Republic of Iraq
Ministry of Higher Education
and Scientific Research
University of Baghdad
College of Education
for Pure Science / Ibn Al-Haitham
Department of Mathematics

A comparison of Different Ridge Regression Methods with Application

A Thesis
Submitted to the Department of Mathematics, College of Education for Pure Science / Ibn Al-Haitham, University of Baghdad as Partial Fulfillment of the Requirements for the Degree of Master of Science in Mathematics

By

Fatima Assim Mahdi

(B.SC. University of Baghdad,2015)

Supervised By

Assist Prof .Dr.Hazim Mansoor Gorgees

2017 A.D **1438 A.H**

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

يَٰٓأَيُّهَا ٱلَّذِينَ ءَامَنُوٓاْ إِذَا قِيلَ لَكُمۡ تَفَسَّحُواْ فِي ٱلۡمَجَٰلِسِ
فَٱفۡسَحُواْ يَفۡسَحِ ٱللَّهُ لَكُمۡۖ وَإِذَا قِيلَ ٱنشُزُواْ فَٱنشُزُواْ يَرۡفَعِ
ٱللَّهُ ٱلَّذِينَ ءَامَنُواْ مِنكُمۡ وَٱلَّذِينَ أُوتُواْ ٱلۡعِلۡمَ دَرَجَٰتٖۚ وَٱللَّهُ بِمَا
تَعۡمَلُونَ خَبِيرٞ ١١

صَدَقَ اللَّهُ الْعَظِيمُ

سورة المجادلة الآية (11)

الإهداء

إلى رسول الإنسانية ،الرسول الأمي، معلم البشرية إلى من أضاء دربنا بنور الهدى إلى من هدانا إلى الحق إلى معلمنا الأول خير الأنام وسيد المرسلين وخاتم النبيين محمد رسول الله (صل الله عليه وعلى اله وصحبه وسلم).

إلى نبع الحنان الحب والأمان إلى من وضعَ ربي جنات خلدهِ تحت قدميها ... من أضاءت طريقي بفيض حنانها وسهرت الليالي لكي أكون (أمي الغالية) أطال الله في عمرها وأدامها الله لي .

إلى نبض قلبي وقرة عيني إلى من رباني وعلمني إلى من أفنى عمره لأجلي إلى من شجعني إلى من أحمل أسمه بكل فخر
(أبي الغالي) أطال الله عمرك وادامك الله لي.

إلى من أشدد بهم أزري في الحياة إلى رفقاء دربي إلى نور حياتي (حسين، وداد، فارس) وجميع أهلي وأحبتي.

إلى كل من مد يد العون لي وساعدني أهدي لكم ثمرة جهدي وخالص دعائي.

فاطمة

شكر وتقدير

احمد الله الذي بنعمته تتم الصالحات ، واشكره على فضله ومنه علي في أتمام رسالتي هذه ، والصلاة والسلام على النبي محمد عليه صل الله عليه وسلم الذي علم البشرية أبجدية الحياة وحث على العلم والتعلم .

كلمة شكر وامتنان وتقدير وعرفاناً بالجميل أقف عندها ويُشرفني أن أوجهها . إلى من كان مثالا علميا يحتذي به فلم يبخل بنصيحة وإرشاد وجهد في أي وقت وكان نبراس دعم وتشجيع مؤثرين وعنوانا لي في إعداد هذه الرسالة أستاذي

(أ.م.د.حازم منصور كوركيس)

كما أتقدم بالشكر والامتنان إلى الأستاذ (أ.عباس نجم سلمان) لتشجيعه المستمر وتعاونه جزاه الله عني خير الجزاء.

أضافه إلى كل من ساعدني في إتمام رسالتي بدعاء أو جهد أو كلمه طيبه لكم مني كل الحب والتقدير (أ.م.نيران صباح جاسم)،(م.م.مريم محمد هلال) والى جميع من لم يسعني ذكرهم.

كما أتقدم بالشكر الجزيل إلى عميد كلية التربية / ابن الهيثم ورئيس قسم الرياضيات والأساتذة المتميزين لمساعدتهم العلمي.وجميع وزملائي وأصدقائي.

فاطمة

Abstract

Multicollinearity is an important problem in regression analysis which produces undesirable effects on the least squares estimation. Ridge regression is one of the most popular solutions to this problem. In this situation, ridge parameter (or biasing constant) has an important effect in parameter estimation. Many statisticians are proposed different methods for selecting the ridge parameter. In this thesis, we attempted to have our own contribution in this field. Accordingly,we have proposed a new method for choosing the ridge parameter. The performance of the suggested method is determined and compared with other methods already proposed by other researchers through simulation and practical study. The proposed method seems to be exactly reasonable in the sense of MSE criterion.

An alternative well known solution to the multicollinearity problem which is also included in the thesis, is the principal components regression. In this approach, instead of using the original non orthogonal explanatory variables in the regression analysis , their principal components are used which are orthogonal to each other. Two main parts are included in this thesis especially, the theoretical part and the experimental and practical part. Statistical programs (MATLAP&SPSS) have been employed to perform the required calculations.

Supervisor Certification

I certify that this thesis was prepared under my supervision at the Department of Mathematics. College of Education for pure sciences, Ibn Al- Haitham, University of Baghdad as a partial fulfillment of the requirements for Master degree of Science in Mathematics.

Signature:

Name: Assist.Prof. Dr. Hazim Mansoor Gorgees

Date: / / 2017

In view of the available recommendations, I forward this thesis for debate by examining committee.

Signature:

Name: Assist Prof. Dr. Majeed Ahmed weli

Head of the Department of Mathematics,
College of Education for pure sciences,
Ibn- Al-Haitham, University of Baghdad

Date: / /2017

Committee Certification

We certify that we have read this thesis entitled"**A comparison of Different Ridge Regression Methods with Application**" and, as an examining Committee, we examined the student (**Fatima assim mahdi**) in its contents and what is connected with it, and that in our opinion it is adequate for the partial fulfillment of the requirement for the degree of master of Science in Mathematics.

(Member and supervisor)	**(Member)**
Signature :	**Signature:**
Name:Assist.Prof. Dr. Hazim M. Gorgees	**Name:Assist.Prof. Dr.Awatif R.Mezaal**
Date: / / 2017	**Date: / / 2017**
(Chairman)	**(Member)**
Signature :	**Signature:**
Name: Assist.Prof. Abbas N.Salman	**Name: Assist.Prof. Dr.Nashaat J.Mohammed**
Date: / / 2017	**Date: / / 2017**

Approved by the University Committee of graduate studies

Signature:
Name: Assist. Prof. Dr. Khalid F. Ali
The dean of Education College for Pure Science/ Ibn AL Haitham
University of Baghdad

Date: / / 2017

Contents

List of Tables

List of Figures

Abbreviations

SYMBOL	DESCRIPTION
ANOVA	Analysis of variance
BLUE	Best Linear Unbiased Estimator.
b_{OLS}	Ordinary Least Squares Estimator
b_{RR}	Ordinary Ridge Regression Estimator
VIF	Variance Inflation Factors
FM	The full Model
RM	The Reduced Model
df	degrees of Freedom
GLM	General Linear Model
GRR	Generalized Ridge Regression
w.r.to	with respect to
SST	Total sum of squares
SSE	Erorr Sum of Squares
SSR	Regression Sum of Squares
TMSE	Total Mean Squares Error
MSE	Mean Squares Error
MSR	Mean Squares Regression
OLS	Ordinary Least Squares
ORR	Ordinary Ridge regression
CN	Condition Number
LW	Lawless and Wang
HKB	Hoerl,Kennard and Baldwin
tr	trace of the matrix
PC	Principal Components
R^2	Coefficient of determination
RR	Ridge Regression Estimator

Chapter One

Introduction & Literature Review

1.1 Introduction

One of the aims of science is to find, describe, and to predict relationships among events in the world in which we live. One way that this is accomplished by finding a formula or equation that relates quantities in the real world. For example, in an industrial situation, it may be known that the tar content in the outlet stream in a chemical process is related to the inlet temperature. It may be of interest to develop a method of prediction, that is a procedure for estimating the tar content for various fuels of the inlet temperature from experimental information. In medical, we may be interested in how a several vaccines affects a certain disease. In the economic,we may be interested in the relationship of supply, demand and the price of certain commodities. Among different models that deal with the real life situations, the most widely used statistical model is refer to linear regression model. The term "regression" literally means "step back towards the average" it was first used by a British biometrician, Sir Francis Galton (1822-1911) in connection with the inheritance of stature. Galton found that the offspring of abnormally tall or short parents, tend to "regress or step back" to the average population height [15].

Although it is desirable to be able to predict one quantity exactly in terms of others, this is in general not possible, and in most instances we have to be concerned with predicting averages or expectation. This problem of predicting the average value of one variable in terms of the known value of another variable (or the known values of other variables) is the problem of regression.

To explain this point of view, let us consider an example of income employing information in income and number of years of formal schooling to estimate the extent which the man's annual income is related to his years of schooling [40].

One possibility would be that a man who had zero years of school, we would anticipate his annual income as a (dollars) and for every year of schooling he had ,we would expect his income to be larger by b (dollars),thus for a man having (x)years of schooling we would expect his annual income to be (a + bx) dollars. The expect means we are thinking of the average of all men who had (x) years at school, if one man was picked at random we would expect his income to be (a + bx). The y denotes to income we write E(y) for expected income and hence E(y) = a + bx.

A general form for the model would be y = $f(x_1, x_2, x_3, \ldots, x_p) + \varepsilon$ where f is some unknown function and ε is the error term. Since we often don't have adequate information to estimate f directly, we have to assume that it has some more restricted form, probably linear as:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \varepsilon \qquad \ldots(1)$$

Where

β_j , j = 1,2,3,...,p are unknown parameters,

β_0 is called the intercept term of the model.

ε is not observable, but something about the distribution of ε is often stated as a part of the model. Hence, the problem is reduced to the estimation of (p + 1) values rather than the complicated function f. The single variable y is called the response (output or dependent variable) and $x_1, x_2, \ldots, x_p$ are known as predictor (input, independent or explanatory variables) when P = 1 ,the model is said to be simple regression model but if P > 1 it is called multiple linear regression model. When is more than one y then the model is called the multivariate multiple regression model which we will not including in our study. In a linear model, the parameters enter linearly but the predictors do not have to be linear [2][5].

However, some relationship can be transformed to linearity as the last equation which can be linearized by taking logarithms.

Recently, a great attention is being focused on biased estimation of regression parameters of a linear regression model. This attention is due to the inability of ordinary least squares to provide reasonable estimates when the matrix of explanatory variables is ill conditioned. Despite possessing the very desirable property of being the best linear unbiased estimator (BLUE) under the usual conditions imposed on the model, the least squares estimators can nevertheless, have extremely large variances, when the data are multicollinear which is one form of ill conditioning. Therefore ,many researches were performed to achieve biased estimators with better overall performance than the ordinary least squares estimators [5][6].

1.2 The Aims of the thesis

In our thesis, we tried to attain different goals that can be summarized as follows:

1. To assess the performance of PC estimators and RR estimators as an alternatives to OLS estimators in the case of existence of the multicollinearity problem.

2. Representing different methods for detecting the multicollinearity problem and determine its probable effect, on the linear regression model.

3. Studying different methods for estimating the ridge parameter (k say) and contains our proposed method for estimating the ridge parameter then specifying the best of these methods, by using simulation some statistical measurements.

1.3 Study Limitations

In our study, we assume that there is no missing data or outliers in the dataset. Moreover, we focus our attention on the multicollinearity problem irrespective of other problems that the analyst may face, such as the autocorrelation and heteroscedasticity problems.

1.4 Literature Review

The problem of multicollinearity among the data set and different topics related with it , were broadly discussed in literature. The following are some instances.

In an article about response surfaces, Hoerl (1962).introduced many concepts which was the basis of ridge regression. This was followed by Hoerl and kennard (1970) article which gave a major impetus to ridge regression the bulk of the article was devoted to the ordinary ridge estimator $b_{RR} = (X'X + kI)^{-1}X'y$ where k is an exogenous parameter have to be determined. The authors stated that there alwayes exists a $k > 0$ such that mean square error b_{RR} less than mean square error b_{OLS} . They also mentioned to the generalized form of ridge regression K where K is a digonal matrix of ridge parameters [30].

Lawrence S . Mayer and Thomas A. Willke (1974) viewed the ridge estimators as a subclass of the class of linear transforms of the least squares estimator. An alternative class of estimators, labeled shrunken estimators was considered. It was shown that these estimators satisfy the admissibilily condition proposed by Hoerl and Kennard [34].

M. Goldstein and A. F . M . Smith (1974) followed a new derivation of the Hoerl-Kennard (1970) ridge estimator and its generalization. Comparison was made with James-stein estimator and with the generalized inverse estimator proposed by Marquardt (1970). Also a Bayesian approach was noted [24].

Donald, W. Marquardit and Ronald D. Snee (1975) discussed the use of biased estimation in data analysis and model building A review of the theory of ridge regression and its relation to generalized inverse regression was presented along with the results of a simulation experiment and three examples of the use of ridge regression in practice .Comments on variable selection procedures were included.

They concluded that when the predictor variables are highly correlated, ridge regression produces coefficients which predict and extrapolate better than least squares and is a safe procedure for selecting variables [32].

Richard F . Gunst and Robert L. Mason (1977) employed mean squared error criterion to compare five estimators of regression coefficients. Specifically, least squares, Principal components, ridge regression, latent root, and a shrunken estimator. Each of the biased estimators was shown to offer improvement in mean squared error over least squares for wide range of choices of the parameters of the model. The results of a simulation involving all five estimators indicated that the principal components and latent root estimators perform best overall, but the ridge regression estimator has the potential of a smaller mean square error than either of these [29].

William, E. Strawderman (1978) used the generalized ridge regression estimator to estimate the vector of regression coefficients where the ridge constant was chosen on the basis of the data. For general quadratic loss he produced such estimators whose risk function dominates that of the least squares procedure provided the number of regressors is at least three. He studied the problem by the usual reduction to estimating the mean vector of a multivariate normal distribution [43].

George Cassela (1980) used an entirely new method of proof to derive conditions that are necessary and sufficient for minimaxity of a large class of ridge regression estimators. The conditions he derived were very similar to those derived for minimaxity of some stein type estimators [13].

George Cassela (1985) mentioned that the ridge regression was originally formulated with two goals in mind : improvement in mean squared error and numerical stability of the coefficient estimates. Conditions were given under which a minimax ridge regression estimator can also improve stability, a quantity that can be measured with the condition number of the matrix to be inverted. The consequences of trading numerical stability for minimaxity were also discussed [14].

Quirino, Paris . (2001) proposed a novel maximum entropy estimator as an alternative to ridge regression estimator. The proposed estimator does not depend upon any additional information and does not affected by any level of multicollinearity and dominates the OLS estimator uniformly as it was shown by Monte Carlo experiments. The same experiments provided evidence that it is asymptotically unbiased and the estimates are normally distributed [38].

Fikri Akdeniz and selahttin Kaciranlar (2001) considered the standard multiple linear regression model where the matrix X was assumed to be of full column rank. They introduced a new biased estimator known as restricted Liu estimator and compared it with restricted least squares estimator in the matrix mean squared error sense [8].

G.R. Pasha, Muhammad Akber Ali shah and Ghosia (2004) adopted an unconventional method of the principal components for the solution of multicollinearity and an attempt was made to show that by using such technique, some fairly precise estimates of the coefficients were obtained. The comparison between the variance of OLS estimates and principal components estimates was made on income and consumption model [39].

Yuzo Maruyama and william, E. Strawderman (2005) considered the standard linear regression model.They considered the estimation of β under general quadratic loss functions. In fact , they extended the work of strawderman (1978) and Cassela (1985) by finding adaptive minimax estimators of β , which have greater numerical stability (i.e., smaller condition number) than the usual least squares estimator. They gave a subclass of such estimators which have a very simple form [33].

Hazim Mansoor Gorgees (2009) viewed the ridge estimators as a subclass of the class of shrinkage estimators. He stated the fact that the shrinkage factor can be chosen will guarantee the ridge estimator to have mean square error smaller than the variance of least squares estimator, [26].

M. Revan Ozkala (2009) Introduced a new estimator when there exist stochastic linear restrictions on the parameter vector. The new estimator introduced by combining the ideas underlying the mixed and ridge regression estimators under the assumption that the errors are not independent and identically distributed. He called his new estimator as the stochastic restricted ridge regression (SRRR) estimator. The performance of (SRRR) estimator over the mixed estimator with respect of the variance and MSE matrices was examined [37] .

A.V. Dorugade and D.N. Kashid (2010) proposed new method for choosing the ridge parameter. The performance of the proposed method was evaluated and compared through simulation study in terms of mean square error. The technique developed seems to be very reasonable because of having smaller MSE [20].

Hazim Mansoor Gorgees (2010) considered a Bayesian formulation of ridge regression problem which derived from a direct specification of prior informations about parameters of general linear regression model when the multicollinearity problem is presented. In addition to the Bayesian estimator of the ridge parameter, he followed entirely a new approach to derive the conventional estimator for the ridge parameter proposed by Horel-Kennard. A numerical example was given in order to compare the performance of such estimators [27].

M. EI - Dereny and N. I. Rashwan (2011) introduced many different methods of ridge regression. Those methods included ordinary ridge regression(ORR), generalized ridge regression (GRR) and directed ridge regression (DRR). Methods of selecting ridge parameter were discussed. They used simulation to compare between such methods with OLS method. They were better than OLS method when the multicollinearity is exist [16].

Feras Sh. M. Batah (2011) proposed a new estimator. Namely, Generalized Jackknife ridge regression estimator (GJR) by generalizing the modified jackknife ridge regression estimator (MJR). He showed that the (GJR) estimator is superior in the MSE sense to the LASSO estimator , Generalized ridge regression estimator, Jackknife ridge regression estimator and modified Jackknifed ridge regression estimator [12].

Hazim mansoor Gorgees and Bushra Abd alrasool Ali (2013) studied two types of ordinary ridge regression estimators according to the choice of ridge parameter as well as the generalized ridge regression estimator. These methods were applied on a data set suffer from a high degree of multicollinearity. It was found that the generalized ridge regression estimator perform better than the other two methods in the sense of MSE and coefficient of determination R^2 [25].

Yasin Asar, Adnan Karaibrahimoglu and Asir Genc (2014) proposed some modified ridge parameters. They compared their estimators with some estimators proposed earlier according to MSE criterion. All results were calculated by a Monte Carlo simulation. They concluded that their estimators perform better than the other in most situations in the sense of MSE [11].

Anwar Fitrianto and Lee CinyYik (2014) conducted some simulation study to compare the performance of ridge regression estimator and the OLS. They found that Hoerl and Kennard ridge regression estimation method has better performance than the other approaches [22].

Ashok V. Dorugade (2014) introduced a new approach to obtain the ridge parameter. Furthermore, he compared the proposed ridge parameter with other well-known ridge parameters in terms of MSE criterion. Finally, a numerical example and simulation study had been conducted to illustrate the optimality of the proposed ridge parameter [18].

Ahlam Abdullah Al somahis, Salwa Mousa and lutfiah ismail ALTurk (2015) proposed new methods for choosing ridge parameter for logistic regression. The performance of the proposed methods were evaluated and compared with other models that having different previously suggested ridge parameter through a simulation study in terms of (MSE). They concluded that their suggested logistic ridge regression estimators were superior in most of the cases [10].

Chapter Two
The Theoretical Side

2.1 Multiple Linear Regression

The most widely used regression models is the multiple linear regression model. In this model the response (dependent) variable y may be regarded as the weighted sum of the explanatory (independent) variables $x_1,x_2,...,x_p$ (say) with unknown weights $\beta_1,\beta_2,...,\beta_p$. In general ,the multiple linear regression model can be written as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i \ , i = 1, 2, ..., n$$

where p is the number of explanatory variables, n is the number of observations.

The matrix representation of the model is $y = X\beta + \varepsilon$ where $y = (y_1, y_2, ..., y_n)'$ is the vector of the response variable, $\beta = (\beta_0, \beta_1, ..., \beta_p)'$ is the vector of the unknown parameters, $X = \begin{pmatrix} 1 & x_{11} \cdots & x_{1p} \\ 1 & x_{21} \cdots & x_{2p} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{np} \end{pmatrix}$ is an $n \times (p + 1)$ matrix of explanatory variables and $\varepsilon' = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)'$ is the vector of random errors.

In this situation many assumptions have to be satisfied, these assumptions can be summarized as follows:

i. The error terms are independent and have a constant variance σ^2.

ii. The error terms are normally distribute with mean vector equal to 0 and variance-covariance matrix $\sigma^2 I_n$ that is $\varepsilon \sim N(0, \sigma^2 I_n)$.

iii. The matrix X is of full column rank (i.e. rank (X) = p + 1 < n).

2.1.1 Ordinary Least Squares Estimators

The ordinary least squares method is the most popular estimation procedure, based on the minimization of the sum of squared deviations $\varepsilon'\varepsilon$ where

$$\varepsilon'\varepsilon = (y - X\beta)'(y - X\beta)$$

$$\varepsilon'\varepsilon = y'y - \beta'X'y - y'X\beta + \beta'X'X\beta$$

$$\varepsilon'\varepsilon = y'y - 2\beta'X'y + \beta'X'X\beta$$

This follows due to the fact that $\beta'X'y$ is a (1×1) matrix or scalar,

whose transpose $(\beta'X'y)' = y'X\beta$ must have the same value.

The least squares estimate of β is the value b_{ols} (say) which when substituted in equation (2.1) minimizes $\varepsilon'\varepsilon$. It can be accomplished by differentiating $\varepsilon'\varepsilon$ with respect to β and setting the resultant matrix equation equal to zero, at the same time replacing β by b_{ols} .This provides the normal equations

$$(X'X)b_{ols} = X'y$$

The solution of this equations is

$$b_{ols} = (X'X)^{-1}X'y \qquad \text{...(2.1)}$$

2.1.2 Properties of Ordinary Least Squares Estimator of Regression Coefficients

The ordinary least squares estimator b_{ols} has the following properties:

1. b_{ols} is an estimate of β which minimizes the error sum of squares $\varepsilon'\varepsilon$ irrespective of any distribution properties of the errors.
2. In fact, an assumption that the error terms ε_i, i =1,2,...,n are normally distributed is not required to obtain the estimators of the unknown parameters, but it is required in order to construct the t and F tests which depends upon the assumption of normality of error terms as well as obtaining the confidence intervals for the estimated coefficients, based on the t and F tests. However, if the error terms are normally distributed with mean vector 0 and variance-covariance matrix $\sigma^2 I_n$,then the OLS estimator and the MLE of β are equivalent since maximizing the likelihood function is equivalent to minimizing the quantity $\varepsilon'\varepsilon$ [19].
3. The elements of b_{ols} are linear functions of the observations $y_1, y_2, \dots, y_n$ and provide unbiased estimates of the elements of β.This can be easily shown as follows:

$$b_{ols} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta+\varepsilon)$$

$$b_{ols} = \beta + (X'X)^{-1}X'\varepsilon$$

$$E(b_{ols}) = E(\beta + (X'X)^{-1}X'\varepsilon) = \beta + (X'X)^{-1}X'E(\varepsilon)$$

Since $E(\varepsilon)=0$ then $E(b_{ols})=\beta$.

4. Irrespective with the distribution of error terms, the fitted values are obtained from $\hat{y} = X b_{ols}$.
5. The residuals vector given by $e = y - \hat{y}$ is orthogonal to all explanatory variables, that is

$$X'e = X'(y - \hat{y})$$

thus $$X'e = X'y - X'X(X'X)^{-1}X'y = X'y - X'y = 0$$

6. The fitted values are orthogonal to the vector of residuals, that is :

$$e'\hat{y} = (y - \hat{y})'\hat{y} = (y - Xb_{ols})'\hat{y}$$

$$e'\hat{y} = (y' - b_{ols}'X')\hat{y} = y'\hat{y} - b_{ols}'X'\hat{y}$$

$$e'\hat{y} = (y'Xb_{ols} - y'X(X'X)^{-1}X'Xb_{ols}$$

$$e'\hat{y} = y'Xb_{ols} - y'Xb_{ols} = 0$$

7. The variance-covariance matrix of b_{ols} denoted as $\text{var}(b_{ols}) = \sigma^2(X'X)^{-1}$

provides the variances (diagonal terms) and covariances (off diagonal terms) of the estimates. $\text{var}(b_{ols})$ can be derived as follows:

$$\text{var}(b_{ols}) = E(b_{ols} - \beta)(b_{ols} - \beta)'$$

Since $$b_{ols} - \beta = (X'X)^{-1}X'y - \beta$$

$$= (X'X)^{-1}X'(X\beta + \varepsilon) - \beta = (X'X)^{-1}X'\varepsilon$$

hence

$$\text{var}(b_{ols}) = E[(X'X)^{-1}X'\varepsilon][(X'X)^{-1}X'\varepsilon]'$$

$$\text{var}(b_{ols}) = E[(X'X)^{-1}X'\varepsilon][\varepsilon'X(X'X)^{-1}]$$

$$\text{var}(b_{ols}) = [(X'X)^{-1}X']E(\varepsilon\varepsilon')[X(X'X)^{-1}]$$

$$\text{var}(b_{ols}) = (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1}$$

$$\text{var}(b_{ols}) = \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

8. Assuming that X_0' is a specified $1 \times p$ vector of elements which are of the same form as a row of x so that $\hat{y}_0 = X_0' b_{ols} = b_{ols}'X_0$ is a fitted value at a specified point X_0. Then $\hat{y}_0$ is the value predicted at x_0 by the regression equation, and has the variance

$$\text{var}(\hat{y}_0) = \text{var}(X_0' b_{ols}) = X_0' \text{var}(b_{ols}) X_0 = \sigma^2 X_0'(X'X)^{-1}X_0$$

9. It can be easily shown that the ordinary least squares estimator is consistent, sufficient as well as unbiased estimator for β and its variance attains the lower bound of Rao-Gramer inequality [28].

   var(T) ≥ $\frac{[1+B(\theta)]^2}{E[\frac{\partial \ln L}{\partial \theta}]^2}$ where $B(\theta)$ is the bias of the estimator

Accordingly, b_{ols} is the best linear unbiased estimator [BLUE] for β .

2.1.3 Testing of Hypothesis in a Linear Regression Model [15]

The different hypothesis concerning the regression parameters may all be examined in a similar way by a unified approach. Let us reffered to the general linear regression model as the full model (FM). When some regression coefficients are specified, the resulting model is said to be the reduced model(RM).The hypothesis to be tested is

H_0 : RM is adequate against H_1: FM is adequate

In the full model, there are (p+1) parameters ($\beta_0, \beta_1, \dots, \beta_P$) to be estimated. Let us assume that for the reduced model there exist k different coefficients. Let $\hat{y}_i$ and $\hat{y}_i*$ be the values predicted for y_i by the FM and RM respectively. The residuals sum of squares obtained when fitting the full and reduced model are respectively

$$\left.\begin{aligned} \text{SSE(FM)} &= \sum (y_i - \hat{y}_i)^2 \\ \text{SSE(RM)} &= \sum (y_i - \hat{y}_i *)^2 \end{aligned}\right\} \quad \dots(2.2)$$

$$\text{F} = \frac{[SSE(RM) - SSE(FM)]/(P+1-K)}{SSE(FM)/(n-P-1)} \quad \dots(2.3)$$

The FM has (p + 1) regression coefficients, thus, SSE(FM) has n-p-1 d.f., similarly, SSE(RM) has n-k d.f. since the reduced model has k regression coefficients. Consequently, SSE(RM) - SSE(FM) has (n-k) - (n-p-1) = p+1-k d.f., the observed F statistic has F distribution with (p+1-k) and (n-p-1) d.f. If the observed F value is larger than the theoretical value of F with (p+1-k) and (n-p-1) d.f. and specified value of significant level α then the null hypothesis H_0 is rejected at level of significant α. In other word H_0 is rejected if

$$F \geq F_{(p+1-k,n-p-1,\alpha)}$$

Where F is the observed value of F test in equation (2.3), $F_{(p+1-k,n-p-1,\alpha)}$ is the tabulated value obtained from the F table , α is the level of significant.

A substantial special case of the F test in equation (2.3) is obtained when the null hypothesis is $H_0: \beta_j = 0, j = 1,2,\dots,p$ which means that all explanatory variables under consideration have no significant effect. In such a case the reduced and full models will be :

$$\text{RM}: H_0: y = \beta_0 + \varepsilon \qquad \text{...(2.4)}$$

$$\text{FM}: H_1: y = \beta_0 + \beta_1 X_1 + \dots + \beta_P X_P + \varepsilon \qquad \text{...(2.5)}$$

Here, SSE(FM) = SSE and SSE(RM) = $\sum (y_i - \bar{y})^2$ = SST since the least square estimate of β_0 in the RM is $\bar{y}$ .The F ratio in (2.4) reduces to

$$F = \frac{(SST - SSE)/P}{SSE/(n-P-1)} \qquad \text{...(2.6)}$$

Hence, the ANOVA table in multiple regression can be arranged as follows:

Source	d.f	Sum of squares	Mean square	F test
Regression	P	SSR	MSR= $\frac{SSR}{P}$	F = $\frac{MSR}{MSE}$
Residual	n-p-1	SSE	MSE= $\frac{SSE}{n-p-1}$	
Total	n-1	SST		

2.1.4 The Coefficient of Determination [15][3]

The ratio $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$ (2.7) is called the coefficient of determination .It can be explained rate as the proportion of total variability in the response variable y which can be accounted by the set of predictors $x_1, x_2, \ldots, x_p$.

Clearly, the value of R^2 is close to one when the model fits the data well. In such a cases the observed and predicted values become close to each other, and $\sum (y_i - \hat{y}_i)^2$ will be small, and R^2 will be near unity. However, the reverse of this term is inaccurate, which implies that a large value of R^2 does not necessarily mean that the model fits the data well. A more detailed analysis is required to in order to ensure that the model described the data properly.

A value related to R^2 is called the adjusted R^2 denoted by R_a^2. It can be obtained from R^2 by dividing SSE and SST by their respective d.f. and it is defined as follows :

$$R_a^2 = 1 - \frac{SSE/(n-P-1)}{SST/(n-1)} \qquad ...(2.8)$$

This measurement is sometimes employed to compare models that have different numbers of explanatory variables.

2.2 Singular Value Decomposition

In this section we introduce and discuss the concept of singular value decomposition of matrices as a powerful technique for analyzing the linear regression model. Accordingly the following definitions are necessary.

Definition 1

The singular values of a square matrix A is defined to be the square root of the eigenvalues of matrix (A'A) [40][26].

Definition 2

The condition number is regarded as a ratio of the largest singular value to the smallest singular value [14][26] .

Definition 3

If the condition number is too large, the matrix is said to be ill conditioned. If the condition number is infinite then the determinant of the matrix is equal to zero and hence the matrix is singular [14][26].

Definition 4

Any symmetric matrix A is said to be positive definite if for each non zero vector y then y'Ay> 0 [28][26].

Definition 5

Any real symmetric matrix $A_{m\times m}$ has a spectral decomposition of the form A= U Λ U'. The columns of U are the normalized eigenvectors of A and Λ is diagonal matrix whose diagonal elements represent the eigenvalues of A [42].

Definition 6

Any real matrix $B_{n\times m}$, (n > m) has singular value decomposition of the form B = U Γ V'.

Where:

U is an $n \times m$ matrix with orthonormal columns (U'U = I) , V is an $m \times m$ orthonormal matrix (V'V = I) and Γ is an $(m \times m)$ diagonal matrix with positive or zero diagonal elements known as the singular values.

From matrix B we can construct two positive definite matrices BB' and B'B where

BB'= U Γ V'V Γ U' = U Γ^2 U'

Similarly B'B = V Γ^2 V'

Using the decomposition above, we can identify the eigenvectors and eigenvalues for B'B as the columns of V and the squared diagonal elements of Γ respectively. The latter show that the eigenvalues of B'B must be non negative [42].

2.3 The Case of Collinear Data [2][4][17]

One of the most important assumptions associated with regression analysis is that the explanatory variables are not strongly interrelated. Usually, the regression coefficient is interpreted as the change in the response variable when the corresponding explanatory variable is increased by one unit while all other explanatory variables are held constant. This explanation will not be useful if there exist strong linear relationships between the explanatory variables. When the linear relationship among explanatory variables is completely absent, they are known to be orthogonal. In most regression problems the explanatory variables are not orthogonal, generally, the absence of orthogonality is not significant enough to abort the analysis. However, in many cases the explanatory variables are very strongly interrelated that the regression results are unclear.

The condition of sever non orthogonality is also referred to as the problem of collinear data or the problem of multicollinearity. Two types of multicollinearity may be faced in regression analysis. specifically perfect and near multicollinearity as an example of perfect multicollinearity, suppose that the four ingredients of a mixture are studies considered by including their percentages of the total, p_1, p_2, p_3, p_4. These variables will have the exact linear relationship $p_1 + p_2 + p_3 + p_4 = 100$ The problem of multicollinearity may be highly difficult to discover. It is not a specification error that can be detected by investigating regression residual. Actually, it is not modeling error. It is a case of imperfect data. However, it is necessary to know when multicollinearity is exist and to be aware of its possible effects. Accordingly, one have to be very careful about any or all substantive conclusions based on regression analysis in the existence of multicollinearity.

2.3.1Consequences of Multicollinearity [5][6][3]

During regression calculations, the exact linear relationship among the explanatory variables implies a division by zero which in turn causes the calculations to be aborted. When the relationship is not exact the division by zero does not happen and the calculations would not aborted but the division by a very small quantity still deform the results. In the case of near multicollinearity it is impossible to estimate the unique effects of individual variables in the regression equation because the multicollinearity can be thought of as a situation where two or more explanatory variables move together, consequently, it is impossible to determine which of the explanatory variables is producing the observed change in the response variable .

The estimated values of the coefficient are very sensitive to inconsiderable variations in the data and addition or deletion of variables in the equation. The regression parameters would have large standard errors which influence both inference and forecasting that is based on the regression equation. Multicollinearity deflate the partial t tests for the regression coefficients and give false non significant P values.

2.3.2 Sources of Multicollinearity [1][7][21]

In order to deal with multicollinearity problem we have to be able to identify its sources because the source of multicollinearity affects the analysis, the correction and interpretation of the linear regression model. The origins of multicollinearity may be summarized as follows:

1. The multicollinearity has been created by the sampling technique . In this case the data have been collected from a narrow subspace of the explanatory variables. Collecting more data on an expanded range would treat this kind of multicollinearity problems.
2. Substantial restrictions of the linear regression model or population. This source of multicollinearity exist whatever the sampling technique is applied. Many industrial or economic processes have restrictions on explanatory variables, either physically, legally or politically which will cause multicollinearity.
3. Model choice or specification. This source of multicollinearity results from using explanatory variables that are powers or interactions of the original variables.
4. Extreme values or outliers in the X space can cause multicollinearity. This should be remove before any treatment is applied.

2.3.3 Detection of Multicollinearity [3][5][15][32]

Different methods of detecting multicollinearity are presented, we review some of them:

1. At the first , we consider pair wise scatter plots of pairs of explanatory variables looking for near exact relationships. Also glance at the correlation matrix for high correlations. However, multicollinearity does not always clear when studying the variables two at a time next, we investigate variance inflation factors (VIF)values.VIF greater than 10 indicates the existence of multicollinearity problem.
2. Eigen values of the correlation matrix of explanatory variables near zero indicate . the existence of multicollinearity problem
3. The large condition numbers is an indicator of the presence of multicollinearity .
4. Explanatory variables whose regression coefficients are opposite in sign from what we believed may reveal multicollinearity problem.
5. Farrar-Glouber test.

This test is based on the chi square statistic .The null hypothesis to be tested is

$$H_0 : X_j \text{ are orthogonal }, j = 1,2, \ldots, P$$

Against the alternative hypothesis

$$H_1 : X_j \text{ are not orthogonal}$$

The test statistic is

$$\chi_0^2 = - [(n-1) - \frac{1}{6}(2P + 5)] \ln |D| \qquad \ldots(2.9)$$

Where

n is the number of observations

P is the number of explanatory variables

$|D|$ is the determinant of the correlation matrix.

Comparing the calculated value χ_0^2 with the tabulated value at P(P -1)/ 2 degrees of freedom and specified level of significant α .

If the calculated value χ_0^2 is greater than the tabulated value, then the null hypothesis H_0 will be rejected which means that the explanatory variables are interrelated. otherwise, the null hypothesis cannot be rejected.

2.3.4 The Class of Shrinkage Estimators [26][34][36]

Assuming that an $(n \times p)$ matrix of explanatory variables X and an $(n \times 1)$ vector of the associated response variable y are known. Moreover let us assume that the sample means have been removed from the dataset so that $\underline{1}'X = 0'$ and $\underline{1}'y = 0$ where $\underline{1}$ is an (n) vector of ones. The singular value decomposition technique will be employed in order to obtain a deeper understanding of our data set. Consequently , we decompose the matrix X as follows :

$$X = H\,\Lambda^{\frac{1}{2}}G' \qquad ...(\,2.10\,)$$

Where H is an $n \times p$ matrix satisfy H'H = I_p , $\Lambda^{\frac{1}{2}}$ is a $p \times p$ diagonal matrix of ordered singular values of X, that is $\lambda^{\frac{1}{2}}{}_1 \geq \lambda^{\frac{1}{2}}{}_2 \geq ... \geq \lambda^{\frac{1}{2}}{}_p > 0$ so that β is estimable G is $(p \times p)$ orthogonal matrix whose columns are the eigenvectors of X'X.

Using this techniques the information matrix X'X can be written as follows:

$$X'X = G\,\Lambda^{\frac{1}{2}}\,H'H\,\Lambda^{\frac{1}{2}}\,G' = G\,\Lambda G' \qquad ...(\,2.11\,)$$

Accordingly , the ordinary least squares estimator can be rewritten as :

$$b_{OLS} = (X\,'X\,)^{-1}Xy = (G\,\Lambda G\,')^{-1}G\,\Lambda^{\frac{1}{2}}H\,'y$$

$$b_{OLS} = G\,\Lambda^{-1}G\,'G\,\Lambda^{\frac{1}{2}}\,H\,'y = G\,\Lambda^{-\frac{1}{2}}\,H\,'y = GC$$

Where the vector $C = \Lambda^{-\frac{1}{2}} H'y$ is the vector of uncorrelated components of b_{OLS} .

Obviously :

E(C) = E (G' b_{OLS}) = G'β = γ (say) and var(C) = var (G' b_{OLS}) = G' var (b_{OLS})G

Hence,

var(C) = σ^2 G' $(G\Lambda G')^{-1}$ G = σ^2 G'G Λ^{-1} G'G = σ^2 Λ^{-1}

Which is a diagonal matrix , therefore the components of C are uncorrelated since the off diagonal elements of var(C) which represent the covariance terms are equal to zero.

The class of shrinkage estimators denoted by b_{SH} will have the general form

$$b_{SH} = G\,\Delta\,C = \sum_{j=1}^{p} \overrightarrow{g_j}\,\delta_j C_j \qquad \text{...(2.12)}$$

Where $\overrightarrow{g_j}$ is the j^{th} column of the matrix G , δ_j is the j^{th} diagonal element of the shrinkage factors matrix Δ, the range of shrinkage factors is usually restricted and it be:

$0 \le \delta_j \le 1$, j=1,2,...,p, C_j is the j^{th} element of vector C.

2.3.5 Properties of Shrinkage Estimators [31][35][41]

In general the shrinkage estimators are biased since

E(b_{SH}) = E(G Δ C)= G Δ E(C) =G Δ γ and this vector is never equal to β = G γ unless $\Delta = I$. Hence, bias(b_{SH})=G(I- Δ) γ

The variance matrix of b_{SH} for non stochastic shrinkage factors is

var(b_{SH})= var(G Δ C)=G Δ var(C) Δ G'= σ^2 G Δ^2 Λ^{-1} G'

The mean square error matrix of b_{SH} is given by

MSE(b_{SH})= E(b_{SH} - β) (b_{SH} - β)'

Equivalently:

MSE(b_{SH}) = MSE(GΔC) = G MCE(ΔC) G'

Here, we focus our attention on MSE(ΔC) which can be derived as :

MSE(ΔC) = var (ΔC) + (bias ΔC) (biasΔC)'

MSE(ΔC) = $\sigma^2\ \Delta^2\ \Lambda^{-1}$ + (I - Δ) γ γ' (I - Δ)

Clearly, MSE (ΔC) is the sum of two matrices, the diagonal variance matrix $\sigma^2\ \Delta^2\ \Lambda^{-1}$ and the matrix (I - Δ) γ γ' (I - Δ) with squared bias terms on the diagonal.

Let us consider the i^{th} diagonal element of the matrix MSE (ΔC) denoted by $\delta_i\ C_i$ Hence,

$$\text{MSE}\ (\delta_i\ C_i\ ) = \frac{\sigma^2\delta_i^2}{\lambda_i} + (1-\delta_i\ )^2\ \gamma_i^2 \qquad \text{...(2.13)}$$

Clearly, MSE ($\delta_i\ C_i$) changes as the i^{th} shrinkage factor δ_i changes. Actually, the first partial derivative of MSE ($\delta_i\ C_i$) with respect to δ_i is

$$\frac{\partial MSE\ (\delta_i\ C_i\ )}{\partial \delta_i} = \frac{2\sigma^2\delta_i}{\lambda_i} - 2(1-\delta_i\ )\gamma_i^2 \qquad \text{...(2.14)}$$

While the second partial derivative is

$$\frac{\partial^2 MSE\ (\delta_i\ C_i\ )}{\partial \delta_i^2} = \frac{2\sigma^2}{\lambda_i} + 2\gamma_i^2 \qquad \text{...(2.15)}$$

To obtain the optimal value of δ_i denoted as δ_i^{MSE} Minimize MSE($\delta_i\ C_i$), we equate the first partial derivative to zero and solve for δ_i as follows:

$$\frac{\partial MSE(\delta_i C_i)}{\partial \delta_i} = 0 \Rightarrow \frac{2\sigma^2\delta_i}{\lambda_i} - 2(1-\delta_i)\gamma_i^2 = 0$$

$$\Rightarrow \frac{\sigma^2\delta_i}{\lambda_i} + \delta_i\gamma_i^2 = \gamma_i^2 \Rightarrow \delta_i\left[\frac{\sigma^2}{\lambda_i} + \gamma_i^2\right] = \gamma_i^2$$

It follows that

$$\delta_i^{MSE} = \frac{\gamma_i^2}{\gamma_i^2 + \frac{\sigma^2}{\lambda_i^2}} = \frac{\gamma_i^2}{\frac{\lambda_i\,\gamma_i^2 + \sigma^2}{\lambda_i}} = \frac{\lambda_i\gamma_i^2}{\lambda_i\gamma_i^2 + \sigma^2}$$

Dividing both the numerator and denominator by γ_i^2 to get

$$\delta_i^{MSE} = \frac{\lambda_i}{\lambda_i + \frac{\sigma^2}{\gamma_i^2}} \qquad \text{...(2.16)}$$

Clearly δ_i^{MSE} of equation (2.16) can never be negative nor larger than 1.

2.4 Ordinary Ridge Regression Estimator

Different methods have been suggested to deal with co-linear data by adjusting the least squares method in order to allow introducing some bias in the estimators of regression parameters. One of the most popular methods is labeled as ridge regression method. The ridge regression estimators depend exactly upon an external parameter (k say) known as the ridge parameter or biasing constant for any k ≥ 0, we define the ridge regression estimator b_{RR} as follows:

$$b_{RR} = (X'X + kI)^{-1}X'y \qquad \text{...(2.17)}$$

Where the value of k is chosen by the analyst according to some reasonable criteria established by Hoerl and kennard [30][21].

It can be easily shown that the ridge regression estimator given in equation (2.17) is a member of the class of shrinkage estimators as follows:

By using singular value decomposition approach and matrix algebra we have

$$b_{RR} = (X'X + kI)^{-1}X'y = [G(\Lambda + kI)G']^{-1}G\Lambda^{\frac{1}{2}}H'y$$

$$b_{RR} = G(\Lambda + kI)^{-1}G'G\Lambda^{\frac{1}{2}}H'y$$

$$b_{RR} = G(\Lambda + kI)^{-1}\Lambda^{\frac{1}{2}}H'y$$

$$b_{RR} = G[(\Lambda + kI)^{-1}\Lambda]\Lambda^{-\frac{1}{2}}H'y$$

$$b_{RR} = G\Delta C \qquad ...(2.18)$$

Where $\Delta = (\Lambda + kI)^{-1}\Lambda$ which is a diagonal matrix . The j^{th} diagonal element of the matrix Δ has the form

$$\delta_j = \frac{\lambda_j}{\lambda_j + k}, \quad \text{j=1,2,...,P}$$

Where λ_j is the j^{th} element (eigenvalue) of the diagonal matrix Λ and k is the ridge parameter [21].

2.4.1 Properties of Ridge Estimator

We proceed our discussion with the following theorem since it is of great importance as we believe to clarity the properties of ridge regression estimator.

Theorem 1 [28]

Let y be an $n \times 1$ random vector. Let E(y) = μ and var(y) = $\sum$. Then

E(y'Ay) = tr A $\sum$ + μ ' A μ ...(2.19)

In order to control the inflation and general instability associated with the least squares estimate, Hoerl, A.E.(1962) suggested using the estimator

$b_{RR} = (X'X + kI)^{-1}X'y$, k $\geq$ 0

Putting W=$(X'X + kI)^{-1}$ then b_{RR} can be rewritten as follows:

$$b_{RR} = W X'y \qquad ...(2.20)$$

The following alternative form describe the relationship of ridge estimate with the least squares estimate

$b_{RR} = (X'X + kI)^{-1}X'Xb_{OLS}$ = $[I_p + k(X'X)^{-1}]^{-1}b_{OLS}$ = Z b_{OLS} ...(2.21)

Where :

Z= $[\mathrm{I} + k(X'X)^{-1}]^{-1}$ From (2.21) it is clear that b_{RR} is a linear transformation of b_{OLS} and that b_{RR} is a biased estimate of β since E(b_{RR}) = E(Z b_{OLS}) =Z E(b_{OLS}) = Z β

The variance-covariance matrix of b_{RR} can be obtained as follows:

var(b_{RR}) = var(W X' y) = WX' [var(y)] XW'

var(b_{RR}) = WX' [$\sigma^2 I_n$] XW' = $\sigma^2 WX'XW'$

var(b_{RR}) = $\sigma^2 (X'X + kI)^{-1}X'X(X'X + kI)^{-1}$ ...(2.22)

The residual sum of squares may be written as

SSE (k) = (y- Xb_{RR})' (y- Xb_{RR})

SSE (k) = (y- Xb_{OLS})' (y- Xb_{OLS}) + (b_{RR} - b_{OLS})' X'X (b_{RR} - b_{OLS}) ...(2.23)

The formula in (2.23) can be demonstrated as follows:

= (y- Xb_{RR})' (y- Xb_{RR})

= (y - Xb_{RR} + Xb_{OLS} - Xb_{OLS})' (y - Xb_{RR} + Xb_{OLS} - Xb_{OLS})

= [(y - Xb_{OLS}) - X(b_{RR} - b_{OLS})]' [(y - Xb_{OLS}) - X(b_{RR} - b_{OLS})]

= (y- Xb_{OLS})' (y- Xb_{OLS}) + (b_{RR} - b_{OLS})' X'X (b_{RR} - b_{OLS}) - (b_{RR} - b_{OLS})' X'(y- Xb_{OLS}) - (y- Xb_{OLS})' X(b_{RR} - b_{OLS})

= (y- Xb_{OLS})' (y- Xb_{OLS}) + (b_{RR} - b_{OLS})' X'X (b_{RR} - b_{OLS}) - (b_{RR} - b_{OLS})'(X'y-X'Xb_{OLS}) - (X'y - X'X b_{OLS})' (b_{RR} - b_{OLS})

Clearly, the last two terms are equal to zero since X'y = X'Xb_{OLS} and hence the result.

the total mean square error is

TMSE (b_{RR}) = E (b_{RR} - β)' (b_{RR} - β)

By applying theorem 1 we get

TMSE (b_{RR}) = $\sigma^2 tr[(X'X+kI)^{-1}X'X(X'X+kI)^{-1}]+\beta'(Z-I)'(Z-I)\beta$ ...(2.24)

Substituting from Z by [I + k$(X'X)^{-1}]^{-1}$ in the second term of the right hand side of equation (2.24) and using matrix algebra for simplification we obtain

$$\text{TMSE}(b_{RR}) = \sigma^2\sum_{i=1}^{P}\frac{\lambda_i}{(\lambda_i+k)^2}+k^2\beta'(X'X+kI)^{-2}\beta \quad ...(2.25)$$

$$\text{TMSE}(b_{RR}) = \psi_1(k)+\psi_2(k)$$

The first term on the right hand side of equation(2.25) $\psi_1(k)$ is the sum of variances (total variance) of b_{RR} components and the second term $\psi_2(k)$ is the square of the bias.

Theorem 2 [30]

The total variance $\psi_1(k)$ is a continuous, monotonically decreasing function of k.

Corollary 2.1

The first derivative w.r.to k of the total variance $\psi_1'(k)$ approaches (- ∞) as k approaches 0^t *and* $\lambda_P \longrightarrow 0$, moreover the matrix X'X becomes singular. The proof of theorem (1) and its corollary (2.1) is readily obtained by expressing $\psi_1(k)$ and $\psi_1'(k)$ in terms of λ_i .

Theorem 3 [28]

The squared bias $\psi_2(k)$ is continuous, monotonically increasing function of k.

Proof:

Recalling that X'X = G Λ G' where G is the orthogonal matrix whose columns are the normalized eigenvectors of X'X and Λ is the diagonal matrix whose elements are the eigenvalues of X'X. Then $\psi_2(k)$ can be rewritten as

$$\psi_2(k)=k^2\beta'(X'X+kI)^{-2}\beta$$

$$\psi_2(k) = k^2 \beta'[G\Lambda G' + kGG']^{-2}\beta$$

$$\psi_2(k) = k^2 (G\beta)'(\Lambda + kI)^{-2} G\beta$$

Putting $\alpha = G\beta$ and expressing $\psi_2(k)$ in terms of its components we get

$$\psi_2(k) = k^2 \sum_{i=1}^{P} \frac{\alpha_i^2}{(\lambda_i + k)^2} \qquad \text{...(2.26)}$$

Each element $\lambda_i + k$ is positive since $\lambda_i > 0$ for all i and $k \geq 0$,hence there are no singularities in the sum. Also it is clear that $\psi_2(0) = 0$.

Then $\psi_2(k)$ is a continuous function for $k \geq 0$. If $k > 0$, $\psi_2(k)$ can be written as follows :

$$\psi_2(k) = \sum_{i=1}^{P} \frac{\alpha_i^2}{(1 + \frac{\lambda_i}{k})^2} \qquad \text{...(2.27)}$$

Since $\lambda_i > 0$ for all i, the functions $\frac{\lambda_i}{k}$ are obviously monotone decreasing as k increase and consequently each term of $\psi_2(k)$ is monotone increasing. This yield $\psi_2(k)$ to be monotone increasing.

Corollary 2.2

The squared bias $\psi_2(k)$ approaches $\beta'\beta$ as an upper limit

Proof:

From (2.27), we have

$$\lim_{k\to\infty} \psi_2(k) = \lim_{k\to\infty} \sum \frac{\alpha_i^2}{(1 + \frac{\lambda_i}{k})^2} = \sum \alpha_i^2 = \alpha'\alpha = \beta' G' G \beta = \beta'\beta$$

Theorem 4 (Existence Theorem) [30]

There alwayes exists a $k > 0$ such that :

TMSE (b_{RR}) < TMSE (b_{OLS})

Proof :

First , we have to note that

$$\psi_1(0) = \sigma^2 \sum_{i=1}^{P} \frac{1}{\lambda_i} \text{ and } \psi_2(0) = 0$$

Moreover

$$\frac{d\,TMSE\,(b_{RR})}{dk} = \frac{d\psi_1(k)}{dk} + \frac{d\psi_2(k)}{dk}$$

$$\frac{d\,TMSE\,(b_{RR})}{dk} = -2\sigma^2 \sum \frac{\lambda_i}{(\lambda_i + k)^3} + 2k \sum \frac{\lambda_i\,\alpha_i^2}{(\lambda_i + k)^3} \qquad ...(\ 2.28\ )$$

Each of $\psi_1(k)$ and $\psi_2(k)$ was established to be monotonically decreasing and increasing respectively. Their first derivatives are alwayes non positive and non negative respectively. The proof is completed whenever there exist, a k > 0 such that:

$$\frac{d\,TMSE\,(b_{RR})}{dk} < 0$$

From equation (2.28) it can be shown that the above inequality have been satisfied

when $k < \dfrac{\sigma^2}{\alpha^2_{MAX}}$ ...(2.29)

2.4.2 Choice of Ridge Parameter

The ordinary ridge regression estimator does not provide a unique solution to the multicollinearity problem , but provide a family of solutions. These solutions depend upon the value of k (the ridge parameter). No explicit optimum value can be found for k. Yet, several stochastic choices have been proposed for this ridge parameter. Some of these choices may be summarized as follows .

Hoerl and Kennard (1970).Proposed graphical method called ridge trace to select the value of the ridge parameter k. When viewing the ridge trace, the analyst picks the value of k for which the regression coefficients have stabilized [18][7].

Often, the regression coefficients will vary widely for small values of k and then stabilize. We have to choose the smallest value of k(which introduces the smallest bias) after which the regression coefficients have seem to remain constant.

Hoerl, Kennard and Baldwin (1975) ,proposed another method to select a single value of k given as[30] .

$$\hat{k}_{HKB} = \frac{p S^2}{b_{OLS}' b_{OLS}} \qquad \text{...(2.30)}$$

Where p is the number of explanatory variables, S^2 is the OLS estimator of σ^2 and b_{OLS} is the OLS estimator of the vector of regression coefficients β.

Lawless and Wang (1976) proposed selecting the value of k by using the following formula

[3].

$$\hat{k}_{LW} = \frac{p S^2}{b_{OLS}' X' X b_{OLs}} \qquad \text{... (2.31)}$$

Assuming that the regression coefficients vector has certain prior distribution srivastava followed Bayesian approach to estimate the ridge parameter . He concluded that [42].

$$\hat{k}_{Bayes} = Max\left[0, \frac{tr(X'X)}{\left[\frac{n-p-3}{n-p-1}\left(\frac{b_{OLS}' X' X b_{OLS}}{S^2}\right) - p\right]}\right. \qquad \text{...(2.32)}$$

Where tr (X'X) denoted to the trace of the matrix X'X.

2.4.3 Proposed Method

Our contribution in this topic represented by utilizing the concept of condition number in order to select the ridge parameter. The condition number is defined to be the ratio of the largest to the smallest singular value of the matrix of the explanatory variables X [14].

The suggested estimator denoted as $\hat{K}_{CN}$ is defined as follows :

$$\hat{k}_{CN} = Max\,[0, \frac{pS^2}{b_{OLS}{}'b_{OLS}} - \frac{1}{CN}] \qquad ...(\,2.33\,)$$

Where CN referred to condition number.

Our proposed estimator is the modification of $\hat{k}_{HKB}$. The small amount $\frac{1}{CN}$ is subtracted from $\hat{k}_{HKB}$.This amount , however ,varies with the strength of multicollinearity in the model.

If the condition number is too large then $\hat{k}_{CN}$ would coincide with $\hat{k}_{HKB}$ since in such case , the fraction $\frac{1}{CN}$ would approach to zero.

On the other hand if the condition number is too small (approximately equal to 1) then the possibility that $(\frac{pS^2}{b_{OLS}{}'b_{OLS}} - \frac{1}{CN})$ be negative is too large.

In this case we choose $\hat{k}_{CN}$ to be equal to zero which means that the ridge regression estimator would coincide with the ordinary least squares estimator and the data set is not influenced by the multicollinearity problem.

2.4.4 Generalized Ridge Regression [21][36][43]

By using the singular value decomposition technique in order to derive the generalized ridge regression , we can rewrite the linear regression model as follows:

$$y = X\beta + \varepsilon = (H\Lambda^{\frac{1}{2}})(G'\beta) + \varepsilon$$

or $$y = Z\alpha + \varepsilon \qquad ...(2.34)$$

Where : $Z = H\Lambda^{\frac{1}{2}}, \alpha = G'\beta$

The model in equation(2.34) is called the canonical model or uncorrelated components model . The OLS estimator of α is given as:

$$\alpha_{OLS} = (Z'Z)^{-1}Z'y = \Lambda^{-1}Z'y \qquad ...(2.35)$$

And $Var(\alpha_{OLS}) = \sigma^2(Z'Z)^{-1} = \sigma^2\Lambda^{-1}$ which is diagonal. This shows the important property of this parameterization since the elements of α_{OLS} , namely, ($\alpha_1, \alpha_2, ..., \alpha_p)_{OLS}$ are uncorrelated.

The generalized ridge estimator for α is given by:

$$\alpha_{GRR} = (Z'Z + K)^{-1}Z'y = (\Lambda + K)^{-1}Z'y \qquad ...(2.36)$$

$$= (\Lambda + K)^{-1}Z'Z\alpha_{OLS} = (I + K\Lambda^{-1})^{-1}\alpha_{OLS}$$

$$= W_K\alpha_{OLS} = diag(\frac{\lambda_i}{\lambda_i + K_i})\alpha_{OLS}\text{ , i=1,2,...,p}$$

Where K= diag$(K_1, K_2, ..., K_p)$ and, $W_K = (I + K\Lambda^{-1})^{-1} =$ diag$(\frac{\lambda_i}{\lambda_i + K_i})$ i= 1,2 ,...,p

The mean square error of α_{GRR} is given by

$$MSE(\alpha_{GRR}) = \text{var}(\alpha_{GRR}) + (bias\ \alpha_{GRR})(bias\ \alpha_{GRR})'$$

$$= \sigma^2 tr(W_K\Lambda^{-1}W_K') + (W_K - I)\alpha_{OLS}\alpha_{OLS}(W_K - I)'$$

$$= \sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i + K_i)} + \sum_{i=1}^{p}\frac{K_i^{\ 2} + \alpha^2_{i(OLS)}}{(\lambda_i + K_i)^2} \qquad ...(2.37)$$

To obtain the value of K_i that minimize $MSE(\alpha_{GRR})$ we differentiate equation (2.37) with respect to K_i and equating the resultant derivative to zero. Thus

$$\frac{\partial MSE(\alpha_{GRR})}{\partial K_i} = -\sigma^2 \sum \frac{\lambda_i}{(\lambda_i + K_i)^3} + \sum \frac{\lambda_i K_i \alpha_i^2{}_{(OLS)}}{(\lambda_i + K_i)^3} = 0$$

By solving for K_i we obtain $K_i = \frac{\sigma^2}{\alpha_i^2{}_{(OLS)}}$ Since the value of σ^2 is usually unknown, we use the estimated value $\hat{\sigma}^2$. Therefore, when the matrix K satisfies:

$$\hat{K}_i = \frac{\hat{\sigma}^2}{\alpha_i^2{}_{(OLS)}} = \text{diag}\left(\frac{\hat{\sigma}^2}{\alpha_1^2{}_{(OLS)}}, \frac{\hat{\sigma}^2}{\alpha_2^2{}_{(OLS)}}, \ldots, \frac{\hat{\sigma}^2}{\alpha_p^2{}_{(OLS)}}\right)$$

Then the MSE of generalized ridge regression attains the minimum value .

The original form of generalized ridge regression estimator can be converted back from the canonical form by $b_{(GRR)} = G\alpha_{(GRR)}$ ...(2.38)

2.5 Principal Components Regression

Ridge regression was offered as a technique which attempted to overcome the multicollinearity problem. An alternative procedure known as principal components approach ,was first proposed by Harold Hoteling (1933).

In order to obtain a good realization of this approach let us proceed our discussion with the case of two predictors $x_1 and\ x_2$.If these predictors are correlated then the matrix X will not be orthogonal consequently, this will complicate the interpretation of the effects of x_1 and x_2 on the response variable y [5][3].
From the geometric point of view, let us rotate the coordinate axis so that in the new system, the independent variables are orthogonal. Moreover, let us make the rotation so that the first axis lies in the direction of the largest variation in the data, the second axis lies in the direction of the second largest variation in the data [44].
These rotated directions ($Z_1$and Z_2 say in our two predictors case) are simply linear combinations of the original predictors.

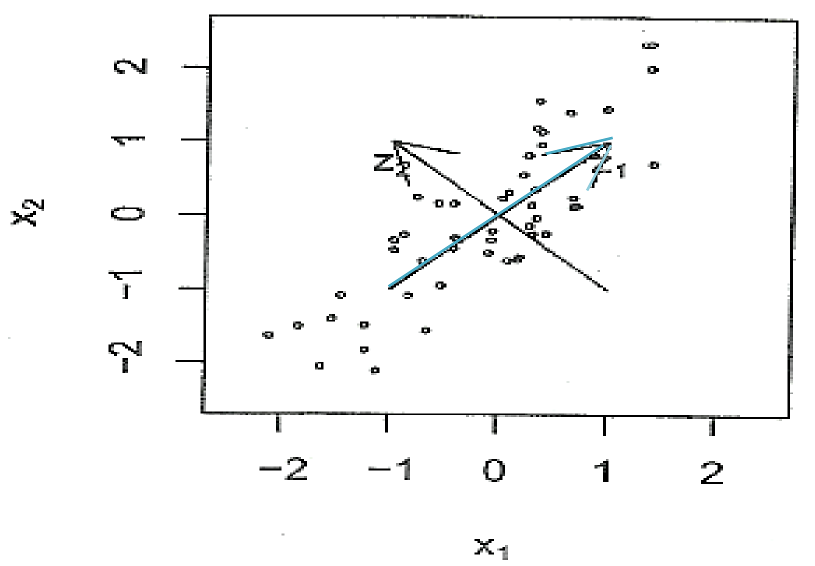

Figure (2-1)

Original independent variable are X_1 and X_2, Principal Components are $Z_1$and Z_2

We now illustrate how these directions can be calculated. Using singular value decomposition then X = H $\Lambda^{\frac{1}{2}}$ G' where each of H, Λ , G is defined earlier

$X'X = G\ \Lambda^{\frac{1}{2}}\ H'H\ \Lambda^{\frac{1}{2}}\ G' = G\ \Lambda G'$

Since G is orthogonal matrix then the general linear regression model y = Xβ + ε can be rewritten as follows:

$$y = X\,GG'\,\beta + \varepsilon = Z\alpha + \varepsilon \qquad ...(2.39)$$

Where Z= XG and $\alpha = G'\beta$

Hence

$$Z'Z = G'X'XG = G'(G\,\Lambda\,G')G = \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_P)$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_P > 0$ are the eigenvalues of X'X. The columns of G are the eigenvectors of X'X and the columns of Z are the principal components of X and these are orthogonal to each other [44][9][23].

Thus, the procedure creats a set of artificial variables $z_j^{'s}$ from the original $X_j{}^{'s}$ via a linear transformation Z = XG in such a way that the Z vectors are orthogonal to each other. The Z_j corresponding to the largest λ_j value is called the principal component and it explains the largest proportion of the variation in the standardized dataset.

Further, $z_j^{'s}$ explain smaller and smaller until all variation is explained. Typically, one does not use all the $z_j^{'s}$ but follows some type of selection rule. No universal rule is presented for selecting the components. Some statisticians use the rule that only eigenvalues greater than 1 are of interest. Other statisticians suggested that the components might be computed until some arbitrarily large proportion (may be 0.75 or more) of the variances has been explained the OLS estimator of α is given as:

$$\hat{\alpha} = (Z'Z)^{-1}Z'y = \Lambda^{-1}G'X'y \qquad ...(2.40)$$

Assuming that the first q (q < p) principal components are selected, then the reduced estimator can be written as follows:

$$\hat{\alpha}_q = (Z_q{}'Z_q)^{-1}Z'_q\,y = \Lambda_q^{-1}G'_q\,X'y \qquad ...(2.41)$$

Where $Z_q = XG_q$, G_q is the matrix of the first q eigenvectors of X'X and Λ_q is the diagonal matrix contains the first q eigenvalues of X'X.

To find the principal component estimator of the regression coefficients in terms of the original variables, by solving $\alpha = G'\beta$ for β to get $\beta = G\alpha$ that is because G is orthogonal matrix. Let b_{PC} denoted to the principal component estimator of β then

$$b_{PC} = G\hat{\alpha}$$

If q principal components are selected then

$$b_{(PC)q} = G_q \hat{\alpha}_q = G_q \Lambda_q^{-1} G'_q X' y \qquad \text{...(2.42)}$$

Chapter Three
The Experimental and The Practical side

The Experimental Side

3.1 Introduction

Whereas, simulation in a narrow sense (also referred to as stochastic simulation) is defined as experimenting with the model over time, it includes sampling stochastic variates from probability distribution. Usually, simulation is regarded as " method of last resort " to be employed when every other approach else has failed. The development of software's and the used techniques have made the simulation one of the most popular and accepted tools for researchers in the system structure and analysis.

simulation approach has a numerous fields of applications. For instance, conducting and analyzing manufacturing systems, evaluating military weapons systems or their logistical requirements and analyzing economic systems.

The main purpose of this chapter is to employ the results obtained from the simulation study which assess the performance of some ridge regression methods as well as the principle component method to the data obtained from Tagi gas distribution plant explained in the practical side of this chapter.

In this section, we will discuss the simulation study that compares the ridge estimators with principal component estimator under several degrees of multicollinearity, especially when correlation level between explanatory variables ($\rho = 0.70, 0.80, 0.90, 0.95$). We consider four different ordinary ridge estimators corresponding to four different values of ridge parameter k .

The values of k are k_{HKB}, k_{LW}, k_{Bayes} as well as our proposed value of the ridge parameter which we denote it by k_{CN} .

The several values of k were calculated by using equations (2.30),(2.31) ,(2.32), and (2.33).

Moreover, the generalized ridge regression estimator and principal component estimators were obtained from equation (2.36) and (2.42) respectively .Since the performance of different estimators is influenced by the sample size , we choose four types of samples, small of size 10 , median of size 40 and large of size 100 and very large of size 200. The error terms were generated at different levels of standard deviations, In particular, at ($\sigma = 5, 10, 20 \, and \, 25$).

The mean square error (MSE) is used as a measure to assess the performance of the stated methods.

3.1.1 Study Design

In our experimental study, we aim to assess and compare the performance between the several estimation methods that was already stated. The simulation technique was used for this situation and the experiment was repeated 1000 time.
Moreover , we assume that the error term is normally distributed as N(0, σ^2).The random variables were generated according to the equations

```
x1=sqrt((1-row^2)).*x(:,1)+row*x(:,1);
x2=sqrt((1-row^2)).*x(:,2)+row*x(:,2);
x3=sqrt((1-row^2)).*x(:,3)+row*x(:,3);
x4=sqrt((1-row^2)).*x(:,4)+row*x(:,4);
xi=[x1 x2 x3 x4]
```

The measurment MSE is used as

MSE = sqrt(sum(y-yhad).^2/(n*r));

when r = 1000

Table (3-1): The values of MSE at ρ = 0.70

n	Method	Standard deviation σ			
		5	10	20	25
10	PC	1.4043e-017	2.1065e-017	2.1065e-017	7.3728e-017
	G_{RR}	0.0197	0.0227	0.0233	0.0233
	$\hat{k}_{HKB}$	0.0229	0.0238	0.0235	0.0233
	$\hat{k}_{LW}$	0.0024	0.0024	0.0024	0.0024
	$\hat{k}_{Bayes}$	0.0039	0.0039	0.0039	0.0039
	$\hat{k}_{CN}$	7.3728e-017	7.3728e-017	7.3728e-017	1.4043e-017
40	PC	1.7115e-017	3.9497e-018	2.1943e-01	6.5828e-018
	G_{RR}	0.0126	0.0125	0.0125	0.0125
	$\hat{k}_{HKB}$	0.0764	0.0747	0.0742	0.0742
	$\hat{k}_{LW}$	0.0031	0.0031	0.0031	0.0031
	$\hat{k}_{Bayes}$	0.1128	0.1134	0.1136	0.1136
	$\hat{k}_{CN}$	0.0037	5.7051e-018	7.5922e-017	1.0050e-016
100	PC	8.8818e-018	6.8834e-017	6.6613e-018	1.1102e-017
	G_{RR}	0.0078	0.0078	0.0078	0.0078
	$\hat{k}_{HKB}$	0.0355	0.0303	0.0273	0.0268
	$\hat{k}_{LW}$	4.3631e-004	4.3625e-004	4.3623e-004	4.3623e-004
	$\hat{k}_{Bayes}$	0.0436	0.0438	0.0439	0.0439
	$\hat{k}_{CN}$	0.0012	3.8272e-004	3.9968e-017	2.6645e-017
200	PC	4.7495e-017	9.0280e-017	8.5963e-017	1.9116e-016
	G_{RR}	0.0054	0.0054	0.0054	0.0054
	$\hat{k}_{HKB}$	0.0525	0.0499	0.0490	0.0489
	$\hat{k}_{LW}$	3.4149e-004	3.4148e-004	3.4148e-004	3.4148e-004
	$\hat{k}_{Bayes}$	0.0732	0.0734	0.0735	0.0735
	$\hat{k}_{CN}$	8.5901e-004	5.1013e-004	7.6542e-017	1.8723e-016

Table (3-2): The values of MSE at ρ = 0.80

n	Method	Standard deviation σ			
		5	10	20	25
10	PC	2.1065e-017	4.9152e-017	3.1598e-017	1.4043e-017
	G_{RR}	0.0197	0.0227	0.0233	0.0233
	$\hat{k}_{HKB}$	0.0228	0.0238	0.0235	0.0233
	$\hat{k}_{LW}$	0.0024	0.0024	0.0024	0.0024
	$\hat{k}_{Bayes}$	0.0039	0.0039	0.0039	0.0039
	$\hat{k}_{CN}$	4.2130e-017	4.2130e-017	4.2130e-017	5.9684e-017
40	PC	7.4605e-018	1.2727e-017	4.8713e-017	2.6770e-017
	G_{RR}	0.0126	0.0125	0.0125	0.0125
	$\hat{k}_{HKB}$	0.0764	0.0747	0.0742	0.0742
	$\hat{k}_{LW}$	0.0031	0.0031	0.0031	0.0031
	$\hat{k}_{Bayes}$	0.1128	0.1134	0.1136	0.1136
	$\hat{k}_{CN}$	0.0038	4.6957e-017	2.1943e-018	1.1279e-016
100	PC	6.6613e-018	6.6613e-018	2.8866e-017	4.8850e-017
	G_{RR}	0.0078	0.0078	0.0078	0.0078
	$\hat{k}_{HKB}$	0.0356	0.0304	0.0273	0.0268
	$\hat{k}_{LW}$	4.3631e-004	4.3625e-004	4.3623e-004	4.3623e-004
	$\hat{k}_{Bayes}$	0.0436	0.0438	0.0439	0.0439
	$\hat{k}_{CN}$	0.0012	3.9946e-004	2.4425e-017	7.3275e-017
200	PC	1.5308e-017	2.2766e-017	7.3009e-017	5.4561e-017
	G_{RR}	0.0054	0.0054	0.0054	0.0054
	$\hat{k}_{HKB}$	0.0525	0.0499	0.0490	0.0489
	$\hat{k}_{LW}$	3.4149e-004	3.4148e-004	3.4148e-004	3.4148e-004
	$\hat{k}_{Bayes}$	0.0732	0.0734	0.0735	0.0735
	$\hat{k}_{CN}$	8.5901e-004	5.1794e-004	3.0617e-017	2.3787e-016

Table (3-3): The values of MSE at ρ = 0.90

n	Method	Standard deviation σ			
		5	10	20	25
10	PC	7.0217e-018	1.0533e-017	2.1065e-017	3.8619e-017
	G_{RR}	0.0192	0.0226	0.0232	0.0233
	$\hat{k}_{HKB}$	0.0227	0.0238	0.0235	0.0234
	$\hat{k}_{LW}$	0.0024	0.0024	0.0024	0.0024
	$\hat{k}_{Bayes}$	0.0039	0.0039	0.0039	0.0039
	$\hat{k}_{CN}$	1.0533e-017	6.6706e-017	6.6706e-017	7.0217e-018
40	PC	2.3259e-017	4.1691e-017	2.1943e-018	4.3447e-017
	G_{RR}	0.0126	0.0125	0.0125	0.0125
	$\hat{k}_{HKB}$	0.0766	0.0748	0.0742	0.0742
	$\hat{k}_{LW}$	0.0031	0.0031	0.0031	0.0031
	$\hat{k}_{Bayes}$	0.1127	0.1134	0.1136	0.1136
	$\hat{k}_{CN}$	0.0042	2.7648e-017	4.9591e-017	6.9778e-017
100	PC	8.8818e-018	4.4409e-018	8.8818e-018	7.3275e-017
	G_{RR}	0.0078	0.0078	0.0078	0.0078
	$\hat{k}_{HKB}$	0.0359	0.0307	0.0274	0.0269
	$\hat{k}_{LW}$	4.3632e-004	4.3625e-004	4.3623e-004	4.3623e-004
	$\hat{k}_{Bayes}$	0.0435	0.0438	0.0439	0.0439
	$\hat{k}_{CN}$	0.0013	4.7300e-004	2.8866e-017	4.8850e-017
200	PC	6.2019e-017	6.2411e-017	3.6112e-017	4.9065e-017
	G_{RR}	0.0054	0.0054	0.0054	0.0054
	$\hat{k}_{HKB}$	0.0528	0.0500	0.0490	0.0489
	$\hat{k}_{LW}$	3.4150e-004	3.4148e-004	3.4148e-004	3.4148e-004
	$\hat{k}_{Bayes}$	0.0731	0.0734	0.0735	0.0735
	$\hat{k}_{CN}$	8.8349e-004	5.5233e-004	1.7153e-016	9.8131e-018

Table (3-4): The values of MSE at ρ = 0.95

n	Method	Standard deviation σ			
		5	10	20	25
10	PC	2.8087e-017	2.1065e-017	5.9684e-017	7.0217e-017
	G_{RR}	0.0185	0.0225	0.0232	0.0233
	$\hat{k}_{HKB}$	0.0225	0.0237	0.0235	0.0235
	$\hat{k}_{LW}$	0.0024	0.0024	0.0024	0.0024
	$\hat{k}_{Bayes}$	0.0039	0.0039	0.0039	0.0039
	$\hat{k}_{CN}$	1.0533e-016	7.6210e-17	7.6120e-17	8.7771e-017
40	PC	1.0094e-017	1.6238e-017	8.3382e-018	1.1849e-017
	G_{RR}	0.0126	0.0125	0.0125	0.0125
	$\hat{k}_{HKB}$	0.0768	0.0749	0.0743	0.0743
	$\hat{k}_{LW}$	0.0031	0.0031	0.0031	0.0031
	$\hat{k}_{Bayes}$	0.1126	0.1134	0.1136	0.1136
	$\hat{k}_{CN}$	0.0047	5.3979e-017	4.8713e-017	3.9058e-017
100	PC	6.6613e-018	1.5543e-017	3.9968e-017	1.3323e-017
	G_{RR}	0.0078	0.0078	0.0078	0.0078
	$\hat{k}_{HKB}$	0.0363	0.0311	0.0276	0.0270
	$\hat{k}_{LW}$	4.3633e-004	4.3625e-004	4.3623e-004	4.3623e-004
	$\hat{k}_{Bayes}$	0.0435	0.0438	0.0439	0.0439
	$\hat{k}_{CN}$	0.0013	5.5274e-004	1.7764e-017	4.4409e-017
200	PC	1.5230e-016	4.9458e-017	8.7140e-017	2.0804e-017
	G_{RR}	0.0054	0.0054	0.0054	0.0054
	$\hat{k}_{HKB}$	0.0531	0.0502	0.0491	0.0489
	$\hat{k}_{LW}$	3.4150e-004	3.4148e-004	3.4148e-004	3.4148e-004
	$\hat{k}_{Bayes}$	0.0731	0.0734	0.0735	0.0735
	$\hat{k}_{CN}$	9.0803e-004	3.8924e-004	6.1626e-017	1.1116e-016

3.2 The Practical side

In the practical side of this chapter, we apply the procedures discussed earlier employing the data obtained from Tagi gas filling company during the period (2008-2016). The company is one of the formations of the oil ministry that had set up in (1967).the company is linked to more than 250 gas filling plants in Baghdad and other provinces. The company produces each of liquid gas from propane gas mixture and liquid butane gas, as well as the production of some solvents such (as hexane) to meet the need for business activity.

In our study we wish to determine the effect of four explanatory variables X_1, X_2, X_3 and X_4 on the response variable Y. Where Y it represents the annual output of liquid gas cylinders, and the explanatory variables $X_1, X_2, X_3 and\ X_4$ refer to craftsmen, administrators, technicians and engineers respectively. We assume that the explanatory variables and the response variable represented according to the linear model from as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \qquad (3.1)$$

Table (3-5)

Values of four explanatory variables X_1, X_2 , $X_3 and\ X_4$ and the response variable Y

Y	X_1	X_2	X_3	X_4
29024876	2186	490	1673	312
29024876	2184	464	1673	325
28259383	2397	510	1836	357
31691496	2552	544	1955	380
32655027	2575	549	1973	383
33691061	2828	604	2166	421
35441678	2787	593	2135	415
36872615	2929	624	2244	436
39256145	3297	702	2524	490

The table below represents descriptive statistics for the explanatory variables and the dependent variable.

Table (3-6)

Descriptive Statistics

variable	N	Minimum	Maximum	Mean	Standard .Deviation
Y	9	28259383	39256145	32879684.11	3813683.112
X_1	9	2184	3297	2637.222	363.51127
X_2	9	464	702	564.4444	74.2460
X_3	9	1673	2524	2019.889	278.35249
X_4	9	312	490	391	56.16939

In table below, the correlation matrix is displayed which involve the correlation coefficients between the explanatory variables themselves and between each explanatory variable and the response variable Y.

Table (3-7)

Matrix of Correlation coefficients

	Y	X_1	X_2	X_3	X_4
Y	1.0000	0.9578	0.9596	0.9578	0.9498
X_1	0.9578	1.0000	0.9951	1.0000	0.9974
X_2	0.9596	0.9951	1.0000	0.9949	0.9855
X_3	0.9578	1.0000	0.9949	1.0000	0.9976
X_4	0.9498	0.9974	0.9855	0.9976	1.0000

The table below ,the eignvalues and condition numbers of the correlation matrix are represented.

Table (3-8)

Analyses of eigenvalues for correlation and matrix condition numbers

Eigen value	Condition numbers
31.882165	1
0.117604	271.0976
0.000219	145580.7
0.000009	3542462.8

The eigenvectors of the correlation matrix are give in table (3-9)

Table (3-9)

Eigenvectors of correlation matrix

X_1	X_2	X_3	X_4
-0.5670	0.0527	1.08514	0.5009
0.4087	-0.7635	-1.11670	0.4988
-0.4183	0.0697	1.10856	0.5009
0.5800	0.6398	-1.11163	0.4994

3.2.1 The Farrar-Glauber test :-

As we mentioned earlier, the Farrer-Glouber test is used to detect the existence of multicollinearty problem. For our dataset the test is applied as follows.

$$D = \begin{bmatrix} 1 & r_{X_1X_2} & r_{X_1X_3} & r_{X_1X_4} \\ r_{X_2X_1} & 1 & r_{X_2X_3} & r_{X_2X_3} \\ r_{X_3X_1} & r_{X_3X_2} & 1 & r_{X_3X_4} \\ r_{X_4X_1} & r_{X_4X_2} & r_{X_4X_3} & 1 \end{bmatrix}$$

Where we have a specific matrix

$$D = \begin{vmatrix} 1.0000 & 0.9951 & 1.0000 & 0.9974 \\ 0.9951 & 1.0000 & 0.9949 & 0.9855 \\ 1.0000 & 0.9949 & 1.0000 & 0.9976 \\ 0.9974 & 0.9855 & 0.9976 & 1.0000 \end{vmatrix}$$

By applying the equation (2.9) ,

the chi-square χ_0^2 = 157.3983

The tabulated value with 6 degrees of freedom and $\alpha = 0.05$ level of significant was found to be (1.64) ,hence, we reject H_0 since $\chi_0^2 = 157.3983 > 1.64$, consequently, the problem of multicollinearity is exist among the explanatory variables.

3.2.2 The Ridge regression analysis

The next step in our practical study represented by applying different types of ridge regression estimators in order to deal with the multicollinearity problem. For this situation we start with the generalized ridge regression estimator, b_{GRR}.
By employing equation(2.36), we found that

$$b_{GRR} = \begin{bmatrix} 0.7776 \\ -4.5126 \\ 12.2599 \\ -7.6087 \end{bmatrix}$$

The analysis of variance calculations are summarized in the following ANOVA table

Table (3-10)
ANOVA in case of G_{RR}

Source	d.f	Sum of squares	Mean square	F test
Regression	4	7.446678	1.861669	13.458148
Residual	4	0.553321	0.138330	
Total	8	8		

Hoerl, Kennard and Baldwin estimator for the ridge parameter $\hat{k}_{HK}$ is obtained by using equation (2.30). It was (0.00000303) . Accordingly, the ridge regression estimator is

$$b_{k_{HKB}} = \begin{bmatrix} -78.3946 \\ -12.6418 \\ 113.7304 \\ -21.8538 \end{bmatrix}$$

The analysis of variance calculations are summarized in the following ANOVA table

Table (3-11)
ANOVA in case of $b_{k_{HKB}}$

Source	d.f	Sum of squares	Mean square	F test
Regression	4	7.714816	1.928704	27.052126
Residual	4	0.285183	0.071295	
Total	8	8		

Lawless and Wang estimator for the ridge parameter $\hat{k}_{LW}$ is obtained by using equation (2.31). It was (0.01325400) . Accordingly, the ridge regression estimator is

$$b_{k_{LW}} = \begin{bmatrix} 0.3264 \\ 0.4550 \\ 0.4513 \\ -0.2740 \end{bmatrix}$$

The analysis of variance calculations are summarized in the following ANOVA table

Table (3-12)
ANOVA in case of $b_{k_{LW}}$

Source	d.f	Sum of squares	Mean square	F test
Regression	4	7.370573	1.842643	11.709980
Residual	4	0.629426	0.157356	
Total	8	8		

Applying the Bayesian approach stated earlier the value of k was $\hat{k}_{Bayes} = 0.21783861$ obtained by using equation (2.32).
Hence, the ridge regression estimator is

$$b_{k_{Bayes}} = \begin{bmatrix} 0.2359 \\ 0.3699 \\ 0.2408 \\ 0.1067 \end{bmatrix}$$

The analysis of variance calculations are summarized in the following ANOVA table

Table (3-13)
ANOVA in case of $b_{k_{Bayes}}$

Source	d.f	Sum of squares	Mean square	F test
Regression	4	7.304121	1.826030	10.496257
Residual	4	0.695878	0.173969	
Total	8	8		

By applying our proposed method and from equation (2.33), the value of k was found $\hat{k}_{CN}$ = 0.000013. Accordingly, the ridge regression estimator equation is

$$b_{k_{CN}} = \begin{bmatrix} -38.8359 \\ -10.1881 \\ 66.9740 \\ -17.0856 \end{bmatrix}$$

The analysis of variance calculations are summarized in the following ANOVA table

Table (3-14)
ANOVA in the case of $b_{k_{CN}}$

Source	d.f	Sum of squares	Mean square	F test
Regression	4	7.603587	1.900896	19.180991
Residual	4	0.396412	0.099103	
Total	8	8		

3.2.3 The Principal component method :-

An alternative popular approach that is widely used to remedy the multicollinearity problem is the principal component approach. For our dataset the following calculations are performed.

Table (3-15)

Principal Component Analysis

Variable	PC1	PC2	PC3	PC4
X_1	-0.567	0.053	1.085	0.510
X_2	0.409	-0.763	-1.117	0.499
X_3	-0.418	0.070	1.109	0.510
X_4	0.580	0.640	-1.112	0.500
Eigen Value	31.882165	0.117604	0.000219	0.000009
Proportion	0.99631765625	0.00367512500	0.00000684375	0.00000028125
Cumulative	0.996	0.999	1.000	1.000

Obviously, the first two components, with larger eigenvalues explain 99.9 % of the total variance. Hence, only the first two principal components are introduced into analysis. We have to find Z= XG* where G* is a (4×2) matrix obtained from the first two columns of principal components matrix, is found :

$$Z = \begin{bmatrix} -0.000013989 & 0.000523899 \\ 0.000503686 & 0.006293600 \\ 0.000996264 & -0.000208627 \\ 0.000199344 & -0.004299124 \\ -0.001374345 & 0.000120307 \\ 0.001018815 & -0.010413387 \\ -0.000878755 & 0.001565974 \\ -0.001711530 & -0.000660762 \\ 0.001260510 & 0.007078122 \end{bmatrix}$$

By applying the principal of ordinary least squares to fit the model Y =Z α+ε ,Where α =G*'β we obtain :

$$\hat{\alpha} = \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = (Z'Z)^{-1}Z'Y = \begin{bmatrix} -183.2070 \\ 21.2568 \end{bmatrix}$$

Thus the regression equation is:

$\hat{y} = \hat{\alpha}_1 pc_1 + \hat{\alpha}_2 pc_2 = -183.2070pc_1 + 21.2568\ pc_2$

In terms of the original values we have:

$\hat{y} = -107.3581X_1 - 14.9086X_2 + 147.1196X_3 - 24.9894X_4$

The Variance-covariance matrix is given as

$$\text{var-cov}(\hat{\alpha}) = S^2 (Z'Z)^{-1} = \begin{bmatrix} 2717.2833750 & 0.00000028854 \\ 0.00000028854 & 117.4586773 \end{bmatrix}$$

The analysis of variance calculations are summarized in the following ANOVA table**.**

Table (3-16)

ANOVA in the case of b_{PC}

Source	d.f	Sum of squares	Mean square	F test
Regression	2	7.581678	3.790839	54.372080
Residual	6	0.418321	0.069720	
Total	8	8		

3.2.4 Coefficient of Determination

The following table shows the coefficient of determination obtained from different methods

Table (3-17)

Method	R- Square
PC	94.770975
GRR	93.083484
k_{HKB}	96.435207
k_{LW}	92.132167
k_{Bayes}	91.301517
k_{CN}	95.044821

3.2.5 Conclusions and discussion

From our theoretical, experimental and practical study, we believe that the following points are considerable.

1- The simulation results displayed that the principal components estimator performs better than almost all types of generalized and ordinary ridge regression estimators that are included in the study,under different conditions of multicollinearity levels, sample sizes and different levels of standard deviations of the error terms.

2- As we stated earlier, our proposed method for estimating the ridge parameter depends upon the level of multicollinearity between the explanatory variables. It shows the importance of the condition number as an indicator of the presence of multicollinearity problem. Moreover, the simulation results imply that the ordinary ridge regression estimator based on the proposed ridge parameter $\hat{k}_{CN}$ performs well in the sense of MSE. It seems to be better all other types of ridge regression estimators included in this study whatever the level of multicollinearity, the sample size , or the value of standard deviation is.

3- In our practical study, the Farrer-Glauber test established the existence of multicollinearity problem in our real data set. Many other indicators ensure the presence of this problem, such as the large values of correlation coefficients between some explanatory variables (close to 1) as it is shown in table (3-8) and a very small eigenvalues(near zero) which imply a very large values of condition numbers as it is displayed in table (3-9).

4-The analysis of variance tables demonstrate that the principal component estimator is superior to all types of ridge regression estimators in the sense of MSE as it is shown in table (3-16).

5- With regard to different types of ridge regression estimators, the ANOVA tables displayed that the ordinary ridge regression estimator based on $\hat{k}_{HKB}$ is the best , followed by the ordinary ridge regression estimator based on our proposed ridge parameter $\hat{k}_{CN}$ in the sense of MSE as it is shown in tables (3-11) and (3-14) respectively.

6- At 0.05 level of significant, it was found that the observed values of the F statistic from all ANOVA tables is greater than the corresponding tabulated values with p and n-p-1 degrees of freedom. This implies that the null hypothesis $H_0: \beta_j = 0 , j = 1,2,3,4$ is rejected. Consequently, a statistically regression equations have been obtained and the studied variables have an explanatory power.

7- As mentioned earlier, the coefficient of determination denoted by R^2 is defined to be the percent of variations in the response variable that can be explained by the regression equation. With regard to our practical study, the regression equation based on $\hat{k}_{HKB}$ explain approximately 96.4 % of variations in the response variable, while the regression equation based on our proposed ridge parameter $\hat{k}_{CN}$ explain about 95.04 % of variations in the response variable, as it is shown in table (3-17).

8- For the purpose of future works, many other estimators can be employed to overcome the multicollinearity problem such as the generalized inverse estimator, Liu estimator, the restricted ridge regression estimator and Jackknife ridge regression estimator.

References

[1] الحيالي، طالب حسن نجم " مقدمة في القياس الاقتصادي " (1991) دار الكتب للطباعة والنشر- جامعة الموصل.

[2]الراوي، خاشع محمود " المدخل الى تحليل الانحدار " (1987) جامعة الموصل - كلية الزراعة والغابات - دار الكتب للطباعة والنشر.

[3]جبريل، محمد سليمان محمد " التعدد الخطي أسبابة وتأثيراته والمعالجة بانحدار الحافة وانحدار المركبات الرئيسية مع التطبيق على البيانات الافتراضية" (2014) أطروحة دكتوراه جامعة السودان للعلوم والتكنولوجيا.

[4]كاظم، أموري هادي، الصفاوي، صفاء يونس، "توظيف البيانات العرضية في معالجة مشكلة التعدد الخطي"، المجلة العراقية للعلوم الاحصائية، المجلد 1، ص 1-11، (2001).

[5]كاظم، اموري هادي، مسلم، باسم شليبه. (2002) "القياس الاقتصادي المتقدم: النظرية والتطبيق"، مطبعة الطيف، بغداد، العراق.

[6]محبوب، عادل عبد الغني " أصول الاقتصاد القياسي - النظرية والتطبيق " (1998) شركة الاعتدال للطباعة بغداد.

[7]محمد، رواء صالح " استخدام انحدار الحرف (Ridge) لدراسة أثر بعض العوامل على المؤشر العام لسوق الاوراق المالية " (2010) الجامعة المستنصرية - كلية الادارة واقتصاد- مجلة القادسية للعلوم الادارية والاقتصادية - المجلد 12.

[8] Akdeniz, F. and Kaciranlar,S.(2001); "More on the new biased estimator in linear regression " The Indian Journal of Statistics Vol.63. SeriesB, pt.3,pp321-325

[9] Al-Hassan, Y.M. and Al-Kassab, M.M.(2009); " A Monte Carlo Comparison between Ridge and Principal Components Regression Methods", Applied Mathematical Sciences, Vol. 3, no. 42, 2085 – 2098

[10] Al- Somahi, A.A., Mousa, S. and Al-Turk,L.I.(2015); "Some New Proposed Ridge Paremeters for the Logistic Regression Model" International Journal of Research in Applied Natural and Social Sciences, Vol.3 lssue 1 pp 67-82

[11] Asar,Y. Karaibrahimoglu, A. and Genco, A.(2014); " Modified Ridge Regression Parameters: A Comparative Monte Carlo Study" Hacettepe Journal of Mathematics and Statistics, Vol.43(5) pp 827-841

[12] Batah, F. Sh. M.(2011); "A New Estimator By Generalized Modified Jackknife Ridge Regression Estimator" Journal of Basrah Researches ((Sciences)) Vol.37, No.4,pp 128-149

[13] Cassela, G.(1980);"Minimax Ridge Regression Estimation "The Annals of Statistics Vol.8.No.5,pp1036-1056

[14] Cassela, G.(1985);"Condition Numbers and Minimax Ridge Regression Estimators" Journal of the American Statistical Association , Vol.80,No,391,pp753-758

[15] Chatterjee, S. and Prices, S.(1977); "Regression Analysis by Example ". John wiley and sons. New York.

[16] Dereny, M. El. and Rashwan, N.I.(2011); " Solving Multicollinearity Problem Using Ridge Regression Models" Int. J. Contemp. Math. Sciences, Vol.6,No.12, 585-600

[17] Dorugade, A.V. (2010); " Alternative Method for Choosing Ridge Parameter for Regression", Applied Mathematical Sciences, Vol. 4, no. 9, 447 - 456.

[18] Dorugade, A.V.(2014); " On Comparison of Some Ridge Parameters in Ridge Regression" Journal of Applied Statistics, Vol (15-1)
[19] Drapper, N. R. and Smith, H.(1981); "Applied Regression Analysis", Second Edition, John Wiley and Sons, New York.

[20] Durugade, A .v. and Kashid ,D.N.(2010); "Alternative Method for Choosing Ridge parameter for Regression" Applied Mathematical Sciences,Vol.4,No.9,447-456

[21] El-Dereny, M. and Rashwan, N.I.(2011); "Solving Multicollinearity Problem: Using Ridge Regression Models", Int. J. Contemp. Math. Sciences, , Vol. 6, no. 12, 585 – 600

[22] Fitrianto, A. and Yik, L.C.(2014); " Performance of Ridge Regression Estimator Methods on Small Sample, Size by Varying Correlation Coefficients Simulation Study" Journal of Mathematics and Statistics, 10(1) pp 25-29

[23] Fujiwara, M. and Minamidani, T. and others (2012); " Principal Components Regression by using Generalized Principal Components Analysis", *Department of Mathematics, Graduate School of Science, Hiroshima University,1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526*
.
[24] Goldstein, M. and Smith, A.F.M.(1974);" Ridge Type Estimation for Regression Analysis" J. Roy . Statist. B Vol.36 pp284-291

[25] Gorgees, H.M. and Ali, B.A.(2013);Ridge Regression Procedure to Remedy the Multicollinearity Problem" Ibn Al- Haitham Journal For pure and Applied Science, Vol.26, No.1,pp 320-327

[26] Gorgees, H.M.(2009);"Using Singular Value Decomposition Approach For Estimating the Ridge parameter" Journal of Economic and Administrative Science,Vol.53,No.15,pp1-10

[27] Gorgees, H.M.(2010); "Bayesian Analysis of Ridge Regression Problems" Ibn Al- Haitham Journal for pure and Applied Science,Vol.23,No.3,pp253-264

[28] Graybill, F.A.(1976);"Theory and Application of the Linear Model", Wadsworth Publishing Company, INC.USA.

[29] Gunst, R.F. and Mason , R.L. (1977); " Biased Estimation Regression: An Evaluation using Mean Squared Error" Journal of the American statistical Association. Vol. 72.No.359pp616-628

[30] Hoerl, A.E. and Kennard, R.W. (1970); " Ridge Regression Biased Estimation for Non orthogonal problems " Techno metrics, Vol.12.pp55-67

[31] Lindley, D.V. and Smith, A. F. M.(1972); "Bayes Estimators for Linear Models (with Discussions)", J. Roy. Statist. Soc. B 34, pp. 1-41.

[32] Marquardt, D. W. and Snee, R.E.(1975);"Ridge Regression in Practice" Amer.Statist.Vol.12.pp3-19

[33] Maruyama, Y, and Strawderman, W.E.(2005);" A New Class of Generalized Bayes MiniMax Ridge Regression Estimators" The Journal of Statistics, Vol.33,No.4,pp1753-1770

[34] Mayer, L.S. and WillKe, T.A.(1974); " On Biased Estimation in Linear Models " Techno metric Vol.15 pp497-508

[35] Obenchain, R.L.(1975); "Ridge analysis following a preliminary of the shrunken hypothesis", Technometrics, Vol. 17, pp. 431- 445.

[36] Obenchain, R.L.(1978) ; "Good and optimal ridge estimators . Ann. statist .6, 1111, 1121.

[37] Ozkala, M. R.(2009); "A Stochastic restricted ridge regression estimator" Journal of Multivariate Analysis 100 pp 1706 - 1716

[38] Paris, Q.(2001)"Multicollinearity and Maximum entropy estimators"

Economics Bolletin Vol.3,No.11,pp1-9

[39] Pasha, G.R., Shah, M.A. and Ghosia (2004);"Estimation and Analysis of Regression Coefficients When Explanatory Variables are Correlated' Journal of Research (Science) ,Vol.15,No.1,pp97-100

[40] Rao, C.R.(1965); "Linear Statistical Inference and its Applications", New York, Wiley.

[41] Salh, S.M.(2014); "Using Rigde Regression Model to solving Multicollinearity Problem", International Journal of Scientific & Engineering Research, Volume 5, Issue 10.

[42] Srivastava, M.S.(2002); "Methods of Multivariate Statistics", Wiley, New York.

[43] Strawderman, W.E. (1978); "Minimax Adaptive Generalized Ridge Regression Estimators "Journal of the American Statistical Association. Vol. 73.,No . 363 pp623-626

[44] Vajargah, K.F.(2013); "Comparing Ridge Regression and Principal Components Regression by Monte Carlo Simulation Based on MSE", Journal of Computer Science & Computational Mathematics, Volume 3, Issue 2.

المستخلص

تعد مشكلة التعدد الخطي أحدى أهم المشاكل في تحليل الانحدار لما لها من أثار غير مرغوبة على مقدرات المربعات الصغرى. أن انحدار الحرف هو احد الحلول الأكثر شيوعا لهذه المشكلة. ففي هذه الحالة فان معلمة الحرف (أو ثابت التحيز) لها دور مهم في تقدير المعلمات. لقد أقترح العديد من علماء الإحصاء طرائق مختلفة لاختيار معلمة الحرف. في هذه الرسالة حاولنا أن يكون لنا أضافه خاصة بنا في هذا المجال. على هذا الأساس اقترحنا طريقه جديدة لإيجاد معلمة الحرف ومقارنتها مع الطرائق المقترحة سابقا من قبل باحثين آخرين من خلال المحاكاة والدراسة التطبيقية وقد تبين لنا أن الطريقة المقترحة هي طريقة مقبولة تماما باعتماد متوسط مربعات الخطأ كمعيار للمفاضلة.

أن الحل البديل الشائع الأخر لمشكلة التعدد الخطي والذي تمت تغطيته أيضا في هذه الرسالة وهو انحدار المركبات الرئيسية. ففي هذا الأسلوب بدلاً من استخدام المتغيرات التوضيحية الأصلية والغير متعامدة في تحليل الانحدار فان مركباتها الرئيسية هي التي تستخدم في التحليل والتي تكون متعامدة الواحدة مع الأخرى.

تشتمل هذه الرسالة على جزئيين رئيسيين وعلى وجه الخصوص، الجزء النظري والجزء العملي الذي يشتمل على الجانب التجريبي والجانب التطبيقي.قد تم استخدام البرامج الإحصائية (MATLAP & SSPS) لإنجاز الحسابات المطلوبة.

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بغداد
كلية التربية للعلوم الصرفة / ابن الهيثم
قسم الرياضيات

مقارنه بين مختلف طرق انحدار الحرف مع التطبيق

رسالة

مقدمة إلى كلية التربية للعلوم الصرفة / ابن الهيثم-جامعة بغداد

وهي جزء من متطلبات نيل شهادة ماجستير علوم

في الرياضيات

من قبل

فاطمة عاصم مهدي

(**بكلوريوس جامعة بغداد** 2015)

بإشراف

أ. م.د. حازم منصور كوركيس

1438 هـ **2017 م**