



مجلة التجارة والتمويل

[/https://caf.journals.ekb.eg](https://caf.journals.ekb.eg)

كلية التجارة – جامعة طنطا

العدد : الرابع

ديسمبر 2023
(الجزء الثاني)

**Review of Classical Methods and Variables Selection in
Case of Multicollinearity: A Case Study with Real-Data**

الدكتور/ شيماء لبيب إبراهيم بركات

مدرس - المعهد العالي للإدارة - بالمحلة الكبرى

Review of Classical Methods and Variables Selection in Case of Multicollinearity: A Case Study with Real-Data

Abstract

The addition of excessive variables to a model can lead to severe consequences. When a model contains numerous variables, it is likely that some of them will exhibit strong correlations. However, explanatory variables should ideally not possess strong relationships among themselves. This issue, known as multicollinearity, can significantly impact the interpretation of results by causing notable variations between models. Variable selection further compounds this problem by introducing uncertainty as to which subset of potential explanatory variables or predictors should be used. This paper presents a succinct overview of ten traditional methods for tackling multicollinearity and variable selection in linear regression models. These methods were assessed using a real-life dataset across various sample sizes. The findings suggest that modified group lasso, group lasso, and adaptive group lasso exhibit particular efficacy in estimating variable selection and addressing collinearity issues in this model.

Keywords: Tolerance, Over fitting, Condition number, Correlation coefficient, Variance inflation factor, Condition index.

1. Introduction:

As for regression problems, the identification of important explanatory factors for predicting the response variable is of great interest. Each explanatory factor may be represented by a group of derived input variables. A common example is the multifactor analysis-of-variance (ANOVA) problem. The objective of ANOVA is typically to select significant main effects and interactions to ensure accurate prediction, which involves selecting groups of derived input variables. In this paper, we propose and investigate several methods that achieve accurate prediction while simultaneously selecting a subset of important factors. (Liu and Li, 2016),(El Sheikh, et al,2021).

Multicollinearity, or near-linear dependence among the explanatory variables in a regression model, is a significant challenge in practical applications. Multicollinearity has severe implications for the ordinary least squares (OLS) estimates of the regression coefficients. If multicollinearity exists, in the method of least squares, the estimates of parameters will be generally poor, the variance of the regression coefficients may be inflated, the absolute values of the estimates will be too large and they will be unstable(Duzan and Shariff, 2015).

Multicollinearity, is widely recognized as one of the most critical issues in linear regression models, introducing various risks to the underlying assumptions (Duzan and Shariff, 2015). these risks include the following mentioned:

1. Difficulty in estimating the population parameters in the multiple linear regression model.
2. Increased variance of the population estimator in multiple linear regression.
3. Reduced quality of estimating the ordinary least squares parameters of the population in multiple linear regression.
4. Impact on the determination of the true linear model's quality.

This multicollinearity may be total, linked with two variables or more from explanatory variables in the model, and maybe partially linked with only one variable of the explanatory variables. There are different measures to check multicollinearity in multiple linear regression:

1. Difficulty in the signs of population parameter which expresses the relation between the explanatory variables and the dependent variable in multiple linear regression of its true value in the economic theory.

2. Increasing the coefficient value of determining when most of the variables of the explanatory variables.
3. The difference between the F test model and from T test model (Duzan and Shariff, 2015).

In social in social science studies, the variables of interest are often categorical, such as race, gender, and nationality. However, it is often difficult to fit a linear model on such data, especially when some or all of the explanatory variables are categorical, we need to overcome the defects of ordinary least squares. However, when the covariates are also categorical, corresponding variables are coded using dummy variables into the design matrix. In this approach, the data matrix becomes sparse; the column dimension increases, and columns might be highly correlated. This might result in a singular data matrix making coefficients of Linear Square Estimation (LSE) impossible. For the purpose of avoiding this pitfall, researchers used the methods to handle the problems of multicollinearity and variable selection (Streib and Dehmer, 2019), (El Sheikh, et al., 2021). The paper is organized as follows: Section two introduces general linear regression, while section three focuses on model selection. In section four, the concept of multicollinearity is discussed. Section five explores classical methods for handling variable selection and multicollinearity, including techniques such as lasso, adaptive lasso, adaptive group lasso, modified group lasso, group lasso, elastic net estimator, ridge regression, principle component regression, and partial least squares. Moving on to section six, real data examples are presented to illustrate the application of these methods. Finally, section seven offers concluding remarks to summarize the key findings and implications of the study.

2. Estimation of General Linear Regression Model:

A model-fitting procedure yields a vector of coefficients, such as the ordinary least squares (OLS) estimates achieved by minimizing the residual sum of squares (RSS). The criteria for assessing the quality of a model vary depending on the context, but typically two aspects are of particular importance:

- Accuracy of prediction on future data: A model that exhibits poor predictive performance is difficult to justify and defend.
- Interpretability of the model: Scientists and researchers often prefer simpler models as they shed more light on the relationship between the response variable and the covariates. Parsimony is especially an important issue when the number of predictors is large (Dunn, et al.,2018).

Consider the following general linear regression model:

$$y_i = x_{ij}\beta_j + \varepsilon_i \quad i = 1,2,\dots,n$$

$$j = 1,2,\dots,p \quad (1)$$

Where x_{ij} is $n \times p$ matrix of non-stochastic predetermined variables standardized, so that $x'_{ij}x_{ij}$ is nonsingular correlation matrix, y_i is a vector of observations the dependent variable, measured in terms of deviations from sample means, β_j is a vector of p parameter and ε_i is a vector of disturbances, with expectation $E(\varepsilon_i) = 0$ and variance -covariance matrix $\text{var}(\varepsilon_i) = \hat{\sigma}^2 I_n$.

3. Model Selection:

Model selection is often purely based on the evidence provided by the data, for example by applying a variable selection algorithm, Model selection can also be achieved by applying least angle selection and shrinkage operator (LASSO) penalties, which are based on subtracting a multiple (λ) of the absolute sum of regression coefficients from the log likelihood and thus setting some regression coefficients to zero (Heinze, 2018).

3.1. Variable selection algorithms:

A variable selection algorithm has one or several tuning parameters that can be fixed to a prespecified value or estimated, for example, by cross-validation or AIC optimization. Note that tenfold cross-validation and selection by AIC are asymptotically equivalent (Heinze,2018).

3.2. Check of Variables Selection:

Choosing the right sensors can hugely affect the outcome of any modeling. For instance, the interest in process analytical technology has led to a number of investigations where near-infrared spectroscopy was chosen for modeling certain properties. We introduced the measurements, which check the issues variables selection (Andersen and Bro, 2010).

Overfitting

If there are many more variables than samples, it is possible, by chance and over fitting, to find a certain number of variables that correlate to the property to be predicted. If such variables are chosen and the model is applied to new samples, Then, the predictions may be very poor, or there may be no relationship at all. Therefore, validation is fundamental (Andersen and Bro, 2010).

Outliers

It is well known that proper handling of outliers is essential in data analysis. This is even more so in variable selection. Many variable selection methods are based on assessing minor differences in model quality or even in assessing statistics such as significance calculated from model parameters. Therefore, the result of variable selection is even more sensitive to outliers than the actual model fit. For this reason, every result during the process of variable selection should therefore be complemented by careful outlier detection. This may be difficult in practice, but at the very least, the resulting model obtained after variable selection should be carefully assessed, and the variable selection may be re-run upon handling new outliers to verify the result (Andersen and Bro, 2010).

Redundancy:

Two or more variables can exhibit a certain degree of similarity. A sound model's loading plot will display these variables clustered closely together. When aiming for a streamlined model, it becomes feasible to eliminate redundant variables. The predictions will probably not improve, but the model will be based on fewer variables, each with a unique appearance, and it may therefore be easier to understand and interpret (Andersen and Bro,2010).

Effect of Variables Selection:

1. Having fewer variables in the model means less computational time and complexity.
2. Simple models with fewer variables are preferred over complex models with many variables.
3. Many variables in the model make the model more dependent on the observed data.
4. Simple models are easier to interpret.
5. Important variables are not excluded from the simple model (Chowdhury and Turin, 2022).

4. Multicollinearity:

Multicollinearity is another significant problem that arises when the explanatory variables are correlated with each other. In such cases, the ridge regression method is utilized to address this issue. Multicollinearity can cause serious problem in estimation and prediction, increasing the variance of least squares of the regression coefficients and tending to produce least squares estimates that are too large in absolute value (Shrestha,2020).

4.1. The Effect of Multicollinearity: -

The effects of multicollinearity can be used to detect multicollinearity problems in multiple regression. However, many other measures can also be used to detect multicollinearity. One commonly used method is to examine the correlation matrix of independent variables. High correlation values in the matrix may indicate multicollinearity. However, multicollinearity can still exist even when all correlations are low. The determinant and eigenvalues of the correlation matrix can be used as simple measures for detecting multicollinearity. If the determinant and the smallest eigenvalue are close to zero, multicollinearity may be present. Moreover, other common and useful diagnostics for multicollinearity include variance inflation factors (VIF) and condition number (Adeboye,et al, 2014; Shrestha, 2020). When the determinant and the smallest eigenvalue is closed to zero, then multicollinearity can exist. Furthermore, some other most common and useful diagnostics of multicollinearity such as variance inflation factors (VIF) and condition number (Adeboye, et al, 2014 ;Shrestha, 2020).

The Tolerance Level:

In multiple regressions, tolerance is used as an indicator of multicollinearity. Tolerance is estimated by $(1-R^2)$, where R^2 is calculated by regressing the independent variables of interest unto the remaining independent variables included in the multiple regression analyses. Researchers desire more levels of tolerance, as low levels are known to adversely affect the result associated with a multiple regression analysis (Adeboye, et al ,2014). The tolerance is calculated as follows:

$$(1-R^2)$$

Variance Inflation Factor (VIF):

Variance inflation factor is a very popular diagnostic for detecting multicollinearity. VIF is the diagonal elements of the inverse of $(x'_{ij}x_{ij})^{-1}$

matrix when the variables are standardized (Shrestha,2020). Otherwise it can be formulate as mentioned below:

$$VIF=(1-R^2)^{-1}$$

Where R^2 is defined as in Tolerance level.

Condition Index:

Another diagnostic for multicollinearity is condition number or condition index, which is derived from the eigenvalues of correlation matrix. The condition number (C_I) is defined as the ratio of the largest eigenvalue to the smallest eigenvalue (Shrestha,2020). Thus, it can be formulated as mentioned:

$$C_I = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

If $C_I < 10$, there is little multicollinearity and a serious problem cannot be observed. Multicollinearity is Medium-leveled in $10 < C_I < 30$, while $30 < C_I$ indicates severe multicollinearity and more than one multicollinearity must be assumed.

Simple Correlation Coefficient:

Another diagnostic for multicollinearity is simple correlation coefficient, when $r \geq 0.5$ that mean high correlation between independent variables (Shrestha ,2020).

5. Methods for Handling Multicollinearity and Variables Selection:

5.1. Least Absolute Shrinkage and Selection Operator (LASSO):

Least absolute shrinkage and selection operator

(LASSO) regression methods are widely used in domains with massive datasets, such as genomics, where efficient and fast algorithms are essential. The LASSO is, however, not robust to high correlations among predictors and will arbitrarily choose one and ignore the others, potentially leading to breakdowns when all predictors are identical. The LASSO penalty expects many coefficients to be close to zero, with only a small subset being larger (and nonzero). The LASSO estimator uses the penalized least squares criterion to obtain a sparse solution to the following optimization problem (El Sheikh, et al,2021). The LASSO

technique is inspired by ridge regression, a standard technique for shrinking coefficients. However, contrarily to the latter, LASSO can set some coefficients to zero, resulting in an easily interpretable model (Epprecht, 2017). The lasso estimator is given by:

$$\frac{1}{2} \left\| y_i - \sum_{j=1}^p x_{ij}' \beta_j \right\|_2^2 + \lambda \sum_{j=1}^p \|\beta_j\| \quad (2)$$

Where y_i and x_{ij} are defined as equation (1), $x_{ij}'x_{ij}$ is defined as equation (1), p is the number of explanatory variables, n is the number of observations, λ_j is tuning parameter determination from the analysis data, dependent on cross validation and Bayesian data, dependent on cross validation and Bayesian information computation.

5.2. Adaptive Lasso:

Adaptive lasso improves their performance as sample size increases and the number of relevant and candidate variables decreases. Regarding parameter estimation, Auto metrics present the lowest absolute average bias and variance, as expected by the definition of OLS estimation when the correct model is selected, Bias of LASSO estimators tends to be larger in absolute value than the bias of adaptive LASSO estimators, which are expected to be close to the ones produced by OLS. The weighting strategy of adaptive LASSO makes the penalty term small for the relevant variables (Epprecht, 2017). This estimator can be defined as below:

$$\frac{1}{2} \left\| y_i - \sum_{j=1}^p x_{ij}' \beta_j \right\|_2^2 + \lambda_j \|\beta_j\| \quad (3)$$

Where y_i and x_{ij} are defined as equation (1), β_j is defined as an equation (1), (p, n) defined as an equation (2), λ_j is a vector $(p \times 1)$ tuning parameter determination from the analysis data dependent on cross validation and Bayesian information computation.

5.3. Adaptive Group Lasso:

Adaptive group lasso is an attractive method that enjoys the oracle property, and it is a convex penalty method, the adaptive group Lasso can be extended to some high-dimensional semiparametric models (Wang and Tian,

$$2017). \quad \frac{1}{2} \left\| y_i - \sum_{j=1}^p x_{ij}' \beta_j \right\|_2^2 + n \lambda_j \|\beta_j\| \quad (4)$$

Where y_i and x_{ij} are defined as an equation β_j is defined as an equation (1), (p, n) is defined as an equation (2), λ_j is defined as equation (3).

5.4. Modified Group Lasso:

Modified grouped lasso is used in high dimensional. It reduces the biased and the variance and is used in large sample size, this method improves the quality of the model, it is used to handle the case of multicollinearity and variable selection. We can use it in categorical data. Its estimators are symbiotic. We cannot use it in polynomial regression. Modified group lasso is especial case of group lasso (El Sheikh, et al,2021), This estimator can be defined as:

$$\frac{1}{2} \left\| y_i - \sum_{j=1}^p x_{ij}' \beta_j \right\|_2^2 + \lambda_j \|\beta_j\|_2 \quad (5)$$

Where y_i and x_{ij} are defined as equation (1) β_j is defined as equation (1) , p and n are defined as equation (2) , λ_j is defined as an equation (3) .

5.5. Group Lasso:

The last modern regression model we are discussing is the group LASSO. The group LASSO is different to the other regression models because it focuses on groups of variables instead of individual variables. The reason for this is that there are many real-world application problems related to, e.g., pathways of genes, portfolios of stocks, or substage disorders of patients, which have substructures, whereas a set of predictors forms a group that either should have nonzero or zero coefficients simultaneously. The various forms of group lasso penalty are designed for such situations (Streib and Dehmer, 2019). This estimator can be defined as:

$$\frac{1}{2} \left\| y_i - \sum_{j=1}^p x_{ij}' \beta_j \right\|_2^2 + \lambda_j \sqrt{p} \|\beta_j\|_2 \quad (6)$$

Where: y_i and x_{ij} are defined as equation (1) β_j is defined as equation (1) , p and n are defined as equation (2) λ_j is defined as equation (3) .

5.6. Elastic Net:

Elastic Net simultaneously does automatic variable selection and continuous shrinkage, and it can select groups of correlated variables. Elastic Net shrinks the regression coefficients by combining L1-norm penalty (lasso) and L2-norm penalty (ridge) together, The L1-norm part

of the penalty generates a sparse model by shrinking some regression coefficients exactly to zero. The L2-norm part of the penalty removes the limitation on the number of selected variables (Liu and Li,2016), the elastic net method overcomes the limitations of the LASSO method which uses a penalty function based on:-

$$\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$$

The use of this penalty function has several limitations. For instance, in the "large p , small n " case the LASSO selects at most P variables before it saturates. Also, if there is a group of highly correlated variables, then the LASSO tends to select one variable from a group and ignore the others. To overcome these limitations, the elastic net adds a quadratic part to the penalty $\|\beta\|^2$, which is used alone in ridge regression. Its estimators are consistency and efficient. It's used in large sample. Its estimators are stable. It is used in high dimensional, polynomial regression, multiple regression and categorical data see 6. It increases the flexibility of the model (El Sheikh, et al, 2021). It can be defined as:

$$\frac{1}{2} \left\| y_i - \sum_{j=1}^p x_{ij}' \beta_j \right\|_2^2 + \lambda \sum_{j=1}^p \|\beta_j\| + |\beta_j \sqrt{k}|^2 \quad (7)$$

Where: y_i and x_{ij} are defined as equation (1), β_j is defined as equation

(1), $k = \frac{p\hat{\sigma}^2}{\hat{B}'_{OLS}\hat{B}_{OLS}}$, σ^2 is the variance covariance matrix of estimator ordinary least square, p , n and λ are defined as equation (2).

5.7. Ridge Regression:

Ridge regression is motivation for improving OLS is the fact that the estimates from such models have often a low bias but a large variance, this is related to the prediction accuracy of a model because it is known that either by shrinking the values of regression coefficients or by setting coefficients to zero the accuracy of a prediction can be improved (Streib and Dehmer, 2019),(El Sheikh, et al ,2021). The ridge regression is defined as follows:

$$\hat{\beta}_{Ridge} = (x'_{ij} x_{ij} + k)^{-1} x'_{ij} y_i \quad (8)$$

where k is defined as equation (7). y_i and x_{ij} are defined as equation (1), $(x'_{ij} x_{ij})$ is defined as equation(1) .

5.8. Principle Component Regression:

Principal component regression is the method which includes ordinary least square and ordinal principal component. It is used to handle correlation when multicollinearity is minimum, median and maximum. It reduces variance inflation factor. It isn't used in the case of outliers (Abdel Salam, 2014), Principal component analysis transforms a data-set of original variables into a new a dataset of uncorrelated derived variables. These new derived variables are called principal components (PCs), which are the results of linear functions of the original variables (Haque, et al,2018). It isn't used in large sample It can be used as:

$$\hat{B}_r = T_r(T_r'x'T_r)^{-1}T_r'x'y, \quad (9)$$

Where $T_r = [t_1, t_2, t_3, \dots, t_r]$ is the remaining columns of T having been deleted $(p - r)$ columns where $r \leq p$. y_i and x_{ij} are defined as equation

(1) $(x_{ij}' x_{ij})$ is defined as equation (1).

5.9. Stepwise Method:

The stepwise selection method combines certain aspects of forward selection and backward elimination methods. Like the forward selection method, it starts with no variable in the model, and variables are added one by one to the model by fulfilling the p criteria ($p < 0.1$). After a variable is added in the model, the stepwise selection method examines all the variables in the model and deletes any variable that show a p-value greater than the critical value. The next variable is added in the model only after checking the model and deleting any variables if necessary. This process continues till none of the variables outside the model have a p-value less than the critical value and every single variable in the model satisfies the p criteria (Haque, et al, 2018), This method like the forward selection except that at each step we consider dropping variables in backward elimination. It improves the accuracy of the model. It isn't used to handle outliers. It isn't used in the case of large samples. Convergence is generated if $f_1 > f_2$. The following steps will show the alliteration of the stepwise method in general point of view (El Sheikh, et al, 2021). **The above mentioned steps are:**

1. Start with no input's variables in model.
2. Select (add) one significant variable x_i , compare the criterion value of all models with one variable.

3. Select one more significant variables x_i , selection method compare criterion value of all models that include the first x_i and one additional x_{ij} if the model with the additional x_{ij} gives the best criterion value when the first x_i are already in the model. If there is no selection go to step (4).
4. Delete (remove) one insignificant variable x_{ij} . Deletion method compare the criterion value of all models that include the x_i ,s without one x_i , delete x_i if the model removing x_j gives the best criterion value. If there is no deletion or no more deletion, go to step (3). Otherwise go to step (5).
5. Stop the stepwise method for variable selection. The different stepwise methods for variable selection are different in their criterion used in selection of step (2). Selection method and deletion method of other steps. So, without loss of generality, we compare the criterion of different stepwise

5.10. Partial Least Square Regression:

Partial least square with regression coefficients named PLS-BETA directly utilizes the regression coefficients estimated by partial least square. The significant variables are selected according to the magnitude of the absolute values of regression coefficients (Liu and Li,2016). Partial least square is a reasonably alternative method developed by Helland (1990) as method for constructing predictive models when the explanatory variables are many and highly collinear. It may be used with any number of explanatory variables, even for more than the number of observations (Abdel Salam,2014).

6. Real-Data:

To compare the performance of the proposed methods, a real data is used about marketing since 2020(1. january_30. November), real data is daily time series, it attached from Central Agency for Public Mobilization and Statistics in Egypt, In marketing data, the predictors are week(x_1), week ID(x_2), month(x_3), month ID (x_4), year(x_5), Day-Name(x_6), visitors (x_7), revenue (x_8) and marketing-spend(x_9). The response is promo variable(y) and the data set has 183 observations.

The standard deviation (STD) values of the predictor variables of the data set are (18.3, 7.8, 4.2, 1.7, 0.4, 503.5, 6297.6, 691.8) which show high outlier and null values among the predictor variables. The data from outlier and null values is cleaned. The level of collinearity among the predictor variables is examined. The collinearity between visitors' variables and revenue variable is equal (-1). The collinearity between marketing- spend variable and visitors variable is equal (0.56). The collinearity between revenue variable and marketing- spend variable is equal (-0.49). This data is used to examine the performance of lasso, group lasso, adaptive lasso, adaptive group lasso, modified group lasso, ridge, elastic net, step wise, principle component and partial least square. The data set is attached with python program. The measurements: mean square error (MSE), coefficient of determination (R square) and variance inflation factor (VIF) are used to fit the model.

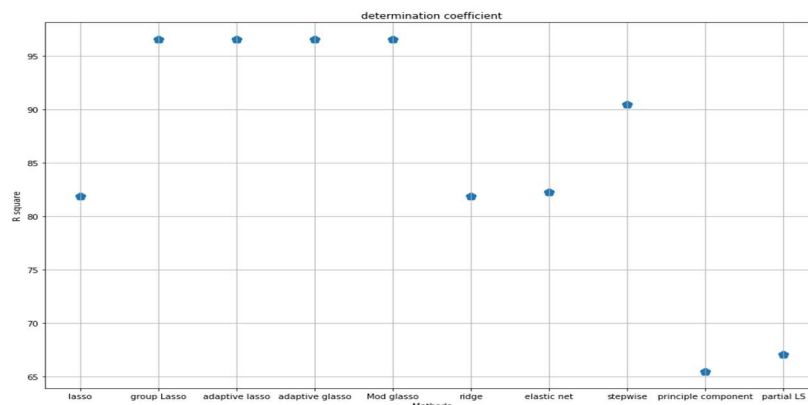


Fig. 1: Coefficient value of determination.

In figure (1), it shows that coefficient of determination for partial least squares (67.1) is more than the coefficient of determination for principle component regression (65.5). On the other hand, the coefficient of determination for lasso and ridge regression (81.9) is more than the coefficient of determination for partial least squares (67.1). On the other hand, coefficient of determination for elastic net (82.3) is more than the coefficient of determination for lasso and ridge regression (81.9). On the other hand, the coefficient of determination for adaptive lasso and

adaptive group lasso (96.582) is more than the coefficient of determination for elastic net (82.3). On the other hand, the coefficient determination for group lasso and modified group lasso (96.586) are more than coefficient of determination for adaptive lasso and adaptive group lasso (96.582), Modified group lasso and group lasso are better than all other estimator, so the modified group lasso and group lasso are the best method.

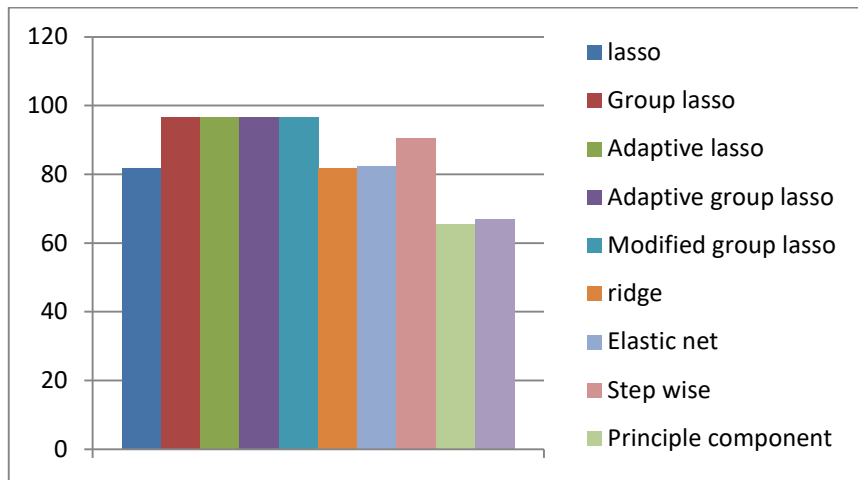


Fig. 2: Presentation coefficient of determination.

In figure (2), it shows that the presentation coefficient of determination for partial least square is more than the presentation coefficient of determination for principle component regression, the presentation coefficient of determination for lasso and ridge regression are more than the presentation coefficient of determination for partial least square, the presentation coefficient of determination for elastic net is more than the presentation coefficient of determination for lasso and ridge regression, the presentation coefficient of determination for adaptive lasso and adaptive group lasso are more than the presentation coefficient of determination for elastic net. The presentation coefficient of determination for group lasso and modified group lasso are more than the presentation coefficient of determination for adaptive lasso and adaptive group lasso, modified group lasso and group lasso are better than all other estimator, so the modified group lasso and group lasso are the best method.

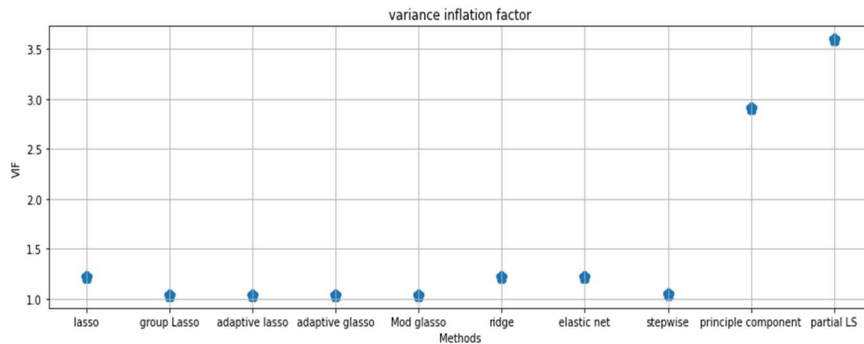


Fig3: the values of variance inflation factors.

In figure (3), it shows that the variance inflation factor of partial least squares (3.6) is less than the variance inflation factor of principle component regression (3.9). On the other hand, the variance inflation factors of lasso and ridge regression (1.22) are less than the variance inflation factor of partial least squares s(3.6), On the other hand, the variance inflation factor of elastic net (1.213) is less than the variance inflation factor of lasso and ridge regression (1.22). On the other hand, the variance inflation factor of adaptive lasso and adaptive group lasso (1.03533) are less than the variance inflation factors of elastic net (1.213). On the other hand, the variance inflation factor of group lasso and modified group lasso (1.03533) are less than the variance inflation factor of adaptive lasso and adaptive group lasso (1.03532). Modified group lasso and group lasso are better than all other estimator, so the modified group lasso and group lasso are the best method.

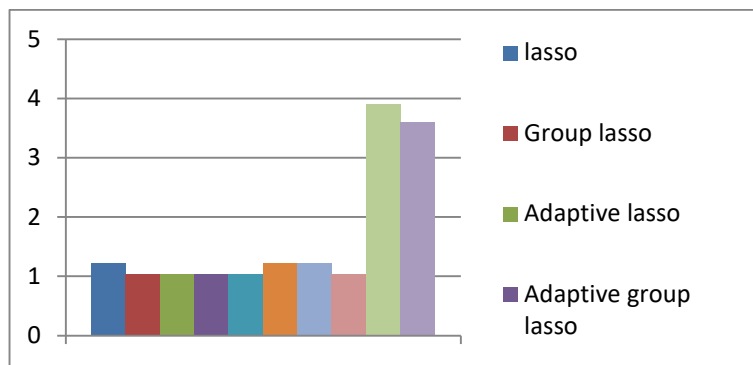


Fig.4 : Presentation of variance inflation factors.

In figure (4), it shows that the presentation of variance inflation factor for partial least squares is less than the presentation of the variance inflation factor for principle component regression. On the other hand, the presentation of variance inflation factors for lasso and ridge regression are less than the presentation of variance inflation factor for partial least squares, On the other hand, the presentation of variance inflation factors for elastic net is less than the presentation of variance inflation factors for lasso and ridge regression. On the other hand, the presentation of variance inflation factor for adaptive lasso and adaptive group lasso are less than the presentation of variance inflation factors for elastic net. On the other hand the presentation of variance inflation factor for group lasso and modified group lasso are less than the presentation of variance inflation factor for adaptive lasso and adaptive group lasso. Modified group lasso and group lasso are better than all other estimator, so the modified group lasso and group lasso are the best method.

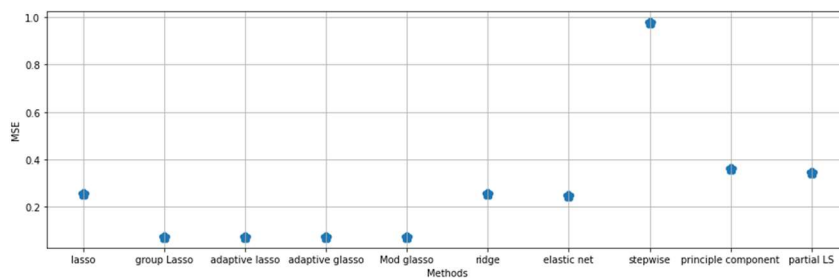


Fig.5: Values of mean square error.

In figure (5), it shows that the mean square error of partial least squares (.345) is less than the mean square error of principle component regression (.869). On the other hand, the mean square error of lasso and ridge regression (.235) is less than the mean square error of partial least square (.345). On the other hand, the mean square error of elastic net (.248) is less than the mean square error of lasso and ridge regression (.253). On the other hand, the mean square error of adaptive lasso and adaptive group lasso (.0738) are less than the mean square error of elastic net (.248). On the other hand the mean square error of group lasso and modified group lasso (.0737) are less than the mean square error of adaptive lasso and adaptive group lasso (.0738) , so the modified group lasso and group lasso are the best method.

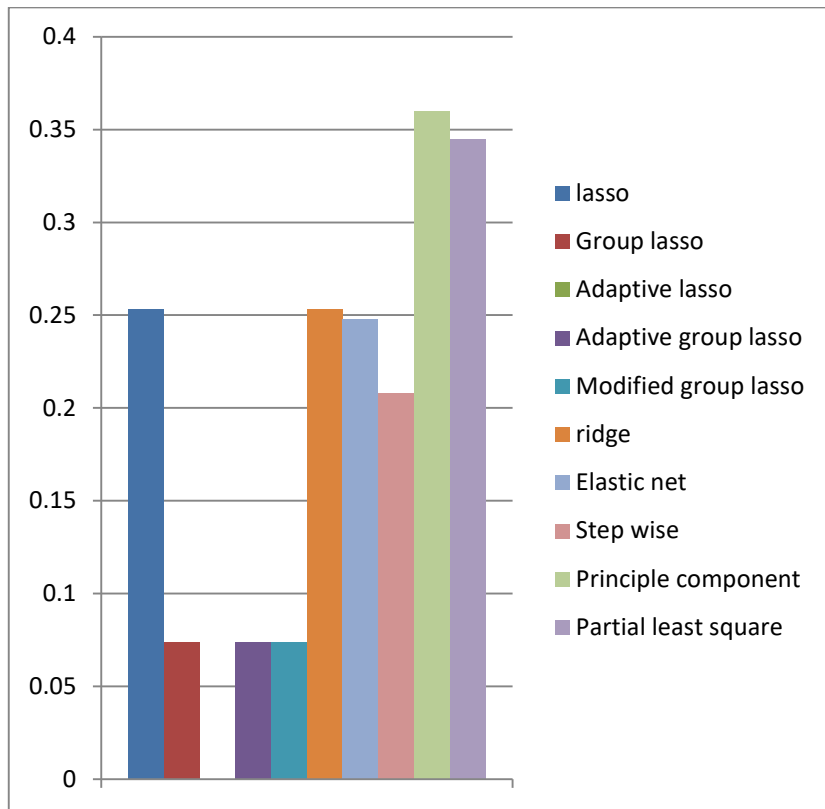


Fig.6: Presentation of mean square error.

In figure (6), it shows that the presentation of mean square error for partial least square is less than the presentation of mean square error for principle component regression. On the other hand, the presentation of mean square error for lasso and ridge regression is less than the presentation of mean square error for partial least square. On the other hand, the presentation of mean square error for elastic net is less than the presentation of mean square error for lasso and ridge regression. On the other hand, the presentation of mean square error for adaptive lasso and adaptive group lasso is less than the presentation of mean square error for elastic net. On the other hand, the presentation of mean square error for group lasso and modified group lasso are less than the presentation of mean square error for adaptive lasso and adaptive group lasso, so modified group lasso and group lasso are the best methods.

Table 1: Comparison between ten imputation method by using (VIF, R square, MSE).

Methods	VIF	R square	MSE
Lasso	1.22	81.9	.253
Group lasso	1.03532	96.588	.0737
Adaptive lasso	1.03533	96.587	..738
Adaptive group lasso	1.03533	96.587	.0738
Modified group lasso	1.03532	96.588	.0737
Ridge	1.22	81.9	.253
Elastic net	1.213	82.3	.248
Step wise	1.04	90.45	.208
Principle component	3.9	65.5	.360
Partial least square	3.6	67.1	.345

In table (1), it shows that the variance inflation factor of partial least squares (3.6) is less than the variance inflation factor of principle component regression (3.9). Since the mean square error of partial least square (.345) is less than the mean square error of principle component regression (.869), the determination coefficient of partial least square (67.1) is more than the determination coefficient of principle component regression (65.5), partial least square is better than principle component regression. On the other hand, the variance inflation factors of lasso and ridge regression (1.22) are less than the variance inflation factor of partial least square (3.6), and the mean square error of lasso and ridge regression (.235) is less than the mean square error of partial least square (.345). Since the determination coefficients of lasso and ridge regression (81.9) are more than the determination coefficients of partial least square (67.1), lasso and ridge regression are better than partial least square. On the other hand, the variance inflation factor of the elastic net (1.213) is less than the variance inflation factor of the lasso and ridge regression (1.22). Since the mean square error of elastic net (.248) is less than the mean square error of lasso and ridge regression (.253). The determination coefficient of elastic net (82.3) is more than the determination coefficient of lasso and ridge regression (81.9); elastic net is better than lasso and

ridge regression. On the other hand, the variance inflation factors of adaptive lasso and adaptive group lasso (1.03533) is less than the variance inflation factors of elastic net (1.213). Since the mean square error of adaptive lasso and adaptive group lasso (.0738) is less than the mean square error of elastic net (.248), The determination coefficient of adaptive lasso and adaptive group lasso (96.582) is more than the determination coefficients of elastic net (82.3); adaptive lasso and adaptive group lasso is better than elastic net. On the other hand, the variance inflation factor of group lasso and modified group lasso (1.03533) is less than the variance inflation factor of adaptive lasso and adaptive group lasso (1.03532). Since the mean square error of group lasso and modified group lasso (.0737) is less than the mean square error of adaptive lasso and adaptive group lasso (.0738), The determination coefficient of group lasso and modified group lasso (96.586) is more than the determination coefficient of adaptive lasso and adaptive group lasso (96.582), Modified group lasso and group lasso are better than all other estimators, so the modified group lasso and group lasso are the best methods.

7. Conclusion:

Multicollinearity and omitted variables are considered two problems in a linear regression model. In this paper, we present the methods that are used to handle multicollinearity and the methods which select the best variables selection to accurately represent the model regression. Generalized Liu two-type estimator and $(r - k)$ class estimator are used to handle multicollinearity and reduce the matrix mean square errors in linear regression models. Least absolute shrinkage and selection operation (LASSO), adaptive LASSO and elastic net are used to handle multicollinearity and select variables to accurate the linear model in large sample but LASSO is used to handle individual variable. Adaptive LASSO is used when the degree of multicollinearity is minimum and medium, and elastic net tends to be accurate when the degree of multicollinearity is maximum. Step-wise selection, back-ward elimination and forward selection are used to select the best explanatory variables in the context of multiple linear regression analysis.

References:

1. Abdel Salam, M.E. (2014). A note of partial least squares regression for multicollinearity. *Journal of Applied Science and Technology Magazine.*, 4(1), 163-170.
2. Adeboye, N.O, Sfagoyinbo, I.S. and Olatayo, T. O. (2014). Estimation of the effected of multicollinearity on the standard error for regression coefficients. *IOSR Journal of Mathematics.*,10, 1-5.
3. Andersen, C.M and Bro, R. (2010). Variable selection in regression _a tutorial, *J. Chemometrics.Wiley Online .*, 24,728-737.
4. Chowdhury.M.Z and Turin.T,(2020),"Variable Selection Strategies and its Importance in Clinical Prediction Model", *Fam Med Community Health*, vol(8),no(1),pp(1-7)
5. Dunn, Peter, Smyth and Gordon. (2018). Generalized linear models with examples in R. *Springer Texts in Statistics.*
6. Duzan,H and Shariff, M,(2015),"simulation study to verify the appropriate k value for ridge regression in two-variable regression model", *Indian Journal of science and technology*,vol(12),no(7),pp(103-110).
7. ElSheikh.A.A,Barakat.SH.LandMohamed.S.M,(2021),"New Aspect on the Modified Group Lasso Using the Least Angle Regression and Shrinkage Algorithm"*InformationScienceLetter*,vol(10),no(3),pp(527-536)
8. Epprecht, C., Guegan, D., Veiga, Á., and Joel, C., (2017), Variable Selection and Forecasting via Automated Methods for Linear Models: LASSO/adaLASSO and Autometrics, <https://halshs.archives-ouvertes.fr/halshs-00917797v2>.
9. Haque, M, Rahman, A, Hagare, D, and Chowdhury, R, (2018), "A Comparative Assessment of Variable Selection Methods in Urban Water Demand Forecasting", *Water*, vol(10), no(4), pp (1-15).
10. Heinze.G, Wallisch.C and Dunkler.D, (2018), "Variable Selection- A review and Recommendations for Practicing Statistician", *WILLY,Biometrical Journal*, vol(60),no(3),pp(431-449).
11. Liu, W and Li, Q.(2016).An efficient elastic net with regression coefficients method for variable selection of spectrum data. *Journal.PONE.*, 12(2), 1-13.

12. Shrestha,N,(2020), "Detecting Multicollinearity in Regression Analysis",vol(8),no(2),pp(39-42).
13. Streib,F.E. and Dehmer, M.(2019).High –dimensional lasso-Based computational regression models: Regularization shrinkage and selection. Machine Learning and Knowledge Extraction, 1, (359-383).
14. Wang.M and Tian.G.L,(2017),"Adaptive Group Lasso for High – Dimensional Generalized Linear Models",Springer,Stat Papers, vol(60),pp(1469-1486).