independent variables can explain this phenomenon. In other words, all control points (X,Y) lie on the estimated line. The explanatory power is very high, which means that there is a Goodness and correlation between the independent variables and the dependent variable. This gives an idea of the generalizability of the model. The difference between R^2 and Adj R^2 is 0. It is the variable X that explains the phenomenon Y. We also pay attention to the significance of the model because the value of p is very small. Therefore the null hypothesis was rejected($H_0: \beta_j = 0$).

Table (5)

ANOVA						
		Sum Sq	df	Mean Sq	F value	Pr(>F)
Regression	CO2	108980	1	108980	2644953147	<2.2e-16 ***
	N2O	3	1	3	60870	<2.2e-16 ***
	CH4	11843	1	11843	287441357	<2.2e-16 ***
	FGAS	4	1	4	106831	<2.2e-16 ***
Residuals		0	25	0		
Dependent Variable:GHG Predictors: (Constant),CO2,N2O,CH4,FGAS						

Table (6) Regression ridge specification

Measure: Mean-Squared Error						
	Lambda	alpha	Index	Measure	SE	RMSE	R^2
min	0.1000	0	51	0.2362	0.1473	0.281673	0.9999803
1se	0.1259	0	50	0.3328	0.2058		

Coefficients Ridge(train)				
(Intercept)	CO2	CH4	N2O	FGAS
1.0832659	0.9838626	0.9746848	1.2474644	1.7085214

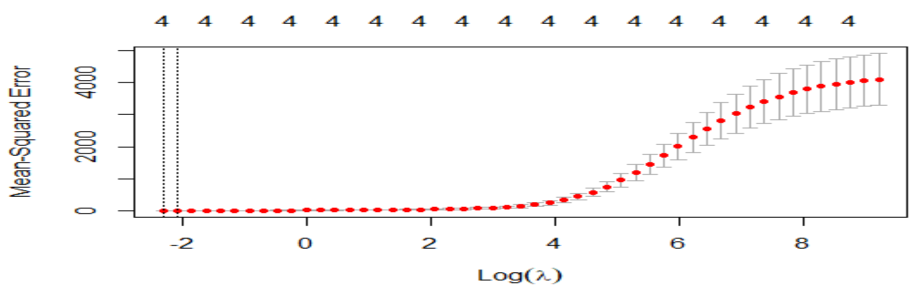

Figure (6) Ridge

Ridge regression is used to deal with the multicollinearity problem. Due to the multiplicity of linear lines, the (least squares) model estimates show high variance, although the Ridge line regression is not BLUE, this method adds a degree of bias to the regression estimates, but the variance is lower. Or rather, as the lambda value increases, the bias increases and the variance decreases. The alpha value for Ridge regression is zero and the optimal value is obtained by (10) Cross validation = 0.1.

Table (7) Regression Lasso specification

Measure: Mean-Squared Error							
	Lambda	alpha	Index	Measure	SE	RMSE	R^2
min	0.1	1	51	0.02892	0.009031	0.1394793	0.9999952

Coefficients lasso				
(Intercept)	CO2	CH4	N2O	FGAS
0.1699167	1.0045777	1.0064697	0.8437845	0.7680273

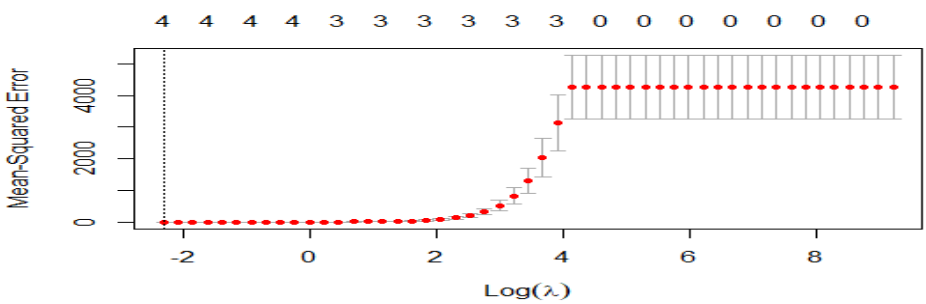

Figure (7) Lasso

shows cross-validation to find the optimal value of lambda in lasso regression. According to this plot, the value of log lambda is about 0.1, which is the optimal point of the intensity of the penalty (cost) function of the model, this is the point where the MSE value increases in the subsequent cross-validation. The alpha value for Lasso regression is 1 and the optimal value was obtained through (10) validation blocks equal to (0.1). We notice very little change in the estimates and change in RMSE value and R-Square value.

Table (8) Regression Elastic net specification

Measure: Mean-Squared Error					
scale	Lambda	alpha	RMSE	R^2	MAE
4	0.137805	0.914438	1.902866	0.9992502	4.534688
Coefficients Elastic net					
(Intercept)	CO2	CH4	N2O	FGAS	
0.3482601	1.0016659	1.0011696	0.8964783	0.9140459	

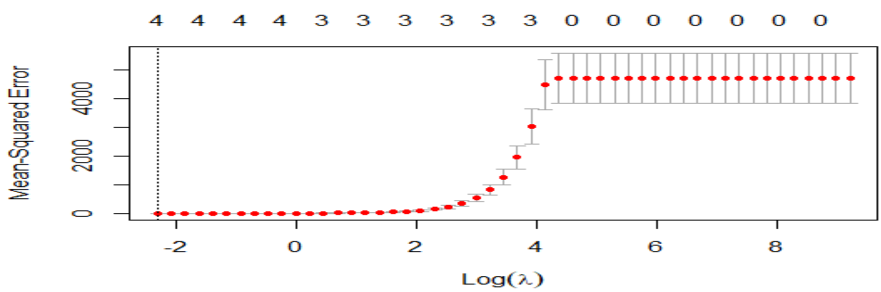

Figure (8) Elastic net

Alpha value for network elastic regression = 0.914438 and the optimal value obtained through 10 validation blocks is equal to 0.137805. which is the optimal point of intensity of the penalty function (cost) of the model, we notice an increase in the value of RMSE and a decrease in the value of R-Square.

14. Evaluation and comparison of regression model

There are many criteria for evaluating the regression model, each of which has its own characteristics. We list them as follows.

Table (9) tests of four regression models

	OLE	Ridge	Lasso	Elasticnet
R^2	1	0.99998	0.9999	0.999884
Adj R^2	1	0.99998	0.9999	0.999865
MAE	0.0047	0.20840	0.1106	1.7237
RMSE	0.0059	0.28167	0.1395	1.9029
MSE	3.4e-05	0.00174	0.0051	3.622

To compare the accuracy of four regression models MSE, RMSE, MAE, R^2, Adj R^2 were compared. Figure (6) shows the values of this index.

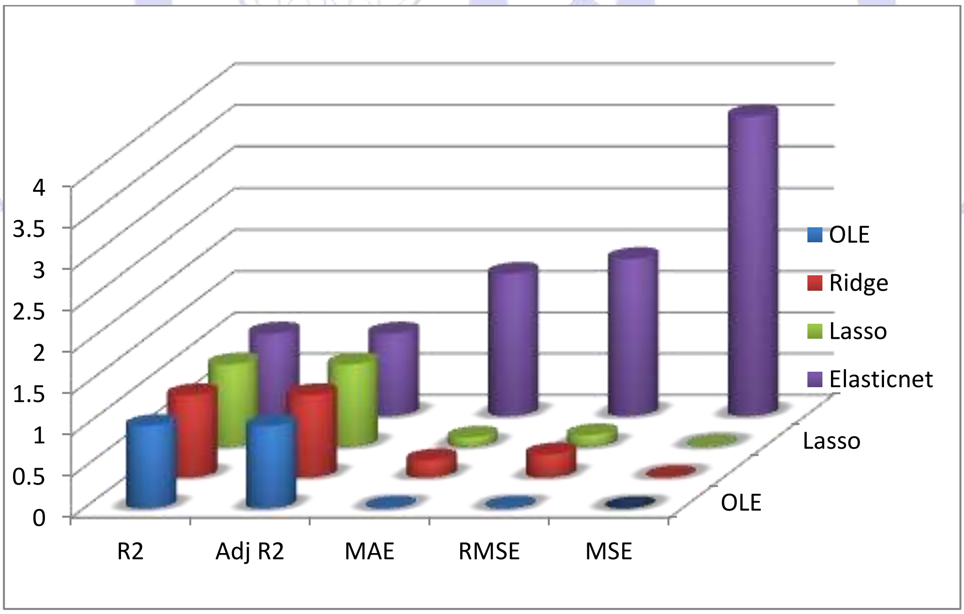

Figure (9) Values of four regression models

15. Conclusion

All indications are that Least Squares (OLS) regression is best, but the presence of the Multicollinearty may distort the results and give the analyst the illusion that they are the best. In some cases, some researchers prefer to ignore the multiple collinearity and keep the results the same. Since each treatment to reduce the severity of the problem will have defects of another kind, and therefore leaving the results as they are is the right way. This view holds that the presence of multicollinearity does not always reduce the t-ratio and make it meaningless, or alter the value of the parameters by making them conflict with what the researcher expects of them. Use mitigation methods only if the problem affects the results. The problem of multiple collinearity is similar in some models to some non-life-threatening diseases. In our model, results are kept the same for several reasons:

- The number of observations n is greater than the number of explanatory variables $n \gg k + 1$.
- The t rate does not decrease, all of which are significant, and the standard deviation value is less than the value of the estimated parameters, and this indicates sufficient and tangible changes in the observations of each explanatory variable.
- It is known that the standard deviation value is high for the model that suffers from Multicollinearty, but our results in the OLS model indicate that the standard deviation is small and accurate, sigma = 0.00641.
- Although the Pearson correlation coefficient for some variables is stronger than 0.90, the parameters of these variables for t as the

relationship between the CH4 variable and the FGAS variable = 0.947 and at the same time the t ratio remains significant for both.

- Collinearity does not technically violate Gaussian Markov assumptions. This is what we found in the OLS model, where the Markov probability value was p-value = 0.9062.
- We note that the coefficients of the OLS model are significant, although the Collinearity tends to show non significant coefficients.
- Model parameters that suffer from multicollinearity have large variances and covariance values. But in the OLS model, we noticed that the variances and covariance value are small.

References

[1] Yang.Q "Regression",Publishing AG: Springer International ,DOI: 10. 1007/ 978-3-319-32001-4_174-1 In book: Encyclopedia of Big Data (pp.1-4),pp1, may 2017.

[2] Hutcheson, G & Sofroniou. N, "The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models",Sage Publication, Thousand Oaks, CA, 1999.

[3] J. T. Kilmer & R. L. Rodriguez "Ordinary least squares regression is indicated for studies of allesnetry", Journal of Evolutionary, Biology, Vol (31), issue 1, Pp: 4-12, 2016.

[4] Matt N. Williams &Carlos Alberto Gómez Grajales ,"Assumptions of Multiple Regression: Correcting Two Misconceptions",t publication to Practical Assessment, Research & Evaluation, Vol 18, No 11, September 2013.

[5] Jason W. Osborne & Elaine Waters,"Four assumptions of multiple regression that researchers should always test" ,Practical Assessment, Research, and Evaluation: Vol. 8 , Article 2, 2002.

[6] Douglas C.Montgomery, Elizabeth A.Pecket," Introduction to Linear Regression Analysis" 5nd ed,pp71-74, Canada, Published by John Wiley &Sons, INC.2012.

[7] Jeffrey M. Wooldridge,"Introductory Econometrics A Modern Approach" 5nd ed, Michigan State University,pp 83-170, 2013, 2009 South-Western.

[8] Carmen R. Wilson VanVoorhis and Betsy L. Morgan,"Understanding Power and Rules of Thumb for Determining Sample Sizes", Tutorials in Quantitative Methods for Psychology, vol. 3 (2), p. 43-50,2007.

[9] Berry, WD. and Feldman, S. Multiple Regression in Practice "Quantitative Applications in the Social Sciences), series no. 07-050", Newbury Park, 1985.

[10] Tabachnick, B. G., & Fidell, L. S. Using Multivariate Statistics (3rd ed.). New York: Harper Collins,College Publishers, 1996.

[11] Pedhazur. E. J, "Multiple Regression in Behavioral Research", 3rd ed,Orlando , FL:Harcourt Brace,1997.

[12] Jibril .M. Suleiman,"Multicollinearity its Causes, Effects and Remedies by Ridge Regression and Principle Components Regression with Application to Hypothetical Data",Sudan University of Science and Technology, PhD research, 2014.

[13] M. Jafari, Y. Dinpashoh, E. Asadi," Introduction to multiple linear regression and its application to simulate daily pan evaporation" Journal of Water and Sustainable Development, Vol.2, No.2, Pages 67 to 76,2016.

[14] Wessel N. van Wieringen,"Lecture notes on ridge regression",Universiteit Amsterdam Version 0.40, May 28, 2021.

[15] Al-Azzawi . Dijlah &Al-Hassani,"Testing Hypotheses in High Dimensional Regression",ISSN(1681 6870), no46 , 2020.

[16] Hojjatollah Farahani,"Comparing Linear Regression to Shrinkage Regression Algorithms (RR, Lasso, El Net)Using PTSD Patients' Data",Journal of Applied Psychological Research,11(3), 193-206,2020, (DOI): 10.22059/japr .2020 .305766 .643568.

[17] David A. Belsley et "Conditioning Diagnostics: Collinearity and Weak Data in Regression" 1st Edition, Publisher Wiley & Sons, Inc., New York, January 22, 1991.

[18] Belinda. B & Peat. J, "Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal ",2nd edition, Wiley, UK, 2014.