Tikrit Journal of Administrative and Economic Sciences, Vol. 19, No. Special Issue, Part (5): 244-263
Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

Tikrit Journal of Administrative and Economics Sciences
مجلة تكريت للعلوم الإدارية والاقتصادية
ISSN: 1813-1719 (Print)

Weighted Robust Ridge Regression Compared with Weighted Ridge Regression to Estimate the Parameters in Presence of Multicollinearity and Outliers

Kafi Dano Pati*
Department of Computer Science, College of Science, University of Duhok

Keywords:
Ridge regression; Multicollinearity; Robust regression; Outliers; Ridge least absolute value; Weighted ridge least absolute value.

Article history:

Received	20 Apr. 2023
Accepted	29 Apr. 2023
Available online	30 Aug. 2023

*Corresponding author:
Kafi Dano Pati
Department of Computer Science, College of Science, University of Duhok

Abstract: The purpose of this research is to look at how to estimate parameters in multiple linear regression models. When multicollinearity and outliers are present in the model, existing techniques such as Ordinary Least Squares (OLS) estimate, Ridge Regression (RR), Ridge Least Absolute Value (RLAV), and Weighted Ridge Regression (WRID) are compared with a proposed new technique, Weighted Ridge Least Absolute Value (WRLAV). Ridge Least Absolute Value (RLAV) regression is a reliable alternative to the OLS, especially when the disturbances follow non-normal distributions with outliers. Using some test scenarios, the (WRLAV) estimator is based on (RLAV), which retains insensitive to inaccurate measurements and highly robust in the presence of these problems. The proposed estimate (WRLAV) is very robust to contaminated data for heavy tailed errors and shows good performance under various situations of errors and different degrees of multicollinearity using the Root Mean Square Error (RMSE) criteria for Simulation study and Standard Error (SE) for numerical example that was worked out when compared to other techniques of estimation.

Tikrit Journal of Administrative and Economic Sciences, Vol. 19, No. Special Issue, Part (5): 244-263
Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

مقارنة بين متانة انحدار الحافة الموزون مع انحدار الحافة المرجح لتقدير المعلمات بوجود الارتباط الذاتي الخطي المتعدد والقيم المتطرفة

كافي دنو بتي
قسم علوم الحاسبات، جامعة دهوك

المستخلص

الغرض من هذا البحث هو النظر في كيفية تقدير المعلمات في نماذج الانحدار الخطي المتعددة. عند وجود القيم المتعددة الخطية والقيم المتطرفة في النموذج، تتم مقارنة التقنيات الحالية مثل تقدير المربعات الصغرى العادية (OLS) ، وانحدار الحافة (RR) ، وقيمة الحافة الأقل المطلقة (RLAV)، وانحدار الحافة المرجحة (WRID) مع تقنية جديدة مقترحة القيم المطلقة للقمة المرجحة .(WRLAV) يعد انحدار(Ridge Least Absolute value (RLAV بديلاً موثوقًا به لـ OLS ، خاصةً عندما تتبع الاضطرابات توزيعات غير طبيعية مع قيم متطرفة. باستخدام بعض سيناريوهات الاختبار، يعتمد مقدر (WRLAV) على (RLAV)، والذي يحتفظ بعدم الحساسية للقياسات غير الدقيقة وقوي للغاية في وجود هذه المشاكل. يعتبر التقدير المقترح (WRLAV) قويًا جدًا للبيانات الملوثة للأخطاء ثقيلة الذيل ويظهر أداءً جيدًا في ظل مقارنات مختلفة من الأخطاء ودرجات مختلفة من التعددية الخطية باستخدام معايير الجذر التربيعي للخطأ (RMSE) لدراسة المحاكاة والخطأ القياسي (SE) للمثال ذات القيم الحقيقية تم وضعه عند مقارنته بتقنيات التقدير الأخرى.

الكلمات المفتاحية: انحدار الحافة. متعدد الارتباط الذاتي الخطي. الانحدار المتين. أقل حافة للقيمة المطلقة؛ أقل حافة للقيمة المطلقة المرجحة.

1. Introduction

Ordinary Least Squares (OLS) are a type of least squares that When OLS is used to estimate regression models, it yields unbiased parameter estimates as well as estimates with the least variance when the disturbances are independent and distributed in the same normal way. OLS regression is also appropriate because of its speed, efficiency, and simplicity. It also provides accurate estimators for the model's unknown parameters, as long as the errors are independent and normally distributed with a mean of zero and a variance of one, Subhash C. Narula et al., (1999). When non-normal errors are present, however, the OLS performance can be seriously impacted, especially if the errors follow a distribution that is sensitive to generating outliers.

Outliers, on the other hand, have a significant impact on OLS estimates since the weight given to each value in the model solution is the same. In other words, the OLS regression is far from optimal in a variety of non-Gaussian cases, especially when the errors follow longer-tails distributions, Rice and White (1964). In terms of regression problems, Peter J. Huber (1973) noted that only one substantially outlying observation can damage the OLS estimate. Furthermore, in comparison to the simple location issue, outliers are significantly more difficult to find within the regression.

This is why outliers resulting from extreme values of explanatory variables can be extremely difficult. Because it is typical to obtain some extreme data points and the presence of multicollinearity in the data distribution, the model does not provide adequate estimations, and the OLS solution is known to be inaccurate.

With this, the OLS has a breakdown point of 0 percent. This indicates that a small amount of contamination data can cause the estimators to take values from-∞ to +∞ Rousseeuw and Leroy (2005). Similarly, when looking at higher-dimensional data, outlying points may be ignored.

As a result, if the observations are not normally distributed or contain outliers, the OLS model is no longer appropriate. This is what requires the use of robust regression procedures to reduce the negative effects of these events. The estimations of the parameters for OLS can be explained as follows:

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

In the following sections of this paper, a number of other estimating processes, such as LAV, will be studied and briefly explained. This technique is similar to the OLS estimate, and it will be demonstrated that it has poor data rejection features by configuration.

Every technique is implemented utilizing real life data sets and the simulation examination. This is done with three various error distributions for three different sample sizes (n=25 and 100).

The goal of this paper is to use Monte Carlo simulation to compare the five regression approaches. The RMSE is used to evaluate the performance of estimators. However, in real life, finding a data set that can satisfy all of the assumptions required for the implementation of this method is quite difficult.

The rest of the paper is organized as follows: The ridge regression model is introduced in section 2; robust regression and robust ridge regression are discussed in section 3 (including descriptions of the test weighted robust ridge regression procedures); the numerical example and simulation study are described in section 4; and finally, the results and conclusions are discussed in section 6.

2. Ridge Regression Estimators:

Multicollinearity is a severe issue in model regression. If other remedy procedures (e.g., removing one or more correlated predictor variables, using OLS regression with Principal Components Analysis (PCA) as the explanatory variables instead of the individual explanatory variables, etc.) fail, the ridge regression is a tool that can help.

Ridge regression made use of the OLS estimator, which is a little biased. When a biased estimator with a small standard error and an unbiased estimate with a big standard error are both available, the biased estimator with the small standard error is usually favored. Ridge regression uses the correlation transformation in combination with a biasing constant in order to obtain the ridge estimators for the transformed model, Kutner et al., (2004). This constant is usually chosen on the basis on the Variance Inflation Factor (VIF), which will decrease with the increasing in the biasing constant before stabilizing. As a result, the appropriate biasing constant is chosen based on who it stabilizes the VIFs. Hoer and Kennard (1970) created a ridge regression method for estimating an appropriate parameter, which a constant k is 0< k <1 added to the $X'X$ matrix and computing the estimated coefficient as:

$$\hat{\beta}_{Rid} = (X'X + kI)^{-1} X'y \tag{2}$$

Where $\hat{\beta}_{Rid}$ is the ridge coefficient, I is the pxp identity matrix and p is the number of explanatory variables. On the basis of the researcher's judgment, the constant *k* based on a trace might be determined subjectively. In accordance with (Dawoud et al., 2022 and Hoerl et al., 1975) used the value of *k* determined in [Hoerl et al., 1975 and Montgomery et al. 2021]. Typically, the estimate is given by

$$K_{HK} = \frac{pS_{LS}^2}{\hat{\boldsymbol{\beta}}_{LS}' \hat{\boldsymbol{\beta}}_{LS}} \tag{3}$$

where

$$S_{LS}^2 = \frac{\left(\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}_{LS}\right)'\left(\mathbf{y}-\mathbf{x}\hat{\boldsymbol{\beta}}_{LS}\right)}{n-p} \tag{4}$$

3. Robust Regression

Because the OLS estimator could be severely affected if the outliers appear to be significant, Robust Regression is an alternative procedure for detecting influential outliers in model regression. How could such outliers be handled with in this case? If there is no evidence of measurement or recording error, the decision whether or not to exclude the outlier observations from the dataset should be carefully considered. The elimination of outlying cases may be supported by the idea that there is no interest in modeling extreme events if a model is built to predict general trends.

However, if the decision is made to keep the outliers, robust regression can be used to reduce the effects of the outliers. As a result, the LAV regression is proposed to be a component of the more generic quantile regression.

Kutner et al., (2004) introduced a number of robust regression methods, but this paper will focus on the robust LAV combined with weighted to be more robust to outlier resistance. The weights in question are computed using residuals, which are evaluations of how observations diverge from its predicted value.

Kutner et al., (2004) explained the Huber and Bisquare weighting functions. Scaled-residuals should be used for each weight. They explain a number of methods for obtaining scaled-residuals, including Median Absolute Deviation (MAD). The robustreg package in R now makes it simple to do robust regression. R includes the LAV estimating method, which was first introduced by Wagner H.M. (1959).

3.1. Least Absolute Values Regression (LAV)

Least absolute value, also known as Least Absolute Errors (LAE), or the L1 norm and Least Absolute Deviations (LAD) problem, is a mathematical optimization technique used to find a function that closely approximates a set of data, similar to the popular least squares technique.

LAV regression overcomes the drawbacks of OLS and provides an interesting alternative. Andrews D.F., (2012) that is less sensitive to extreme errors than OLS regression. It also has implicit mechanics for refusing inaccurate data and does not require a normal distribution of data, which is quite unrealistic in practice, Claerbout and Muir (1973).

Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

Some studies wanted to design estimation methods that are resistant to outlier-producing error distributions, and LAV was viewed as the estimation method in the case where the errors are dependent. Dawoud, et al., (2022) looked into the accuracy of model coefficient estimation using LAV regression with autocorrelation correction, whereas Dielman and Rose (1994) looked also into the accuracy of forecasts using LAV estimated regressions with autocorrelation correction.

In fact, the LAV regression is one of the most extensively used robust regression techniques. In comparison to its OLS match, extreme observations have a lower impact on LAV estimations. Some investigations have shown that the LAV regression criterion improves estimates Dielman and Rose (1994), Dielman and Pfaffenberger (1988) and Efron, (1988).

LAV, which predates OLS by around 50 years, is the simplest and earliest technique to bounded effect robust regression. Furthermore, it is insensitive to changes shown in a small percentage of observations.

An estimator is considered robust if it is completely efficient under an assumed model but indicates efficiency to a high degree for possible alternatives.

When disturbances are not dependent but not essentially normal, Dielman and Pfaffenberger (1988), Gabriela and Subhash (1991), Dawoud and Abonazel (2021), Dawoud et al.(2022), Dielman and Rose (1996) have investigated inference for regression by employing LAV estimation.

To address the problem [10] defined the LAD estimators.

$$\min \sum_{i=1}^{n} |\varepsilon_i| = \min \sum_{i=1}^{n} \left| y_i - \sum_{i=1}^{m} x_{ij} \beta_j \right|, \quad i=1, 2\ldots n, j=1, 2\ldots m \qquad (5)$$

The goal of the LAV technique is to estimate the parameter by minimizing the sum of the residual absolute values, and the resulting estimators are known as median estimators, which are extensively used robust location estimators. This method is very robust to *y* values that are out of the normal. The estimator LAV's breakdown point is 1/n.

Let's identify the linear regression model with n observations as given $\hat{y}_i = \sum_{i=1}^{n} x_{ij} \beta_j$.

Step 1:- compute the correctly fitting regression function from the data set.

Step 2:- Let $(x_{i1}, x_{i2}, ..., x_{im}, y_i)$ be the ith observations and let $\beta_o, \beta_1, ..., \beta_m$ estimated by minimizing the overall absolute values of the differences between the values of $\hat{y}$ and y be the estimates of $\beta_o, \beta_1, ..., \beta_m$ is $\varepsilon_i = |y_i - \hat{y}_i|$

$$\hat{\beta}_{LAV} = \min \sum_{i=1}^{n} |\varepsilon_i|$$

Step 3:- The LAV method is given as,

As the name proposes, this method minimizes the sum of absolute values of deviations rather than the sum of squared deviations. As a result, unlike the OLS, it does not place as much emphasis on significantly deviating observations, producing in more robust outlier estimators.

3.2. Ridge least absolute value estimator (RLAV)

This estimator combines the properties of LAV estimators with ridge regression estimators known as RLAV estimators, as explained by Dielman and Pfaffenberger (1988), Dawoud and Abonazel (2021), Dawoud et al., (2022), Dielman T. E. and Pfaffenberger (1990).

$$\hat{\boldsymbol{\beta}}_{RLAV} = (\mathbf{X'X} + k_{LAV}\,\mathbf{I})^{-1}\,\mathbf{X'Y} \qquad (6)$$

The following is how the value of k_{LAV} is calculated using data similar to that of Hoerl and Kennard (1970).

Replacing k with k_{LAV} in Equation (4) and (5), one can get

$$k_{LAV} = \frac{p\,S^2_{LAV}}{\hat{\boldsymbol{\beta}}'_{LAV}\,\hat{\boldsymbol{\beta}}_{LAV}} \quad \text{where} \qquad (7)$$

$$S^2_{LAV} = \frac{\sum_{i=1}^{n} \varepsilon^2_{LAV}}{n-p} \qquad (8)$$

where n is the sample size, p is the number of explanatory variables and, $\hat{\beta}_{LAV}$ is the estimates of β and ε_{LAV} is the residuals from LAV estimator identified as the solution to $\hat{\beta}_{LAV} = \min \sum_{i=1}^{n} |\varepsilon_i|$

3.3 Weighted Ridge Estimators (WRID)

It is proposed that Bisquare weighted function be utilized as follows:

$$w_{B}(u)=\begin{cases}\left[1-\left(\frac{u}{c}\right)^{2}\right]^{2} & |u|\le c\\ 0 & |u|>c\end{cases}$$

, to estimate the (9)

$$\hat{\boldsymbol{\beta}}_{WRID}=(\mathbf{X'WX}+k\mathbf{I})^{-1}\mathbf{X'WY} \quad (10)$$

$\hat{\boldsymbol{\beta}}_{WRID}$ is the estimator with the biasing parameter k and weights $w_{\beta}(u)$ determined from the data is denoted as the weighted ridge WRID estimator.

3.4. Weighted Ridge LAV Estimators (WRLAV)

On the basis of the weighted ridge LAV estimator WRLAV, modifications to robust ridge regression are suggested. These tools are essential for parameter estimation when the generated data set contains multicollinearity and outliers, and Maximum Likelihood Estimators (MLEs), such as the OLS, cannot handle them efficiently.

The robust LAV estimator is represented by equation (5), while robust ridge estimators are computed by equation (6). This formula is used to calculate the weighted robust ridge regression parameters:

$$\hat{\boldsymbol{\beta}}_{WRLAV}=(\mathbf{X'WX}+k_{WLAV}\mathbf{I})^{-1}(\mathbf{X'WY}) \quad (11)$$

where k_{WLAV} is achieved from the weighted robust estimator. Therefore, if this estimator's resistance is to be calculated, the numerical example's results are compared by the use of the SE as well as the simulated data's results by the use of the Bias and RMSE to get the maximum parameter estimation.

4. Applications

4.1 Numerical Example

For the purpose of testing the estimators' performance, 252 observations of body fat are used from, Penrose (1985). There are fourteen explanatory variables in the data available. The response variable, y= PCTBF (percentage of body fat), which was regressed to the following variables: x1= Density, x2= Age, x3= Weight, x4= Height, x5= Neck, x6= Chest, x7=Abdomen, x8= Hip, x9= Thigh, x10= Knee, x11= Ankle, x12= Biceps, x13= Forearm and x14= Wrist. The following table shows the VIF of the parameters, is shown in the table below. The VIF may be used to define whether explanatory variables are involved in multicollinearity, this

Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

condition can be used as, $VIF = \frac{1}{1-R^2}$, where R^2 is the determinant of the matrix $X'X$.

Similarly, by computing the residuals connected with LMS regression, it is possible to identify outliers in the data. $s = 1.4826\left(1+\frac{5}{n-p}\right)\sqrt{med\left(\varepsilon_i^2\right)}$, where i=1, 2… n, and med is the median of the squared residuals, the number of explanatory variables is denoted by the symbol *p*. When the equivalent standardized residual is big, the points $\left(y_i, x_{i1},..., x_{ip}\right)$ are referred to be regression outliers.

In specifically, Rousseeuw and Zomeren (1990) referred to the i^{th} vector as a regression outlier if $|r_i|/s > 2.5$, implying that the value is outlier. The ordinary or simple residuals are, in fact, the most extensively used measures for detecting outliers (observed- predicted values).

Table (1): Variance Inflation Factor VIF for the body fat data set

Var.	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**
VIF	3.780	2.252	33.691	1.661	4.353	9.377	17.939
Var.	X8	X9	X10	X11	X12	X13	X14
VIF	14.811	7.809	4.566	1.901	3.615	2.215	3.486

The outcomes of the analysis are presented in Table 1 and illustrate when using real data from the body fat application, the VIF for some variables has exceeded the value of 10 since the calculation of VIF is highly dependent on the calculation of. As a result, it is apparent that the explanatory variables have a multicollinearity problem. As a result, the maximum value in the X3 variable in Table 1 is 33.691, as well as the data that included 13 outliers, which impacted the data and resulted in poor estimations.

The standard errors SE of the statistical criteria for this data, as given in Table 2, imply that WRLD is less than other existing methods when performed (OLS, RIDGE and RLAV). In addition, the proposed method WRLAV is better than any other methods and has the least SE. Table 2 summarizes the parameter estimations as well as the standard error for proposed and existing techniques.

Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

Table (2): Bisquare weighted function used to estimated parameters and standard error SE of $\hat{\beta}_1,\ldots,\hat{\beta}_{14}$ for the existing and proposed methods WRLAV estimators for the body fat data set.

Coef.	Estimate	OLS	RIDGE	RLAV	WRID	WRLAV
$\hat{\beta}_1$	parameter	-407.09	-0.9203	-0.9204	-0.9927	-0.9963
	S.E.	8.1754	0.0201	0.0200	0.0056	0.0031
$\hat{\beta}_2$	parameter	0.0125	0.0192	0.0192	0.0006	0.0011
	S.E.	0.0156	0.0155	0.0151	0.0041	0.0022
$\hat{\beta}_3$	parameter	0.0100	0.0318	0.0318	0.0076	0.0176
	S.E.	0.0572	0.0569	0.0556	0.0163	0.0107
$\hat{\beta}_4$	parameter	-0.0079	-0.0036	-0.0036	-0.0051	-0.0100
	S.E.	0.0281	0.0134	0.0133	0.0038	0.0034
$\hat{\beta}_5$	parameter	-0.0282	-0.0087	-0.0087	0.0021	-0.0007
	S.E.	0.0687	0.0217	0.0216	0.0059	0.0031
$\hat{\beta}_6$	parameter	0.0265	0.0266	0.0266	0.0002	-0.0020
	S.E.	0.0304	0.0313	0.0311	0.0089	0.0050
$\hat{\beta}_7$	parameter	0.0184	0.0309	0.0307	-0.0025	-0.0020
	S.E.	0.0429	0.0427	0.0417	0.0116	0.0062
$\hat{\beta}_8$	parameter	0.0190	0.0153	0.0154	0.0098	0.0057
	S.E.	0.0430	0.0388	0.0386	0.0119	0.0060
$\hat{\beta}_9$	parameter	-0.0166	-0.0093	-0.0093	-0.0137	-0.0078
	S.E.	0.0426	0.0289	0.0287	0.0080	0.0043
$\hat{\beta}_{10}$	parameter	-0.0046	-0.0012	-0.0012	0.0017	-0.0007
	S.E.	0.0709	0.0222	0.0221	0.0063	0.0033
$\hat{\beta}_{11}$	parameter	-0.0848	-0.0167	-0.0168	0.0004	0.0009
	S.E.	0.0651	0.0145	0.0144	0.0037	0.0022
$\hat{\beta}_{12}$	parameter	-0.0545	-0.0190	-0.0190	-0.0004	-0.0026
	S.E.	0.0504	0.0198	0.0197	0.0051	0.0029
$\hat{\beta}_{13}$	parameter	0.0336	0.0086	0.0086	-0.0009	-0.0003
	S.E.	0.0589	0.0155	0.0154	0.0040	0.0022
$\hat{\beta}_{14}$	parameter	0.0072	-0.0002	-0.0002	-0.0015	-0.0021
	S.E.	0.1601	0.0195	0.0194	0.0053	0.0028

Because of the presence of autocorrelation and outliers in the data, the values of the standard error SE for the WRID are better than OLS, not only for OLS but also for existing Ridge regression and RLAV, as shown in Table 2. If the data contains a single observation outlier, it is evident that the OLS will be affected. On the other hand, the WRLVA's value is superior to all existing methods for all parameters.

Finally, a numerical example is provided to show how this procedure behaves with data with multicollinearity and outliers.

4.2 Monte Carlo Study

To confirm the performance of the proposed estimate technique WRLAV, a simulation is used. Due to computer time limitations, this comparison is performed with sample sizes of n = 25, 100, with 0 percent, 5%, 10%, and 20% of outliers and p= 0.0, 0.5 and 0.99 of correlation coefficients, respectively. A total of 500 simulation runs were used to arrive at the results. The values generated were held steady throughout the experiment by observing that if *n* is norm (0, 1), and then such generated data were used to generate estimates of the unknown parameters.

A Monte Carlo study was used to compare the performance of the WRLAV algorithms. Simulations are carried out, according to Dawoud et al., (2022) in order to provide data that will be used to evaluate the proposed method WRLAV. OLS, RR, WRID, and weighted ridge least absolute value WRLAV estimator were all compared.

The simulation processes generate the following data for the multivariate stochastic model:

$$y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \qquad (12)$$

with each parameter assigned the value 1 i.e. $\beta_0, \beta_1, \beta_2$ *and* β_3 equal to one. The explanatory variables x_{i1}, x_{i2} *and* x_{i3} are generated as:

$$X_{ij} = (1 - p^2) Z_{ij} + p_{ij} \text{ where i=1, 2, …, n j =1, 2, …, m} \qquad (13)$$

where Z_{ij} standard normal random variables that are independent of each other. In order to investigate model performance in a wide range of conditions, three different distributional forms for the ε_i disturbances are considered. The following are the distributional forms:

1. Normal, with mean 0 and variance 1; i.e. N (0, 1).
2. t-Student distribution with degrees of freedom three.
3. Cauchy, with median 0 and scale parameter 1.

The non-normal distribution, as well as Cauchy and Student-t distributions, are all heavy-tailed distributions with a tendency for producing extreme values.

Tikrit Journal of Administrative and Economic Sciences, Vol. 19, No. Special Issue, Part (5): 244-263
Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

The goal of this paper is to examine least absolute value parameter estimate in regression analysis and investigate its implementation from a programming viewpoint.

5. Results

Table (3): The values of RMSE for the parameters $\hat{\beta}_1, \hat{\beta}_2$ *and* $\hat{\beta}_3$ used Bisquare weighted function generated normal disturbance distribution, with different sample sizes, ρ=0.0 and 0% of outliers.

Sample size	Par.	OLS	RIDGE	RLAV	WRID
25	$\hat{\beta}_1$	0.2245	0.5123	0.5072	0.5332
	$\hat{\beta}_2$	0.2189	0.5096	0.5085	0.5277
	$\hat{\beta}_3$	0.2265	0.5194	0.5137	0.5376
100	$\hat{\beta}_1$	0.0986	0.4997	0.4991	0.5005
	$\hat{\beta}_2$	0.1081	0.4987	0.4992	0.4993
	$\hat{\beta}_3$	0.1018	0.4972	0.4977	0.5036

Table (4): The values of RMSE for the parameters $\hat{\beta}_1, \hat{\beta}_2$ *and* $\hat{\beta}_3$ used Bisquare weighted function generated normal disturbance distribution, with different sample sizes, ρ=0.5 and 5%, 10%, and 20% of outliers.

Sample size	Par.	OLS	RIDGE	RLAV	WRID	WRLAV
25	$\hat{\beta}_1$	7.5895	1.0045	0.9724	0.9493	0.9472
	$\hat{\beta}_2$	7.3614	1.0033	0.9691	0.9493	0.9490
	$\hat{\beta}_3$	7.7133	0.9990	0.9665	0.9489	0.9480
100	$\hat{\beta}_1$	3.5751	0.9724	0.9620	0.9519	0.9508
	$\hat{\beta}_2$	3.7258	0.9601	0.9588	0.9521	0.9517
	$\hat{\beta}_3$	3.7352	0.9698	0.9661	0.9521	0.9518
25	$\hat{\beta}_1$	10.2802	1.0104	0.9757	0.9615	0.9594
	$\hat{\beta}_2$	10.2638	0.9959	0.9771	0.9608	0.9606
	$\hat{\beta}_3$	10.2251	1.0199	0.9770	0.9648	0.9596
100	$\hat{\beta}_1$	5.1701	0.9824	0.9768	0.9656	0.9615
	$\hat{\beta}_2$	5.1229	0.9768	0.9735	0.9705	0.9619
	$\hat{\beta}_3$	5.0642	0.9740	0.9704	0.9626	0.9625

Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

Sample size	Par.	OLS	RIDGE	RLAV	WRID	WRLAV
25	$\hat{\beta}_1$	14.4999	1.0219	0.9821	1.0138	0.9696
	$\hat{\beta}_2$	15.5207	0.9946	0.9804	0.9826	0.9697
	$\hat{\beta}_3$	15.0177	1.0156	0.9792	1.0116	0.9723
100	$\hat{\beta}_1$	6.7126	0.9915	0.9840	0.9824	0.9689
	$\hat{\beta}_2$	6.8573	0.9757	0.9743	0.9757	0.9691
	$\hat{\beta}_3$	6.8244	0.9822	0.9776	0.9731	0.9697

Table (5): The values of RMSE for the parameters $\hat{\beta}_1, \hat{\beta}_2$ *and* $\hat{\beta}_3$ used Bisquare weighted function generated normal disturbance distribution, with different sample sizes, ρ =0.99 and 5%, 10%, and 20% of outliers.

Sample size	Par.	OLS	RIDGE	RLAV	WRID	WRLAV
25	$\hat{\beta}_1$	35.9365	16.7776	2.9478	2.0511	1.9886
	$\hat{\beta}_2$	36.1573	16.9173	2.2912	2.2063	2.2038
	$\hat{\beta}_3$	36.6688	17.0441	2.7863	2.1411	1.9680
100	$\hat{\beta}_1$	17.0669	7.6768	3.0205	2.0629	1.1568
	$\hat{\beta}_2$	17.6185	7.8292	3.1472	2.0982	1.1337
	$\hat{\beta}_3$	18.1962	8.0991	3.2043	2.0807	1.1342
25	$\hat{\beta}_1$	48.0805	16.2096	1.5721	3.5149	1.5721
	$\hat{\beta}_2$	49.7334	16.7865	1.7133	3.1362	1.7133
	$\hat{\beta}_3$	48.5524	16.0901	1.7882	3.3948	1.7882
100	$\hat{\beta}_1$	25.1543	8.1136	2.2068	4.0152	1.2601
	$\hat{\beta}_2$	24.5205	7.9333	2.1754	4.1327	1.2291
	$\hat{\beta}_3$	24.5872	7.9869	2.0761	3.8942	1.2724
25	$\hat{\beta}_1$	71.0188	15.7615	1.7698	14.0931	1.1758
	$\hat{\beta}_2$	73.7451	16.7818	1.7472	14.9280	1.1532
	$\hat{\beta}_3$	72.8173	16.3704	1.8650	14.6707	1.1720
100	$\hat{\beta}_1$	33.1962	8.0506	1.6586	7.8960	1.3526
	$\hat{\beta}_2$	32.0578	7.9185	1.5797	7.2164	1.3754
	$\hat{\beta}_3$	33.3515	8.0777	1.5127	7.4497	1.3583

Table (6). The values of RMSE for the parameters $\hat{\beta}_1, \hat{\beta}_2$ *and* $\hat{\beta}_3$ used Bisquare weighted function generated Cauchy disturbance distribution, with different sample sizes, ρ =0.0 and 0% of outliers.

Sample size	Par.	OLS	RIDGE	RLAV	WRID
25	$\hat{\beta}_1$	75.2735	0.8672	0.8667	0.8656
	$\hat{\beta}_2$	19.3696	0.8578	0.8562	0.8485
	$\hat{\beta}_3$	31.8473	0.8692	0.8593	0.8573
100	$\hat{\beta}_1$	24.8413	0.9087	0.9078	0.9007
	$\hat{\beta}_2$	30.4384	0.9056	0.9048	0.9024
	$\hat{\beta}_3$	21.6401	0.9042	0.9028	0.9004

Table (7): The values of RMSE for the parameters $\hat{\beta}_1, \hat{\beta}_2$ *and* $\hat{\beta}_3$ used Bisquare weighted function generated Cauchy disturbance distribution, with different sample sizes, ρ =0.5 and 5%, 10%, and 20% of outliers.

Sample size	Par.	OLS	RIDGE	RLAV	WRID	WRLAV
25	$\hat{\beta}_1$	13.3098	1.0145	0.9874	0.9722	0.9570
	$\hat{\beta}_2$	55.1691	0.9963	0.9733	0.9586	0.9547
	$\hat{\beta}_3$	64.8247	1.0055	0.9923	0.9648	0.9608
100	$\hat{\beta}_1$	53.4601	0.9723	0.9721	0.9627	0.9609
	$\hat{\beta}_2$	32.7594	0.9747	0.9734	0.9624	0.9606
	$\hat{\beta}_3$	28.9825	0.9694	0.9689	0.9601	0.9599
25	$\hat{\beta}_1$	35.1460	1.0300	0.9976	0.9755	0.9705
	$\hat{\beta}_2$	55.2222	1.0057	0.9863	0.9690	0.9650
	$\hat{\beta}_3$	65.0266	1.0057	0.9929	0.9716	0.9708
100	$\hat{\beta}_1$	34.1402	0.9890	0.9826	0.9690	0.9673
	$\hat{\beta}_2$	20.6583	0.9832	0.9807	0.9723	0.9692
	$\hat{\beta}_3$	98.6807	0.9811	0.9801	0.9666	0.9664
25	$\hat{\beta}_1$	13.2081	1.0312	0.9955	1.0219	0.9842
	$\hat{\beta}_2$	52.8313	1.0004	0.9896	0.9885	0.9800
	$\hat{\beta}_3$	65.5664	1.0309	0.9970	1.0207	0.9805

Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

Sample size	Par.	OLS	RIDGE	RLAV	WRID	WRLAV
100	$\hat{\beta}_1$	34.0129	0.9792	0.9787	0.9760	0.9749
	$\hat{\beta}_2$	21.7161	0.9852	0.9826	0.9789	0.9717
	$\hat{\beta}_3$	28.3628	0.9845	0.9842	0.9736	0.9718

Table (8): The values of RMSE for the parameters $\hat{\beta}_1, \hat{\beta}_2$ *and* $\hat{\beta}_3$ used Bisquare weighted function generated Cauchy disturbance distribution, with different sample sizes, ρ=0.99 and 5%, 10%, and 20% of outliers.

Sample size	Par.	OLS	RIDGE	RLAV	WRID	WRLAV
25	$\hat{\beta}_1$	533.9535	17.4457	10.8617	7.3027	4.2407
	$\hat{\beta}_2$	306.7416	16.7671	11.0739	7.9074	3.9128
	$\hat{\beta}_3$	395.8723	16.8982	11.5783	8.3031	4.3479
100	$\hat{\beta}_1$	23.7610	7.7643	7.1568	3.6584	2.0102
	$\hat{\beta}_2$	202.8857	8.1447	7.6190	3.6016	1.9599
	$\hat{\beta}_3$	127.4180	7.9230	7.3542	3.3386	1.9566
25	$\hat{\beta}_1$	535.1287	17.6952	10.0992	6.7466	6.0109
	$\hat{\beta}_2$	305.6981	16.1171	9.3314	6.6378	5.3983
	$\hat{\beta}_3$	396.6405	17.0741	10.5737	7.1043	6.4838
100	$\hat{\beta}_1$	245.0834	7.9065	7.0949	3.5252	2.9020
	$\hat{\beta}_2$	204.1796	7.9547	7.2031	3.2981	2.8233
	$\hat{\beta}_3$	129.2237	8.1316	7.3630	2.9495	2.8870
25	$\hat{\beta}_1$	532.8477	16.8161	8.0062	15.3638	5.9805
	$\hat{\beta}_2$	294.1369	16.4734	7.7864	14.9052	6.7213
	$\hat{\beta}_3$	398.6997	17.0857	8.4902	15.6016	7.2659
100	$\hat{\beta}_1$	176.8833	8.1317	6.9079	5.8913	3.7897
	$\hat{\beta}_2$	117.4754	7.7541	6.6365	6.0833	3.1571
	$\hat{\beta}_3$	130.4582	7.9429	6.9430	6.1820	3.6424

Tikrit Journal of Administrative and Economic Sciences, Vol. 19, No. Special Issue, Part (5): 244-263
Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

Table (9): The values of RMSE for the parameters $\hat{\beta}_1, \hat{\beta}_2$ *and* $\hat{\beta}_3$ used Bisquare weighted function generated Student-t disturbance distribution, with different sample sizes, ρ =0.0 and 0% of outliers.

Sample size	Par.	OLS	RIDGE	RLAV	WRID
25	$\hat{\beta}_1$	0.3804	0.6103	0.5934	0.6077
	$\hat{\beta}_2$	0.3858	0.6060	0.5888	0.6065
	$\hat{\beta}_3$	0.3668	0.6027	0.5954	0.6088
100	$\hat{\beta}_1$	0.1669	0.5873	0.5843	0.5843
	$\hat{\beta}_2$	0.1674	0.5846	0.5818	0.5830
	$\hat{\beta}_3$	0.1695	0.5836	0.5811	0.5879

Table (10). The values of RMSE for the parameters $\hat{\beta}_1, \hat{\beta}_2$ *and* $\hat{\beta}_3$ used Bisquare weighted function generated Student-t disturbance distribution, with different sample sizes, ρ =0.5 and 5%, 10%, and 20% of outliers.

Sample size	Par.	OLS	RIDGE	RLAV	WRID	WRLAV
25	$\hat{\beta}_1$	7.6786	1.0086	0.9691	0.9501	0.9499
	$\hat{\beta}_2$	7.4330	1.0071	0.9778	0.9500	0.9481
	$\hat{\beta}_3$	7.6790	0.9913	0.9727	0.9500	0.9497
100	$\hat{\beta}_1$	3.7678	0.9643	0.9622	0.9530	0.9517
	$\hat{\beta}_2$	3.7496	0.9642	0.9626	0.9528	0.9519
	$\hat{\beta}_3$	3.8127	0.9629	0.9660	0.9529	0.9518
25	$\hat{\beta}_1$	10.4723	1.0125	0.9800	0.9684	0.9612
	$\hat{\beta}_2$	10.2496	1.0254	0.9841	0.9634	0.9603
	$\hat{\beta}_3$	10.2897	1.0033	0.9773	0.9605	0.9592
100	$\hat{\beta}_1$	5.3210	0.9796	0.9752	0.9645	0.9631
	$\hat{\beta}_2$	5.1595	0.9685	0.9675	0.9638	0.9621
	$\hat{\beta}_3$	5.3121	0.9747	0.9720	0.9661	0.9622
25	$\hat{\beta}_1$	15.7839	1.0147	0.9829	1.0060	0.9739
	$\hat{\beta}_2$	15.5025	1.0471	0.9921	1.0282	0.9726
	$\hat{\beta}_3$	15.3365	1.0022	0.9858	0.9954	0.9689

Doi: www.doi.org/10.25130/tjaes.19.SP3.5.16

Sample size	Par.	OLS	RIDGE	RLAV	WRID	WRLAV
100	$\hat{\beta}_1$	6.9710	0.9810	0.9803	0.9768	0.9699
	$\hat{\beta}_2$	7.2950	0.9807	0.9776	0.9726	0.9713
	$\hat{\beta}_3$	7.0346	0.9860	0.9810	0.9852	0.9699

Table (11). The values of RMSE for the parameters $\hat{\beta}_1, \hat{\beta}_2$ *and* $\hat{\beta}_3$ used Bisquare weighted function generated Student-t disturbance distribution, with different sample sizes, ρ=0.99 and 5%, 10%, and 20% of outliers.

Sample size	Par.	OLS	RIDGE	RLAV	WRID	WRLAV
25	$\hat{\beta}_1$	35.7909	16.9097	4.8740	2.1211	2.0392
	$\hat{\beta}_2$	35.7881	16.6886	5.2141	2.0650	1.9496
	$\hat{\beta}_3$	37.4684	17.1436	4.7739	2.0753	2.0444
100	$\hat{\beta}_1$	18.7748	8.1367	5.8711	1.9110	1.2865
	$\hat{\beta}_2$	18.2211	8.0609	5.5965	1.8869	1.3464
	$\hat{\beta}_3$	18.0355	8.1939	5.7841	1.9842	1.2541
25	$\hat{\beta}_1$	50.6871	16.7896	3.8586	4.8435	1.9512
	$\hat{\beta}_2$	50.1287	16.5573	3.3905	4.4809	1.9321
	$\hat{\beta}_3$	49.7171	16.6716	3.6346	4.4749	1.8527
100	$\hat{\beta}_1$	24.9702	8.4091	4.9818	3.5450	1.5262
	$\hat{\beta}_2$	24.9386	8.1083	4.6778	3.5537	1.6115
	$\hat{\beta}_3$	26.0116	8.4156	5.0253	3.5172	1.5022
25	$\hat{\beta}_1$	76.8565	17.1798	2.4207	15.2161	2.1436
	$\hat{\beta}_2$	73.6097	16.6150	2.4231	14.7406	2.0971
	$\hat{\beta}_3$	74.9847	17.1175	1.9474	15.2240	1.9691
100	$\hat{\beta}_1$	33.7089	8.3943	3.5631	7.0242	1.6815
	$\hat{\beta}_2$	34.9399	8.4871	3.7719	6.9805	1.8238
	$\hat{\beta}_3$	34.5884	8.3546	3.6453	6.9442	1.7090

6. Conclusion

Earlier to data analysis, existing outliers should be examined, and then the necessary tests should be performed to determine whether the underlying assumptions are met or not. Following that, the appropriate estimate approaches should be carried out.

As a result, the outcomes obtained in this paper from the application of the proposed method WRLAV for real data and simulation studies. Show that from the numerical example in Table (2) to the results obtained from the simulation studies, listed in Tables (4), (5), (7), (8), (10) and (11) the WRLAV estimates of parameters are too close to the true values. And performed best estimate when compared to the methods WRID, RLAV, Ridge regression, and OLS, followed by WRID, RLAV, RIDGE and OLS estimates.

When probability disturbance distributions of errors are normal, Cauchy, or Student-t, the WRLAV estimator is found to be more successful than the other estimators. However, in Tables 3 and 9, the OLS estimator outperforms other estimators in terms of having the least RMSE when there is no multicollinearity among the explanatory variables and no outliers in the data set.

Except when the error disturbance is a Cauchy distribution, the RMSE values in Table 6 for WRID are less than OLS.

When such abnormalities occur, however, OLS completely fails. The poor performance of OLS estimators when dealing with non-normal error distributions demonstrates the necessity for other techniques. Furthermore, the RMSE Values for RLAV in Table 9 are smaller than the WRID and Ridge regression RMSE Values. When the disturbance distribution is Student-t, however, it is higher than OLS.

Finally, because the RMSE for WRLAV is so small, it is preferred, as it has the best performance and most reliable results.

When comparing the robust ridge estimators RLAV, WRID, and WRLAV, it is clear that the WRLAV outperforms the WRID estimator for a large variety of error distribution types and degrees of multicollinearity. When multicollinearity and outliers are present, the simulation experiments clearly show that the WRLAV estimator is the preferred option in compared to the other estimators.

References

1. Subhash C. Narula, Paulo H. N. Saldiva, Carmen D. S. Andre, Silvia N. Elian, Aurea Favero Ferreira, Vera Capelozzi. The minimum sum of absolute errors regression: a robust alternative to the least squares regression, https://doi.org/10.1002/(SICI), 1999.
2. Rice J.R., White J.S. Norms for smoothing and estimation. SIAM review,
3. https://doi.org/10.1137/1006061. Vol. 6, Iss. 3, 1964. Pp. 243–256.
4. Peter J. Huber. Robust regression: Asymptotic, conjectures and Monte Carlo, the Annals of Statistics, Vol. 1, No. 5, 1973. Pp. 799-821.
5. Rousseeuw P.J. and Leroy A.M. Robust regression and outlier detection. Books. 2005 New York: Wiley.
6. Kutner M.H. Nachtsheim C.J. Neter J and Wasserman W. Applied Linear Regression Models. 4th edition. McGraw-Hill/ Irwin series Operations and decision sciences. Avenue of the Americas, 2004. ISBN 9780071232524.
7. Hoerl A.E. and Kennard R.W. Ridge regression: Biased estimation for non-orthogonal problems. Technometric, Vol 12 issue 1. 1970. Published online: 09 Apr 2012 PP. 55-67.
8. Dawoud, I, Abonazel, MR., Azazy, AR., (2022), Quantile-Based Estimation of Biasing Parameter for Kibria-Lukman Regression Estimator: Simulation and Application to Portland Cement Data. International Journal of Mechanical Engineering 7 (1), 6276-6284.
9. Hoerl A.E.; Kannard R.W. and Baldwin K.F. Ridge regression: some simulations. Communications in Statistics. Taylor and Francis Volume 4, Issue 2, 1975. Published online: 27 Jun 2007, Pp 105-123.
10. Montgomery D.C., Peck E. A. and Vining G.G. Introduction to Linear Regression Analysis. Canada: 2021. John Wiley and Sons fourth edition.
11. Wagner H.M. Linear programming techniques for regression analysis. Journal of the American Statistical Association, 1959. Taylor & Francis. Volume 54, Issue 285, Published online: 11 Apr 2012, Pp 206-212.
12. Andrews D.F. A robust method for multiple linear regression. Technometrics, Taylor & Francis, Volume 16, Issue 4, 1974. Published online: 09 Apr 2012. Pp. 523–531.
13. Claerbout J.F. and Muir F. Robust modeling with erratic data. Geophysics, Volume 38, Issue 5, 1973. pp. 826–844, https://doi.org/10.1190/11440378.
14. Dawoud, I, Abonazel, MR., Awwad, FA. (2022). Generalized Kibria-Lukman Estimator: Method, Simulation, and Application. Frontiers in Applied Mathematics and Statistics. https://doi.org/10.3389/fams-2022.880086.
15. Dielman T. E. and Rose E.L. Forecasting in least absolute value regression with autocorrelated errors. A small-sample study. International Journal of Forecasting Volume 10, Issue 4, Pages 539-547, https://doi.org/10.1016/0169-2070-1994.
16. Dielman T. and Pfaffenberger R. Least absolute value regression: Necessary sample sizes to use normal theory inference procedures. Decision Sciences, Volume 19, Issue 4, PP 734-743, https://doi.org/10.1111/j.1540-5915.1988.tb00298-1988.

17. Efron B. Computer-intensive methods in statistical regression. SIAM Review, Vol. 30, Issue 3, Pp 421-449, https://doi.org/10.1137/1030092-1988.
18. Gabriela Stangenhaus; Subhash C. Narula Inference procedures for the L1 regression, Computational Statistics & Data Analysis. Volume 12, Issue 1, 1991, Pages 79-85.
19. Dawoud, I, Abonazel, MR. (2021). Robust Dawoud-Kibria estimator for handling multicollinearity and outliers in the linear regression model. Journal of Statistical Computation and Simulation 91 (17), 3678-3692.
20. Dawoud, I, Abonazel, MR, Awwad FA. (2022). Modified Liu estimator to address the multicollinearity problem in regression models: A new biased estimation class, Scientific African, 17, e01372.
21. Dielman T.E. and Rose E.L. A note on hypothesis testing in LAV multiple regression: A small sample comparison. Computational Statistics and Data Analysis. Volume 21, Issue 4, Pp 463-470, https://doi.org/10.1016/0167-9473-1996.
22. Dawoud, I., Lukman AF., Haadi, AR. (2022). A new biased regression estimator: Theory, simulation and application, Scientific African, 15, e01100. Dielman T. E. and Pfaffenberger R.C. Tests of linear hypotheses in LAV regression. Communications in Statistics -Simulation and Computation. Volume 19, Issue 4, Pp 1179-1199, 1990. Published online: 27 Jun 2007, https://doi.org/10.1080/0361-0919008812911.
23. Penrose K.W., Nelson A.G. and Fisher A.G. Generalized Body Composition Prediction Equation for Men Using Simple Measurement Techniques. Medicine and Science in Sports and Exercise. 17(2), 189 1985.
24. Rousseeuw P.J. and Van Zomeren B.C. Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association. Volume 85, Issue 411 Pp 633-639. 1990. Published online: 27 Feb 2012.