On the Assessment Restricted Liu Estimator for Dealing with Multi-Collinearity Problem

Noor Edin Rabeh
Baghdad University, College of Science, Baghdad, Iraq
Email: Edinnoori45@gmail.com

Abstract

In this paper, we concentrate on comparing the restricted least squares ($\hat{\beta}_{RLIU}$)with restricted Liu estimator ($\hat{\beta}_{RLS}$) based on (*MSE*) criterion in the existence of multi-collinearity. In addition, we find the best estimation in many different cases with some related numerical examples.

Keywords: Least squares; Liu estimator; Multi-collinearity

1. Introduction

Many statistical researchers, as well as the knowledge of its statistical consequences on the estimated model parameters about its desired scientific characteristics in explaining the scientific phenomenon in the correct scientific method, so this problem should be avoided by developing appropriate solutions to it have known the problem of multilinearity (the multiplicity of the linear relationship). It is known that the usual least squares method (OLS) is one of the most efficient known unbiased estimation methods and requires the availability of analysis assumptions, the most important of which is the independence of explanatory variables. Sometimes economic theory provides us with some theoretical information in the form of constraints (one or more identical constraints about the parameters to be estimated) from outside the scope of the research sample, in this case, the usual least squares method turns into another formula, the restricted least squares formula (RLS), so that it is possible (in some behavioral functions) to employ these constraints to address the problem of multilinearity. In 1993, the researcher (Liu) put forward a new model to address this problem by combining the Stein estimators and the usual character regression estimators ($\hat{\beta}_{ORP}$) with a special estimator called the $\left(\hat{\beta}_{OLiu}\right)$estimators, and when placing constraints on the latter estimators, they become the $\left(\hat{\beta}_{RLiu}\right)$ constrained estimators. This study was divided into two parts, the first part included the theoretical side, in which the method of finding the least squares estimators and the (Liu) restricted estimators was explained, and the second part included the experimental side, in which the ready-made application program (Minitab under Windows) was used to generate data using the Montecarlo method of simulation, developing correlation formulas and finding the average error squares for both methods.

2. Theoretical aspect

2.1. OLs and RLS capabilities:

The usual least squares method is one of the important and most common methods for estimating the parameters of a linear regression model and is characterized by more effective qualities than other methods, the ease of calculating the estimation of its parameters, the logic of the results obtained and the ease of understanding the mechanics of its operation. It is characterized by the fact that its estimates have the lowest variances of all other unbiased methods if the analysis assumptions of the linear model are available, represented by the error being a normally distributed random variable with a rate of (0) and a variance of (σ_u^2)Which means that the errors are not self-related, that the

DOI: https://doi.org/10.54216/PMTCS.040105
Received: December 20, 2023 Revised: March 24, 2024 Accepted: June 04, 2024

response variable (Y) is a random variable with a normal distribution with a rate of ($\propto$) and a variance of ($\sigma_u^2\ I$), as well as that The columns of the matrix of explanatory variables (X) are linearly independent and that (Rank(X)= P < n). In this research, the following linear model was taken:

$\underline{Y} = X\underline{\beta} + \underline{U}$ ……………………………….(1)

Where the:

$\underline{Y}_{n*1}$: response variable vector.

X_{n*p+1} :the Matrix of explanatory variables.

$\underline{\beta}_{p+1*1}$: vector of regression parameters.

$\underline{U}_{n*1}$: random error vector where E(U)=0, and $V(U) = \sigma_u^2 I_n$

(P): represents the number of explanatory variables, (n) represents the sample size.

I_{n*n}: a monoidal Matrix.

It is known that the least squares estimators are obtained by applying the following formula:

$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$ ……………….(2)

And that the matrix of the mean squares of the error for the estimator $\hat{\beta}_{OLS}$ of β_{OLS} is:

$MSE(\hat{\beta}_{OLS}) = E(\hat{\beta}_{OLS} - \beta_{OLS})(\hat{\beta}_{OLS} - \beta_{OLS})' = V(\hat{\beta}_{OLS}) + [(bias(\hat{\beta}_{OLS}))(bias(\hat{\beta}_{OLS}))']$ ….(3)

Where:

$$bias(\hat{\beta}_{OLS}) = E(\hat{\beta}_{OLS}) - \beta_{OLS}$$

When assuming the existence of (q) of the constraints imposed on the linear model, which can be in the following form:

$R\beta_{OLS} = r$ ………(4)

Where:

R_{q*p}: a matrix, the rows of which are of full rank and p≻q, which is a known Matrix.

r_{q*1}: vector, He is also known.

So, the ($\hat{\beta}_{RLS}$)estimators, according to the constraints of Equation (4) above, have the following form:

$\hat{\beta}_{RLS} = \hat{\beta}_{OLS} + S^{-1}R'(RS^{-1}R')^{-1}(r - R\hat{\beta}_{OLS})$ …….(5)

Hence:

$\hat{\beta}_{OLS} = S^{-1}X'Y \quad and \quad S = X'X$

If we assume that:

$A = S^{-1} - S^{-1}R'(R\,S^{-1}\,R')^{-1}RS^{-1} \quad and \quad \delta = r - R\,\hat{\beta}_{OLS}$ ..(6)

We get:

$V(\hat{\beta}_{RLS}) = \sigma^2\,A$ ……(7)

We assume that the Liu estimator is ($\hat{\beta}_{LIU}$)for the parameter ($0 \prec d \prec 1$) in the following form:

$\hat{\beta}_{LIU} = (S + I)^{-1}(X'Y + d\hat{\beta}_{OLS})$ …..(8)

If we assume that:

$F_d = (S + I)^{-1}(S + d\,I)$ ….(9)

If the $\hat{\beta}_{LIU}$ estimator can be written as:

$\hat{\beta}_{LIU} = F_d\,\hat{\beta}_{OLS}$ ….(10)

So that:

$E(\hat{\beta}_{LIU}) = F_d\,\beta_{OLS}$ ….(11)

And:

$V(\hat{\beta}_{LIU}) = \sigma^2 F_d S^{-1}\,F'_d$ …..(12)

From the above definition and according to the constraints in Equation (4), we find that the bound Liu estimator has the following formula:

$\hat{\beta}_{RLIU} = F_d\,\hat{\beta}_{RLS}$ …..(13)

So that:

$E(\hat{\beta}_{RLIU}) = F_d\,\beta_{OLS} + F_d S^{-1}R'(R\,S^{-1}\,R')^{-1}\delta$ …….(14)

And:

$V(\hat{\beta}_{RLIU}) = \sigma^2\,F_d\,A\,\,F'_d$ …..(15)

2.2. Comparison of $\hat{\beta}_{RLS}$ estimators and $\hat{\beta}_{RLIU}$ estimators:

When taking the difference between the variances of both estimators, we find that:

$V(\hat{\beta}_{RLS}) - V(\hat{\beta}_{RLIU}) = \sigma^2(A - \,F_d\,A\,\,F'_d)$

DOI: https://doi.org/10.54216/PMTCS.040105
Received: December 20, 2023 Revised: March 24, 2024 Accepted: June 04, 2024

$= \sigma^2(1-d)\,(S+I)^{-1}[AS\ +\ SA +\ (1+d)A](S+I)^{-1}$
$= \sigma^2(1-d)\,(S+I)^{-1}[G-H](S+I)^{-1}$ ….(16)
Where: $G = 2I + (1+d)S^{-1}$is a definite and positive matrix (p.d)), and:

$$H = S^{-1}R'(R\,S^{-1}\,R')^{-1}R + R'(R\,S^{-1}\,R')^{-1}R\,S^{-1} + (1+d)S^{-1}\,R'(R\,S^{-1}\,R')^{-1}R\,S^{-1}$$

It is a symmetric matrix. Since G is a definite and positive Matrix, there exists a monoidal Matrix like Q such that: $Q'\,G\,Q = I$, and $Q'\,H\,Q = \Lambda$, since Λ is a diagonal matrix whose diagonal elements are the characteristic roots of $|H-\lambda G| = 0$
If (*G-H*) is a non-negative definite Matrix (n.n.d))
And $Q'\,G\,Q - Q'\,H\,Q = I - \Lambda$ is also a non-negative definite matrix if:

$$1-\lambda_i \geq 0$$

We also find that:

$$\lambda_{max}(G^{-1}H) \leq 1$$

From the above we find that:

$$\lambda_p \leq \frac{x'Hx}{x'Gx} \leq \lambda_1$$

Where:

$$\lambda_1(G^{-1}H) \geq \lambda_2(G^{-1}H) \geq \ldots \geq \lambda_p(G^{-1}H)$$

From here we find that: $x'H\,x \leq x'G\,x$, and that: (G-H) is also (n.n.d)
From this information we find that it is clear that:
$V(\hat{\beta}_{RLS}) - V(\hat{\beta}_{RLIU})$ *is* (*n.n.d*) $\forall\ \ 0 \prec d \prec 1$ ..(17)
So that, the variance of the constrained Liu estimators is less than the variance of the constrained least squares estimators if and only if:

$$\lambda_{max}(G^{-1}H) \leq 1$$

3. The experimental side

3.1. Model description and simulation:

In this research, the Monte Carlo simulation method was used to generate the views of the regression model in Equation (1) for five samples of each sample with a size of (100) views and each view repeated a hundred times, then the rate of these views was taken, and finally these views were divided into five columns, each column contains (100) views
$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + U_i$, since the values of $(\beta_1, \beta_2, \beta_3, \beta_4)$ are taken corresponding to the characteristic roots of the standard (*X'X*) matrix with the assumption that the value of $\beta_0 = 0$, and then a linear relationship or lack thereof will be established between the explanatory variables X_{ij} generated, using the following relationship:

$X_{ij} = \sqrt{(1-\alpha^2)}Z_{ij} + \alpha\,Z_{i5}$, $j = 1{,}2$, $i = 1{,}2{,}\ldots{,}100$
$X_{ij} = \sqrt{(1-\alpha^{*2})}Z_{ij} + \alpha *\ Z_{i5}$, $j = 3{,}4$, $i = 1{,}2{,}\ldots{,}100$
Whereas the value of (α) is represented by the degree of correlation between the two variables (X_1, X_2), and (α*) is represented by the degree of correlation between the two variables (X_3, X_4).In this research, three correlations were postulated:

- High correlation of the two groups represented by the values (α=0.9, α*=0.9)
- A low correlation of the two groups represented by the values (α=0.3, α*=0.3)
- High correlation for the first group and low for the second group represented by the values (α=0.9, α*=0.1)

Finally, the explanatory variables are converted to their standard form using the constant formula of length one and as follows:
$W_i *= \frac{W_i - \overline{W}}{\sqrt{\sum_{i=1}^{n}(W_i - \overline{W})^2}}$ …………(19)
Where:
W_i is any random variable, and that W* represents the standard form of that variable.

3.2. The stage of estimation of parameters:

In this research, the following two identical constraints were assumed:
1. The final sum of all the estimated parameters is equal to the correct one, that is:

DOI: https://doi.org/10.54216/PMTCS.040105
Received: December 20, 2023 Revised: March 24, 2024 Accepted: June 04, 2024

$$\sum_{j=1}^{p} \beta_j = 1$$

2. The first limit slope is equal to the second limit slope, that is:

$\beta_1 = \beta_2$ or $\beta_1 - \beta_2 = 0$

The employment of such previous constraints necessitates the application of the restricted least squares method, which requires reformulating the above constraints using matrices and vectors and as follows:

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, \quad r = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

At this stage, the parameters were estimated by the restricted least squares method according to Equation (5) and for each correlation case, and the parameters of the (Liu) restricted method were also estimated according to equation (13) and also for each correlation case with the choice of four values for (d), that (0.1 ‹0.4 ‹0.6 ‹0.9) is the results were obtained as in the following table (A):

After the estimators of both methods have been found, the mean squares of the error for both methods and for all correlation cases are now found by the following equations:

For the $(\hat{\beta}_{RLS})$ estimators, the:

$$MSE(\hat{\beta}_{RLS}) = \frac{\sum_{j=1}^{4}(\hat{\beta}_{RLS} - \beta_{eign(S)})^2}{n}$$

As for the $(\hat{\beta}_{RLS})$ estimators, the:

$$MSE(\hat{\beta}_{RLIU}) = \frac{\sum_{j=1}^{4}(\hat{\beta}_{RLIU} - \beta_{eign(S)})^2}{n}$$

Where: $\beta_{eign(S)}$ represents the characteristic roots of the harsh Matrix S.

Table 1: The (*MSE*) results for both methods.

	$\alpha = 0.3$, $\alpha *= 0.3$	$\alpha = 0.1$, $\alpha *= 0.9$	$\alpha = 0.9$, $\alpha *= 0.9$
$MSE\ \hat{\beta}_{RLS}$	**0.121**	**1.202**	**0.085**
$MSE\ \hat{\beta}_{RLIU}$ **d=0.1**	**0.042**	**0.209**	**0.074**
$MSE\ \hat{\beta}_{RLIU}$ **d=0.4**	**0.088**	**0.433**	**0.073**
$MSE\ \hat{\beta}_{RLIU}$ **d=0.6**	**0.098**	**0.641**	**0.084**
$MSE\ \hat{\beta}_{RLIU}$ **d=0.9**	**0.063**	**1.044**	**0.084**

4. Conclusions

- From the observation of Table 1 we find that, Liu's method is more accurate than the restricted least squares method.
- This method can be used even in the absence of multilinearity between explanatory variables.

References

[1] Kazim, Amouri Hadi and Muslim, Bassem shalaiba, (2002) "advanced economic measurement theory and practice", House of books and documents, University of Baghdad.
[2] Akdeniz, F and Kaciranlar, S., (2001)" More on the new biased estimator in linear regression ", Sankhya, Vol.63, Ser. B, Pt.3, pp.321-325.
[3] Ghrishtopher, Z .M. (1997)"Monte Carlo Simulation", SAGE Publications, Delhi.
[4] Gunst, R. F. and Mason, R. L. and Webster, J. F., (1975) "Regression Analysis and problems of Multicollinerity "Communication in statistics, Vol. 22, No.3."
[5] Liu, Communication in statistics, Ser. A, Vol. 22, pp. 393-402

DOI: https://doi.org/10.54216/PMTCS.040105
Received: December 20, 2023 Revised: March 24, 2024 Accepted: June 04, 2024