Journal of Modern Applied Statistical Methods

Volume 14 | Issue 2 Article 8

11-1-2015

Two Stage Robust Ridge Method in a Linear Regression Model

Adewale Folaranmi Lukman
Ladoke Akintola University of Technology, wale3005@yahoo.com

Oyedeji Isola Osowole
University of Ibadan, dosowole@yahoo.com

Kayode Ayinde
Ladoke Akintola University of Technology, kayinde@lautech.edu.ng

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
Lukman, Adewale Folaranmi; Osowole, Oyedeji Isola; and Ayinde, Kayode (2015) "Two Stage Robust Ridge Method in a Linear Regression Model," *Journal of Modern Applied Statistical Methods*: Vol. 14 : Iss. 2 , Article 8.
DOI: 10.22237/jmasm/1446350820

Two Stage Robust Ridge Method in a Linear Regression Model

Cover Page Footnote
I want to acknowledge the wonderful contribution of my supervisor Dr Osowole O.I and the following people: Dr Kayode Ayinde and Professor Hussein towards the successful completion of this study.

Journal of Modern Applied Statistical Methods
November 2015, Vol. 14, No. 2, 53-67.
Copyright © 2013 JMASM, Inc.
ISSN 1538 − 9472

Two Stage Robust Ridge Method in a Linear Regression Model

Adewale Folaranmi Lukman
Ladoke Akintola
University of Technology
Ogbomoso, Nigeria

Oyedeji Isola Osowole
University of Ibadan
Ibadan, Nigeria

Kayode Ayinde
Ladoke Akintola
University of Technology
Ogbomoso, Nigeria

Two Stage Robust Ridge Estimators based on robust estimators M, MM, S, LTS are examined in the presence of autocorrelation, multicollinearity and outliers as alternative to Ordinary Least Square Estimator (OLS). The estimator based on S estimator performs better. Mean square error was used as a criterion for examining the performances of these estimators.

Keywords: Two Stage Least Square, Ridge Estimator, Ordinary Least Square, Robust Estimators, Two Stage Robust Ridge Estimator.

Introduction

Multiple regressions routinely assess the degree of relationship between one dependent variable and a set of independent variables. The Ordinary Least Squares (OLS) Estimator is most popularly used to estimate the parameters of regression model. Under certain assumptions, the estimator has some very attractive statistical properties which have made it one of the most powerful and popular estimators of regression model. A common violation in the assumption of classical linear regression model is the non-normal error terms. OLS estimator produces unstable prediction estimates when the assumption of normality of errors is not met (Ryan, 1996). Multiple regression methods also yield unstable results in the presence of outlier data points. When outliers occur in the data, the assumption of normally distributed errors is violated. An alternative strategy to deal with outliers is to accommodate them. Accommodation is accomplished by using any one of several robust regression estimation methods.

Adewale Folaranmi Lukman is a postgraduate student in the Department of Pure and Applied Statistics. Email at wale3005@yahoo.com. Dr. Oyedeji Isola Osowole is a Lecturer in the Department of Statistics. Email at dosowole@yahoo.com. Prof. Kayode Ayinde is a lecturer in the Department of Pure and Applied Mathematics. Email at: kayinde@lautech.edu.ng.

Also, the problem of autocorrelated error is another violation to the assumption of independence of error terms in classical linear regression model. The term autocorrelation may be defined as correlation between members of series of observations ordered in time as in time series data (Gujarati 1995). In the regression context, the classical linear regression model assumes that such autocorrelation does not exist in the disturbances ε_i. Symbolically

$$E\left(\varepsilon_i \varepsilon_j\right) = 0 \forall i \neq j \tag{1}$$

When this assumption breaks down, this is autocorrelation problem. A number of remedial procedures that rely on transformations of the variables have been developed. In order to correct for autocorrelation, one often uses Feasible Generalized Least Square (FGLS) procedures such as the Cochrane-Orcutt or Prais-Winsten two-step or the Maximum Likelihood Procedure or Two stage least Squares which are based on a particular estimator for the correlation coefficient (Green, 1993; Gujarati, 2003).

Another serious problem in regression estimation is multicollinearity. It is the term used to describe cases in which the explanatory variables are correlated. The regression coefficients possess large standard errors and some even have the wrong sign (Gujarati, 1995). In literature, there are various methods existing to solve this problem. Among them is the ridge regression estimator first introduced by Hoerl and Kennard (1970). Keijan (1993) proposed an estimator that is similar in form but different from the ridge regression estimator of Hoerl and Kennard. Ayinde and Lukman (2014) proposed some generalized linear estimator (CORC and ML) and principal components (PCs) estimator as alternative to multicollinearity estimation methods.

Inevitably, these problems can exist together in a data set. Holland (1973) proposed robust M-estimator for ridge regression to handle the problem of multicollinearity and outliers. Askin and Montgomery (1980) proposed ridge regression based on the M-estimates. Midi and Zahari (2007) proposed Ridge MM estimator (RMM) by combining the MM estimator and ridge regression. Samkar and Alpu (2010) proposed robust ridge regression methods based on M, S, MM and GM estimators. Maronna (2011) proposed robust MM estimator in ridge regression for high dimensional data. Eledum and Alkhaklifa (2012) proposed Generalized Two Stages Ridge Estimator (GTR) for the multiple linear model which suffers from both problem of autocorrelation AR (1) and multicollinearity.

The main objective of this study is to re-examine the study of Eledum and Alkhaklifa (2012). Efforts are made to correct the various assumptions violations of classical regression model which could have led into misleading conclusions. In this study, Two Stage Robust Ridge methods based on M, S, MM, LTS estimators are examined in the presence of outliers, autocorrelated errors and multicollinearity. A real life data considered in the study of Eledum and Alkhaklifa (2012) was used.

Outliers in least square regression

Barnett and Lewis (1994) define an outlier as an observation that appears inconsistent with the remainder of the data set. Outlier identification is important in OLS not only due to their impact on the OLS model, but also to provide insight into the process. These outlying cases may arise from a distribution different from the remaining data set. The distribution of the full dataset is contaminated in this instance. To statisticians, unusual observations are generally either outliers or 'influential' data points. In regression analysis, generally they categorize unusual observation (outliers) into three: outliers, high leverage points and influential observations. In other words, Hawkins (1980) pointed out that, an outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism.

Outliers are classified in three ways:

i. the change in the direction of response (Y) variable
ii. the deviation in the space of explanatory variable(s), deviated points in X-direction called leverage points and are also referred to as exterior X-space observation in this research, and
iii. The other is change in both directions (direction of the explanatory variable(s) and the response variable). According to Belsley, Kuh, and Welsch (1980), influential observations is one which either individual or together with several other observations have a demonstrably larger impact on the calculated values of various estimates than is the case for most of the other observations. Chatterjee and Hadi (1986) pointed out that, as with outliers, high leverage points need not be influential and influential observations are not necessarily high-leverage points. When an observation is considered to be both an outlier and influential, regression results are usually reported with and without the observation. When

observations are not outliers but are influential, it is less clear what should be done.

Robustness ideas in regression

One idea to deal with this problem is to identify outliers, remove them, and then to proceed as before assuming we now have an appropriate data set for the standard methods. If the true coefficients were known, then outliers would not be hard to detect. Look for the points corresponding to the largest residuals. The field of regression diagnostics attempts to address the issue of how to identify influential points and outliers, in the general case when we do not know the true coefficient values. When there is only have one outlier, some diagnostic methods work very well by looking at the effect of one at a time deletion of data points. Unfortunately it is much more difficult to diagnose outliers when there are many of them, especially if the outliers appear in groups. In these situations, it is necessary to deal with the phenomena of outlier masking. Outlier masking occurs when a set of outliers goes undetected because of the presence of another set of outliers. Often when outliers are used to fit the parameter values, the estimates are badly biased, leaving residuals on the true outliers that do not indicate that they actually are outliers. Once there are several outliers, deletion methods are no longer computationally feasible. Then it is necessary to look at the deletion of all subsets of data points below a suitably chosen maximum number of outliers.

Another approach to dealing with outliers is robust regression, which tries to come up with estimators that are resistant or at least not strongly affected by the outliers. In studying the residuals of a robust regression, perhaps true outliers can be found. In this field many different ideas have been proposed, including Least Trimmed Squares (LTS), Least Median of Squares (LMS), M-estimators, and GM-estimators or bounded-influence estimators and S-estimators.

Robust regression and outlier diagnostic methods end up being very similar. They both involve trying to find outliers and trying to estimate coefficients in a manner that is not overly influenced by outliers. What is different is the order in which these two steps are performed. When using diagnostics, look for the outliers first and then once they have been removed use OLS on this clean data set for better estimates. Robust regression instead looks to find better robust estimates first and given these estimates, we can discover the outliers by analyzing the residuals.

Methodology

The data set was extracted from the study of Eledum and Alkhaklifa (2012); it represents the product in the manufacturing sector, the imported intermediate, the capital commodities and imported raw materials, in Iraq in the period from 1960 to 1990. An econometric model for this study is specified as follows:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon_t, t = 1, 2, \ldots, 31 \tag{2}$$

Where

Y = Product value in the manufacturing sector
X_1 = The value of the imported intermediate
X_2 = Imported capital commodities
X_3 = Value of imported raw materials
β_1, β_2, β_3 are the regression coefficients.

M-estimation procedure

The most common general method of robust regression is M-estimation, introduced by Huber (1964) that is nearly as efficient as OLS. Rather than minimize the sum of squared errors as the objective, the M-estimate minimizes a function ρ of the errors. The M-estimate objective function is,

$$\min \sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \min \sum_{i=1}^{n} \rho\left(\frac{y_i - X'\hat{\beta}_i}{s}\right) \tag{3}$$

where s is an estimate of scale often formed from linear combination of the residuals. The function ρ gives the contribution of each residual to the objective function. A reasonable ρ should have the following properties:
$\rho(e) \geq 0, \rho(0) = 0, \rho(e) = \rho(-e)$, and $\rho(\mathrm{e}_i) \geq \rho(e_i')$ for $|e_i| \geq |e_i'|$
the system of normal equations to solve this minimization problem is found by taking partial derivatives with respect to β and setting them equal to 0, yielding,

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - X'\hat{\beta}_i}{s}\right) X_i = 0 \tag{4}$$

where ψ is a derivative of ρ. The choice of the ψ function is based on the preference of how much weight to assign outliers. Newton-Raphson and iteratively reweighted Least Squares (IRLS) are the two methods to solve the M-estimates nonlinear normal equations. IRLS expresses the normal equations as,

$$X'WX\hat{\beta} = X'Wy \tag{5}$$

MM estimator

MM-estimation is special type of M-estimation developed by Yohai (1987). MM--estimators combine the high asymptotic relative efficiency of M-estimators with the high breakdown of class of estimators called S-estimators. It was among the first robust estimators to have these two properties simultaneously. The 'MM' refers to the fact that multiple M-estimation procedures are carried out in the computation of the estimator. Yohai (1987) describes the three stages that define an MM-estimator:

1. A high breakdown estimator is used to find an initial estimate, which we denote $\tilde{\beta}$ the estimator need to be efficient. Using this estimate the residuals, $r_i(\beta) = y_i - x_i^T\tilde{\beta}$ are computed.
2. Using these residuals from the robust fit and $\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i}{s}\right) = k$ where k is a constant and the objective function ρ, an M-estimate of scale with 50% BDP is computed. This $s\left(r_i\left(\tilde{\beta}\right),\ldots,r_n\left(\tilde{\beta}\right)\right)$ is denoted s_n. The objective function used in this stage is labeled ρ_0.
3. The MM-estimator is now defined as an M-estimator of β using a redescending score function, $\varphi_1(u) = \frac{\partial\rho_1(u)}{\partial u}$, and the scale estimate s_n obtained from stage 2. So an MM-estimator $\hat{\beta}$ defined as a solution to

$$\sum_{i=1}^{n} x_{ij}\varphi_1\left(\frac{y_i - x_i^T\tilde{\beta}}{s_n}\right) = 0,\ j = 1,\ldots,p. \tag{6}$$

S estimator

Rousseeuw and Yohai (1984) introduced S estimator, which is derived from a scale statistics in an implicit way, corresponding to $s(\theta)$ where $s(\theta)$ is a certain type of robust M-estimate of the scale of the residuals $e_1(\theta), \ldots, e_n(\theta)$. They are defined by minimization of the dispersion of the residuals: minimize $S\left(e_1(\theta),\ldots,e_n\left(\hat{\theta}\right)\right)$ with final scale estimate $\hat{\sigma} = S\left(e_1(\theta),\ldots,e_n\left(\hat{\theta}\right)\right)$. The dispersion $\left(e_1(\theta),\ldots,e_n\left(\hat{\theta}\right)\right)$ is defined as the solution of

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{e_i}{s}\right) = k \tag{7}$$

K is a constant and $\rho\left(\frac{e_i}{s}\right)$ is the residual function. Rousseeuw and Yohai (1984) suggest Tukey's biweight function given by:

$$\rho(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} + \frac{x^6}{6c^4} & \text{for } |x| \le c \\ \frac{c^2}{6} & \text{for } |x| > c \end{cases} \tag{8}$$

Setting $c = 1.5476$ and $K = 0.1995$ gives 50% breakdown point (Rousseeuw & Leroy, 1987).

LTS estimator

Rousseeuw (1984) developed the least trimmed squares estimation method. Extending from the trimmed mean, LTS regression minimizes the sum of trimmed squared residuals. This method is given by,

$$\hat{\beta}_{LTS} = \arg\min Q_{LTS}(\beta) \tag{9}$$

where $Q_{LTS}(\beta) = \sum_{i=1}^{h} e_i^2$ such that $e_{(1)}^2 \le e_{(2)}^2 \le e_{(3)}^2 \le \ldots \le e_{(n)}^2$ are the ordered squares residuals and h is defined in the range $\frac{n}{2} + 1 \le h \le \frac{3n+p+1}{4}$, with n and p being

sample size and number of parameters respectively. The largest squared residuals are excluded from the summation in this method, which allows those outlier data points to be excluded completely. Depending on the value of h and the outlier data configuration. LTS can be very efficient. In fact, if the exact numbers of outlying data points are trimmed, this method is computationally equivalent to OLS.

Two Stage Robust Ridge Estimator

Two Stage Ridge Regression approach used by Eledum and Alkhaklifa (2012) and Robust Ridge Regression Methods adopted by Samkar and Alpu (2010) are combined in this study to obtain Two Stage Robust Ridge Regression. This method is adopted to deal with the problem of autocorrelated error, outliers and, multicollinearity sequentially. Consider the Linear regression model:

$$Y = X\beta + u_t \tag{10}$$

X is an $n \times p$ matrix with full rank, Y is a $n \times 1$ vector of dependent variable, β is a $p \times 1$ vector of unknown parameters, and ε is the error term such that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I$ and assume that the error term follows the AR(1) scheme, namely,

$$u_t = \rho u_{t-1} + \varepsilon_t, -1 < \rho < 1 \tag{11}$$

ε_t is a white noise error term such that $\varepsilon_t \sim N(0, \sigma^2 I)$
Premultiply equation (10) by P we obtain:

$$PY = PX\beta + PU \tag{12}$$

Equivalently, equation (12) becomes:

$$Y^* = X^*\beta + U^* \tag{13}$$

P is a non-singular matrix such that $P\Omega P' = I$ which implies $PP' = \Omega^{-1}$, $U^* \sim N(0, \sigma^2 I)$, $Y^* = PY$, $X^* = PX$, and $U^* = PU$.

Therefore, we can apply Robust Estimators to the transformed model (5) and obtain Two Stage Robust Estimator.

$$\hat{\beta}_{TRE} = \left(X^{*\prime} X^{*} \right)^{-1} X^{*\prime} Y^{*} = \left(X'P'PX \right)^{-1} X'P'PY$$

$$\hat{\beta}_{TRE} = \left(X'\Omega^{-1} X \right)^{-1} X'\Omega^{-1} Y \tag{14}$$

The variance-covariance matrix becomes:

$$V\left(\hat{\beta}_{TRE} \right) = \sigma^{2} \left(X'\Omega^{-1} X \right)^{-1} (3.6) \tag{15}$$

where

$$E\left(UU' \right) = \sigma^{2}\Omega = \sigma^{2} \begin{bmatrix} 1 & \rho & \rho^{2} & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^{2} & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}$$

$$\sigma^{2} = \frac{\sigma_{\varepsilon}^{2}}{1-\rho^{2}}$$

and the inverse of Ω is

$$\Omega^{-1} = \frac{1}{1-\rho^{2}} \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^{2} & -\rho & \cdots & 0 \\ 0 & -\rho & 1+\rho^{2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Consider, $(n-1) \times n$ matrix P^{*} for transformation.

$$P^* = \begin{bmatrix} -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ 0 & 0 & -\rho & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Therefore, $P^{*}{}'P^{*} = P$ by adding a new row with $\sqrt{1-\rho^2}$ in the first position and zero elsewhere.

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 1 & 1 & -\rho & 1 \end{bmatrix}$$

Then

$$Y^* = PY = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

$$X^* = PX = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ 1 & X_{31} & X_{32} & \cdots & X_{3p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

$$\Omega = P'P = \begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

However, the estimate obtained from applying Robust Estimators to the transformed model is used to obtain the ridge parameter K which is used in the Ridge Estimator since the estimates obtain from OLS will be inefficient when we have the problem of outliers or non-normal error term.

Results

From Table 1, it can be seen that estimation based on the OLS estimator produces residuals that reveals the problem of autocorrelation (DW p-value=0.0005) and multicollinearity (VIF>10) simultaneously. The problem of multicollinearity might be the reason for the wrong sign in the value of imported raw materials. We handle the problem of autocorrelation in Table 2 by transforming the data set. The original data set is transformed using $\hat{\rho} = 0.547$ (from Table 1) to correct the problem of autocorrelation by applying Two Stage Least Squares. Table 2 shows that the new data set obtain through transformation suffered the problem of non-normal error term using Jarque-Bera Statistic and Table 3 also shows the presence of bad leverages using robust diagnostics which might be the reason for the non-normality of the error term. The data set still suffered the problem of multicollinearity (VIF>10) as revealed in Table 2. Due to the presence of bad leverages OLS will not correctly estimate the parameters in the model. This prompts the use of the Two Stage Robust Estimators in Table 4. LTS and S estimators perform better than other estimators when we have leverages and outliers in y axis (bad leverages) in terms of the MSE (B). But the coefficient of LTS seems to be much different from the class of other estimators. We then prefer to consider S estimator in its stead. Due to the occurrence of both problem of multicollinearity and bad leverages in the new data set, we then use the Ridge combined with S estimator adopted from the concept of Samkar and Alpu (2010) to compute the ridge parameter. Geometric version of the ridge parameter

proposed by Kibria (2003) was used $\hat{K}_{GM} = \dfrac{\hat{\sigma}^2}{\left(\prod_{i=1}^{p} \alpha_i\right)^{\frac{1}{p}}}$ where $\hat{\sigma}^2$ is the variance obtained from S estimator and α_i is the obtained coefficient.

Table 1. Ordinary Least Square (OLS)

Variable	Coefficient	Std. Error	*p*-value	VIF
*X*1	0.208	0.218	0.348	128.26
*X*2	0.921	0.196	0.000	103.43
*X*3	-1.34	0.162	0.415	70.87
R-squared	0.9896	DW	0.0005	
Jarque-Bera *p*-value	0.2493	σ^2	0.0111	
RHO	0.547			

Table 2. Two Stage Least Square (TS)

Variable	Coefficient	Std. Error	*p*-value	VIF
*X*1T	0.200	0.160	0.2211	26.839
*X*2T	0.963	0.191	0.0000	38.358
*X*3T	-0.1790	0.127	0.1687	16.904
R-squared	0.9735	DW *p*-value	0.2332	
Jarque-Bera *p*-value	0.0732	σ^2	0.028	
RHO	0.11			

Table 3. Robust Diagnostics

Observation	Mahalanobis	Robust MCD Distance	Leverage	Standardized Robust Residual	Outlier
12	1.5024	5.8641	*	4.7737	*
14	0.9716	3.0421		4.9055	*
15	4.6559	29.4708	*	9.1178	*
16	1.0615	8.2135	*	11.2653	*
17	1.6992	8.6846	*	1.4033	
18	2.2534	19.0971	*	-0.5591	
20	3.0865	24.3649	*	-2.4415	
21	3.8595	26.6181	*	0.4649	
22	1.2315	8.8886	*	0.4301	
30	3.421	3.0381		16.2649	*
31	1.2827	1.1007		-8.5191	*

Table 4. Two Stage Robust Estimators and OLS

Variables	OLS	TS	M	MM	S	LTS
*X*1T	0.208	0.200	0.329	0.328	0.346	0.032
*X*2T	0.921	0.963	0.976	0.976	0.963	1.723
*X*3T	-1.34	-0.1790	-0.228	-0.228	-0.221	-0.648
R-squared	0.9896	0.9735	0.7918	0.7939	0.8023	0.9951
σ^2	0.0111	0.028	0.0102	0.019	0.017	0.003
MSE(*B*)	0.1122	0.0782	0.0324	0.0303	0.0272	0.029

Table 5. Two Stage Robust Ridge Estimators

Variables	Coefficient	VIF
*X*1	0.3443	1.2972
*X*2	0.4278	1.0011
*X*3	0.1836	1.5526
MSE(β)	0.071687	
K	0.097	

Conclusion

OLS performs better than other estimators when there is no violation of assumptions in Classical Linear Regression Model. In this study the problem of autocorrelation was handled using Two Stage Least Square. The problem of multicollinearity and outlier are still presents. OLS will not be efficient because of the present of both problem therefore we apply Robust Methods to the transformed data. *S* and LTS estimators perform better than other Robust Methods in terms of the MSE. *S* estimator was chosen because LTS does not correctly estimate the model when compared with other estimators. Ridge parameter *K* is then obtained using the estimates obtain from *S* estimation. Robust ridge estimates was computed. Two stage robust ridge estimator performs better than the Generalized Two stage ridge regression proposed by Hussein et al (2012). This is because after the problem of autocorrelation was corrected in the study of Hussein et al (2012), the data sets still suffered the problem of multicollinearity and outlier. This was corrected in this study by obtaining the ridge parameter using a robust estimator instead of OLS.

Authors' note

We acknowledge the assistance of Osowole O.I, Kayode Ayinde, and X Hussein.

References

Askin, G. R., & Montgomery, D. C. (1980). Augmented robust estimators. *Technometrics*, *22*(3), 333-341. doi:10.1080/00401706.1980.10486164

Ayinde, K. & Lukman, A. F. (2014). Combined estimators as alternative to multicollinearity estimation methods. *International Journal of Current Research*, *6*(1), 4505-4510.

Barnett, V. & Lewis, T. (1994). *Outliers in statistical data* (Vol. 3). New York: Wiley.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. New York: Wiley.

Chatterjee, S., Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science*, *1*(3), 379-416. doi:10.1214/ss/1177013622

Eledum, H. Y. A., Alkhalifa, A. A. (2012). Generalized two stages ridge regression estimator for multicollinearity and autocorrelated errors. *Canadian Journal on Science and Engineering Mathematics*, *3*(3), 79-85.

Green, W.H. (1993). *Econometric analysis* (2nd ed.). New York: MacMillan.

Gujarati, D. N. (1995). *Basic econometrics* (3rd ed.). New York: McGraw-Hill.

Gujarati, D. N. (2003). *Basic econometrics* (4th ed.) (pp. 748, 807). New Delhi: Tata McGraw-Hill.

Hawkins, D. M. (1980), *Identification of Outliers*, London: Chapman & Hall.

Hoerl, A. E. & Kennard, R.W. (1970). Ridge regression biased estimation for nonorthognal problems. *Technometrics*, *12*(1), 55-67. doi:10.1080/00401706.1970.10488634

Holland, P. W. (1973). *Weighted ridge regression: Combining ridge and robust regression methods* (NBER Working Paper No.11). Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w0011

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, *35*(1), 73-101.doi:10.1214/aoms/1177703732

Keijan, L. (1993). A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods*, *22*(2), 393-402. doi:10.1080/03610929308831027

Maronna, R. A. (2011). Robust ridge regression for high-dimensional data. *Technometrics*, *53*(1), 44-53. doi:10.1198/TECH.2010.09114

Midi, H., & Zahari, M. (2007). A simulation study on ridge regression estimators in the presence of outliers and multicollinearity. *Jurnal Teknologi*, *47*(C), 59-74. doi:10.11113/jt.v47.261

Rousseeuw, P.J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, 79, 871–880. doi:10.1080/01621459.1984.10477105

Rousseeuw P. J., & Leroy A. M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley-Interscience.

Rousseeuw, P. J. & Van Driessen, K. (1998). Computing LTS regression for large data sets. *Technical Report*, University of Antwerp, submitted.

Rousseeuw P. J., & Yohai, V. (1984). Robust regression by means of S-estimators. In J. Franke, W. Härdle & D. Martin (Eds.), *Robust and Nonlinear Time Series Analysis* (Vol. 26, pp. 256-272). New York: Springer-Verlag. doi:10.1007/978-1-4615-7821-5_15

Ryan, T.P. (1996). *Modern Regression Method*. Canada: John Wiley & Sons.

Samkar, H. & Alpu, O. (2010). Ridge regression based on some robust estimators. *Journal of Modern Applied Statistical Methods*, *9*(2), 17.

Yohai, V. J. (1987). High breakdown point and high breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, *15*(2), 642-656. doi:10.1214/aos/1176350366