Journal of Advance Research in Computer science and Engineering (ISSN: 2456-3552)

DOI: https://doi.org/10.53555/nncse.v4i12.393 **Publication URL: http://nnpub.org/index.php/CSE/article/view/393**

APPLICATION OF LATENT ROOTS REGRESSION TO MULTICOLLINEAR DATA

Mowafaq Muhammed Al-Kassab[1]*, **Dilnas Salah Adeen Younis[2]**

*[*1,2]Department of Statistics & Informatics, College of Mathematics and Computer Sciences, University of Mosul, Mosul, Iraq*

*[*1]Email: Mowafaq2002@yahoo.co.uk, [2]Dilnasyonis@gmail.com*

Corresponding Author: -

Email ID - Mowafaq2002@yahoo.co.uk

Abstract: -

Several applications are based on the assessment of a linear model including a variable y to Predictors $x_1, x_2,..,x_p$. It often occurs that the predictors are collinear which results in a high instability of the model obtained by means of multiple linear regression using least squares estimation. Several alternative methods have been proposed in order to tackle this problem. Among these methods Ridge Regression, Principal Component Regression. We discuss a third method called Latent Root Regression. This method depends on the Eigen values and Eigen vectors of the matrix $A'A$, where A is the matrix of y and $x_1, x_2,..,x_p$. We introduce some properties of latent root regression which give new insight into the determination of a prediction model. Thus, a model may be determined by combining latent root estimators in such a way that the associated mean squared error is minimized. The method is illustrated using three real data sets. Namely: Economical, Medical and Environmental data . According to applications, our new estimators depending on the Latent Root Regression have better performances in the sense of MSE in most of the situations.

Keywords: - *Regression; Multicollinearity; Latent Root; Eigen values; Least Squares; Ridge Regression; Principal Component Regression.*

1- Multiple Linear Regression Model

Any standard study of a particular phenomenon requires the identification of the factors influencing this phenomenon and the formulation of the relationship between these factors in the form of a model that expresses them. The model may represent one or several equations. In terms of a single equation, it may be simple or may be multiple. Common forms of use include a linear form that takes a mathematical form in writing and includes more than one explanatory variables. This model will be used in this research and the general formula for this the model is (Yan & Gang Su, 2009):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1} + \ldots + \beta_p x_{ik} + u_i$$

Where:

y_i:The value of the response variable, i=1,2,3,…,n; n is the size of sample.
β_j:Regression parameters, j=0,1,2,…,p, p is number of explanatory variables.
x_{ij}:The explanatory variable j for observation i, i=1,2,3,…,n; j=1,2,…,p.
u_i: The random error of observation i, i=1,2,3,…,n. The above equation can be expressed in matrix form as:

$$\underline{y} = X\underline{\beta} + \underline{u} \qquad (1)$$

Where:

y: is (n×1) vector of observations of the response variable.

$\bar{X}$: is (n×k); k=p+1, matrix of observations of the explanatory variables whose first column contains the values of the correct one.
$\underline{\beta}$:is (k×1) vector of the parameters to be estimated.
$\underline{u}$:is (n×1) vector of random errors.

In order to estimate the parameters of the model and to ensure that the estimations have desirable properties, there are certain hypotheses that must be met. These assumptions are(Chatterjee& Price, 2000):

1. The relationship between the variables involved is a linear relationship and this means that the variable y is determined as a linear structure of independent variables, X and random variable u.
2. The variance of the random error is σ^2, and
$$\operatorname{cov}(u_i, u_j) = 0 \quad \text{for } i \neq j \text{ Where } i, j = 1,2,3, \ldots, n$$
3. The random error is normal with mean zero and variance σ^2.
4. The matrix X is distributed separately from u, so the covariance is zero, ie:
$$\operatorname{cov}(u_i, X_j) = 0 \quad \text{Where } i = 1,2,3, \ldots, n \ , j = 1,2,3, \ldots, p$$
5. This assumption is concerned with the rank of matrix X where we assume that:
Rank(X) = p > n

This assumption includes:
- The number of observations should be greater than the number of estimated parameters.
- No multicollinearity between the values of the explanatory variables.

2- Least Squares Method

This method is one of the most widely used methods for estimating the parameters of the linear regression model. The least squares estimate of the regression parameters in this method is (Kutner et al., 2005):

$$\underline{\hat{\beta}}_{OLS} = (X'X)^{-1}\left(X'\underline{y}\right) \ldots (2)$$

Here are the properties of this method (Draper & Smith, 1981) (Fisher, 1981&Mason) and (Gunst& Mason, 1980):

1. Linearity: the estimated parameters in this method are linear in terms of the response variables:
$$\underline{\hat{\beta}}_{OLS} = (X'X)^{-1}\left(X'\underline{y}\right) = [(X'X)^{-1}X']\underline{y}$$
2. Unbiased: That is, the expected value of the estimated parameters is equal to its real value:
$$E\left[\underline{\hat{\beta}}_{OLS}\right] = \underline{\beta}$$
3. Variance: The variance of the estimated parameters is minimum
$$\operatorname{Var}\left(\underline{\hat{\beta}}_{OLS}\right) = (X'X)^{-1}\sigma^2 \quad \ldots (3)$$

Where $\sigma^2 = \frac{\underline{y}'\underline{y} - \underline{\hat{\beta}}'_{OLS}X'\underline{y}}{n-p-1}$

3-The Concept of the Multicollinearity in the Regression Model

Multicollinearity is one of the problems that occur in many cases due to the existence of a relationship between the explanatory variables. The existence of the complete multicollinearity between the variables leads to making the matrix

$(X'X)$ not of full rank, ie, its determinant is zero. Thus, it is difficult to find the inverse of this matrix, which means that the regression parameters cannot be estimated using the Least Squares method. The existence of an incomplete but powerful multicollinearity leads to the amplification of the variance and thus the acquisition of inaccurate capabilities (Dounald, 1987) and (Chatterjee et al., 2000).

4- Detecting Multicollinearity

Multicollinearity between the explanatory variables can be detected by many methods (Draper & Smith, 1981) and (Fisher, 1981&Mason) and (Gunst& Mason, 1980):

1. The correlation coefficients matrix: Notice the elements are not diagonal in the correlation matrix$(X'X)$ it is a simple indicator of inference on the existence of multicollinearity between explanatory variables, where the greater the value of these elements than zero this indicates the existence of relations between these variables, whether those relations are positive or negative.
2. Determinant of matrix: The determinant value of the matrix ranges from zero to one, when it is zero ($|X'X| = 0$) this indicates the existence of complete multicollinearity between the variables. On the other hand, if it is one ($|X'X| = 1$) this indicates the nonexistence of multicollinearity. When the value of the determinant is between zero and one there is a multicollinearity of certain degree, that is when ($1>|X'X|>0$), the multicollinearity exists and increases whenever the determinant close to zero (Mason, 1986). $|X'X|$ can be calculated using its latent values and through the following equation:

$$|X'X| = \prod_{i=1}^{p} \lambda_i \quad ; \qquad i = 1,2, \ldots . p$$

 Where λ_i is latent values for $(X'X)$ matrix, when there is a multicollinearity a number of latent values will be small or close to zero leading to a small value for the matrix determinant.

3. Latent values for $(X'X)$ matrix: The calculation of the latent values and vectors is a multicollinearity detection method, if the value of one of the latent values is equal to zero this indicates the presence of complete multicollinearity which indicates the difficulty or impossibility of finding the estimators of the Least Squares because of impossibility of finding the inverse matrix since the determinant will be zero. Latent values are also used to calculate two types of auxiliary statistics in the detection of multicollinearity, they are(Kutner et al., 2005):

- Condition index (CI): It is calculated according to the relationship:

$$CI_j = \sqrt{\frac{\lambda_{max}}{\lambda_j}} \quad ; j = 1,2, \ldots , p$$

Where:
λ_{max}: Is the maximum latent value.
λ_j: *is the* jth latent value.

- Condition number: It is calculated according to the relationship:

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

Where:
λ_{max}: Is the maximum latent value.
λ_{min}: Is the minimum latent value (Belsley et al., 1980) proposed that if CN values between 30 and 100 this was a sign of very high multicollinearity .

5- Solution of Multicollinearity

There are several methods proposed to minimize the effect of multicollinearity, Such as (Fisher & Mason, 1981):

1. Delete the explanatory variables that are associated with other variables in order to get rid of the effects of this link and this deletion process according to certain criteria proposed to delete the specific variables.
2. Add new data to the original data.
3. Use biased estimation methods.

6- Latent Roots Method

This method was proposed in 1973 by Hawkins, the idea of this method is to find the latent roots of the correlation matrix and then to exclude the roots that are not important in the prediction process. The following is a detailed explanation of this method (Mason, 1986):
Correlation matrix is obtained by multiplying the transpose matrix (A) and the same matrix ie:
$R = A'A$
Where:
A: Is the standardized information matrix which contains the standardized values of the response variable and the standardized values of the explanatory variables:

$$A = [\underline{y}^* \quad X^*]$$

Where

$$y_i^* = \frac{y_i - \bar{y}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad ; \quad X_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{\sum_{i=1}^{n}(x_{ik} - \bar{x}_k)^2}}$$

R: the correlation matrix between all variables, it is as follows:

$$R = \begin{bmatrix} 1 & r_{y1} & r_{y2} & r_{y3} & \cdots & r_{yp} \\ r_{1y} & 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{2y} & r_{21} & 1 & r_{23} & \cdots & r_{2p} \\ r_{3y} & r_{31} & r_{32} & 1 & \dots & r_{3p} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{py} & r_{p1} & r_{p2} & r_{p3} & \dots & 1 \end{bmatrix}$$

Latent roots latent values and latent vectors, are obtained according to the following formula:

$$\Lambda = \begin{bmatrix} \lambda_0 & 0 & \dots & 0 \\ 0 & \lambda_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} \gamma_{00} & \gamma_{01} & \cdots & \gamma_{0p} \\ \gamma_{10} & \gamma_{11} & \cdots & \gamma_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p0} & \gamma_{p1} & \cdots & \gamma_{pp} \end{bmatrix} = \begin{bmatrix} \gamma_{00} & \gamma_{01} & \cdots & \gamma_{0p} \\ \underline{\gamma}_0 & \underline{\gamma}_1 & \cdots & \underline{\gamma}_p \end{bmatrix}$$

The estimation of the regression parameters vector by Least square Method that based on the

$$\underline{\hat{\beta}}_{OLS} = -\frac{\sum_{j=0}^{p} \frac{\gamma_{0j}\underline{\gamma}_j}{\lambda_j}}{\sum_{j=0}^{p} \frac{\gamma_{0j}^2}{\lambda_j}} \qquad \text{……(4)}$$

To find the Latent Root Estimators, all the values and vectors that are not significant in the prediction are deleted from the equation (4), the roots that meet the following conditions are deleted:

$$\lambda_j < 1 \quad and \quad |\gamma_{0j}| < 0.5 \quad for \quad j = 0,1,2,\dots,p$$

The remaining latent roots are less thanor equal to p, denoted by q, and estimated by the Latent Root Method are as follows:

$$\underline{\hat{\beta}}_{LR} = -\frac{\sum_{j=0}^{q} \frac{\gamma_{0j}\underline{\gamma}_j}{\lambda_j}}{\sum_{j=0}^{q} \frac{\gamma_{0j}^2}{\lambda_j}} \qquad \text{……(5)}$$

Latent Root estimators have the following properties:

1.Bias: The Latent Root estimator is biased and its bias is:

$$\text{Bais}\left(\underline{\hat{\beta}}_{LR}\right) = \frac{\sum_{j=q+1}^{p} \frac{\gamma_{0j}\underline{\gamma}_j}{\lambda_j}}{\sum_{j=q+1}^{p} \frac{\gamma_{0j}^2}{\lambda_j}} \qquad \text{…..(6)}$$

2. The variance: The variance of the Least Squares estimators in terms of latent roots is:

$$\text{Var}\left(\underline{\hat{\beta}}_{OLS}\right) = \sigma_{OLS}^2 \left[\sum_{j=0}^{p} \frac{\underline{\gamma}_j\underline{\gamma}_j'}{\lambda_j} - \frac{\left(\sum_{j=0}^{p} \frac{\gamma_{0j}\underline{\gamma}_j}{\lambda_j}\right)\left(\sum_{j=0}^{p} \frac{\gamma_{0j}\underline{\gamma}_j'}{\lambda_j}\right)}{\sum_{j=0}^{p} \frac{\gamma_{0j}^2}{\lambda_j}}\right] \qquad \text{……(7)}$$

The variance of the latent Root estimators is

$$\text{Var}\left(\underline{\hat{\beta}}_{LR}\right) = \sigma_{LR}^2 \left[\sum_{j=0}^{q} \frac{\underline{\gamma}_j\underline{\gamma}_j'}{\lambda_j} - \frac{\left(\sum_{j=0}^{q} \frac{\gamma_{0j}\underline{\gamma}_j}{\lambda_j}\right)\left(\sum_{j=0}^{q} \frac{\gamma_{0j}\underline{\gamma}_j'}{\lambda_j}\right)}{\sum_{j=0}^{q} \frac{\gamma_{0j}^2}{\lambda_j}}\right] \qquad \text{……(8)}$$

The variance of the ith estimators is

$$\operatorname{Var}\left(\hat{\underline{\beta}}_{iLR}\right) = \sigma_{LR}^2\left[\sum_{j=0}^{q}\frac{\gamma_{ij}^2}{\lambda_j} - \frac{\left(\sum_{j=0}^{q}\frac{\gamma_{oj}\gamma_{ij}}{\lambda_j}\right)^2}{\sum_{j=0}^{q}\frac{\gamma_{ij}^2}{\lambda_j}}\right]$$

3.Mean Squares Error: Since the Latent Root estimator is biased, this makes the mean squares error as follows:

$$\operatorname{MSE}\left(\hat{\underline{\beta}}_{LR}\right) = \sigma_{LR}^2\sum_{i=1}^{p}\left[\sum_{j=0}^{q}\frac{\gamma_{ij}^2}{\lambda_j} - \frac{\left(\sum_{j=0}^{q}\frac{\gamma_{oj}\gamma_{ij}}{\lambda_j}\right)^2}{\sum_{j=0}^{q}\frac{\gamma_{oj}^2}{\lambda_j}}\right] + \sum_{i=1}^{p}\left[\frac{\sum_{j=q+1}^{p}\frac{\gamma_{oj}\gamma_{ij}}{\lambda_j}}{\sum_{j=q+1}^{p}\frac{\gamma_{oj}^2}{\lambda_j}}\right]^2$$

Basilevsky in 1994 suggested a approximation of the mean squares error for the Latent Root estimator as:

$$\operatorname{MSE}\left(\hat{\underline{\beta}}_{LR}\right) \cong \sigma_{LR}^2\sum_{i=1}^{p}\left[\sum_{j=0}^{q}\frac{\gamma_{ij}^2}{\lambda_j} - \frac{\left(\sum_{j=0}^{q}\frac{\gamma_{oj}\gamma_{ij}}{\lambda_j}\right)^2}{\sum_{j=0}^{q}\frac{\gamma_{oj}^2}{\lambda_j}}\right] + \left(\underline{\gamma}_0'\hat{\underline{\beta}}_{LR}\right)^2 \qquad \ldots\ldots (9)$$

7- Application Part

Two sets of data are analyzed using the theoretical part. These two sets are: (Pasha & Ali Shah, 2004). This group of set consists of five variables that will be explained later. The second group of set (Al-Saffar, 2016) Consists of four explanatory variables.

7-1 Group One

These data were used by (Pasha & Ali Shah, 2004), where the number of observations are 150 , the independent variable (y) is the number of persons employed (Million) and five dependent variables that are believed to affect them were:

X_1 Land Cultivated (Million Hectors)

X_2 Inflation Rate (%)

X_3 Number of Establishment

X_4 Population (Million)

Literacy Rate (%)

X_5

7-1-1 Correlation Matrix

The correlation matrix was found as follows:

Table (1): The simple correlation coefficient between the explanatory variables and the independent variable

	Y	**X_1**	**X_2**	**X_3**	**X_4**	**X_5**
Y	1	0.9731	0.7118	0.9437	0.993	0.9572
X_1	0.9731	1	0.6635	0.9427	0.9761	0.9564
X_2	0.7118	0.6635	1	0.659	0.729	0.6813
X_3	0.9437	0.9427	0.659	1	0.9633	0.8672
X_4	0.993	0.9761	0.729	0.9633	1	0.9506
X_5	0.9572	0.9564	0.6813	0.8672	0.9506	1

7-1-2 Finding Latent Roots and Latent Vectors:

The Latent Roots and Latent Vectors of the Correlation Matrix were found using a program written in the Matlab language, as follows:

$$\Lambda = \begin{bmatrix} 0.0037 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0188 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0277 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1346 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.4597 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5.3555 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} 0.5111 & -0.4235 & 0.5999 & 0.0549 & 0.1169 & 0.4276 \\ 0.0057 & -0.5069 & -0.7141 & 0.0827 & 0.2174 & 0.423 \\ 0.0407 & -0.0024 & -0.0889 & -0.0274 & -0.9372 & 0.3336 \\ 0.218 & 0.474 & -0.1162 & -0.7137 & 0.1869 & 0.4124 \\ -0.827 & -0.0883 & 0.3271 & -0.0928 & 0.089 & 0.4299 \\ 0.0749 & 0.5755 & -0.0421 & 0.6866 & 0.1333 & 0.4149 \end{bmatrix}$$

7-1-3 Detection of Multicollinearity

The following table (table 2) gives the values of the latent roots and vectors for this data set .It also checks the multicollinearity between the variables according to the following conditions:

$\lambda_j < 1$ *and* $|\gamma_{0j}| < 0.5 \quad for \quad j = 0,1,2, \ldots , p$

Table (2): Test results

i	λ_i	γ_{0i}	conditions
0	0.0037	0.5111	One holds
1	0.0188	-0.4235	Two holds
2	0.0277	0.5999	One holds
3	0.1346	0.0549	Two holds
4	0.4597	0.1169	Two holds
5	5.3555	0.4276	One holds

The above table shows that three values satisfy the two conditions. This means that there is a multicollinearity between these variables, so the Latent Root estimators and its variances will depend only on the remaining three values i.e q = 3.

7-1-4 Estimation of the Regression Parameters

The values of the regression parameters are estimated using Ordinary Least Squares equation (2). The following table (Table 3) gives the values of the estimated regression parameters, variances and their significances:

Table (3): Estimators, variances and the t- test values of the regression coefficients in the Least Squares

i	$\hat{\beta}_{iOLS}$	$V(\hat{\beta}_{iOLS})$	$t(\hat{\beta}_{iOLS})$
1	0.0332	0.0155	$0.2665^{(N.S)}$
2	-0.0382	0.0012	$-1.0903^{(N.S)}$
3	-0.1814	0.0127	$-1.6115^{(N.S)}$
4	1.1323	0.0342	6.1201
5	0.0325	0.011	$0.3106^{(N.S)}$

We see from the above table that only variable four is significant. Using Latent Root method equation (5), the following table (Table 4) gives the values of the estimated regression parameters ,variances and their significances:

Table (4): Estimators, variances and the t- test values of the regression coefficients in the Latent Root

i	$\hat{\beta}_{iLR}$	$V(\hat{\beta}_{iLR})$	$t(\hat{\beta}_{iLR})$
1	0.1732	0.0086	$1.8721^{(N.S)}$
2	-0.0449	0.0003	-2.5203
3	-0.3313	0.0023	-6.9351
4	1.2842	0.0278	7.7054
5	-0.1135	0.0003	-6.6548

We see from the above table that only variable one is not significant.

7-1-5 Estimating the Mean Squares Error (MSE)

MSE is estimated for the two methods: Ordinary Least Squares according to equation (4) and Latent Root method according to equation (9) as in table 5:

Table (5): MSE for Ordinary Least Squares and the Latent Roots before deletion

The method	σ	Part1	Part2	MSE	R2
Least Squares	0.022	0.0746	0	0.0746	98.94%
Latent Roots	0.0232	0.0392	0.0137	0.0529	98.82%

From the above table, we note that the value of the MSE in the Latent Root method is lower than that in the Least Squares method as well as the value of R^2.

The following table shows the MSE values and the determination coefficient for each of the two methods after ignoring the non-significant variables:

Table (6): MSE for the Least Squares and Latent Roots after deletion

The method	σ	MSE	R2
Least Squares	0.0232	0.0005386	98.6%
Latent Roots	0.0216	0.0509	98.93%

From the above table it can be said that the estimated model using the Latent Root method is better than the estimated model in the Least Squares method taking in consideration the number of significant variables for both methods although the MSE is larger than the MSE for the Least Squares method .

7-1-6 The Estimated Regression Equation

After deleting the variables that are not important in the prediction process, the estimated regression equation in the Latent Roots method is:

$$\hat{y}_i^* = -0.0447x_2^* - 0.1756\, x_3^* + 1.1451x_4^* + 0.0524\, x_5^*$$

7-2 Second Group

These data were used by (Al-Saffar, 2016), where the number of observations 26 and the independent variable (y) in the study is the size of Carbon Dioxide Emissions (Metric tons) and four dependent variables that are believed to affect them were :

X_1 Industrial Growth.
X_2 Total Population.
X_3 Value of Foreign Direct Investment.
X_4 Industrial Energy Used

7-2-1 Correlation Matrix

The correlation matrix was found as follows:

Table (7): The simple correlation coefficient between the explanatory variables and the independent variable

	Y	**X_1**	**X_2**	**X_3**	**X_4**
Y	1	0.723	-0.7753	-0.6956	-0.6145
X_1	0.723	1	-0.8697	-0.4858	-0.5746
X_2	-0.7753	-0.8697	1	0.6301	0.8448
X_3	-0.6956	-0.4858	0.6301	1	0.5167
X_4	-0.6145	-0.5746	0.8448	0.5167	1

7-2-2 Finding Latent Roots and Latent Vectors:

The Latent Roots and Latent Vectors of the Correlation Matrix were found using a program written in the Matlab language, as follows:

$$\Lambda = \begin{bmatrix} 0.0358 & 0 & 0 & 0 & 0 \\ 0 & 0.2219 & 0 & 0 & 0 \\ 0 & 0 & 0.4508 & 0 & 0 \\ 0 & 0 & 0 & 0.5791 & 0 \\ 0 & 0 & 0 & 0 & 3.7124 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} -0.0105 & 0.8058 & 0.2844 & 0.2409 & -0.4601 \\ 0.4494 & -0.3508 & 0.5923 & -0.3555 & -0.4447 \\ 0.799 & 0.2008 & 0.046 & 0.2634 & 0.4997 \\ -0.0955 & 0.4166 & 0.135 & -0.8016 & 0.3955 \\ -0.0955 & 0.4166 & 0.135 & -0.8016 & 0.3955 \end{bmatrix}$$

7-2-3 Detection of Multicollinearity

The following table (table 8) gives the values of the latent roots and vectors for this data set .It also checks the multicollinearity between the variables according to the following conditions:

$\lambda_j < 1$ and $|\gamma_{0j}| < 0.5$ for $j = 0,1,2, \dots , p$

Table (8): Test results for Latent Roots

i	λ_i	$\gamma 0i$	conditions
0	0.0358	-0.0105	Two holds
1	0.2219	0.8058	One holds
2	0.4508	0.2844	Two holds
3	0.5791	0.2409	Two holds
4	3.7124	-0.4601	One holds

The above table shows that three values satisfy the two conditions. This means that there is a multicollinearity between these variables, so the Latent Root estimators and its variances will depend only on the remaining two values i.e q = 2. **7-**

2-4 Estimation of Regression Parameters Values

The values of the regression parameters are estimated using Ordinary Least Squares equation (2). The following table (Table 9) gives the values of the estimated regression parameters, variances and their significances:

Table (9): Estimators, variances and the t- test values of the regression coefficients in the Least Squares

i	$\hat{\beta}_{iOLS}$	$V(\hat{\beta}_{iOLS})$	$t(\hat{\beta}_{iOLS})$
1	0.3451	0.1001	1.0909$^{(N.S)}$
2	-0.1740	0.2644	-0.3384$^{(N.S)}$
3	-0.3809	0.0256	-2.3802
4	-0.0724	0.0831	-0.2511$^{(N.S)}$

We see from the above table that only variable three is significant. Using Latent Root method equation (5), the following table (Table 10) gives the values of the estimated regression parameters, variances and their significances:

Table (10): Estimators, variances and the t- test values of the regression coefficients in the Latent Root

i	$\hat{\beta}_{iLR}$	$V(\hat{\beta}_{iLR})$	$t(\hat{\beta}_{iLR})$
1	0.4088	0.0018	9.7572
2	-0.2238	0.0016	-5.6074
3	-0.4907	0.0017	-11.9276
4	0.161	0.0006	6.8445

We see from this table (table 10) that all variables are significant.

7-2-5 Estimating the Mean Squares Error (MSE)

MSE is estimated for the two methods: Ordinary Least Squares according to equation (4) and Latent Root method according to equation (9) as in table 11:

Table (11): MSE for Ordinary Least Squares and the Latent Roots before deletion

The method	σ	Part1	Part2	MSE	R2
Least Squares	0.1208	0.4732	0	0.4732	69.38%
Latent Roots	0.1264	0.0056	0.0001133	0.0057	66.47%

From the above table, we note that the value of the MSE in the Latent Root method is lower than that in the Least Squares method as well as the value of R^2.

The following table shows the MSE values and the determination coefficient for each of the two methods after ignoring the non-significant variables:

Table (12): MSE for the Least Squares and Latent Roots after deletion

The method	σ	MSE	R2
Least Squares	0.1467	0.0215	48.38%
Latent Roots	0.1264	0.0057	66.47%

From the above table it can be said that the estimated model using the Latent Root method is better than the estimated model using Least Squares method depending on the MSE criterion.

7-2-6 The Estimated Regression Equation

After deleting all the variables that are not important in the prediction process, the estimated regression equation in the Latent Roots method is:

$$\hat{y}_i^* = 0.4088x_1^* - 0.2238\, x_2^* - 0.4907x_3^* + 0.161\, x_4^*$$

8 -References

[1].Bastlevsky, Alexander, (1994), "**Statistical Factor Analysis and Related Methods**", John Wilay& Sons, INC.

[2].Belsley, D. A., Kuh, E. and Welsch, R. E., (1980), "**Regression Diagnostics Identifying Influential Data and Sources of Collinearity**", Wiley, New York.

[3].Dounald F.M., (1978), "**Multivariate statistical Methods**", 2nd Ed., McGraw-Hill Book Company, Tokyo.

[4].Chatterjee, Samprit and Price, Bertram , (2000), "**Regression Analysis by Example**", 3rd edition, John Wiley and Sons.

[5].Draper N.R., and Smith H., (1981), "**Applied Regression Analysis**", 2n Ed. John Wiley and Sons, Canada.

[6].Fisher J.C., and Mason R.L., (1981), "**The Analysis of Multicollinear Data in Criminology**", John Wiley and Sons.

[7].Gunst R.F., and Mason R.L., (1980), "**Regression Analysis and It's Application**", Marcel Dekker, New York, U.S.A.

[8].Hawkins, D. M., (1973), "**On the Investigation of Alternative Regression by Principal Component Analysis**", Applied Statistics, Vol. 22, No. 3, pp. 275-286.

[9].Kutner, Michael H.; Nachtsheim, Christopher J. and Neter, John; Li, William, (2005), "**Applied Linear Statistical Models**", 5th edition, McGraw- Hill Irwin, New York, USA.

[10]. Mason R.L., (1986), "**Latent Root Regression: A Biased Regression Methodology for Use with Collinear Predictor Variables**", Commun. Statist. theor. Math., Vol.15 No. 9, pp. 26512678.
[11]. Pasha, G. R and Ali Shah, Muhammad Akbar, (2004), "**Application of Ridge Regression to Multicollinear Data**", Journal of Research (Science), Bahauddin Zakariya University, Multan, Pakistan, Vol. 15, No. 1, pp. 97-106.
[12]. Al-Saffar, Azhar S. A., (2016), "**Diagnosis and treatment of Outlier values in the models Linear with the application**",Master Thesis, University of Mosul, Iraq.
[13]. Yan, Xin and Gang Su, Xiao, (2009), "**Linear Regression Analysis: Theory and Computing**", World Scientific Publishing, USA.