Sains Malaysiana 53(4)(2024): 907-920
http://doi.org/10.17576/jsm-2024-5304-14

A Remedial Measure of Multicollinearity in Multiple Linear Regression in the Presence of High Leverage Points

(Pemulihan Ukuran Multikolinearan dalam Model Regresi Linear Berganda dengan Kehadiran Titik Terpencil)

SHELAN SAIED ISMAEEL[1], HABSHAH MIDI[2,*] & KURDISTAN M. TAHER OMAR[1]

[1]*Department of Mathematics, Faculty of Science, University of Zakho, Iraq*
[2]*Faculty of Science and Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

Received: 14 March 2023/Accepted: 5 March 2024

ABSTRACT

The ordinary least squares (OLS) is the widely used method in multiple linear regression model due to tradition and its optimal properties. Nonetheless, in the presence of multicollinearity, the OLS method is inefficient because the standard errors of its estimates become inflated. Many methods have been proposed to remedy this problem that include the Jackknife Ridge Regression (JAK). However, the performance of JAK is poor when multicollinearity and high leverage points (HLPs) which are outlying observations in the *X*- direction are present in the data. As a solution to this problem, Robust Jackknife Ridge MM (RJMM) and Robust Jackknife Ridge GM2 (RJGM2) estimators are put forward. Nevertheless, they are still not very efficient because they suffer from long computational running time, some elements of biased and do not have bounded influence property. This paper proposes a robust Jackknife ridge regression that integrates a generalized M estimator and fast improvised Gt (GM-FIMGT) estimator, in its establishment. We name this method the robust Jackknife ridge regression based on GM-FIMGT, denoted as RJFIMGT. The numerical results show that the proposed RJFIMGT method was found to be the best method as it has the least values of RMSE and bias compared to other methods in this study.

Keywords: High leverage points; jackknife; MM-estimator; multicollinearity; ridge regression

ABSTRAK

Kaedah kuasadua terkecil sering digunakan dalam model linear regresi berganda kerana tradisi dan sifatnya yang optimal. Walau bagaimanapun, dalam kehadiran multikolinearan, kaedah *OLS* tidak cekap disebabkan penganggar ralat piawai menjadi besar. Banyak kaedah telah dicadangkan bagi mengatasi masalah ini termasuk kaedah *Jackknife Ridge Regression* (*JAK*). Namun, prestasi kaedah *JAK* sangat lemah dengan kehadiran multikolinearan dan titik tuasan tinggi iaitu cerapan terpencil dalam arah *X*. Sebagai penyelesaian bagi masalah ini, penganggar *Robust Jackknife Ridge MM* (*RJMM*) dan penganggar *Jackknife Ridge GM2* (*RJGM2*) di ketengahkan. Walau bagaimanapun, kaedah ini masih tidak cukup cekap kerana mereka mengambil masa pengiraan yang panjang, mempunyai unsur kepincangan dan tidak mempunyai sifat pengaruh terbatas. Kertas ini mencadangkan kaedah *robust Jackknife ridge regression* yang menggabungkan penganggar- M teritlak (GM) dan penganggar pantas terubah suai GT (GM-FIMGT) dalam membangunkannya. Kaedah ini dinamakan *robust Jackknife ridge regression* berdasarkan GM-FIMGT, ditandakan dengan RJFIMGT. Keputusan berangka menunjukkan bahawa kaedah RJFIMGT yang dicadangkan adalah yang terbaik kerana ia mempunyai nilai *RMSE* dan pincang terkecil berbanding dengan kaedah lain dalam kajian ini.

Kata kunci: Jackknife; multikolinearan; penganggar MM; regresi ridge; titik tuasan tinggi

INTRODUCTION

The multiple linear regression (MLR) model is one of the most widely used statistical approaches in applied and social sciences. The OLS estimation method is the most commonly used method to estimate the parameters of a multiple linear regression model. This is because it has excellent properties and makes computation simple. However, in the presence of multicollinearity and outliers, the OLS estimates become very unstable and may have large variance (Pison et al. 2003), which leads to poor predictions. Multicollinearity occurs when two or more predictor variables are highly correlated. Outliers are noted to have significant impacts on the OLS estimates, and they cause model failure and misleading conclusions.

In addressing the problem of multicollinearity, one of the suggested approaches is the ridge regression (RR) approach, as introduced by Hoerl and Kenard (1970). However, while the RR is noted to have optimal properties that enable it to manage the presence of multicollinearity, its estimators are significantly biased (Akdeniz Duran & Akdeniz 2012; Batah, Ramanathan & Gore 2008). In attempting to address the issue of bias in the RR approach, Singh, Chaubey and Dwivedi (1986) suggested an almost unbiased ridge estimator based on the Jackknife technique.

Interestingly, Singh, Chaubey and Dwivedi (1986) showed that the Jackknife ridge estimator has a smaller bias and lower mean square error (MSE) than the classical RR under certain conditions. Shah et al. (2021) suggested using Jackknife ridge estimator by introducing different values of biasing constant, k. However, the RR and the Jackknife techniques alone are not completely robust to outliers and leverage points. In addressing the problem of outliers, a number of robust methods have been proposed (Huber 2004; Maronna, Martin & Yohai 2016). The methods are the least median squares (LMS), the M-estimator, the MM-estimator and the generalized M (GM-estimator). Unfortunately, the robust methods alone and the RR techniques alone are not adequate enough to address the complicated problems of multicollinearity and outliers (Alguraibawi, Midi & Rana 2015; Midi & Zahari 2007; Zahariah, Midi & Mustafa 2021). However, significant efforts aimed at addressing the inadequacies have been done. Prominent among such efforts is the integration of the RR with the robust method to get an estimator that is much less influenced by multicollinearity and outliers (Alguraibawi, Midi & Rana 2015).

Evidently, Arskin and Montgomery (1980) suggested the use of Weighted Ridge Regression (WRR). Midi and Zahari (2007) proposed the use of robust ridge regression, which is a combination of RR and the MM-estimator. Relatedly, Jadhav and Kashid (2011) suggested the use of Jackknife ridge M-based estimator (RJRM) to overcome multicollinearity and outliers in the Y direction or referred as vertical outliers. Regrettably, the suggested methods do not focus on the combined problem of multicollinearity and high leverage points (HLPs) which are outlying observations in the X-direction. As noted by Dhhan, Rana and Midi (2016), the M estimator can only handle outliers in the Y direction, but cannot cope with HLPs. According to Bagheri and Midi (2015) and Rashid et al. (2021), among the three types of outliers (vertical outliers, HLPs and residual outliers, i.e., observations that have large residuals), HLPs has the most detrimental effect on the computed values of various estimates; hence their effects should be minimized (Zahariah, Midi & Mustafa 2021). Due to this, Alguraibawi, Midi and Rana (2015) proposed a combination of the Jackknife Ridge Regression (JRR) with the MM-estimator and the Jackknife Ridge Regression with the GM2-estimator, which are assumed to be able to handle HLPs. Nonetheless, the GM2-estimator which is almost identical to GM6 estimator is very time consuming due to using minimum volume ellipsoid (MVE) in its establishment (Lim & Midi 2016; Zahariah & Midi 2023). Moreover, the initial weight function of GM6 or GM2 tend to swamp some low leverage points (Bagheri & Midi 2015; Midi et al. 2021). 'Masking' refers to a situation where outliers are incorrectly declared as inliers, while 'swamping' refers to normal observations incorrectly declared as outliers. Hence, integrating the GM2 or GM6 estimator with the Jackknife Ridge Regression is still not very efficient with regard to its parameter estimations and computational running time. In addressing these capabilities, we proposed RJFIMGT method which is based on GM-FIMGT (Midi et al. 2021) which has been proven to be more efficient, less biased, and less time-consuming than the commonly used GM6 estimator.

MATERIALS AND METHODS

REMEDIAL MEASURE OF MULTICOLLINEARITY IN THE PRESENCE OF OUTLIERS IN LINEAR REGRESSION

The linear regression model can be written as follow:

$$y = X\beta + \varepsilon_i \quad (1)$$

where y is a vector of the response vector, X is a $n \times p$ matrix of predictor variables, β is $p \times 1$ vector of unknown regression coefficients, ε is an $n \times 1$ vector of random errors with mean 0 and variance σ^2. For convenience, it is assumed that the X variables are standardized so that $X'X$ has the form of correlation matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of $X'X$ and $q_1, .., q_p$ be the corresponding eigenvectors. Let $QX'Q'X = \Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$ and $Q = (q_1, q_2, ..,$

q_p) such that X' X = $Q \Lambda Q'$. The regression model in Equation (1) can written in a canonical form as:

$$y = Z\alpha + \varepsilon \tag{2}$$

where Z = XQ, and $\alpha = Q'\beta$, hence, Λ = Z'Z = Q' X' X Q. The OLS estimator of (2) is given by

$$\alpha_{ols} = (Z'Z)^{-1}Z'y = \Lambda^{-1} Z'y \tag{3}$$

Since $\alpha = Q'\beta$ and $Q'Q = I$, the OLS estimator is given by

$$\hat{\beta}_{OLS} = Q\hat{\alpha}_{ols} \tag{4}$$

Note that, because of the relation $\alpha = Q'\beta$, any estimator $\hat{\alpha}$ of α has a corresponding $\hat{\beta} = Q\hat{\alpha}$ and $MSE(\hat{\alpha}) = MSE(\hat{\beta}) = \sigma^2 \Lambda^{-1}$. Hence, it is sufficient to consider only the canonical form. Some parameter estimation methods will be discussed in the following sections.

THE GENERALIZED RIDGE REGRESSION

The generalized ridge regression (GRR) estimate is obtained by minimizing the penalized sum of squares:

$$\sum_{i=1}^{n}(y_i - x_i^t\beta)^2 + k_i\sum_{i=1}^{p}\beta_j^2 \tag{5}$$

In Equation (5) we can see that GRR penalizes the size of the regression coefficients to be more resistant to multicollinearity. When $k_1 = k_2 = \cdots = k_p = k$, $k > 0$, k is fixed, the solution of GRR, namely the ordinary ridge regression (RR), is given by

$$\hat{\beta}_{RR} = (X^tX + kI_p)^{-1}X^ty \tag{6}$$

Different technique of identifying k has been proposed in the literature. The most widely used technique for choosing the optimal k was developed by Hoerl and Kennard (1970) given as:

$$k = k_{HK} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p}\hat{\alpha}_i^2} \tag{7}$$

where k_{HK} is Hoerl and Kennard shrinkage parameter and $\hat{\sigma}^2$and $\hat{\alpha}$ are obtained by using the Ordinary Least Squares (OLS) method. If k is equal to zero, $\hat{\beta}_{OLS}$ and $\hat{\beta}_{RR}$ are equivalent. The shrinkage parameter k has the impact of shrinking the estimates toward zero, which leads to the introduction of bias but reduces the variance of the estimate (Belsley, Kuh & Welsch 2004; Groβ 2003). The canonical form of the RR estimator in (6) is given by Kutner et al. (2005), Montgomery, Peck and Viving (2001), and Singh, Chaubey and Dwivedi (1986)

$$\begin{aligned}\hat{\alpha}_{RR(k)} &= (\Lambda + kI_p)^{-1}Z'y \\ &= B^{-1}Z'y \\ &= (I_p - kB^{-1})\hat{\alpha}\end{aligned} \tag{8}$$

where Z'Z = Λ, $B = (\Lambda + kI_p)$, $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$ be the matrix of eigenvalues, and hence the RR coefficients can be formulated as:

$$\hat{\beta}_{RR} = Q\hat{\alpha}_{RR}(k) \tag{9}$$

The bias and variance of the RR estimator are given as follows:

$$\begin{aligned}Bias(\hat{\alpha}_{RR}(k)) &= Q\hat{\alpha} \\ \mathrm{var}(\hat{\alpha}_{RR}(k)) &= \sigma^2(I_p - kB^{-1})\Lambda^{-1}(I_p - kB)^t\end{aligned} \tag{10}$$

ROBUST RIDGE REGRESSION

The robust ridge regression (RRR) combines ridge and robust regression in order to handle the problem of multicollinearity and outliers simultaneously. This will dampen the effects of both problems in a classical linear regression model. The robust ridge regression estimate is given as follows:

$$\hat{\beta}_{RRR} = (X'X + kI)'X'Y \tag{11}$$

where k is called the robust ridge parameter by employing Hoerl and Kennard technique,

$$\hat{k} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \tag{12}$$

It is obtained from robust regression methods instead of using the OLS. Equation (12) is written as in Equation (7), the only different is that $\hat{\sigma}^2$, β are replaced by $\hat{\sigma}^2{}_{robust}$, $\hat{\beta}_{robust}$ respectively, where

$$\tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})'(y - X\tilde{\beta})}{\tilde{\beta}'\tilde{\beta}}$$

JACKKNIFE RIDGE REGRESSION

The Jackknife ridge regression (JRR) technique was originally proposed by Quenouille (1956) in order to reduce the bias of an estimation (Asrkin & Montgomery 1980; Lawrence & Arthur 1990; Li & Chen 1985). Singh, Chaubey and Dwivedi (1986) developed an approach to circumvent the biasing in ridge regression which depends upon Jackknife technique, which is formulated as:

$$y_{(i)} = X_{(i)}\beta + \varepsilon_{(i)} \tag{13}$$

where $\varepsilon_{(i)}$ is an error term with ith coordinate deleted and $y_{(i)}$, $X_{(i)}$ denote the vector y with its ith value deleted and the matrix X with its ith row deleted, respectively. The matrix $X_{(i)}$ is not necessarily to be full column rank. Hence, the ridge regression estimator in this reduced model is given as,

$$\hat{\beta} = (X^t X + k I_p)^{-1} X^t y \tag{14}$$

Following Hinkley (1977) and Singh, Chaubey and Dwivedi (1986), the JRR solution is the same as given in Equation (14) in which y is replaced with $y_{(i)}$and X is replaced with $Z_{(i)}$. Let $B = (\Lambda + kI_p)$, $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_p)$ be the matrix of eigenvalues, then the coefficients of JRR are given by

$$\hat{\alpha}_{RR(i)} = (Z_{(i)}^t Z_{(i)} + kI_p)^{-1} Z_{(i)}^t y_{(i)} \tag{15}$$

where k is stated as in Equation (12). Let z_i and y_i being the ith column vector of Z and the ith coordinate of y, respectively, then:

$$\hat{\alpha}_{RR(i)} = \left(Z^t Z - z_i z_i^t + kI_p\right)^{-1}(Z^t y - z_i y_i) \tag{16}$$

The simplified model for JRR estimates can be written as

$$\hat{\alpha}_{JRR}(k) = (I + kB^{-1})\hat{\alpha}_{RR(k)} = (I - k^2 B^{-2})\hat{\alpha}_{OLS} \tag{17}$$

From Equation (17) we can obviously see that the JRR estimators are obtained by shrinking the $\hat{\alpha}_{OLS}$ by the amount ($k^2 B^{-2}$), where $B = (\Lambda + kI_p)$, $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_p)$ be the matrix of eigenvalues. The Jackknife ridge regression coefficients ($\hat{\beta}_{JRR}$) is obtained by:

$$\hat{\beta}_{JRR} = Q\hat{\alpha}_{JRR}(k) \tag{18}$$

The bias and variance for Jackknife ridge regression estimates are,

$$Bias\left(\hat{\alpha}_{JRR}(k)\right) = -k^2 B^{-2}\hat{\alpha} \tag{19}$$

$$Var\left(\hat{\alpha}_{JRR}(k)\right) = \sigma^2\left(I_p - k^2 B^{-2}\right)\Lambda^{-1}\left(I_p - k^2 B^{-2}\right)^t \tag{20}$$

where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_p)$ are the matrix of eigenvalues.

ROBUST JACKKNIFE RIDGE REGRESSION

The robust Jack-knife ridge regression (RJRR) general form of the robust Jack-knife estimator based on the robust estimation method is discussed by Alguraibawi, Midi and Rana (2015), Batah, Ramanathan and Gore (2008), and Jadhav and Kashid (2011). If $\tilde{\alpha}$ is the vector of robust ridge regression based on some robust methods such as M-estimator or MM-estimator, the coefficient of RJRR is given as,

$$\begin{aligned}\hat{\alpha}_{RJRR}(k) &= [I + kB^{-1}]\tilde{\alpha}_{RRR} \\ &= [I + kB^{-1}][I - kB^{-1}]\tilde{\alpha} \\ &= (I - k^2 B^{-2})\tilde{\alpha}\end{aligned} \tag{21}$$

where $\tilde{\alpha}$ is the vector of robust coefficients and $B = (\Lambda + kI_p)$, $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_p)$ be the matrix of eigenvalues. The coefficient of RJRR is given as,

$$\hat{\beta}_{RJRR} = Q\hat{\alpha}_{RJRR}(k) \tag{22}$$

The bias is defined as follows:

$$\begin{aligned}Bias(\hat{\alpha}_{RJRR}(k)) &= E[\hat{\alpha}_{RJRR}(k)] - \alpha \\ &= E[(I - k^2 B^{-2})\tilde{\alpha}] - \alpha \\ &= (I - k^2 B^{-2})E[\tilde{\alpha}] - \alpha \\ &= -k^2 B^{-2}\,\alpha\end{aligned} \tag{23}$$

where α is the vector of true coefficients. The variance of RJRR estimator is given by

$$\begin{aligned}&Var(\alpha_{RJRR}(k)) \\ &= E[(\hat{\alpha}_{RJRR}(k) - E(\hat{\alpha}_{RJRR}(k))][(\hat{\alpha}_{RJRR}(k) - E(\hat{\alpha}_{RJRR}(k))]^t \\ &= E[\hat{\alpha}_{RJRR}(k) - \alpha][\hat{\alpha}_{RJRR}(k) - \alpha]^t \\ &= (I - k^2 B^{-2})\Omega(I - k^2 B^{-2})^t\end{aligned} \tag{24}$$

THE GENERALZED M ESTIMATOR BASED ON FAST IMPROVISED GENERALIZED MT

Midi et al. (2021) summarized the algorithm of Generalized M estimator based on Fast Improvised Generalized (GM-FIGMT) as follows:

Step 1: Calculate the residuals (r_i) based on S estimator developed by Rousseeuw (1984).

Step 2: Calculate the estimated scale (σ) of the residuals s = (1.4826) (the median of the largest (n-p) of the | ri |), where r_i is obtained from Step 1.

Step 3: Compute the standardized residuals (e_i), where, $e_i = \frac{r_i}{s}$.

Step 4: Calculate the initial weight, denoted as w_i,, where $w_i = \min[1, (\frac{CP_{FIMGT}}{FIMGT})]$ where FIMGT is proposed by Midi et al. (2021) by incorporating ISE in its establishment.

Step 5: Compute the bounded influence function for bad leverage points, $t_i = \frac{e_i}{w_i}$.

Step 6: Employ the weighted least squares (WLS) to estimate the parameters of the regression, $\hat{\beta} = (X^T W X)^{-1} X^T W Y$, where the weight w_i is reduced for large residuals to get good efficiency (In this paper, Tukey weight function is employed).

Step 7: Calculate the new residuals (r_i) from WLS and repeat steps (2-6) until convergence.

THE PROPOSE ROBUST JACKKNIFE RIDGE REGRESSION BASED ON GM-FIMGT

Alguraibawi, Midi and Rana (2015) proposed Robust Jack-knife Ridge Regression based on MM estimator (RJMM) and Robust Jack-knife Ridge Regression based on GM2 (RJGM2) estimator to address the combined problems of multicollinearity and HLPs. However, using the RJMM and RJGM2 methods have shortcomings. The RJMM is based on MM estimator which does not have bounded influence property in the sense that it is not robust in the *X*-directions (Simpson, Ruppert & Carroll 1992). The RJGM2 is based on GM2 or GM6 estimator. As noted by Lim and Midi (2016), and Midi et al. (2021, 2020), any estimator that is based on the Minimum Volume Ellipsoid (MVE) such as RJGM2 has longer computational running time and less efficient since the MVE still suffers from swamping effect. The weakness of these methods has inspired us to incorporate the robust Jack-knife Ridge Regression with robust method that has bounded influence property. Midi et al. (2021) developed GM-FIMGT which is proven to have bounded influence property and the algorithm is very fast to compute compared to other estimators in their study. Motivated by the fact that integrating bounded influence property estimator in the algorithm of robust Jack-knife ridge regression can remedy both problems of multicollinearity and outliers, our main aim was to incorporate the GM-FIMGT in the algorithm of RJRR to produce more efficient estimates with less computation running times. The proposed method is call the robust Jack-knife ridge regression based on GM-FIMGT denoted as RJFIMGT. The RJFIMGT is expected to be more efficient than other methods considered in this study as multicollinearity is already handled by RJRR and HLPs are handled by GM-FIMGT.

The algorithm for the proposed RJFIMGT can be summarized by the following steps:

Step 1: Compute the GM-FIMGT estimates following Midi et al. (2021).

Step 2: Calculate the correlation matrix r_{xy} as follows:

$$r^w_{y,xj} = \frac{\sum_{i=1}^n w_i(y_i - \bar{y}^w)(x_{ij} - \bar{x}^w_j)}{\sqrt{(\sum_{i=1}^n w_i(y_i - \bar{y}^w)^2)(\sum_{i=1}^n w_i(x_{ij} - \bar{x}^w_j)^2)}}, i = 1,2,\ldots,n; j = 1,2,\ldots,p$$

where $\bar{x} = \frac{\sum w_i x_i}{\sum w_i}$ and $\bar{y} = \frac{\sum w_i y_i}{\sum w_i}$, where w is robust weight obtain from step (1)

Step 3: Calculate eigenvalue Λ (root of correlation matrix) and eigenvector, Q from correlation matrix obtained in Step (2).

Step 4: Calculate the estimated scale $(\hat{s})$ of the residuals obtained from Step (1) .

$$\hat{s} = 1.4826 * median\ (abs\ (e_i - median\,(e_i))$$

Step 5: Using the estimated parameter $(\hat{\beta})$ from Step (1) and the estimated residuals $(\hat{S})$ from Step (4), compute the constant k by using Equation (12) as

$$\hat{k} = \frac{p\hat{s}^2}{\hat{\beta}'\hat{\beta}}$$

Step 6: Calculate robust Jack-knife estimator as follows:

$$\hat{\alpha}_{RJFIMGT.}(k) = (I - k^2 B^{-2})\,\hat{\alpha}_{GM-FIMGT} \tag{25}$$

where $B = (\Lambda + kI_p)$, $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_p)$ be the matrix of eigenvalues and where $\hat{\alpha}$ is the vector of robust coefficients obtained by using our proposed GM-FIMGT method.

$$\hat{\beta}_{RJFIMGT} = Q\hat{\alpha}_{RJFIMGT}(k) \tag{26}$$

MONTE CARLO SIMULATION STUDY

A Monte Carlo simulation study that focused on the simultaneous presence of multicollinearity and HLPs in a data set is thus used to test the credibility of the RJFIMGT approach proposed in the present study. To generate simulated data with a different degree of multicollinearity, we applied a simulation approach similar to Alguraibawi, Midi and Rana (2015) and Lawrence and Arthur (1990). First, a linear regression model with three explanatory variables (p=3) are generated according to the following relation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + r_i$$

where $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$ and r_i is the error term distributed as $N(0,1)$. The explanatory variables are generated by,

$$x_{ij} = \rho v_{i4} + (1-\rho^2)^{\frac{1}{2}} v_{ij}, i = 1,2,\cdots,n;\ j = 1,2, \text{and } 3 \tag{27}$$

where v_{i1},v_{i2}, v_{i3} and v_{i4}, are independent standard normal pseudo random numbers. The explanatory variables are standardized so that the design matrix $(X^t X)$ is in the canonical form (Brown 1977). The character ρ^2 denotes the degree of collinearity between X_{ij} and X_{im} for $i \neq m$. In addition, three different values of high collinearity were selected that corresponds to ρ = 0.5, 0.9 and 0.99, while four different sets of observations were considered corresponding to n = 30, 60, 100 and 200. In order to generate HLPs, the first $100(\frac{\alpha}{2})$ percent observations for both x_1 and x_2 variables were replaced by observations generated from $N(20,10)$. Thereafter, y is replaced by$N(20,10)$. Different percentages level of HLPs denoted as τ are used, i.e., τ = 0.05, 0.10 and 0.15. Basically, seven estimation methods were applied in this study. The methods are OLS, JAK, GM-FIMGT, RJRM, RJMM, RJGM2, and RJFIMGT. As per Dhhan, Rana and Midi (2016) and Midi et al. (2021), the performances of the estimators are evaluated based on the Root Mean Squared Error (RMSE) and Loss values. According to Groβ (2003), since the original parameters are known, the loss criterion which is based on the sum of squared biases of parameters, may be used. The Loss criterion or simply refers to as overall biases is defined as follows:

$$\begin{aligned}\text{Loss}(\beta,\hat{\beta}) &= (\hat{\beta}-\beta)^t(\hat{\beta}-\beta)\\ &= \textstyle\sum_{j=1}^{p}(\hat{\beta}_j-\beta)^2\end{aligned} \tag{28}$$

where p is the number of parameters. The MSE is given as follows:

$$MSE = \frac{1}{n}\left\|y - X\hat{\beta}\right\|^2 \tag{29}$$

where || . || indicates the Euclidean norm. The Root Mean Squared Error (RMSE) is given by $[MSE]^{1/2}$. The simulation experiments are repeated 1000 times for all possible combinations of n, p and τ. It can be shown that the MSE consists of two components; one measures the variability (precision) and the other measures its bias (accuracy). A good method or the most efficient method is one that has the smallest value of RMSE and Bias. The term 'efficient' here refers to unbiased (very small bias) estimator that has smaller variance. An estimator with low variance with some bias is more desirable than an unbiased estimator with high variance. In this regard, the MSE or RMSE is very useful criteria to evaluate the performance of the estimators (Dhhan, Rana & Midi 2016; Midi et al. 2021). All results (RMSE and Loss or overall bias) are averaged over 1,000 replications and are exhibited in Tables 1-4. Let us first focus on Table 1, where the data have multicollinearity but no outlier. In this situation, the OLS gives the poor results evident by having the largest value of RMSE particularly for higher degree of multicollinearity ($\rho^2 = 0.9$ and 0.99), even though its bias is rather small. The performances of other estimators that are associated with Jackknife estimator are fairly close to each other and their RMSEs are rather small because those estimators can remedy multicollinearity alone. However, as can be seen from Tables 2-4, the presence of HLPs for data having multicollinearity changes things dramatically. We can clearly observe that the values of RMSEs for all estimators increase significantly, as the percentage of HLPs and degree of multicollinearity increases. It appears that the variances make up most of the MSEs because the biases are rather small. In this situation, again the OLS gives very poor results since it has the largest value of RMSE despite of having small bias. This results are as expected because the OLS is easily affected by outliers and their standard error of the estimates become inflated in the presence of multicollinearity. The JAK estimator seems to be better than the OLS but cannot outperform the GM-FIMGT. The RMSE and bias of GM-FIMGT are larger than the other estimators that are associated with Jackknife estimator combined with robust estimators.

TABLE 1. RMSE and LOSS for estimation methods with $\tau = 0\%$

Method	n	30		60		100		200	
	ρ	RMSE	LOSS	RMSE	LOSS	RMSE	LOSS	RMSE	LOSS
OLS	0.5	1.726	0.003	1.723	0.003	1.722	0.0028	1.7256	0.0031
JAK		1.718	0.0029	1.7097	0.0029	1.6086	0.0026	1.6137	0.0026
GM-FIMGT		1.7137	0.0029	1.7034	0.0029	1.6764	0.0025	1.6151	0.0027
RJRM		1.7302	0.0029	1.7315	0.003	1.7305	0.003	1.7305	0.003
RJMM		1.7264	0.003	1.726	0.003	1.7058	0.0029	1.7209	0.003
RJGM2		1.7281	0.003	1.7271	0.003	1.7091	0.0029	1.7218	0.003
RJFIMGT		1.7267	0.003	1.7261	0.003	1.706	0.0029	1.7209	0.003
OLS	0.9	1.792	0.003	1.759	0.0033	1.766	0.0032	1.7303	0.0032
JAK		1.7172	0.0029	1.5592	0.0024	1.7159	0.0029	1.6321	0.0026
GM-FIMGT		1.7198	0.0029	1.5801	0.0023	1.7171	0.0029	1.6301	0.0026
RJRM		1.7337	0.003	1.74	0.003	1.7308	0.003	1.7279	0.003
RJMM		1.7209	0.0029	1.6277	0.0026	1.724	0.003	1.6951	0.0029
RJGM2		1.7231	0.003	1.6382	0.0026	1.725	0.003	1.696	0.0029
RJFIMGT		1.721	0.003	1.6282	0.0026	1.724	0.003	1.695	0.0029
OLS	0.99	2.1798	0.0023	1.7455	0.0059	1.865	0.0025	1.7705	0.0026
JAK		1.6046	0.0022	1.727	0.0059	1.6721	0.0025	1.6682	0.0026
GM-FIMGT		2.1747	0.0024	1.7494	0.0059	1.8814	0.0025	1.7703	0.0026
RJRM		1.6437	0.0026	1.7322	0.006	1.7175	0.0029	1.739	0.003
RJMM		1.492	0.0022	1.722	0.0059	1.6089	0.0026	1.6447	0.0027
RJGM2		1.5828	0.0022	1.7258	0.0059	1.6503	0.0026	1.6505	0.0027
RJFIMGT		1.4885	0.0022	1.7219	0.0059	1.6095	0.0026	1.6446	0.0027

This happens because the GM-FIMGT can only remedy the problem of outlier alone. It can be observed that the values of RMSE and bias for RJM, RJMM are larger than the RJGM2 and RJFIMGT because they depend on the M and MM-estimator, which are known to be less efficient in the presence of HLPs. The RJGM2 and RJFIMGT which are based on GM2 and GM-FIMGT estimators, respectively can do well in the presence of HLPs (Midi et al. 2021). Nonetheless, the computational time of RJGM2 is much longer than the RJFIMGT because GM2 depend on the minimum volume ellipsoid (MVE), while GM-FIMGT is based on Index Set Equality (ISE) which has been shown by Lim and Midi (2016) that its computation running time is much faster than the MVE. Subsequently, the computation running time of RJFIMGT is faster than the RJGM2. Due to space constraint, the computational running times for RJFIMGT are not reported here, but one immediately understands that it should be very fast because it is based on ISE. On the other hand, the RMSE and the bias of the RJFIMGT estimates are consistently the smallest among the other estimators [can be seen clearly for sample size of less than one hundred (100) and higher degree of multicollinearity (ρ^2 = 0.9 and 0.99), irrespective of the percentage of HLPs]. As was to be

TABLE 2. RMSE and LOSS for estimation methods with τ = 0.05

Method	n	30		60		100		200	
	ρ	RMSE	LOSS	RMSE	LOSS	RMSE	LOSS	RMSE	LOSS
OLS	0.5	2.9513	0.0041	2.1545	0.0038	2.2822	0.0043	2.1391	0.0045
JAK		2.3728	0.0041	1.9955	0.0038	2.139	0.0042	2.0349	0.0041
GM-FIMGT		2.5652	0.0039	1.791	0.003	1.8401	0.0032	1.6077	0.0026
RJRM		2.3948	0.0041	1.975	0.0038	1.9412	0.0037	1.8598	0.0035
RJMM		2.4382	0.0039	1.7514	0.003	1.8292	0.0032	1.6652	0.0028
RJGM2		2.4358	0.0037	1.7515	0.002	1.7922	0.0031	1.6705	0.0026
RJFIMGT		2.0573	0.0039	1.74	0.003	1.7736	0.0031	1.6704	0.0028
OLS	0.9	4.1394	0.0042	2.5994	0.0039	2.7111	0.0044	2.4364	0.0059
JAK		2.7374	0.004	2.0868	0.0038	2.2764	0.0041	2.0613	0.0042
GM-FIMGT		2.9513	0.0037	1.8117	0.0029	1.9138	0.0032	1.6306	0.0027
RJRM		2.0861	0.004	1.9763	0.0038	1.9649	0.0036	1.8585	0.0035
RJMM		2.04	0.0036	1.7455	0.0029	1.8418	0.0032	1.6466	0.0027
RJGM2		2.0361	0.0036	1.7506	0.0029	1.8509	0.0032	1.6466	0.0027
RJFIMGT		1.8536	0.0033	1.7125	0.0028	1.804	0.0031	1.6381	0.0025
OLS	0.99	8.4391	0.0063	4.7764	0.0047	4.7896	0.0053	3.9745	0.0158
JAK		4.4728	0.0044	2.6851	0.004	3.1417	0.0042	2.3842	0.0057
GM-FIMGT		6.9888	0.0044	2.1957	0.0031	2.3652	0.0035	1.9757	0.0039
RJRM		4.0583	0.0045	2.0451	0.0037	2.1495	0.0035	1.9159	0.0037
RJMM		6.2001	0.0088	1.895	0.0028	1.9785	0.0031	1.7959	0.0031
RJGM2		6.2529	0.0088	1.8418	0.0028	1.9801	0.0031	1.7859	0.0031
RJFIMGT		2.8254	0.0069	1.693	0.0025	1.7861	0.0028	1.759	0.003

TABLE 3. RMSE and LOSS for estimation methods with $\tau = 0.10$

Method	n	30		60		100		200	
	ρ	RMSE	LOSS	RMSE	LOSS	RMSE	LOSS	RMSE	LOSS
OLS	0.5	2.7275	0.0043	2.5412	0.0043	2.2876	0.0043	2.0903	0.0044
JAK		2.2449	0.0043	2.2347	0.0043	2.0927	0.0042	2.0522	0.0042
GM-FIMGT		1.9556	0.0033	1.8353	0.0032	1.8133	0.0032	1.7267	0.003
RJRM		2.1098	0.004	2.0778	0.004	2.066	0.0042	2.0016	0.0041
RJMM		1.8604	0.0032	1.8536	0.0034	1.8005	0.0032	1.6694	0.0028
RJGM2		1.8574	0.0032	1.8332	0.0033	1.8114	0.0032	1.6693	0.0028
RJFIMGT		1.8124	0.0031	1.8223	0.0032	1.8015	0.0031	1.6690	0.0027
OLS	0.9	4.0301	0.0043	2.4712	0.006	2.7925	0.0043	2.2005	0.0048
JAK		2.6212	0.0042	2.0893	0.0043	2.1721	0.0042	2.0459	0.0042
GM-FIMGT		2.1997	0.0033	1.8178	0.0032	1.8992	0.0033	1.9855	0.0039
RJRM		2.1467	0.0042	2.0569	0.0042	2.0598	0.0042	2.1808	0.0048
RJMM		1.9276	0.0032	1.7705	0.0031	1.8443	0.0033	1.9364	0.0037
RJGM2		1.9382	0.0033	1.7824	0.0031	1.8261	0.0032	1.9342	0.0037
RJFIMGT		1.8381	0.0031	1.7694	0.0031	1.8035	0.0032	1.9344	0.0035
OLS	0.99	8.8865	0.0047	4.008	0.016	5.0821	0.005	4.1142	0.0042
JAK		4.6112	0.0042	2.1046	0.0044	2.671	0.0042	2.7179	0.0042
GM-FIMGT		3.3462	0.003	2.1127	0.0038	2.1857	0.0035	2.1279	0.0039
RJRM		2.4169	0.0042	2.1079	0.0044	2.0566	0.0041	2.0705	0.0043
RJMM		2.2435	0.003	1.8736	0.0032	1.9249	0.0034	2.0004	0.0039
RJGM2		2.1134	0.0029	1.6855	0.0027	1.9057	0.0034	1.9906	0.0039
RJFIMGT		1.8454	0.0022	1.6465	0.0027	1.8397	0.0030	1.9687	0.0030

expected, the RJFIMGT gives the best results followed by the RJGM2, RJMM, RJRM, GM-FIMGT, JAK and OLS. The results seem to be consistent irrespective of the sample size, percentage of HLPs and degree of multicollinearity.

As already discussed, since RJFIMGT clearly outperformed the other estimators, its' performance is further investigated by using the mean squared error ratios of the RJFIMGT to other estimators (Jadhav & Kashid 2011). According to Alguraibawi, Midi and Rana (2015) and Lawrence and Arthur (1990), if the ratio is less than one, the numerator (RJFIMGT) is more efficient than the denominator, while if the ratio is greater than one, the denominator is more efficient than the numerator. If the ratio is exactly one, the numerator and denominator have the same efficiency. The MSE ratios of RJFIMGT to each of the estimators, i.e., RJGM2, RJMM, RJRM, GM-FIMGT, JAK and OLS are shown in Tables 5-7. It is interesting to observe the results of the

MSE ratios from Tables 5-7. Irrespective of sample size, degree of multicollinearity and percentage of HLPs, the MSE ratios of the RJFIMGT to other estimators consistently less than one. This indicates that RJFIMGT is more efficient than RJGM2, RJMM, RJRM, GM-FIMGT, JAK and OLS estimates which is in agreement with the results of using the RMSE and bias criterions as exhibited in Tables 1-4. Summarizing the findings from Tables 1-7, it can be concluded that the RJFIMGT is the best method to overcome the multicollinearity problem in the presence of HLPs. Moreover, this method is very appealing because it is less time consuming than other methods since it is based on ISE. Due to space limitation, the computation running time of the RJFIMGT is not reported here but one may refer to Lim and Midi (2016) for detail discussion on the computational running of the ISE.

TABLE 4. RMSE and LOSS for estimation methods with $\tau = 0.15$

Method	n	30		60		100		200	
	ρ	RMSE	LOSS	RMSE	LOSS	RMSE	LOSS	RMSE	LOSS
OLS		3.1577	0.0042	2.6551	0.0044	2.5042	0.0042	2.2072	0.0042
JAK		2.4145	0.0041	2.2287	0.0043	2.1663	0.0042	2.0745	0.0042
GM-FIMGT		1.923	0.0033	1.8541	0.0032	1.848	0.0033	1.8607	0.0034
RJRM	0.5	1.9761	0.0038	2.1018	0.004	2.0685	0.004	2.0576	0.004
RJMM		1.827	0.0032	1.8504	0.0033	1.8376	0.0033	1.8573	0.0033
RJGM2		1.8395	0.0033	1.8299	0.0032	1.8312	0.0033	1.8488	0.0033
RJFIMGT		1.8155	0.0031	1.8033	0.0032	1.821	0.0033	1.8328	0.0031
OLS		5.3615	0.0045	4.0407	0.0152	3.3888	0.0042	2.5662	0.0042
JAK		3.2144	0.041	2.4152	0.0056	2.3972	0.0042	2.117	0.0042
GM-FIMGT		2.1314	0.0035	2.2488	0.005	1.9475	0.0035	1.9556	0.0037
RJRM	0.9	2.0167	0.0039	2.2124	0.0045	2.0705	0.0043	2.0543	0.0042
RJMM		1.939	0.0036	2.2103	0.0049	1.9237	0.0036	1.9158	0.0036
RJGM2		1.9216	0.0035	2.214	0.0048	1.8918	0.0034	1.9161	0.0036
RJFIMGT		1.876	0.0031	2.2094	0.0046	1.8657	0.0031	1.9047	0.0032
OLS		12.1555	0.0062	8.4956	0.0047	6.8864	0.0043	4.3358	0.0043
JAK		6.2049	0.0043	3.5511	0.0086	3.6199	0.0042	2.451	0.0041
GM-FIMGT		3.011	0.0036	2.2434	0.0045	2.2603	0.0038	2.1607	0.0041
RJRM	0.99	2.1153	0.0039	2.1232	0.0044	2.0867	0.004	2.0598	0.004
RJMM		2.0529	0.0037	2.0915	0.0042	2.0398	0.0039	2.017	0.004
RJGM2		2.0419	0.0036	2.0525	0.0042	2.0388	0.0038	2.0178	0.004
RJFIMGT		1.9262	0.0031	2.0814	0.004	1.9587	0.0035	2.0008	0.0033

TABLE 5. MSE Ratios of RJFIMGT to other estimates when $\tau = 0.05$

N	ρ	$\frac{RJFIMGT}{OLS}$	$\frac{RJFIMGT}{JAK}$	$\frac{RJFIMGT}{GM-FIMGT}$	$\frac{RJFIMGT}{RJRM}$	$\frac{RJFIMGT}{RJMM}$	$\frac{RJFIMGT}{RJGM2}$
30	0.5	0.697083	0.867035	0.802004	0.85907	0.843778	0.84461
60		0.807612	0.871962	0.971524	0.881013	0.993491	0.993434
100		0.777145	0.829173	0.963861	0.913662	0.969604	0.989622
200		0.749848	0.788245	0.997699	0.862458	0.963248	0.960192
30		0.447794	0.677139	0.628062	0.888548	0.908627	0.910368
60	0.9	0.658806	0.820634	0.945245	0.866518	0.981094	0.978236
100		0.665413	0.792479	0.942627	0.918113	0.979477	0.974661
200		0.672344	0.794693	0.978788	0.88141	0.994838	0.994838
30		0.334799	0.631685	0.404275	0.696203	0.455702	0.451854
60	0.99	0.354451	0.630517	0.771053	0.827832	0.893404	0.919209
100		0.372912	0.568514	0.755158	0.830937	0.902755	0.902025
200		0.442571	0.737774	0.890317	0.918106	0.979453	0.984938

TABLE 6. MSE Ratios of RJFIMGT to other estimates when $\tau = 0.10$

N	ρ	$\frac{RJFIMGT}{OLS}$	$\frac{RJFIMGT}{JAK}$	$\frac{RJFIMGT}{GM-FIMGT}$	$\frac{RJFIMGT}{RJRM}$	$\frac{RJFIMGT}{RJMM}$	$\frac{RJFIMGT}{RJGM2}$
30	0.5	0.664491	0.807341	0.926774	0.859039	0.974199	0.975773
60		0.717102	0.815456	0.992917	0.877033	0.983114	0.994054
100		0.787507	0.86085	0.993493	0.871975	0.995276	0.994535
200		0.798641	0.813468	0.966815	0.834033	0.995231	0.999701
30		0.456093	0.701244	0.835614	0.856244	0.953569	0.948354
60	0.9	0.716008	0.846887	0.973374	0.860227	0.999379	0.992706
100		0.645837	0.830302	0.94961	0.87557	0.977878	0.987624
200		0.879073	0.945501	0.974263	0.887014	0.998967	0.997525
30		0.207663	0.4002	0.551491	0.76354	0.822554	0.87319
60	0.99	0.410803	0.782334	0.779335	0.781109	0.878789	0.976861
100		0.361996	0.688768	0.841698	0.894535	0.955738	0.965367
200		0.478513	0.724346	0.925184	0.950833	0.984153	0.988998

REAL EXAMPLE

A real dataset taken from Penrose, Nelson and Fisher (1985) was used to evaluate the performance of our proposed RJFIMGT method. It represents the relationship between response variable (the percentage of body fat for fifty men) and seven independent variables namely, Age (years), Weight (lbs), Height (inches), Neck circumference (cm), Chest circumference (cm), Abdomen 2 circumference (cm) and Hip circumference (cm). Since the values of VIF exceed 10, it indicates that this data set has serious multicollinearity problem. Moreover, this data set has six HLPs. The performance of the RJFIMGT is evaluated based on the standard error of the estimates (SE) and RMSE. Since the distribution of the RJFIMGT is intractable, bootstrap method is employed to obtain its' SE. Table 8 presents the estimates of coefficient, standard errors, VIF and RMSE of RJFIMGT, RJGM2, GM-FIGMT, and OLS estimates. Other estimators except the OLS were not included because their performances were not good as shown in the simulation study. The OLS and GM-FIMGT are included because we want to show that both are not reliable when multicollinearity and HLPs are present in the data. It can be seen from Table 8 that the OLS gives very poor results and the GM-FIMGT is not any better either. The OLS has the largest values of SE and RMSE. The GM-FIMGT is a good method when only HLPs are present in a data. The performance of RJGM2 is good but it cannot outperform the RJFIMGT. It is interesting to observe from Table 8 that the values of standard errors and RMSE for RJFIMGT is the smallest followed by the RJGM2, GM-FIMGT and the OLS. The results of the numerical examples agree reasonably well with the results of the simulation study that the RJFIMGT is the most efficient method compared to other methods in this study when both multicollinearity and HLPs are present in a data.

TABLE 7. MSE Ratios of RJFIMGT to other estimates when $\tau = 0.15$

N	ρ	$\frac{RJFIMGT}{OLS}$	$\frac{RJFIMGT}{JAK}$	$\frac{RJFIMGT}{GM-FIMGT}$	$\frac{RJFIMGT}{RJRM}$	$\frac{RJFIMGT}{RJMM}$	$\frac{RJFIMGT}{RJGM2}$
30	0.5	0.574944	0.751916	0.944098	0.918729	0.993706	0.986953
60		0.679183	0.809126	0.972601	0.857979	0.974546	0.985464
100		0.727178	0.840604	0.98539	0.880348	0.990966	0.99443
200		0.830373	0.88349	0.985006	0.890747	0.986809	0.991346
30		0.349902	0.583624	0.880173	0.930233	0.967509	0.97627
60	0.9	0.546786	0.91479	0.98248	0.998644	0.999593	0.997922
100		0.550549	0.778283	0.957997	0.901087	0.96985	0.986204
200		0.742226	0.899717	0.973972	0.927177	0.994206	0.99405
30		0.158463	0.310432	0.639721	0.910604	0.938282	0.943337
60	0.99	0.244997	0.586128	0.927788	0.980313	0.995171	0.999472
100		0.28443	0.541092	0.866566	0.938659	0.960241	0.960712
200		0.46146	0.81632	0.925996	0.971356	0.991968	0.991575

TABLE 8. The VIF, parameter estimates, SE and RMSE of OLS, GM-FIMGT, RJGM2 and RJFIMGT for body fat data

parameters	VIF	OLS		GM-FIMGT		RJGM2		RJFIMGT	
		Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
$\hat{\beta}_1$	1.56276	0.02581	0.09638	0.03226	0.088325	0.02887	0.09716	0.002970	0.010015
$\hat{\beta}_2$	**27.351**	-0.13269	0.08415	-0.1341	0.09324	0.006656	0.04571	0.004934	0.043722
$\hat{\beta}_3$	1.5131	-0.17263	0.12578	-0.16990	0.12448	-0.008264	0.04947	-0.00105	0.024881
$\hat{\beta}_4$	**5.343**	-0.20133	0.45558	-0.18369	0.323501	0.0080726	0.0791	0.004218	0.041124
$\hat{\beta}_5$	**10.891**	-0.06386	0.25141	-0.06832	0.17594	0.04401	0.06945	0.005106	0.039559
$\hat{\beta}_6$	**16.943**	1.09215	0.18378	1.08793	0.12694	0.004568	0.03969	0.005259	0.038780
$\hat{\beta}_7$	**25.032**	-0.18669	0.30970	-0.16771	0.21116	0.006037	0.0554	0.005128	0.041089
RMSE		4.481382		3.3121		1.12113		0.9314	

CONCLUSION

The purpose of this paper was to develop a new robust Jackknife ridge regression based on GM-FIMGT denoted as JRFIMGT to remedy the combined problem of multicollinearity and HLPs in linear regression model. The numerical results show that the OLS method performed poorly in the presence of multicollinearity and outliers. The performance of all methods except the OLS are fairly closed to each other when only multicollinearity exists in the data. As the degree of multicollinearity closed to 0.99, the RMSE values of GM-FIMGT also increased because GM-FIMGT can only handles the problem of HLPs. The GM-FIMGT and Jackknife (JAK) also perform poorly when both multicollinearity and HLPs are present in the data. This happens because JAK can only rectify multicollinearity problem but not HLPs. It can be concluded that the RJFIMGT is the most efficient estimator compared to other estimators when multicollinearity comes together with the existence of HLPs as it has the smallest RMSE and bias irrespective of the sample size, degree of multicollinearity and percentage of HLPs. However, there are limitations to the current study. The RJFIMGT incorporates GM-FIMGT estimator which uses Index Set Equality (ISE), in its establishment. Although the ISE had tremendously sped up the computation of location and scatter estimator (subsequently sped up the computation of RJFIMGT), even much faster than fast Minimum Covariance Determinant (MCD) (Rousseeuw & Driessen 1999), it is computationally not that stable and still suffers from small swamping effect. However, the effect is very small and is not serious.

In the future work we are looking forward to extend the robust Jackknife ridge regression estimation method by incorporating a new weight function constructed from Diagnostic Robust Generalized Potential based on Reweighted Fast Consistent and High Breakdown (RFCH) estimators. The weight will be integrated in the proposed method with the main aim of reducing the effect of vertical outliers and HLPs. We will also consider using more robust biasing constant, k in the algorithm of the proposed method. In the current study, we only consider the effect of bad HLPs (outlying observations in both X-space and Y-space) on the FIMGT estimates. Hence, the performance of the proposed method will be extensively investigated when vertical outliers as well as both good and bad HLPs are present in a data set.

REFERENCES

Alguraibawi, M., Midi, H. & Rana, S. 2015. Robust jackknife ridge regression to combat multicollinearity and high leverage points. *Economics Computation and Economic Cybernatics Studies and Research* 49(4): 305-322.

Akdeniz Duran, E. & Akdeniz, F. 2012. Efficiency of the modified jackknifed Liu-type estimator. *Statistical Papers* 53(2): 265-280.

Arskin, R.G. & Montgomery, D.C. 1980. Augmented robust estimators. *Technometrics* 22: 333-341.

Bagheri, A. & Midi, H. 2015. Diagnostic plot for the identification of high leverage collinearity-influential observations. *Statistics and Operation Research Transaction Journal* 39: 51-70.

Batah, F.S., Ramanathan, T.V. & Gore, S.D. 2008. The efficiency of modified jackknife and ridge type regression estimators: A comparison. *Surv. Math. Appl.* 3: 111-122.

Belsley, D.A., Kuh, E. & Welsch, R.E. 2004. *Regression Diagnostics, Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons Inc.

Brown, P.J. 1977. Centering and scaling in ridge regression. *Technometrics* 19(1): 35-36.

Dhhan, W., Rana, S. & Midi, H. 2016. A high breakdown, high efficiency and bounded influence modified GM estimator based on support vector regression. *Journal of Applied Statistics* 44(4): 700-714. https://doi.org/10.1080/02664763.2016.1182133

Groβ, J. 2003. *Linear Regression (Lecture Notes in Statistics).* Verlag Berlin Heidelberg: Springer.

Hinkley, D.V. 1977. Jackknifing in unbalanced situations. *Technometrics* 19(3): 285-292.

Hoerl, A.E. & Kennard, R.W. 1970. Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics* 12(1): 55-67.

Huber, P.J. 2004. *Robust Statistics.* New York: John Wiley & Sons.

Jadhav, N.H. & Kashid, D.N. 2011. A jackknifed ridge M-estimator for regression model with multicollinearity and outliers. *Journal of Statistical Theory and Practice* 5: 659-673.

Kutner, M.H., Nachtsheim, C.J., Neter, J. & Li, W. 2005. *Applied Linear Regression Models.* 5th ed. New York: McGraw-Hill.

Lawrence, K. & Arthur, J. 1990. *Robust Regression Analysis and Applications*. New York: Marcel Dekker Inc. pp. 59-86.

Lim, H.A. & Midi, H. 2016. Diagnostic robust generalized potential based on index set equality (DRGP (ISE)) for the identification of high leverage points in linear model. *Computational Statistics* 31(3): 859-877.

Li, G. & Chen, Z. 1985. Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. *Journal of the American Statistical Association* 80: 759-766.

Maronna, R.A., Martin, R.D. & Yohai, V.J. 2006. *Robust Statistics Theory and Methods*. New York: Wiley.

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.

Midi, H., Hendi, T.H., Arasan, J. & Uraibi, H. 2020. Fast and robust diagnostic technique for the detection of high leverage points. *Pertanika J. Sci. & Tech*. 28(4): 1203-1220.

Midi, H., Ismaeel, S.S., Arasan, J. & Mohammad, A.M. 2021. Simple and fast generalized - M (GM) estimator and its application to real data. *Sains Malaysiana* 50(3): 859-867.

Montgomery, D.C., Peck, E.A. & Viving, G.G. 2001. *Introduction to Linear Regression Analysis*. 3rd ed. New York: John Wiley and Sons.

Penrose, K.W., Nelson, A. & Fisher, A. 1985. Generalized body composition prediction equation for men using simple measurement techniques. *Medicine & Science in Sports & Exercise* 17(2): 189.

Pison, G., Rousseeuw, P.J., Filzmoser, P. & Croux, C. 2003. Robust factor analysis. *Journal of Multivariate Analysis* 84(1): 145-172.

Quenouille, M.H. 1956. Notes on bias in estimation. *Biometrika* 43(3-4): 353-360.

Rashid, A.M., Midi, H., Dhnn, W. & Arasan, J. 2021. An efficient estimation and classification methods for high dimensional data using robust iteratively reweighted SIMPLS algorithm based on nu-support vector regression. *IEEE Access* 9: 45955-45967.

Rousseeuw, P.J. 1984. Least median of squares regression. *Journal of the American Statistical Association* 79: 871-880.

Rousseeuw, P.J. & Van Driessen, K. 1999. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41: 212-223.

Shah, I., Sajid, F., Ali, S., Rehman, A., Bahaj, A.A. & Fati, S.M. 2021. On the performance of jackknife based on estimators for ridge regression. *IEEE Access* 9: 68044-68053.

Simpson, D.G., Ruppert, D. & Carroll, R.J. 1992. On one-step GM estimates and stability of influences in linear regression. *Journal of the American Statistical Association* 87: 439-450.

Singh, B., Chaubey, Y.P. & Dwivedi, T.D. 1986. An almost unbiased ridge estimator. *The Indian Journal of Statistics* 48: 342-346.

Zahariah, S. & Midi, H. 2023. Minimum regularized covariance determinant and principal component analysis – based method for the identification of high leverage points in high dimensional sparse data. *Journal of Applied Statistics* 50(13): 2817-2835.

Zahariah, S., Midi, H. & Mustafa, M.S. 2021. An improvised SIMPLS estimator based on MRCD-PCA weighting function and its application to real data. *Symmetry* 13(11): 2211.

*Corresponding author; email: habshah@upm.edu.my