International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 24(2) Special, December, 2024, pp 25-48
DOI: https://doi.org/10.3329/ijss.v24i20.78212

Comparison among Ordinary Least Squares, Ridge, Lasso, and Elastic Net Estimators in the Presence of Outliers: Simulation and Application

H. M Nayem, Sinha Aziz and B. M. Golam Kibria*

Department of Mathematics and Statistics
Florida International University
Miami, FL 33199, USA
*Correspondence should be addressed to B. M. Golam Kibria
(Email: kibriag@fiu.edu)

[Received September 22, 2024; Accepted December 1, 2024]

Abstract

In linear regression models, multicollinearity often results in unstable and unreliable parameter estimates. Ridge regression, a biased estimation technique, is commonly used to mitigate this issue and produce more reliable estimates of regression coefficients. Several estimators have been developed to select the optimal ridge parameter. This study focuses on the top 16 estimators from the 366 evaluated by Mermi et al. (2024), along with seven additional estimators introduced over time. These 23 estimators were compared to Ordinary Least Squares (OLS), Elastic-Net (EN), Lasso, and generalized ridge (GR) regression, to evaluate their performance across different levels of multicollinearity in multiple regression settings. Simulated data, both with and without outliers, and various parametric conditions were used for the comparisons. The results indicated that certain ridge regression estimators perform reliably with small sample sizes and high correlations (around 0.95) in the absence of outliers. However, when outliers were present, some estimators performed better due to small sample sizes and increased variance. Furthermore, GR, EN, and Lasso exhibited robustness with large datasets, except in cases with substantial outliers and high variance.

Keywords: Elastic-Net, Lasso, MSE, Multicollinearity, OLS, Outliers, Ridge regression, Simulation

AMS Classification: 62J07.

1. Introduction

Linear regression is a frequently utilized methodology for analyzing data to construct a model depicting the relationship between the dependent variable and one or more independent variables (Herawati et al., 2024). During this type of analysis, it is common to encounter specific issues and hypotheses regarding the model. One important assumption in regression analysis is the absence of multicollinearity, as violating this assumption can render the model unreliable for estimating population parameters (Damodar N. Gujarati, 2013, Herawati et al., 2018). In multiple regression analysis, multicollinearity occurs when one independent variable is correlated with another independent variable. This can result in inaccurate estimates of the regression coefficients, increased standard errors, reduced partial t-tests, statistically insignificant p-values, and diminished predictability of the model (Gujarati, 2013).

Addressing high multicollinearity is crucial as it can lead to inaccurate decisions and an increased likelihood of accepting the wrong hypothesis. It is essential to find the most suitable method to deal with multicollinearity (Gibbons, 1981). One effective approach to overcome this issue is by using the shrinkage method to reduce the estimated coefficients, also known as regularization. Common regularization methods include Ridge Regression, Least Absolute Shrinkage and Selection Operator (Lasso), and Elastic-Net (EN) (Herawati et al., 2024). Ridge Regression stabilizes the regression coefficient in the presence of multicollinearity issues by introducing a level of bias to the regression estimate (Li et al., 2010). This reduces the standard error and provides a more precise estimation of the regression coefficient compared to the Ordinary Least Squares (OLS) method (Kibria & Lukman, 2020). On the other hand, Lasso and EN address multicollinearity problems by shrinking the regression coefficients of highly correlated or perfectly correlated independent variables toward zero (Emmert-Streib & Dehmer, 2019; Liu & Li, 2017).

Hoerl and Kennard (1970) were the first to propose the concept of ridge regression to address the issue of multicollinearity in engineering data. They discovered that there exists a non-zero value of k (ridge parameter) for which the MSE of the ridge regression estimator is smaller than the variance of the OLS estimator (Hoerl & Kennard, 1970). Over the years, numerous researchers have delved into this area of study, devising and suggesting various estimators for k. To cite a few, Dempster et al., 1977; Gibbons, 1981; Hoerl et al., 1975; Hoerl & Kennard, 1970; J. F. & P, 1976; Khalaf & Shukur, 2005; Kibria, 2003; Kibria et al., 2012a; Månsson et al., 2010; McDonald & Galarneau, 1975; McDonald & Schwing, 1973; Muniz & Kibria, 2009; Walker & Birch, 1988and very recently Arashi & Valizadeh, 2015a, 2015b; Aslam, 2014a, 2014b; Ayinde & Lukman, 2016; A. V. Dorugade, 2014; Emmert-Streib & Dehmer, 2019; Golam Kibria & Lukman, 2020; Hefnawy & Farag, 2014; Herawati et al., 2018, 2024; Kibria, 2023; Lukman et al., 2018, 2019; Lukman & Ayindez, 2017; Melkumova & Shatskikh, 2017; Mermi et al., 2024; Yüzbaşı et al., 2020 Hoque & Golam Kibria, 2023; Hoque & Kibria, 2024, among others.

Mermi et al. (2024) compared 366 different ridge parameter estimators that were proposed at different times in terms of the number of independent variables (p), sample size (n), the correlation coefficient between independent variables (ρ), and standard deviation of errors (σ) (Mermi et al., 2024). Here, we have incorporated a total of 16 best estimators out of 366, with the exclusion of the robust estimator. Furthermore, we have applied seven estimators recommended by Kibria et al. (2003, 2023) tailored for asymmetric populations.

The objective of this paper is to provide a thorough analysis to compare 23 ridge parameter estimators with OLS, Lasso, EN, and GR. Our analysis not only focused on MSE but also examined how the models behaved in the presence of multicollinearity and outliers. Our goal is to contribute to the existing literature on ridge regression by providing a comprehensive analysis of estimator performance, particularly in simulation scenarios where outliers have a significant impact on model estimates. Additionally, we presented real-life examples to validate the simulation results. The R programming language is utilized for the purpose of data analysis.

The organization of the paper is as follows: The statistical methodology is described in Section 2. A simulation study for comparing the performance of the estimators is presented in Section 3. Two real-life datasets are analyzed in Section 4. The paper ends with some concluding remarks in Section 5.

2. Statistical Methodology

2.1 Ordinary least squares method

Consider the following multiple linear regression model,

$$Y = X\beta + \varepsilon,$$

where Y is a column vector representing the dependent variable; X is a matrix of independent variables, and β is a parameter vector. The ε column vector represents the error terms. The ordinary least squares estimator of β is obtained as,

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$$

2.2 Ridge Regression Method

Hoerl and Kennard (1970) first introduced the ridge regression estimator as a solution to the multicollinearity issue. This approach minimizes the sum of squared residuals while applying a constraint on the total sum of the coefficient squares,

$$\hat{\beta}_{Ridge} = \underbrace{\operatorname{argmin}}_{\beta} (Y - X\beta)^T (Y - X\beta) + k \parallel \beta \parallel^2$$

where $\parallel \beta \parallel^2 = \sum_{i=1}^{p} \beta_i^2$. Taking the above equation and finding its derivative with respect to β and setting the outcome to zero, we get

$$\hat{\beta}_{Ridge} = (X^T X + kI)^{-1} X^T Y, k \geq 0$$

where k is known as ridge or shrinkage estimator and need to be estimated from observed data. The MSE of $\hat{\beta}_{\text{Ridge}}$ is obtained as follows:

$$\begin{aligned} \text{MSE}(\hat{\beta}_{\text{Ridge}}) &= \text{Var}(\hat{\beta}_{\text{Ridge}}) + [\text{Bias}(\hat{\beta}_{\text{Ridge}})]^2 \\ &= \sigma^2 \text{tr}\{(X^T X + kI)^{-2} X^T X\} + k^2 \beta^T (X^T X + kI)^{-2} \beta \\ &= \sigma^2 \sum_{i=1}^{p} \frac{\lambda_j}{(\lambda_i + k)^2} + k^2 \beta' (X'X + kI_p)^{-2} \beta \end{aligned}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are eigenvalues of $X^T X$ matrix.

2.3 Generalized Ridge (GR) Regression

Yang and Emura (2017) proposed GR regression by changing from uniform shrinkage to non-uniform shrinkage that replaced the identity matrix I_p with the diagonal matrix $\widehat{M}(\Delta)$ (Yang & Emura, 2017). The following is the proposed GR method.

$$\hat{\beta}_{GR} = \{X^\mathrm{T} X + k\widehat{M}(\Delta)\}^{-1} X^\mathrm{T} y$$

where $\Delta \geq 0$ is the threshold parameter and $k > 0$ is the shrinkage parameter. The diagonal components of $\widehat{M}(\Delta) = \text{diag}\{\widehat{m}_1(\Delta), \ldots, \widehat{m}_p(\Delta)\}$, where (Emura et al., 2024)

$$\widehat{m}_i(\Delta) = \begin{cases} 1/2 & \text{if} \quad z_i \geq \Delta \\ 1 & \text{if} \quad z_i < \Delta \end{cases}$$

where $z_i = \hat{\beta}_i^0/\mathrm{SD}(\hat{\beta}^0)$, and $\mathrm{SD}(\hat{\beta}^0) = \left\{\sum_{i=1}^p \left(\hat{\beta}_i^0 - \sum_{i=1}^p \hat{\beta}_i^0/p\right)^2/(p-1)\right\}^{1/2}$ for $i = 1, \dots, p$, and $\hat{\beta}^0 = \left(\hat{\beta}_1^0, \dots, \hat{\beta}_p^0\right)^{\mathrm{T}}$, defined as $\hat{\beta}_i^0 = X_i^{\mathrm{T}} y / X_i^{\mathrm{T}} X_i$, where X_i is the i-th column of X. And $\hat{\beta}^0$ is a compound covariate estimator (Chen & Emura, 2017).

2.4 Least Absolute Shrinkage and Selection Operator (Lasso)

LASSO (Tibshirani, 1996) is a regression analysis technique in statistics and machine learning that performs both variable selection and regularization to improve the predictive accuracy and interpretability of the statistical model it produces. In LASSO, most of the coefficients of irrelevant variables are set to zero, while the other coefficients are shrunk. Lasso estimators, which include only the best subset of regressors in their final model, have been used by many researchers to address the issue of multicollinearity (Fu & Knight, 2000; Lounici, 2008; Yuan & Lin, 2007). The process for estimating Lasso follows a similar procedure as the ridge estimator, but the difference is that the squared ℓ_2 norm ($\| \beta \|_2^2$) in the ridge has been replaced by ℓ_1 norm ($\| \beta \|_1$).The Lasso technique seeks to minimize the sum of the squared residuals while imposing a constraint on the total absolute value of the coefficients, ensuring that it remains below a specified constant.

$$\hat{\beta}_{\text{Lasso}} = \underbrace{\text{argmin}}_{\beta} (Y - X\beta)^T (Y - X\beta) + \lambda \| \beta \|_1$$

2.5 Elastic net (EN)

Introduced by Zou and Hastie (2005), the EN extends the Lasso and addresses limitations, particularly in variable selection (Zou & Hastie, 2005). It creates regression model that undergoes penalty from both the ridge (L_2 norm) and the Lasso (L_1 norm). The coefficients are effectively reduced, similar to ridge regression, and some are forced to zero, similar to Lasso regression. The L1 norm produces a sparse model, reducing certain coefficients to zero, while the L2 norm removes the constraint on the number of selected variables (Park & Konishi, 2016). Liu and Li (2017) utilized an effective EN with regression coefficients technique to identify significant variables within the spectrum data (Liu & Li, 2017). When this happens, Lasso cannot select more than the specified number of predictors, whereas the EN has the ability to do so. (Emmert-Streib & Dehmer, 2019). The loss function is minimized by EN, and the estimated parameter vector is derived accordingly.

$$\hat{\beta}_{\text{EN}} = \arg\min \left[\sum_{l=1}^{n} \left(y_l - \sum_{i=1}^{p} \beta_i X_{li} \right)^2 + \lambda \left[(1-\alpha) \sum_{i=1}^{p} \beta_i^2 + \alpha \sum_{i=1}^{p} |\beta_i| \right] \right]$$

where $\lambda =$ tuning parameter and $\alpha =$ weight that should be used to determine the allocation for Lasso or ridge such that $0 \le \alpha \le 1$, where $\alpha = 0, \hat{\beta}_{EN}$ becomes $\hat{\beta}_{Ridge}$ and where $\alpha = 1$, $\hat{\beta}_{EN}$ becomes $\hat{\beta}_{Lasso}$.

The MSE for Lasso and EN was calculated by comparing the estimated coefficients of each model with the true coefficient (assumed $\boldsymbol{\beta} = \boldsymbol{I}_p$), and then averaging the results over all simulations to determine the model's performance.

2.6 Ridge parameters

There are numerous ridge estimators available in the literature for different types of models, mostly mentioned in the introduction, to estimate the ridge parameter k. A study compared 80 ridge estimators using Monte Carlo simulation techniques (Mermi et al., 2021). In a recent study, Mermi et al. (2024) conducted a comparison of 366 different ridge parameter estimators that were proposed at different times. For our study, we selected the top 16 estimators from Mermi et al. (2024), excluding the robust estimator. These selected estimators are k_1 through k_4 and k_{10} through k_{21} (Table 2.1). Additionally, we utilized seven estimators, which are k_5 through k_9 proposed by Kibria (2023) and k_{22} and k_{23} by Kibria and Lukman (2020), specifically tailored for asymmetric populations. The 23 ridge parameters used in our study, including both simulation-based and empirical examples, are presented in Table 2.1.

Table 2.1: Existing best estimators based on the Monte Carlo Simulation

Estimators	Citation
$k_1 : max\left(\frac{1}{q_i}\right)$	(Alkhamisi & Shukur, 2007)
$k_2 : max\left(\frac{s^2}{\hat{\beta}_i{}^2}\right)$	(Alkhamisi et al., 2006)
$k_3 : \frac{\hat{\sigma}^2}{min(\frac{\hat{\sigma}^2}{\hat{\alpha}_i{}^2}+\frac{1}{\lambda_i})}$	(Dorugade, 2014)
k_4: $\sqrt{p\sum_{i=1}^{p}\left(\frac{\max(\lambda_i)\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\max(\lambda_i)\hat{\alpha}_i^2}\right)}$	(Lukman & Ayindez, 2017)
k_5: $\text{median}\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2}+\frac{1}{\lambda_i}\right)$	(Kibria, 2023)
k_6: $\max\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2}+\frac{1}{\lambda_i}\right)$	(Kibria, 2023)
k_7: $\min\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2}+\frac{1}{\lambda_i}\right)$	(Kibria, 2023)
k_8: $\text{geometric.mean}\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2}+\frac{1}{\lambda_i}\right)$	(Kibria, 2023)
k_9: $\text{harmonic.mean}\left(\frac{\hat{\sigma}^2}{\hat{\beta}_i^2}+\frac{1}{\lambda_i}\right)$	(Kibria, 2023)
k_{10}: $\text{median}\left(\frac{max(\lambda_i)\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_i\hat{\alpha}_i^2}\right)$	(Lukman & Ayindez, 2017)
k_{11}: $max\left(\frac{max(\lambda_i)\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+max(\lambda_i)\hat{\alpha}_i^2}\right)$	(Lukman et al., 2018)
k_{12}: $\sqrt{p\sum_{i=1}^{p}\left(\frac{\lambda_i\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_i\hat{\alpha}_i^2}\right)}$	(Lukman & Ayindez, 2017)

Estimators	Citation
$k_{13}: \frac{1}{p}\sum_{i=1}^{p} \frac{\max(\lambda_i)\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_i\hat{\alpha}_i^2}$	(Alkhamisi et al., 2006)
$k_{14}: max\left(\frac{1}{\frac{\lambda_i\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_i\hat{\alpha}_i^2}}\right)$	(Muniz et al., 2012)
$k_{15}: \left(\prod_{i=1}^{p} \frac{\max(\lambda_i)\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_i\hat{\alpha}_i^2}\right)^{\frac{1}{p}}$	(Muniz et al., 2012)
$k_{16}: max\left(1/\sqrt{\frac{\lambda_i\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+[max(\lambda_i)]\hat{\alpha}_i^2}}\right)$	(Lukman et al., 2018)
$k_{17}: max\left(1/\frac{\lambda_i\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_i max(\hat{\alpha}_i^2)}\right)$	(Lukman et al., 2018)
$k_{18}: \prod_{i=0}^{p}\left(\frac{1}{q_i}\right)^{\frac{1}{p}}$	(Kibria et al., 2012)
$k_{19}: max\left(1/\frac{\lambda_i\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_i\left(max(\hat{\alpha}_i)\right)^2}\right)$	(Lukman et al., 2018)
$k_{20}: \frac{1}{\sqrt{\frac{2\hat{\sigma}^2}{\max(\lambda_i)\max(\hat{\alpha}_i^2)}}}$	(Ayinde & Lukman, 2016)
$k_{21}: \frac{1}{\sqrt{\frac{2p\hat{\sigma}^2}{\max(\lambda_i)\sum_{i=1}^{p}\hat{\alpha}_i^2}}}$	(Ayinde & Lukman, 2016)
$k_{22}: \hat{\sigma}p^{\left(1+\frac{p}{n}\right)}$	(Kibria & Lukman, 2020)
$k_{23}: \hat{\sigma} \times \max\left(p^{\left(1+\frac{p}{n}\right)}, p^{\left(1+\frac{1}{p}\right)}\right)$	(Kibria & Lukman, 2020)

3. Simulation study

As the theoretical comparison is not feasible, we conducted a Monte Carlo simulation study using the R program in this section.

3.1 Simulation Technique

The data generation process for the models was carried out according to a widely used method as described below (Gibbons, 1981).

$$x_{li} = (1-\rho^2)^{1/2}u_{li} + \rho u_{l,p+1}, l = 1,2,\cdots,n; i = 1,2,\ldots,p$$

where ρ denotes the correlation coefficient between any two explanatory variables, u_{jli} denotes the independent pseudo-random variable obtained from standard normal distribution and p denotes the number of explanatory variables. Moreover, the dependent variable Y was obtained through the following equation.

$$Y_l = \beta_0 + \beta_1 X_{l1} + \beta_2 x_{l2} + \cdots + \beta_p x_{lp} + \varepsilon_l, l = 1,2, \cdots, n$$

where, ε_l is the error term, i.e. $\varepsilon_l \sim N(0, \sigma^2)$.

For modelling, two different values of explanatory variables, $\mathrm{p} = 5$ and 10, three different values of the correlation coefficients between the explanatory variables, $= 0.9, 0.95$, and, 0.99, four different values of the sample sizes, $n = 20, 30, 50$, and 100, and two different values of the error variance $\sigma^2 = 1$ and 5 were taken. The data generation process of the explanatory variables was carried out with the help of the values determined for p, ρ, n, and σ. The experiment was repeated, $\mathrm{N} = 5000$ times.

The MSE was calculated by comparing the estimated coefficients of each model with the true coefficients (assumed $\boldsymbol{\beta} = \boldsymbol{I}_p$) and averaged MSE was calculated based on the following formula over all simulations.

$$AMSE(\hat{\beta}_i^*) = \frac{1}{\mathrm{N}} \sum_{i=1}^{P} \left(\hat{\beta}_i - \beta\right)' \left(\hat{\beta}_i - \beta\right)$$

where $\hat{\beta}_i^*$ is any of the OLE, Ridge, Lasso, or EN estimators, β is the true parameter (assumed $\boldsymbol{\beta} = \boldsymbol{I}_p$), and p is the number of predictors.

3.2 Results and Discussion

In this section, we thoroughly examined a simulation study comparing the performance of current best estimators with traditional OLS, EN, Lasso, and GR. In section 3.2.1, the simulation results in the absence of outliers in the data are discussed, where Tables 3.1.1 to 3.1.3 represent the MSE scenario of the estimators in the presence of the correlation coefficients 0.90, 0.95, and 0.99, respectively. Section 3.2.2 discusses the simulation results in the presence of 10% and 25% outliers in the dataset. The simulation results, showcased in Tables 3.2.1 to 3.2.6, present outcomes for various correlation coefficients, numbers of independent variables, variance numbers, and sample sizes in the presence of outliers of 10% and 25%.

3.2.1 Without Outlier

Table 3.1.1 reveals that, for a correlation coefficient of 0.90 and no outliers, the MSE of the estimators behaves differently based on sample size, number of predictors, and variance. In smaller samples (e.g., 20 or 30 observations) with five predictors and low variance, estimators such as k_3, k_{10}, k_{11}, k_{13}, k_{15}, and GR tend to show relatively low MSE. However, as the number of predictors or variance increases, the MSE also rises. Specifically, with higher variance and a fixed number of predictors, the MSE escalates, and a similar trend is observed when the number of predictors grows. For small samples (e.g., 20 or 30 observations) with high variance, estimators like k_4, k_{12}, and k_{21} are more effective when the number of predictors is unchanged. When both the variance is high and the number of predictors is large, estimators k_4, k_9, k_{12}, and k_{20} are recommended. In contrast, for larger samples, regardless of the variance level, methods such as Lasso, EN, and GR are generally preferred due to their efficiency, particularly in the presence of a correlation of 0.90. It's important to note that OLS does not perform well when collinearity is present among the independent variables. Figure 3.1 illustrates how estimator performance varies with changes in sample size, number of predictors, and variance.

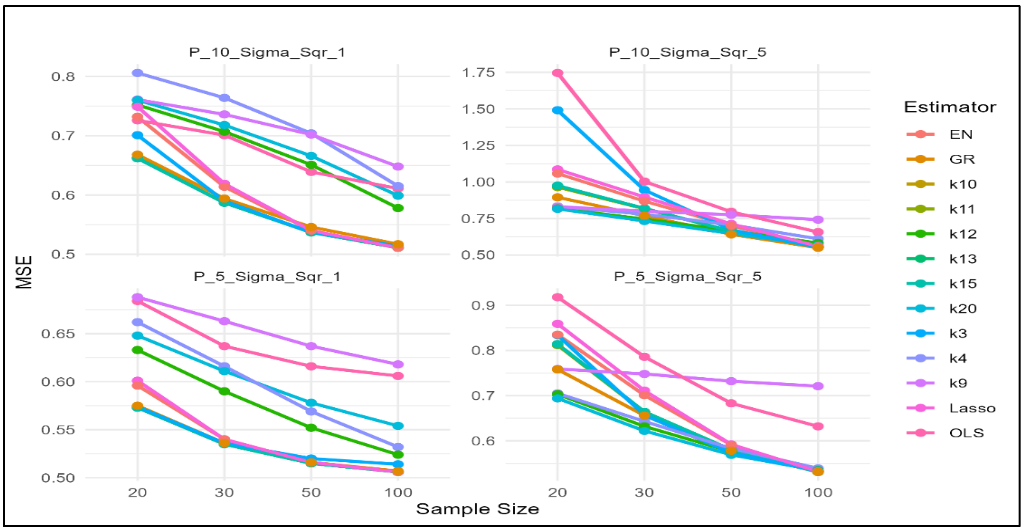

Figure 3.1: MSE of the Estimators when $\rho = 0.9$ and no outlier

Table 3.1.1: MSE values of the estimators with correlation = 0.9

Estimator	P = 5, σ^2 =1				P = 5, σ^2 =5				P = 10, σ^2 =1				P = 10, σ^2 =5			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
OLS	0.684	0.637	0.616	0.606	0.918	0.786	0.683	0.632	0.726	0.701	0.639	0.611	1.746	1.003	0.796	0.657
k_1	0.798	0.851	0.910	0.960	0.902	0.944	0.972	0.989	0.806	0.807	0.843	0.917	0.885	0.892	0.936	0.976
k_2	0.806	0.841	0.884	0.935	0.943	0.969	0.986	0.995	0.853	0.852	0.868	0.915	0.935	0.947	0.973	0.992
k_3	0.574	0.536	0.520	0.514	0.834	0.657	0.573	0.533	0.701	0.590	0.537	0.514	1.491	0.945	0.673	0.553
k_4	0.662	0.616	0.569	0.532	0.705	0.643	0.583	0.539	0.806	0.764	0.704	0.615	0.833	0.779	0.707	0.611
k_5	0.704	0.670	0.640	0.619	0.782	0.766	0.742	0.726	0.793	0.768	0.729	0.653	0.849	0.828	0.800	0.754
k_6	0.640	0.617	0.600	0.591	0.726	0.695	0.675	0.672	0.692	0.653	0.627	0.594	0.866	0.755	0.701	0.670
k_7	0.813	0.768	0.710	0.662	0.913	0.893	0.854	0.803	0.925	0.885	0.833	0.764	0.956	0.962	0.945	0.875
k_8	0.712	0.678	0.645	0.621	0.804	0.788	0.759	0.732	0.805	0.771	0.726	0.661	0.859	0.845	0.820	0.763
k_9	0.688	0.663	0.637	0.618	0.759	0.748	0.732	0.721	0.761	0.736	0.702	0.648	0.831	0.800	0.776	0.742
k_{10}	0.573	0.535	0.515	0.506	0.812	0.664	0.580	0.531	0.663	0.587	0.537	0.511	0.966	0.818	0.663	0.555
k_{11}	0.573	0.535	0.515	0.506	0.812	0.663	0.580	0.531	0.662	0.587	0.537	0.511	0.965	0.817	0.663	0.555
k_{12}	0.633	0.590	0.552	0.524	0.703	0.632	0.576	0.535	0.752	0.707	0.651	0.578	0.819	0.746	0.666	0.582
k_{13}	0.573	0.535	0.515	0.506	0.814	0.664	0.580	0.531	0.662	0.587	0.537	0.511	0.973	0.819	0.663	0.555
k_{14}	0.961	0.972	0.978	0.986	0.953	0.965	0.973	0.982	0.969	0.980	0.989	0.993	0.968	0.978	0.987	0.992
k_{15}	0.573	0.535	0.515	0.506	0.814	0.664	0.580	0.531	0.662	0.587	0.537	0.511	0.975	0.819	0.663	0.555
k_{16}	0.821	0.830	0.836	0.856	0.801	0.806	0.807	0.826	0.881	0.871	0.876	0.885	0.877	0.862	0.864	0.871
k_{17}	0.962	0.973	0.979	0.986	0.953	0.965	0.973	0.982	0.969	0.980	0.989	0.993	0.968	0.978	0.987	0.992
k_{18}	0.785	0.846	0.909	0.960	0.896	0.942	0.972	0.989	0.761	0.788	0.837	0.916	0.861	0.882	0.934	0.976
k_{19}	0.962	0.973	0.979	0.986	0.953	0.965	0.973	0.982	0.969	0.980	0.989	0.993	0.968	0.978	0.987	0.992
k_{20}	0.648	0.611	0.578	0.554	0.694	0.622	0.569	0.534	0.760	0.718	0.666	0.599	0.816	0.733	0.647	0.567
k_{21}	0.628	0.595	0.567	0.547	0.710	0.623	0.567	0.531	0.724	0.685	0.640	0.585	0.843	0.732	0.632	0.557
k_{22}	0.785	0.779	0.775	0.772	0.842	0.841	0.839	0.840	0.904	0.879	0.855	0.827	0.934	0.918	0.905	0.888
k_{23}	0.785	0.773	0.755	0.739	0.842	0.836	0.821	0.812	0.904	0.879	0.855	0.827	0.934	0.918	0.905	0.888
Lasso	0.601	0.540	0.516	0.506	0.859	0.711	0.592	0.532	0.749	0.619	0.540	0.511	1.086	0.897	0.713	0.560
EN	0.596	0.540	0.516	0.506	0.835	0.701	0.591	0.532	0.732	0.614	0.540	0.511	1.058	0.869	0.700	0.559
GR	0.575	0.536	0.516	0.507	0.758	0.656	0.578	0.531	0.668	0.594	0.546	0.517	0.895	0.769	0.642	0.550

Table 3.1.2 shows that the performance of OLS deteriorates significantly as the correlation coefficient rises from 0.90 to 0.95. Consistent with the findings in Table 3.1.1, for small samples with low variance and five predictors, estimators such as k_{10}, k_{11}, k_{13}, k_{15}, and GR exhibit relatively low MSE. While an increase in the number of predictors leads to higher MSE, these estimators maintain stable performance for small sample sizes. Additionally, for small samples with high

variance and a correlation coefficient of 0.95, estimators k_4, k_{12}, and k_{20} are recommended. For larger samples, regardless of variance, Lasso, EN, and GR remain the preferred choices due to their operational efficiency.

Table 3.1.2: MSE values of the estimators with correlation = 0.95

Estimator	P = 5, σ^2 =1				P = 5, σ^2 =5				P = 10, σ^2 =1				P = 10, σ^2 =5			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
OLS	0.731	0.653	0.618	0.606	1.197	0.865	0.695	0.633	0.963	0.792	0.671	0.616	2.799	1.458	0.857	0.684
k_1	0.808	0.845	0.908	0.959	0.900	0.937	0.971	0.988	0.863	0.844	0.847	0.911	0.912	0.903	0.933	0.974
k_2	0.816	0.836	0.884	0.938	0.938	0.964	0.986	0.995	0.884	0.874	0.870	0.908	0.938	0.944	0.969	0.991
k_3	0.610	0.547	0.522	0.515	0.999	0.713	0.582	0.535	0.905	0.669	0.563	0.517	2.203	1.281	0.802	0.574
k_4	0.690	0.631	0.577	0.536	0.742	0.661	0.594	0.541	0.855	0.813	0.742	0.636	0.886	0.832	0.749	0.634
k_5	0.741	0.687	0.649	0.626	0.811	0.776	0.753	0.727	0.860	0.838	0.789	0.674	0.893	0.869	0.840	0.774
k_6	0.663	0.628	0.607	0.597	0.767	0.702	0.684	0.673	0.765	0.711	0.667	0.609	1.036	0.831	0.733	0.680
k_7	0.867	0.808	0.736	0.677	0.926	0.905	0.863	0.806	0.951	0.933	0.891	0.822	0.957	0.963	0.962	0.906
k_8	0.752	0.700	0.657	0.630	0.826	0.797	0.769	0.734	0.867	0.840	0.781	0.689	0.895	0.876	0.852	0.786
k_9	0.715	0.677	0.647	0.626	0.783	0.752	0.741	0.723	0.823	0.798	0.745	0.668	0.897	0.843	0.804	0.755
k_{10}	0.598	0.546	0.517	0.506	0.919	0.718	0.589	0.533	0.750	0.646	0.563	0.515	1.135	0.955	0.751	0.577
k_{11}	0.597	0.546	0.517	0.506	0.919	0.718	0.589	0.533	0.749	0.645	0.563	0.515	1.135	0.955	0.751	0.577
k_{12}	0.660	0.604	0.558	0.527	0.745	0.653	0.584	0.536	0.808	0.762	0.694	0.597	0.904	0.811	0.712	0.603
k_{13}	0.598	0.546	0.517	0.506	0.924	0.719	0.589	0.533	0.749	0.645	0.563	0.515	1.150	0.959	0.752	0.577
k_{14}	0.975	0.979	0.982	0.987	0.970	0.974	0.977	0.983	0.977	0.987	0.993	0.996	0.974	0.986	0.992	0.995
k_{15}	0.598	0.546	0.517	0.506	0.924	0.719	0.589	0.533	0.749	0.645	0.563	0.515	1.155	0.959	0.752	0.577
k_{16}	0.853	0.849	0.850	0.864	0.837	0.823	0.823	0.832	0.915	0.903	0.900	0.906	0.910	0.896	0.892	0.896
k_{17}	0.975	0.980	0.982	0.987	0.970	0.974	0.977	0.983	0.977	0.987	0.993	0.996	0.974	0.986	0.993	0.995
k_{18}	0.794	0.839	0.906	0.959	0.893	0.935	0.971	0.988	0.816	0.823	0.839	0.910	0.894	0.891	0.929	0.974
k_{19}	0.975	0.980	0.982	0.987	0.970	0.974	0.977	0.983	0.977	0.987	0.993	0.996	0.974	0.986	0.993	0.995
k_{20}	0.682	0.629	0.587	0.560	0.731	0.643	0.578	0.535	0.833	0.787	0.717	0.623	0.893	0.806	0.703	0.591
k_{21}	0.661	0.612	0.575	0.552	0.750	0.648	0.575	0.533	0.801	0.753	0.690	0.606	0.935	0.807	0.684	0.577
k_{22}	0.798	0.780	0.778	0.777	0.849	0.836	0.843	0.841	0.919	0.893	0.861	0.829	0.935	0.921	0.907	0.888
k_{23}	0.798	0.775	0.759	0.745	0.849	0.831	0.825	0.813	0.919	0.893	0.861	0.829	0.935	0.921	0.907	0.888
Lasso	0.652	0.559	0.518	0.506	0.971	0.765	0.610	0.534	0.880	0.715	0.579	0.516	1.272	1.032	0.816	0.593
EN	0.642	0.558	0.518	0.506	0.939	0.748	0.605	0.533	0.854	0.701	0.578	0.516	1.234	0.991	0.790	0.588
GR	0.601	0.548	0.519	0.508	0.812	0.687	0.586	0.532	0.761	0.658	0.581	0.526	0.983	0.855	0.698	0.567

Table 3.1.3: MSE values of the estimators with correlation = 0.99

Estimator	P = 5, σ^2 =1				P = 5, σ^2 =5				P = 10, σ^2 =1				P = 10, σ^2 =5			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
OLS	0.941	0.700	0.623	0.607	2.771	0.979	0.715	0.635	2.687	1.436	0.827	0.640	11.705	5.223	2.178	0.806
k_1	0.827	0.846	0.907	0.957	0.909	0.937	0.970	0.988	0.986	0.925	0.872	0.912	0.983	0.942	0.936	0.972
k_2	0.830	0.834	0.885	0.934	0.942	0.965	0.986	0.995	0.988	0.931	0.892	0.910	0.985	0.959	0.968	0.989
k_3	0.816	0.582	0.525	0.515	1.609	0.819	0.596	0.537	2.304	1.283	0.772	0.533	7.544	3.469	1.613	0.656
k_4	0.732	0.658	0.593	0.538	0.792	0.688	0.603	0.545	0.987	0.919	0.810	0.669	1.082	0.947	0.819	0.665
k_5	0.797	0.717	0.665	0.627	0.856	0.798	0.764	0.736	0.983	0.933	0.897	0.710	1.022	0.941	0.905	0.807
k_6	0.708	0.651	0.619	0.597	0.826	0.721	0.688	0.683	1.033	0.867	0.762	0.640	2.073	1.107	0.808	0.691
k_7	0.943	0.886	0.772	0.686	0.953	0.929	0.876	0.814	0.976	0.971	0.969	0.933	0.978	0.973	0.975	0.957
k_8	0.817	0.741	0.677	0.632	0.859	0.820	0.780	0.743	0.984	0.930	0.875	0.742	1.022	0.939	0.894	0.825
k_9	0.755	0.702	0.663	0.628	0.815	0.769	0.749	0.732	0.983	0.914	0.818	0.701	1.202	0.952	0.841	0.772
k_{10}	0.665	0.572	0.521	0.507	1.085	0.789	0.605	0.534	0.985	0.839	0.677	0.532	1.810	1.342	1.028	0.647
k_{11}	0.664	0.572	0.521	0.507	1.084	0.788	0.605	0.534	0.987	0.839	0.677	0.532	1.840	1.346	1.028	0.647
k_{12}	0.709	0.632	0.571	0.528	0.804	0.680	0.593	0.540	0.974	0.899	0.785	0.632	1.283	0.973	0.804	0.636
k_{13}	0.666	0.572	0.521	0.507	1.105	0.790	0.605	0.534	0.990	0.839	0.677	0.532	1.880	1.364	1.033	0.647
k_{14}	0.993	0.990	0.986	0.988	0.991	0.987	0.982	0.985	0.989	0.996	0.999	0.999	0.988	0.996	0.998	0.999
k_{15}	0.667	0.572	0.521	0.507	1.111	0.790	0.605	0.534	0.994	0.840	0.677	0.532	1.929	1.371	1.034	0.647
k_{16}	0.908	0.891	0.866	0.867	0.898	0.871	0.841	0.842	0.971	0.949	0.945	0.949	0.977	0.947	0.937	0.940
k_{17}	0.993	0.990	0.986	0.988	0.991	0.987	0.982	0.985	0.989	0.996	0.999	0.999	0.988	0.996	0.998	0.999
k_{18}	0.807	0.839	0.905	0.957	0.898	0.935	0.970	0.988	0.978	0.920	0.860	0.910	1.015	0.936	0.928	0.971
k_{19}	0.993	0.990	0.986	0.988	0.991	0.987	0.982	0.985	0.989	0.996	0.999	0.999	0.988	0.996	0.998	0.999
k_{20}	0.737	0.661	0.605	0.563	0.780	0.672	0.587	0.539	0.987	0.918	0.808	0.663	1.064	0.948	0.811	0.634
k_{21}	0.718	0.642	0.591	0.555	0.794	0.674	0.583	0.535	0.979	0.907	0.789	0.643	1.149	0.964	0.800	0.616
k_{22}	0.813	0.789	0.786	0.775	0.862	0.845	0.846	0.845	0.979	0.930	0.879	0.838	0.979	0.943	0.912	0.888
k_{23}	0.813	0.784	0.767	0.744	0.862	0.841	0.829	0.817	0.979	0.930	0.879	0.838	0.979	0.943	0.912	0.888
Lasso	0.857	0.615	0.524	0.507	1.504	0.866	0.629	0.536	1.322	1.037	0.806	0.545	2.586	1.644	1.175	0.699
EN	0.822	0.607	0.524	0.507	1.453	0.841	0.621	0.535	1.251	0.985	0.779	0.542	2.455	1.543	1.099	0.682
GR	0.670	0.574	0.524	0.508	0.911	0.716	0.596	0.534	0.983	0.852	0.708	0.557	1.376	1.075	0.824	0.602

From Table 3.1.3, it is evident that conventional methods exhibit lower performance as the correlation among predictors increases (from 0.90 to 0.95 to 0.99). Notably, with the same number of predictors, an increase in variance results in higher MSE. For example, when number of predictors is five and error variance is one for small samples, the models k_{10}, k_{11}, k_{13}, and k_{15} demonstrate low MSE; however, their performance diminishes as the number of predictors and variance increases. Conversely, with small predictors and high variance, models k_4, k_{12}, k_{20}, and k_{21} outperform others, while for high variance and a high number of predictors, models k_7, k_{16}, k_{22}, and k_{23} are recommended. Furthermore, for small samples and low variance, the robust performance of k_{10}, k_{11}, k_{13}, and k_{15} is noticeable, especially when the correlation coefficient is around 0.99. In the case of large samples, traditional models such as Lasso, EN, and GR are preferred due to their ease of use. Figure 3.2 illustrates the performance of the top estimators as the sample size, predictors, and variance increase.

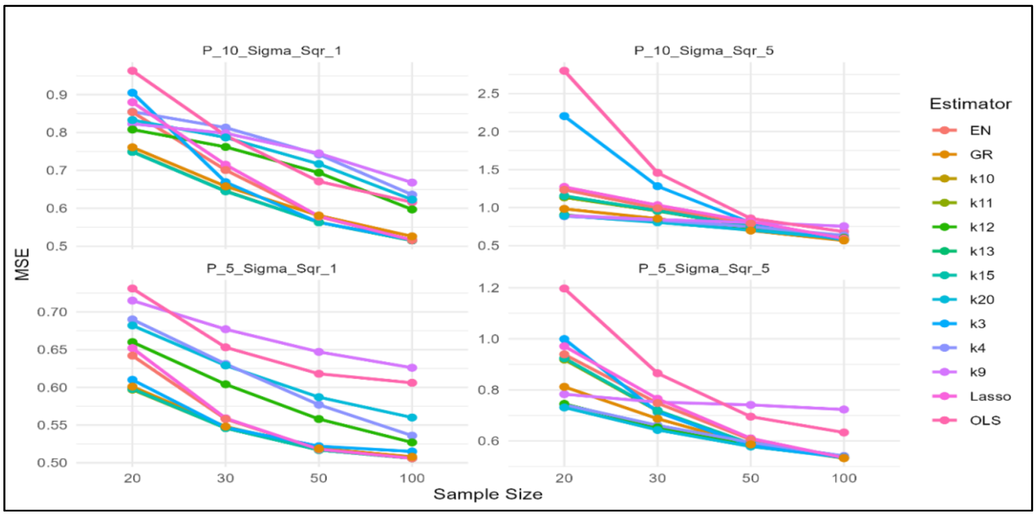

Figure 3.2: MSE of the Estimators when $\rho = 0.99$ and no outlier

3.2.2 With outlier

It is noted from Table 3.2.1 is that the MSE of the estimators is reported for correlated predictors at 0.90 and a 10% outlier present in the data. The addition of outliers is seen to result in an elevated MSE in comparison to the values outlined in Table 3.1.1. It is noteworthy that an increase in variance leads to a higher MSE when the number of predictors remains constant. Notably, in the case of small sample sizes and low variance, it is evident that k_{20}, k_{22}, and k_{23} demonstrate low MSE. However, as the number of predictors and variance increase, k_{22} and k_{23} perform well. Furthermore, for small predictors and high variance, the estimators k_{16}, k_{22}, and k_{23} demonstrate better performance. For high variance and a high number of predictors, k_{16}, k_{22}, and k_{23} are recommended. Additionally, in the case of large samples, traditional models such as Lasso, EN, and GR are recommended due to their ease of operation. Figure 3.3 illustrates the behavior of the top-performing estimators when the sample size, predictors, and variance increase in the presence of a 10% outlier. With an increase in sample size, predictors, and variance, Figure 3.3 portrays the situation for the best-performing estimators.

Table 3.2.1: MSE values of the estimators with correlation = 0.90 and 10% outlier

Estimator	P = 5, σ^2 =1				P = 5, σ^2 =5				P = 10, σ^2 =1				P = 10, σ^2 =5			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
OLS	1.635	1.048	0.856	0.706	1.951	1.178	0.908	0.824	3.548	1.804	1.030	0.887	4.435	2.225	1.159	0.925
k_1	0.829	0.831	0.878	0.935	0.893	0.927	0.962	0.984	0.893	0.862	0.847	0.896	0.911	0.902	0.925	0.969
k_2	0.852	0.817	0.829	0.877	0.911	0.948	0.976	0.990	0.924	0.879	0.857	0.877	0.939	0.943	0.961	0.986
k_3	1.496	0.996	0.736	0.596	1.610	1.075	0.778	0.610	3.231	1.700	0.998	0.676	3.582	1.963	1.097	0.710
k_4	0.928	0.784	0.678	0.588	0.986	0.837	0.707	0.601	0.966	0.845	0.745	0.639	0.992	0.878	0.760	0.650
k_5	0.878	0.762	0.687	0.617	0.871	0.798	0.745	0.694	1.011	0.847	0.765	0.674	1.015	0.870	0.800	0.741
k_6	1.072	0.820	0.684	0.598	1.088	0.842	0.711	0.639	1.832	1.116	0.788	0.635	1.828	1.130	0.789	0.667
k_7	0.871	0.840	0.801	0.712	0.903	0.901	0.891	0.838	0.927	0.914	0.901	0.856	0.948	0.948	0.953	0.937
k_8	0.867	0.766	0.699	0.627	0.866	0.809	0.770	0.717	0.993	0.845	0.769	0.686	0.988	0.871	0.818	0.770
k_9	0.918	0.768	0.682	0.613	0.913	0.794	0.729	0.683	1.179	0.882	0.749	0.657	1.163	0.900	0.779	0.719
k_{10}	1.419	1.000	0.748	0.605	1.654	1.123	0.800	0.623	1.864	1.357	0.956	0.680	2.120	1.642	1.067	0.718
k_{11}	1.407	0.999	0.748	0.605	1.650	1.122	0.800	0.623	1.808	1.351	0.956	0.680	2.102	1.640	1.067	0.718
k_{12}	0.995	0.811	0.686	0.589	1.083	0.878	0.719	0.603	1.106	0.887	0.745	0.630	1.174	0.955	0.773	0.647
k_{13}	1.430	1.002	0.748	0.605	1.662	1.123	0.800	0.623	1.921	1.376	0.958	0.680	2.175	1.653	1.068	0.718
k_{14}	0.946	0.955	0.965	0.976	0.940	0.951	0.962	0.973	0.963	0.975	0.984	0.990	0.962	0.973	0.983	0.989
k_{15}	1.438	1.003	0.748	0.605	1.664	1.123	0.800	0.623	1.996	1.384	0.959	0.680	2.202	1.655	1.068	0.718
k_{16}	0.829	0.799	0.792	0.795	0.830	0.797	0.782	0.781	0.904	0.878	0.858	0.854	0.909	0.873	0.850	0.845
k_{17}	0.948	0.956	0.966	0.976	0.940	0.951	0.963	0.974	0.965	0.975	0.984	0.990	0.962	0.974	0.983	0.989
k_{18}	0.812	0.814	0.871	0.934	0.882	0.923	0.961	0.983	0.911	0.836	0.826	0.891	0.904	0.886	0.920	0.969
k_{19}	0.948	0.956	0.966	0.976	0.940	0.951	0.963	0.974	0.965	0.975	0.984	0.990	0.962	0.974	0.983	0.989
k_{20}	0.814	0.734	0.660	0.587	0.977	0.832	0.702	0.596	0.897	0.829	0.744	0.646	1.043	0.908	0.762	0.643
k_{21}	0.858	0.756	0.664	0.584	1.092	0.891	0.724	0.601	0.953	0.846	0.736	0.631	1.236	1.016	0.801	0.652
k_{22}	0.816	0.765	0.738	0.709	0.841	0.820	0.812	0.796	0.905	0.875	0.837	0.793	0.929	0.912	0.888	0.864
k_{23}	0.816	0.762	0.724	0.681	0.841	0.816	0.795	0.766	0.905	0.875	0.837	0.793	0.929	0.912	0.888	0.864
Lasso	1.033	0.885	0.749	0.613	1.151	0.963	0.804	0.636	1.201	1.046	0.879	0.688	1.417	1.153	0.935	0.726
EN	1.008	0.864	0.743	0.612	1.127	0.941	0.790	0.634	1.173	1.013	0.854	0.679	1.367	1.101	0.905	0.714
GR	0.891	0.797	0.699	0.598	0.993	0.852	0.736	0.614	1.000	0.897	0.772	0.640	1.091	0.983	0.807	0.663

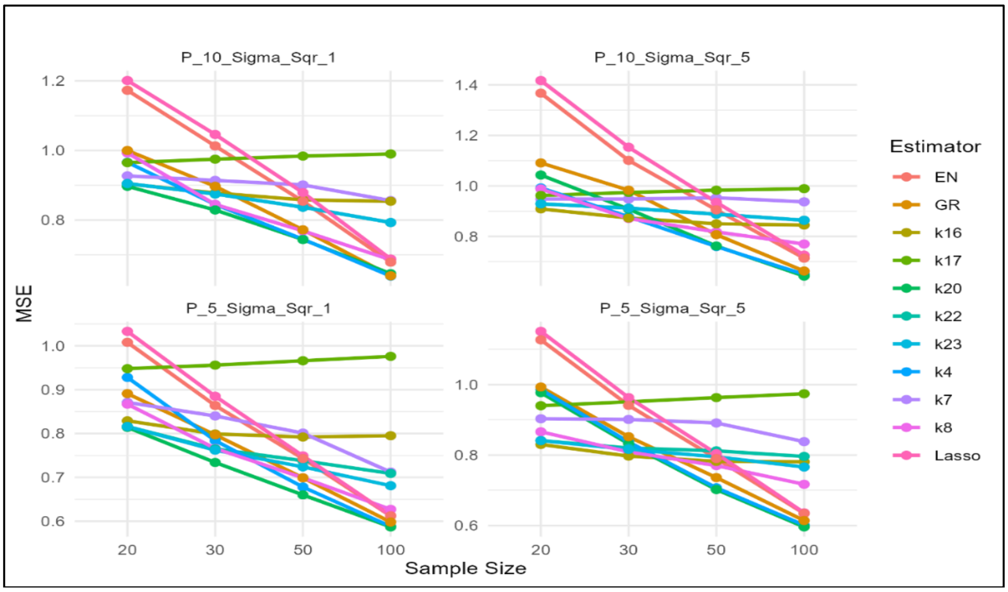

Figure 3.3: MSE of the Estimators when $\rho = 0.90$ and 10% outlier

Table 3.2.2 illustrates the MSE of the estimators when there is a 0.95 correlation among predictors and a 10% outlier in the data. It is evident that the introduction of outliers leads to a higher MSE compared to the findings in Table 3.1.2. In scenarios with the number of predictors is five and error variance is one in small samples, it is observed that k_{18}, k_{20}, k_{22}, and k_{23} yield low MSE, but

k_{18} is noted to be an inconsistent estimator. However, as the number of predictors and variance increases, k_{22} and k_{23} exhibit strong performance. As variance increases with the same number of predictors, the MSE also increases. In situations with fewer predictors and high variance, k_{16}, k_{22}, and k_{23} show superior performance. Similarly, in scenarios with high variance and a high number of predictors, k_{16}, k_{22}, and k_{23} are recommended. In cases of small samples and low variance, k_{20}, k_{22}, and k_{23} demonstrate robust performance when the correlation coefficient is approximately 0.95. For large samples, traditional models such as Lasso, EN, and GR are recommended due to their user-friendly operation.

Table 3.2.2: MSE values of the estimators when correlation = 0.95 and 10% outlier

Estimator	P = 5, σ^2 =1				P = 5, σ^2 =5				P = 10, σ^2 =1				P = 10, σ^2 =5			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
OLS	2.335	1.204	0.806	0.710	2.943	1.484	0.967	0.834	6.649	2.966	1.467	0.871	8.349	3.740	1.766	0.837
k_1	0.852	0.840	0.873	0.935	0.903	0.926	0.959	0.983	0.932	0.903	0.864	0.894	0.936	0.923	0.927	0.968
k_2	0.873	0.826	0.823	0.879	0.927	0.945	0.973	0.990	0.971	0.913	0.870	0.879	0.951	0.949	0.960	0.986
k_3	2.077	1.125	0.778	0.601	2.159	1.277	0.822	0.619	5.939	2.711	1.385	0.753	6.210	3.051	1.559	0.807
k_4	1.009	0.806	0.687	0.591	1.079	0.893	0.726	0.607	1.108	0.919	0.795	0.665	1.137	0.957	0.816	0.679
k_5	0.920	0.780	0.691	0.624	0.927	0.822	0.745	0.702	1.111	0.903	0.816	0.706	1.136	0.924	0.839	0.770
k_6	1.246	0.857	0.693	0.602	1.279	0.910	0.720	0.644	2.756	1.442	0.902	0.663	2.826	1.450	0.903	0.692
k_7	0.888	0.858	0.809	0.728	0.918	0.907	0.889	0.850	0.947	0.934	0.920	0.890	0.955	0.952	0.954	0.949
k_8	0.901	0.780	0.701	0.636	0.914	0.823	0.769	0.728	1.088	0.906	0.814	0.717	1.103	0.922	0.847	0.794
k_9	0.991	0.787	0.684	0.620	0.995	0.825	0.729	0.690	1.459	0.986	0.802	0.682	1.479	0.997	0.824	0.738
k_{10}	1.795	1.110	0.792	0.609	2.179	1.354	0.854	0.632	2.420	1.711	1.224	0.754	2.864	2.138	1.448	0.817
k_{11}	1.776	1.108	0.792	0.609	2.167	1.352	0.854	0.632	2.388	1.709	1.223	0.753	2.847	2.136	1.448	0.817
k_{12}	1.112	0.841	0.698	0.592	1.220	0.955	0.743	0.610	1.386	1.002	0.809	0.656	1.505	1.087	0.848	0.678
k_{13}	1.823	1.114	0.793	0.609	2.204	1.356	0.854	0.632	2.524	1.762	1.233	0.754	2.985	2.170	1.452	0.817
k_{14}	0.965	0.969	0.971	0.978	0.962	0.966	0.967	0.976	0.974	0.985	0.991	0.994	0.971	0.984	0.990	0.994
k_{15}	1.848	1.115	0.793	0.609	2.212	1.356	0.854	0.632	2.720	1.792	1.235	0.754	3.065	2.179	1.452	0.817
k_{16}	0.860	0.827	0.802	0.804	0.865	0.825	0.790	0.791	0.936	0.917	0.889	0.881	0.938	0.912	0.884	0.877
k_{17}	0.967	0.970	0.971	0.978	0.962	0.966	0.967	0.976	0.975	0.985	0.991	0.994	0.971	0.984	0.990	0.994
k_{18}	0.837	0.821	0.865	0.934	0.891	0.921	0.958	0.983	0.979	0.889	0.840	0.887	0.947	0.910	0.920	0.967
k_{19}	0.967	0.970	0.971	0.978	0.962	0.966	0.967	0.976	0.975	0.985	0.991	0.994	0.971	0.984	0.990	0.994
k_{20}	0.848	0.752	0.665	0.592	1.006	0.877	0.718	0.602	0.950	0.888	0.795	0.676	1.105	0.967	0.818	0.672
k_{21}	0.897	0.772	0.670	0.588	1.135	0.950	0.745	0.607	1.014	0.906	0.789	0.658	1.345	1.085	0.866	0.684
k_{22}	0.844	0.779	0.737	0.714	0.863	0.831	0.808	0.800	0.933	0.903	0.851	0.801	0.938	0.924	0.896	0.869
k_{23}	0.844	0.777	0.723	0.687	0.863	0.827	0.792	0.771	0.933	0.903	0.851	0.801	0.938	0.924	0.896	0.869
Lasso	1.162	0.932	0.781	0.619	1.330	1.048	0.826	0.646	1.553	1.202	1.003	0.747	1.863	1.400	1.076	0.793
EN	1.122	0.903	0.770	0.615	1.290	1.005	0.810	0.643	1.445	1.132	0.966	0.733	1.724	1.339	1.039	0.775
GR	0.968	0.825	0.713	0.602	1.073	0.913	0.746	0.623	1.129	0.985	0.849	0.674	1.277	1.103	0.906	0.702

Table 3.2.3 illustrates the MSE of the estimators in the presence of a high predictor correlation (0.99) and a 10% outlier. Compared to Table 3.1.3, the introduction of outliers resulted in increased MSE. Notably, when the number of predictors (p) is five and the error variance is one for small samples, $k_{18,}$ k_{20}, k_{22}, and k_{23} demonstrate low MSE. However, k_{18} is an inconsistent estimator. On the contrary, k_{22} and k_{23} show good performance as both the number of predictors and variance increase. Additionally, an increase in variance with a constant number of predictors leads to an increase in MSE. For small predictors and high variance, k_7, k_8, k_{22}, and k_{23} perform well, while for high variance and a high number of predictors, k_7, k_{16}, k_{22}, and k_{23} are recommended. When dealing with small sample sizes and low variance, k_{20}, k_{22}, and k_{23} prove to be robust when the correlation coefficient is approximately 0.99. Lastly, for large samples, traditional models such as Lasso, EN, and GR are recommended due to their user-friendly operation.

Table 3.2.3: MSE values of the estimators when correlation = 0.99 and 10% outlier

Estimator	P = 5, σ^2 =1				P = 5, σ^2 =5				P = 10, σ^2 =1				P = 10, σ^2 =5			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
OLS	6.527	1.885	0.863	0.718	8.745	2.370	0.981	0.745	28.97	12.31	4.705	1.182	39.34	16.38	6.127	1.339
k_1	0.894	0.853	0.872	0.933	0.924	0.928	0.958	0.983	0.984	0.953	0.908	0.897	0.984	0.963	0.942	0.965
k_2	0.949	0.838	0.826	0.876	0.968	0.945	0.972	0.990	1.057	0.964	0.904	0.883	0.988	0.969	0.965	0.984
k_3	5.328	1.684	0.827	0.607	4.926	1.745	0.887	0.627	25.35	10.79	4.096	1.101	26.06	10.68	4.084	1.159
k_4	1.145	0.865	0.704	0.595	1.220	0.948	0.743	0.612	1.839	1.094	0.872	0.701	1.814	1.126	0.892	0.713
k_5	1.003	0.819	0.707	0.627	1.015	0.853	0.760	0.704	1.257	0.970	0.892	0.741	1.336	0.984	0.900	0.793
k_6	1.563	0.966	0.713	0.605	1.705	0.997	0.738	0.647	7.771	2.853	1.130	0.713	8.729	2.869	1.129	0.724
k_7	0.938	0.893	0.827	0.735	0.943	0.922	0.899	0.854	0.979	0.965	0.957	0.941	0.985	0.972	0.964	0.964
k_8	0.946	0.812	0.718	0.639	0.961	0.847	0.781	0.730	1.257	0.982	0.875	0.756	1.322	0.997	0.889	0.812
k_9	1.112	0.839	0.700	0.622	1.155	0.867	0.741	0.692	2.804	1.261	0.887	0.711	2.953	1.274	0.896	0.749
k_{10}	2.788	1.471	0.838	0.616	3.235	1.776	0.944	0.643	4.704	2.773	1.934	1.023	5.448	3.407	2.480	1.150
k_{11}	2.824	1.460	0.838	0.616	3.239	1.772	0.944	0.643	5.738	2.957	1.944	1.022	5.995	3.492	2.485	1.150
k_{12}	1.311	0.922	0.717	0.596	1.438	1.030	0.766	0.615	3.056	1.353	0.904	0.699	3.242	1.406	0.937	0.721
k_{13}	2.875	1.489	0.839	0.616	3.391	1.791	0.944	0.643	4.997	3.015	2.033	1.026	5.898	3.606	2.540	1.151
k_{14}	0.989	0.984	0.976	0.980	0.988	0.983	0.974	0.978	0.987	0.995	0.998	0.999	0.987	0.995	0.998	0.998
k_{15}	3.154	1.500	0.839	0.616	3.536	1.794	0.944	0.643	5.970	3.312	2.077	1.026	6.482	3.734	2.552	1.151
k_{16}	0.920	0.869	0.820	0.810	0.913	0.865	0.812	0.798	0.978	0.961	0.943	0.931	0.989	0.961	0.933	0.925
k_{17}	0.989	0.985	0.977	0.980	0.988	0.983	0.974	0.978	0.988	0.995	0.998	0.999	0.988	0.995	0.998	0.998
k_{18}	0.882	0.831	0.863	0.931	0.911	0.921	0.957	0.983	1.078	0.967	0.878	0.886	1.049	0.962	0.927	0.964
k_{19}	0.989	0.985	0.977	0.980	0.988	0.983	0.974	0.978	0.988	0.995	0.998	0.999	0.988	0.995	0.998	0.998
k_{20}	0.897	0.781	0.681	0.596	1.037	0.897	0.731	0.606	1.048	0.966	0.862	0.717	1.213	1.029	0.882	0.710
k_{21}	0.935	0.803	0.686	0.591	1.148	0.975	0.763	0.612	1.129	0.993	0.862	0.700	1.491	1.112	0.906	0.724
k_{22}	0.900	0.799	0.746	0.715	0.901	0.842	0.815	0.801	1.030	0.957	0.880	0.810	1.005	0.960	0.909	0.869
k_{23}	0.900	0.797	0.734	0.688	0.901	0.839	0.800	0.772	1.030	0.957	0.880	0.810	1.005	0.960	0.909	0.869
Lasso	1.687	1.122	0.815	0.625	2.330	1.235	0.873	0.657	3.134	2.118	1.437	0.915	4.359	2.689	1.613	0.952
EN	1.598	1.107	0.801	0.622	2.226	1.200	0.856	0.652	3.256	2.036	1.363	0.878	4.400	2.618	1.551	0.915
GR	1.081	0.920	0.739	0.606	1.272	1.008	0.780	0.628	1.616	1.266	0.972	0.737	1.826	1.431	1.047	0.765

Table 3.2.4 presents the MSE of the estimators when there is a 0.90 correlation among predictors and a 25% outlier in the data. A comparison with Table 3.2.1 indicates that as the proportion of outliers in the data set increases, the MSE also increases. Specifically, when the number of predictors (p) is five and the variance is one for small sample sizes, estimators k_7, k_{16}, k_{22}, and k_{23} display small MSE. However, some estimators, such as k_1, k_{14}, k_{18}, and k_{19} show inconsistency under similar conditions. Additionally, it is observed that the MSE of k_{22} and k_{23} increase as the variance and number of predictors increase. Importantly, with an increase in variance, the MSE grows while the number of predictors remains constant. The performance of k_7, k_{16}, k_{22}, and k_{23} are found to be better for small sample sizes and high variance. Furthermore, for a high number of predictors and high variance, these estimators are recommended. However, for large sample sizes, traditional models such as Lasso, EN, and GR are suggested due to their ease of use. From Table 3.2.5, the MSE of the estimators is presented for predictors with a correlation of 0.95 and a 25% outlier in the data. A comparison to Table 3.2.2 reveals that as the outlier in the dataset increased, the MSE also increased. Among the estimators, k_7, k_{20}, k_{22}, and k_{23} demonstrate low MSE for the lower number of predictors and lower error variance in small samples. However, some estimators, such as k_1, k_{14}, k_{16}, k_{17}, and k_{19}, exhibit low MSE for small samples but lack consistency. With an increase in variance and an unchanged number of predictors, the MSE also increases. Notably, for small predictors and high variance, k_7, k_{22}, and k_{23} exhibit favourable performance. Furthermore, for high variance and a high number of predictors, k_7, k_{16}, k_{22}, and k_{23} are recommended. It is observed that traditional regularization methods result in lower MSE when dealing with outliers in

large samples. From Table 3.2.6, the MSE of the estimators is presented, with predictors exhibiting a high correlation of 0.99 and a 25% outlier in the data. A comparison to Table 3.2.3 indicates that an increase in the outlier in the dataset led to a corresponding increase in the MSE. Notably, for a lower number of predictors (e.g. five) and reduced error variance (e.g. one) in small samples, estimators k_7, k_{20}, k_{22}, and k_{23} demonstrate a minimal MSE. However, unlike the consistent estimators such as k_1, k_{17}, and k_{19} shown in Table 2.8, these estimators are not consistent. Moreover, as variance increases, the MSE also increases for consistent estimators. In scenarios with a low number of predictors and high variance, k_7, k_{22}, and k_{23} perform better. Similarly, for high variance and a high number of predictors, k_7, k_{16}, k_{22}, and k_{23} are recommended. Traditional regularization methods are observed to perform poorly when outliers are present in the data, especially for large samples. For situations with a high number of predictors and high variance, k_5, k_9, and k_{20} are recommended. As the sample size, number of predictors, and variance increase, Figure 3.4 visually represents the performance of the most effective estimators. The analysis reveals that k_7, k_{22}, and k_{23} exhibit enhanced performance in the presence of heightened correlation, increased variance, and a greater number of outliers.

Table 3.2.4: MSE values of the estimators when correlation = 0.90 and 25% outlier

Estimator	P = 5, σ^2 =1				P = 5, σ^2 =5				P = 10, σ^2 =1				P = 10, σ^2 =5			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
OLS	3.243	2.104	1.281	0.913	3.699	2.256	1.336	0.947	7.119	3.515	1.818	1.057	8.134	3.881	2.022	1.117
k_1	0.910	0.877	0.865	0.915	0.921	0.926	0.950	0.978	0.953	0.931	0.880	0.886	0.954	0.932	0.925	0.963
k_2	1.046	0.992	0.842	0.834	0.996	0.962	0.954	0.983	1.082	0.983	0.896	0.861	0.968	0.949	0.949	0.979
k_3	2.985	1.984	1.232	0.798	3.046	1.966	1.255	0.822	6.547	3.315	1.753	1.040	6.445	3.395	1.893	1.092
k_4	1.675	1.389	1.029	0.754	1.799	1.436	1.068	0.781	1.447	1.170	0.970	0.806	1.389	1.194	1.004	0.829
k_5	1.424	1.193	0.904	0.710	1.394	1.100	0.863	0.730	1.588	1.148	0.920	0.772	1.493	1.115	0.908	0.782
k_6	2.148	1.542	1.033	0.733	2.180	1.480	0.994	0.731	3.913	2.224	1.297	0.874	3.641	2.054	1.235	0.844
k_7	1.020	0.948	0.866	0.791	1.002	0.943	0.900	0.872	0.985	0.954	0.916	0.889	0.981	0.959	0.946	0.947
k_8	1.371	1.139	0.879	0.710	1.295	1.049	0.848	0.746	1.522	1.142	0.919	0.771	1.402	1.085	0.901	0.796
k_9	1.633	1.282	0.926	0.709	1.607	1.189	0.881	0.720	2.158	1.412	1.009	0.787	1.971	1.312	0.965	0.780
k_{10}	2.875	2.011	1.264	0.811	3.242	2.150	1.319	0.844	3.896	2.705	1.686	1.045	4.065	2.980	1.868	1.103
k_{11}	2.849	2.005	1.263	0.811	3.230	2.148	1.318	0.844	3.692	2.679	1.684	1.045	3.998	2.973	1.868	1.103
k_{12}	1.893	1.500	1.071	0.764	2.068	1.571	1.115	0.792	1.920	1.418	1.073	0.845	1.905	1.478	1.132	0.877
k_{13}	2.898	2.016	1.264	0.811	3.261	2.152	1.319	0.844	4.016	2.756	1.692	1.045	4.201	3.008	1.870	1.104
k_{14}	0.942	0.945	0.952	0.966	0.943	0.945	0.950	0.964	0.963	0.971	0.979	0.986	0.964	0.969	0.978	0.986
k_{15}	2.923	2.018	1.264	0.811	3.267	2.153	1.319	0.844	4.279	2.787	1.693	1.045	4.291	3.015	1.871	1.104
k_{16}	0.947	0.891	0.804	0.769	1.030	0.918	0.808	0.764	0.960	0.932	0.878	0.841	0.996	0.948	0.879	0.839
k_{17}	0.945	0.948	0.953	0.967	0.944	0.946	0.950	0.964	0.965	0.972	0.979	0.986	0.964	0.969	0.979	0.986
k_{18}	0.958	0.877	0.854	0.913	0.918	0.920	0.948	0.978	1.112	0.979	0.868	0.876	1.002	0.934	0.917	0.962
k_{19}	0.945	0.948	0.953	0.967	0.944	0.946	0.950	0.964	0.965	0.972	0.979	0.986	0.964	0.969	0.979	0.986
k_{20}	1.149	1.054	0.866	0.698	1.589	1.339	1.019	0.761	1.071	1.005	0.897	0.766	1.362	1.229	1.035	0.838
k_{21}	1.315	1.180	0.932	0.719	1.880	1.510	1.097	0.784	1.249	1.124	0.964	0.800	1.776	1.504	1.183	0.904
k_{22}	1.040	0.944	0.799	0.711	0.962	0.892	0.816	0.776	0.963	0.935	0.869	0.795	0.952	0.932	0.894	0.854
k_{23}	1.040	0.951	0.804	0.700	0.962	0.895	0.810	0.753	0.963	0.935	0.869	0.795	0.952	0.932	0.894	0.854
Lasso	1.635	1.374	1.068	0.804	1.776	1.374	1.095	0.828	1.875	1.433	1.191	0.932	2.005	1.514	1.229	0.959
EN	1.539	1.327	1.031	0.795	1.686	1.325	1.066	0.817	1.759	1.362	1.150	0.915	2.001	1.468	1.179	0.944
GR	1.324	1.191	0.943	0.745	1.435	1.175	0.970	0.761	1.392	1.176	1.027	0.838	1.500	1.260	1.051	0.862

Table 3.2.5: MSE values of the estimators when correlation = 0.95 and 25% outlier

Estimator	P = 5, σ^2 =1				P = 5, σ^2 =5				P = 10, σ^2 =1				P = 10, σ^2 =5			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
OLS	4.718	2.750	1.461	0.945	5.482	2.995	1.532	0.969	13.045	6.190	2.821	1.337	15.111	6.664	3.113	1.385
k_1	0.943	0.900	0.871	0.915	0.936	0.927	0.948	0.978	0.985	0.966	0.911	0.890	0.979	0.957	0.934	0.958
k_2	1.058	1.009	0.881	0.897	0.991	0.997	0.993	0.982	1.146	1.032	0.932	0.872	0.995	0.966	0.954	0.974
k_3	4.275	2.554	1.403	0.826	4.119	2.467	1.397	0.840	11.875	5.723	2.651	1.298	11.289	5.347	2.728	1.322
k_4	1.965	1.581	1.116	0.773	1.995	1.598	1.141	0.795	1.857	1.347	1.049	0.864	1.751	1.305	1.085	0.875
k_5	1.588	1.304	0.957	0.718	1.492	1.164	0.889	0.740	1.803	1.208	0.957	0.813	1.763	1.163	0.952	0.800
k_6	2.707	1.847	1.133	0.749	2.662	1.734	1.069	0.744	6.338	3.263	1.616	0.986	5.962	2.835	1.496	0.913
k_7	1.037	0.977	0.882	0.794	1.029	0.954	0.897	0.874	1.018	0.975	0.932	0.898	1.010	0.974	0.952	0.944
k_8	1.515	1.240	0.920	0.719	1.402	1.103	0.866	0.757	1.743	1.228	0.969	0.809	1.674	1.155	0.950	0.808
k_9	1.924	1.449	0.991	0.720	1.811	1.310	0.918	0.729	2.931	1.704	1.114	0.843	2.721	1.504	1.056	0.806
k_{10}	3.832	2.559	1.435	0.843	4.276	2.754	1.502	0.866	5.469	3.820	2.360	1.303	5.583	4.006	2.612	1.349
k_{11}	3.793	2.548	1.434	0.843	4.255	2.750	1.501	0.866	5.335	3.796	2.355	1.302	5.566	4.003	2.610	1.349
k_{12}	2.296	1.750	1.175	0.785	2.375	1.799	1.209	0.808	2.731	1.745	1.196	0.927	2.668	1.709	1.259	0.946
k_{13}	3.892	2.571	1.436	0.843	4.336	2.761	1.502	0.866	5.632	3.965	2.385	1.303	5.873	4.091	2.621	1.349
k_{14}	0.961	0.962	0.960	0.970	0.960	0.959	0.957	0.968	0.978	0.981	0.989	0.992	0.976	0.980	0.988	0.991
k_{15}	3.970	2.577	1.436	0.843	4.362	2.762	1.502	0.866	6.288	4.085	2.393	1.303	6.135	4.122	2.622	1.349
k_{16}	0.979	0.917	0.824	0.778	1.043	0.929	0.822	0.775	0.990	0.963	0.913	0.872	1.020	0.977	0.913	0.862
k_{17}	0.963	0.964	0.962	0.970	0.960	0.959	0.957	0.968	0.979	0.982	0.989	0.992	0.976	0.980	0.988	0.991
k_{18}	1.026	0.907	0.859	0.912	0.940	0.920	0.946	0.977	1.195	1.047	0.907	0.878	1.066	0.976	0.926	0.957
k_{19}	0.963	0.964	0.962	0.970	0.960	0.959	0.957	0.968	0.979	0.982	0.989	0.992	0.976	0.980	0.988	0.991
k_{20}	1.214	1.103	0.901	0.707	1.613	1.401	1.067	0.771	1.123	1.058	0.942	0.805	1.416	1.239	1.077	0.876
k_{21}	1.406	1.250	0.980	0.730	1.929	1.619	1.166	0.797	1.321	1.175	1.008	0.847	1.895	1.512	1.239	0.963
k_{22}	1.146	1.003	0.826	0.717	1.016	0.916	0.825	0.782	1.032	0.991	0.906	0.815	0.988	0.961	0.915	0.855
k_{23}	1.146	1.013	0.835	0.706	1.016	0.920	0.820	0.760	1.032	0.991	0.906	0.815	0.988	0.961	0.915	0.855
Lasso	1.995	1.611	1.154	0.829	2.212	1.741	1.182	0.846	2.601	1.819	1.402	1.074	2.877	1.898	1.463	1.093
EN	1.868	1.566	1.118	0.815	2.093	1.652	1.142	0.836	2.478	1.726	1.343	1.038	2.682	1.784	1.398	1.063
GR	1.531	1.351	1.018	0.762	1.682	1.405	1.039	0.776	1.699	1.426	1.156	0.947	1.708	1.446	1.206	0.946

Table 3.2.6: MSE values of the estimators when correlation = 0.99 and 25% outlier

Estimator	P = 5, σ^2 =1				P = 5, σ^2 =5				P = 10, σ^2 =1				P = 10, σ^2 =5			
	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
OLS	14.38	4.801	1.746	0.986	16.59	5.654	1.874	1.235	60.99	26.21	9.718	2.594	69.70	30.19	11.00	2.743
k_1	0.979	0.917	0.876	0.916	0.966	0.940	0.948	0.977	1.019	0.99	0.942	0.901	1.015	0.995	0.955	0.961
k_2	1.015	0.999	0.868	0.841	0.993	0.972	0.953	0.982	1.233	1.075	0.964	0.883	1.041	1.000	0.964	0.976
k_3	12.34	4.353	1.647	0.866	10.01	4.122	1.656	0.897	54.62	23.64	8.644	2.406	49.08	21.19	7.657	2.255
k_4	2.520	1.800	1.192	0.800	2.393	1.859	1.256	0.836	3.969	1.734	1.183	0.948	3.267	1.664	1.189	0.959
k_5	1.824	1.428	1.000	0.737	1.678	1.318	0.936	0.754	1.893	1.151	0.961	0.865	1.877	1.170	0.967	0.844
k_6	3.940	2.280	1.237	0.775	3.854	2.256	1.193	0.772	19.09	7.188	2.316	1.212	18.68	6.908	2.159	1.096
k_7	0.993	0.944	0.885	0.812	0.994	0.961	0.904	0.884	1.038	0.996	0.954	0.926	1.034	1.000	0.964	0.948
k_8	1.528	1.276	0.958	0.735	1.461	1.195	0.904	0.768	1.939	1.223	1.001	0.844	1.905	1.225	0.994	0.841
k_9	2.390	1.630	1.047	0.739	2.273	1.540	0.985	0.745	6.256	2.359	1.285	0.931	5.932	2.216	1.225	0.876
k_{10}	7.553	3.943	1.690	0.883	7.237	4.433	1.815	0.931	11.56	6.689	4.403	2.258	11.03	7.397	4.969	2.373
k_{11}	7.374	3.872	1.688	0.883	7.228	4.410	1.814	0.931	14.73	7.439	4.450	2.254	12.46	7.632	4.989	2.372
k_{12}	3.185	2.085	1.275	0.814	3.028	2.186	1.355	0.853	7.229	2.772	1.381	1.053	6.53	2.566	1.398	1.068
k_{13}	7.496	3.983	1.692	0.883	7.575	4.486	1.816	0.931	11.82	7.385	4.720	2.268	11.99	7.985	5.138	2.376
k_{14}	0.986	0.980	0.968	0.973	0.987	0.981	0.966	0.971	0.991	0.994	0.997	0.998	0.990	0.994	0.997	0.998
k_{15}	8.491	4.049	1.693	0.883	8.000	4.508	1.816	0.931	15.19	8.530	4.919	2.270	13.64	8.488	5.184	2.376
k_{16}	0.982	0.922	0.839	0.791	1.020	0.950	0.842	0.785	1.013	0.989	0.950	0.921	1.047	1.013	0.948	0.918
k_{17}	0.987	0.981	0.969	0.973	0.987	0.981	0.966	0.971	0.991	0.994	0.997	0.998	0.990	0.994	0.997	0.998
k_{18}	1.074	0.928	0.863	0.913	0.987	0.933	0.945	0.977	1.233	1.111	0.948	0.882	1.168	1.059	0.945	0.958
k_{19}	0.987	0.981	0.969	0.973	0.987	0.981	0.966	0.971	0.991	0.994	0.997	0.998	0.990	0.994	0.997	0.998
k_{20}	1.231	1.108	0.917	0.723	1.599	1.445	1.120	0.803	1.222	1.120	0.982	0.840	1.478	1.259	1.081	0.918
k_{21}	1.406	1.259	1.003	0.749	1.895	1.693	1.251	0.836	1.396	1.209	1.042	0.887	1.954	1.443	1.189	1.024
k_{22}	1.292	1.046	0.843	0.730	1.119	0.967	0.840	0.787	1.250	1.119	0.958	0.832	1.125	1.057	0.945	0.870
k_{23}	1.292	1.057	0.854	0.720	1.119	0.973	0.838	0.766	1.250	1.119	0.958	0.832	1.125	1.057	0.945	0.870
Lasso	4.547	2.357	1.312	0.852	4.842	2.715	1.335	0.891	7.452	4.343	2.412	1.569	8.863	4.312	2.563	1.613
EN	4.365	2.196	1.266	0.840	4.724	2.584	1.311	0.882	7.398	4.136	2.255	1.509	8.396	4.235	2.407	1.521
GR	2.147	1.665	1.141	0.786	2.205	1.909	1.162	0.816	2.798	2.071	1.421	1.214	3.098	2.034	1.512	1.229

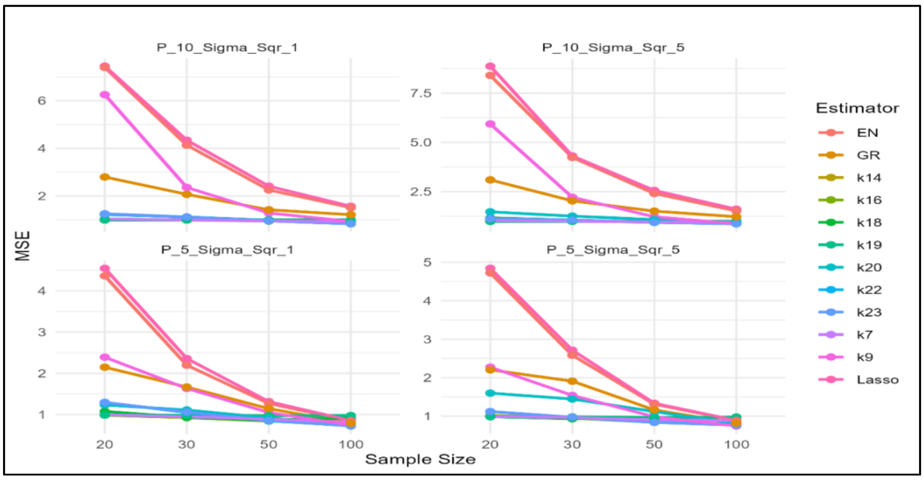

Figure 3.4: MSE of the Estimators when $\rho = 0.99$ and 25% outlier

4. Applications

To illustrate the findings of this paper, we will analyze two real-life datasets in this section.

4.1 Example 1: Portland Cement Data

The initial numerical illustration utilized in this investigation pertains to the Portland cement dataset, which has been extensively referenced in previous scholarly works (Dorugade, 2016; Hadi & Ling, 1998; Lukman et al., 2019; Trenkler & Toutenburg, 1990). The dataset includes 13 data points and four predictors. The independent variables examined are the clinker compounds. The outcome variable (Y) is the heat released after 180 days of curing, measured in calories per gram of cement with 40% water at 35^0 C. The independent variables under consideration are the clinker compounds. Figure 4.1 represents the correlation matrix of the following independent variables.

X_1: Tricalcium aluminate
X_2: Tricalcium silicate
X_3: Tetracalcium aluminoferrite
X_4: Dicalcium silicate

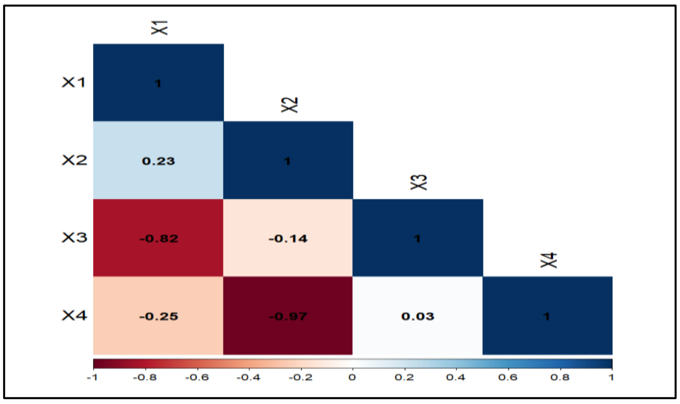

Figure 4.1: Correlation Matrix of Portland Cement Data

In Table 4.1, the Variance Inflation Factor (VIF) of the predictors for the Portland cement data is presented. A VIF value exceeding 10 indicates high multicollinearity. It is noteworthy that all predictors exhibit exceptionally high VIF values, with particular concern surrounding X_2 and X_4, which both exceed a VIF of 250. These findings suggest a high degree of correlation among the variables, resulting in inflated variances due to multicollinearity. Consequently, this scenario may lead to unstable coefficient estimates within the regression model, rendering it challenging to discern the individual predictors' effects.

Moreover, the observed condition number of 423.73 indicates that strong multicollinearity exists among the regressors. However, it is essential to note that the extremely high VIF values for individual variables imply that the overall condition number may not entirely capture the seriousness of the multicollinearity present in specific predictors.

Table 4.1: VIF of the predictors of Portland Cement Data

Predictors	**VIF**
X_1	38.496
X_2	254.423
X_3	46.868
X_4	282.513

Upon examining the dataset, it is evident that there is a moderate level of multicollinearity present. In Figure 4.2, the bar plot displays the estimated MSE of the estimators specifically for the Portland Cement Data. Notably, among the estimators, k_1, k_2, k_5, k_6, k_7, k_8, and k_9 exhibit smaller MSE values. Furthermore, it is worth mentioning that k_{22} and k_{23} outperform Lasso, EN, and GR, which aligns with the findings from the simulation study. Moreover, upon closer inspection, it is apparent that OLS performs significantly worse in terms of MSE, further confirming the outcomes observed in the simulation study.

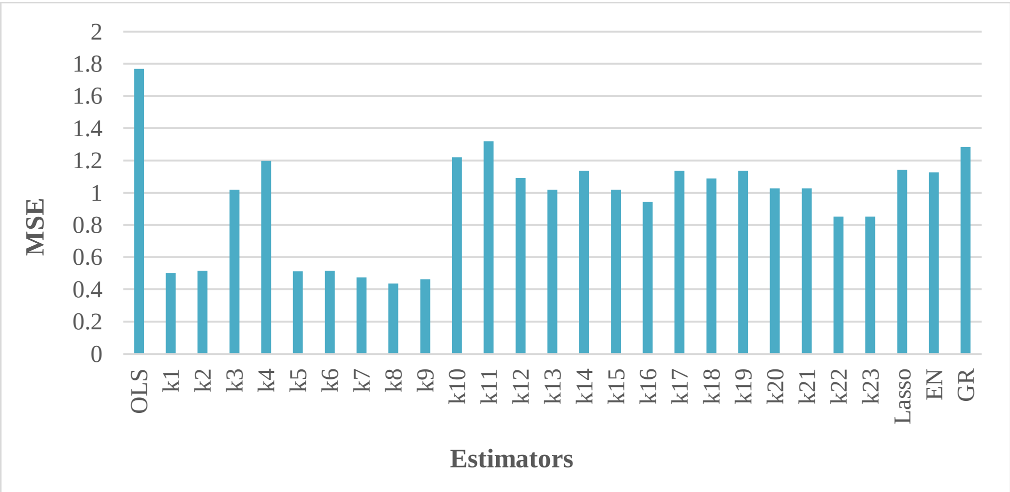

Figure 4.2: MSE of the Estimators for Portland Cement Data

4.2 Example 2: Longley data

We also examined the Longley data to forecast the total derived employment, which depends on a linear combination of several factors: the gross national product implicit price deflator, gross national product, unemployment rate, size of the armed forces, and the non-institutional population 14 years of age. This dataset has been utilized in various studies (Longley, 1967; McDonald & Schwing, 1973; Walker & Birch, 1988; Yousif et al., 2012; Yüzbaşı et al., 2020). The following figure 4.3 represents multicollinearity among the predictors.

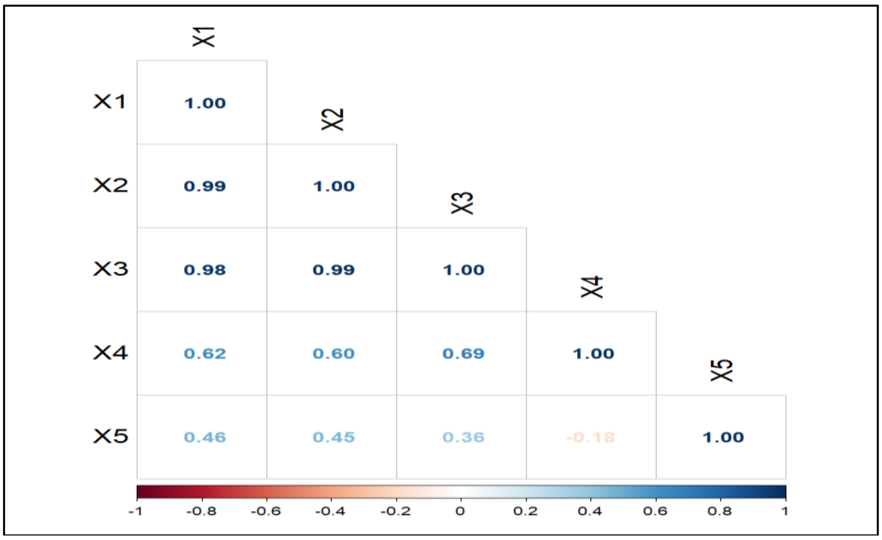

Figure 4.3: Correlation Matrix of Longley Data

Table 4.2 displays the VIF of the predictors for the Longley data. The predictors X_1, X_2, and X_3 demonstrate extremely high VIF values (all well over 100), indicating a high degree of collinearity with other predictors in the model. This can lead to unstable coefficient estimates and make it challenging to determine the individual effect of these predictors. A VIF of 10.787 for X_4 suggests moderate multicollinearity, while a VIF of 2.506 for X_5 indicates relatively low multicollinearity with the other predictors.

This dataset demonstrates extremely severe multicollinearity, as indicated by a condition number of 293682.548. This condition number implies that the design matrix (the set of predictors) is nearly singular, suggesting strong linear dependence among the predictors. Consequently, this could result in significant numerical instability in any regression model applied to this data.

Table 4.2: VIF of the predictors of Longley Data

Predictors	**VIF**
X_1	130.829
X_2	639.049
X_3	339.012
X_4	10.787
X_5	2.506

In Figure 4.4, the analysis revealed the presence of outliers in the Longley data. Specifically, the residual versus fitted plot highlighted data points 6, 16, and 10 as outliers, and this was further corroborated by the normal Q-Q plot, which also identified the same data points as outliers. Out of 16 observations, 3 of them were outliers, which means 18.75% outliers are present in the data.

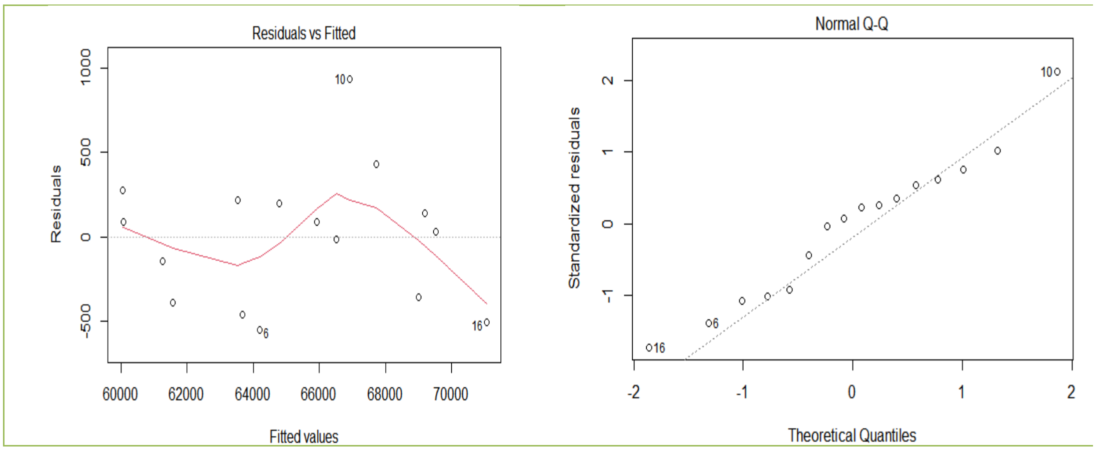

Figure 4.4: Presence of Outlier in Longley Data

Upon conducting a thorough analysis of the Longley dataset, it has become clear that there is a high degree of multicollinearity present, as well as the presence of outliers, accounting for 18.75% of the data points. In Figure 4.5, the bar plot depicts the MSE values for the estimators for the Longley Data. It is noteworthy that among the estimators, k_4, k_5, k_7, k_{10}, k_{12}, k_{13}, k_{16}, k_{22}, and k_{23} exhibit comparatively smaller MSE values. Additionally, it is pertinent to highlight that k_7, k_{16}, k_{22}, and k_{23} outperform all other estimators, which corroborates the findings from the simulation study. Furthermore, upon closer examination, it is evident that the OLS method exhibits notably inferior performance in terms of MSE, as do the traditional Lasso, EN, and GR methods when outliers are present along with severe multicollinearity in terms of small samples. These observations further validate the outcomes identified in the simulation study.

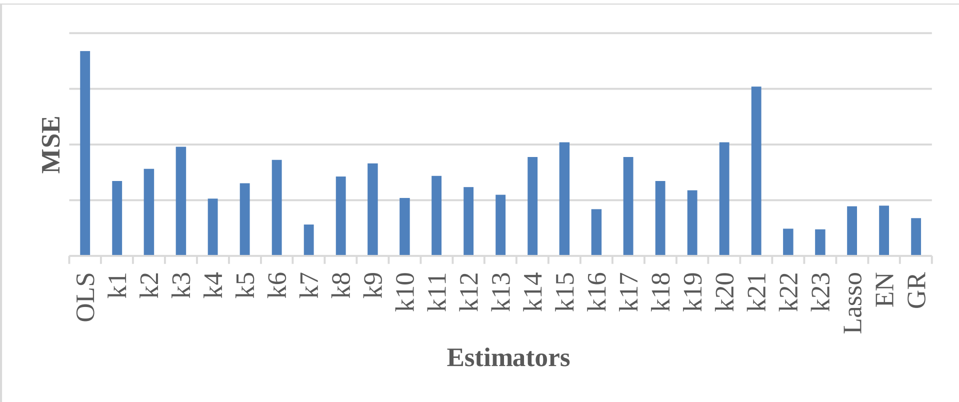

Figure 4.5: MSE of the Estimators for Longley Data

5. Concluding Remarks

In our study, we conducted a comprehensive evaluation of multiple linear regression models in the presence of multicollinearity. We employed ridge regression as a biased estimation technique to obtain more precise estimates of the regression coefficients. From the 366 proposed estimators for the ridge parameter, we focused on the top 16, along with seven estimators suggested by other researchers, and compared their performance against traditional OLS and modern regularization techniques such as Elastic Net (EN), Lasso, and Generalized Ridge (GR) regression. To simulate real-world conditions, we introduced outliers at 10% and 25% levels to assess their impact on estimator performance. The primary comparison criterion was the mean squared error (MSE), a standard metric for evaluating estimator accuracy.

Our findings revealed nuanced insights into the selection of ridge parameter estimators under various parametric conditions. Specifically, with small sample sizes and a high degree of multicollinearity (correlation close to 0.95) in the absence of outliers, the estimators k_{10}, k_{11}, k_{13}, and k_{15} proved to be reliable, balancing bias and variance to produce lower MSEs. However, in the presence of outliers, particularly with small sample sizes and high variance, the estimators k_7, k_{16}, k_{22}, and k_{23} performed better, making them the preferred choices in such situations. We also analyzed two real-world cases, which further supported the simulation results.

For larger sample sizes, the performance dynamics shifted, with GR, EN, and Lasso emerging as robust options. These methods consistently delivered lower MSEs across different levels of multicollinearity, except in cases where significant outliers combined with large variances. Under such challenging conditions, even these robust methods saw performance declines, indicating the need for careful estimator selection in the presence of extreme outliers.

Acknowledgements

The authors would like to express their gratitude to the editorial board of this special issue for the opportunity to contribute this article to the special issue of *International Journal of Statistical Sciences*, dedicated to the memory of the legendary statistician, the late Professor C. R. Rao. His immense contributions to the development of both theoretical and applied statistics have left a lasting impact, and he remains an esteemed figure and icon in the field of statistics.

References

[1] Alin, A. (2010). Multicollinearity. Wiley Interdisciplinary Reviews: Computational Statistics, 2(3), 370–374. https://doi.org/10.1002/wics.84

[2] Alkhamisi, M. A., and Shukur, G. (2007). A Monte Carlo Study of Recent Ridge Parameters. Communications in Statistics - Simulation and Computation, 36(3), 535–547. https://doi.org/10.1080/03610910701208619

[3] Alkhamisi, M., Khalaf, G., and Shukur, G. (2006). Some Modifications for Choosing Ridge Parameters. Communications in Statistics - Theory and Methods, 35(11), 2005–2020. https://doi.org/10.1080/03610920600762905

[4] Arashi, M., and Valizadeh, T. (2015a). Performance of Kibria's methods in partial linear ridge regression model. Statistical Papers, 56(1), 231–246. https://doi.org/10.1007/s00362-014-0578-6

[5] Arashi, M., and Valizadeh, T. (2015b). Performance of Kibria's methods in partial linear ridge regression model. Statistical Papers, 56(1), 231–246. https://doi.org/10.1007/s00362-014-0578-6

[6] Aslam, M. (2014a). Performance of Kibria's Method for the Heteroscedastic Ridge Regression Model: Some Monte Carlo Evidence. Communications in Statistics - Simulation and Computation, 43(4), 673–686. https://doi.org/10.1080/03610918.2012.712185

[7] Aslam, M. (2014b). Performance of Kibria's Method for the Heteroscedastic Ridge Regression Model: Some Monte Carlo Evidence. Communications in Statistics - Simulation and Computation, 43(4), 673–686. https://doi.org/10.1080/03610918.2012.712185

[8] Ayinde, K., and Lukman, A. F. (2016). REVIEW AND CLASSIFICATIONS OF THE RIDGE PARAMETER ESTIMATION TECHNIQUES. Hacettepe Journal of Mathematics and Statistics, 46(113), 1–1. https://doi.org/10.15672/HJMS.201815671

[9] Chen, A.-C., and Emura, T. (2017). A modified Liu-type estimator with an intercept term under mixture experiments. Communications in Statistics - Theory and Methods, 46(13), 6645–6667. https://doi.org/10.1080/03610926.2015.1132327

[10] Damodar, N. and Gujarati, D. C. P. (2013). Basic Econometrics. In Introductory Econometrics: A Practical Approach.

[11] Dempster, A. P., Schatzoff, M., and Wermuth, N. (1977). A Simulation Study of Alternatives to Ordinary Least Squares: Rejoinder. Journal of the American Statistical Association, 72(357), 104. https://doi.org/10.2307/2286916

[12] Dorugade, A. V. (2016). Adjusted ridge estimator and comparison with Kibria's method in linear regression. Journal of the Association of Arab Universities for Basic and Applied Sciences, 21(1), 96–102. https://doi.org/10.1016/j.jaubas.2015.04.002

[13] Dorugade, A. V. (2014). On Comparison of Some Ridge Parameters in Ridge Regression. Sri Lankan Journal of Applied Statistics, 15(1), 31. https://doi.org/10.4038/sljastats.v15i1.6792

[14] Emmert-Streib, F., and Dehmer, M. (2019). High-Dimensional LASSO-Based Computational Regression Models: Regularization, Shrinkage, and Selection. Machine Learning and Knowledge Extraction, 1(1), 359–383. https://doi.org/10.3390/make1010021

[15] Emura, T., Matsumoto, K., Uozumi, R., and Michimae, H. (2024). g.ridge: An R Package for Generalized Ridge Regression for Sparse and High-Dimensional Linear Models. Symmetry, 16(2), 223. https://doi.org/10.3390/sym16020223

[16] Fu, W., and Knight, K. (2000). Asymptotics for lasso-type estimators. The Annals of Statistics, 28(5). https://doi.org/10.1214/aos/1015957397

[17] Gibbons, D. G. (1981). A Simulation Study of Some Ridge Estimators. Journal of the American Statistical Association, 76(373), 131–139. https://doi.org/10.1080/01621459.1981.10477619

[18] Grewal, R., Cote, J. A., and Baumgartner, H. (2004). Multicollinearity and measurement error in structural equation models: Implications for theory testing. Marketing Science, 23(4). https://doi.org/10.1287/mksc.1040.0070

[19 Hadi, A. S., and Ling, R. F. (1998). Some Cautionary Notes on the Use of Principal Components Regression. The American Statistician, 52(1), 15–19. https://doi.org/10.1080/00031305.1998.10480530

[20] Hefnawy, A. El, and Farag, A. (2014). A Combined Nonlinear Programming Model and Kibria Method for Choosing Ridge Parameter Regression. Communications in Statistics - Simulation and Computation, 43(6), 1442–1470. https://doi.org/10.1080/03610918.2012.735317

[21] Herawati, N., Nisa, K., and Setiawan, E. (2018). Regularized Multiple Regression Methods to Deal with Severe Multicollinearity. International Journal of Statistics and Applications, 8(4), 167–172. https://doi.org/10.5923/j.statistics.20180804.02

[22] Herawati, N., Wijayanti, A., Sutrisno, A., Nusyirwan, and Misgiyati. (2024). The Performance of Ridge Regression, LASSO, and Elastic-Net in Controlling Multicollinearity: A Simulation and Application. Journal of Modern Applied Statistical Methods, 23(2), 2–12. https://doi.org/10.56801/Jmasm.V23.i2.4

[23] Hoerl, A. E., Kannard, R. W., and Baldwin, K. F. (1975). Ridge regression:some simulations. Communications in Statistics, 4(2), 105–123. https://doi.org/10.1080/03610927508827232

[24] Hoerl, A. E., and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

[25] Hoque, M. A., and Golam Kibria, B. M. (2023). Some One and Two Parameter Estimators for the Multicollinear Gaussian Linear Regression Model: Simulations and Applications. Surveys in Mathematics and Its Applications, 18, 183–221.

[26] Hoque, M. A., and Kibria, B. M. G. (2024). Performance of some estimators for the multicollinear logistic regression model: theory, simulation, and applications. Research in Statistics, 2(1). https://doi.org/10.1080/27684520.2024.2364747

[27] J. F., L., and P, W. (1976). A simulation study of ridge and other regression estimators. Communications in Statistics - Theory and Methods, 5(4), 307–323. https://doi.org/10.1080/03610927608827353

[28] Khalaf, G., and Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. Communications in Statistics - Theory and Methods, 34(5), 1177–1182. https://doi.org/10.1081/STA-200056836

[29] Kibria, B. M. G. (2003). Performance of Some New Ridge Regression Estimators. Communications in Statistics - Simulation and Computation, 32(2), 419–435. https://doi.org/10.1081/SAC-120017499

[30] Kibria, B. M. G. (2023). More than hundred (100) estimators for estimating the shrinkage parameter in a linear and generalized linear ridge regression models. Journal of Econometrics and Statistics. 2(2), 233-252.

[31] Kibria, B. M. G, and Lukman, A. F. (2020). A new ridge-type estimator for the linear regression model: Simulations and applications. Scientifica, 2020. https://doi.org/10.1155/2020/9758378

[32] Kibria, B. M. G., Månsson, K., and Shukur, G. (2012). Performance of Some Logistic Ridge Regression Estimators. Computational Economics, 40(4), 401–414. https://doi.org/10.1007/s10614-011-9275-x

[33] Li, Y. F., Xie, M., and Goh, T. N. (2010). Adaptive ridge regression system for software cost estimating on multi-collinear datasets. Journal of Systems and Software, 83(11), 2332–2343. https://doi.org/10.1016/j.jss.2010.07.032

[34] Liu, W., and Li, Q. (2017). An efficient elastic net with regression coefficients method for variable selection of spectrum data. PLoS ONE, 12(2), 1–13. https://doi.org/10.1371/journal.pone.0171122

[35] Longley, J. W. (1967). An Appraisal of Least Squares Programs for the Electronic Computer from the Point of View of the User. Journal of the American Statistical Association, 62(319), 819–841. https://doi.org/10.1080/01621459.1967.10500896

[36] Lounici, K. (2008). Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. Electronic Journal of Statistics, 2(none). https://doi.org/10.1214/08-EJS177

[37] Lukman, A. F., Ayinde, K., Binuomote, S., and Clement, O. A. (2019). Modified ridge-type estimator to combat multicollinearity: Application to chemical data. Journal of Chemometrics, 33(5). https://doi.org/10.1002/cem.3125

[38] Lukman, A. F., and Ayindez, K. (2017). Review and classifications of the ridge parameter estimation techniques. Hacettepe Journal of Mathematics and Statistics, 46(5), 953–968. https://doi.org/10.15672/HJMS.201815671

[39] Lukman, A. F., Oluyemi, O. A., Akanbi, O. B., and Clement, O. A. (2018). Classification-Based Ridge Estimation Techniques of Alkhamisi Methods. Journal of Probability and Statistical Science, 16(2), 165–181.

[40] Månsson, K., Shukur, G., and Golam Kibria, B. M. G. (2010). A Simulation Study of Some Ridge Regression Estimators under Different Distributional Assumptions. Communications in Statistics - Simulation and Computation, 39(8), 1639–1670. https://doi.org/10.1080/03610918.2010.508862

[41] McDonald, G. C., and Galarneau, D. I. (1975). A Monte Carlo Evaluation of Some Ridge-Type Estimators. Journal of the American Statistical Association, 70(350), 407–416. https://doi.org/10.1080/01621459.1975.10479882

[42] McDonald, G. C., and Schwing, R. C. (1973). Instabilities of Regression Estimates Relating Air Pollution to Mortality. Technometrics, 15(3), 463–481. https://doi.org/10.1080/00401706.1973.10489073

[43] Melkumova, L. E., and Shatskikh, S. Y. (2017). Comparing Ridge and LASSO estimators for data analysis. Procedia Engineering, 201, 746–755. https://doi.org/10.1016/j.proeng.2017.09.615

[44] Mermi, S., Akkuş, Ö., Göktaş, A., and Gündüz, N. (2024). A new robust ridge parameter estimator having no outlier and ensuring normality for linear regression model. Journal of Radiation Research and Applied Sciences, 17(1), 100788. https://doi.org/10.1016/j.jrras.2023.100788

[45] Mermi, S., Göktaş, A., and Akkuş, Ö. (2021). Are most proposed ridge parameter estimators skewed and do they have any effect on MSE values? Journal of Statistical Computation and Simulation, 91(10), 2074–2093. https://doi.org/10.1080/00949655.2021.1883611

[46] Muniz, G., and Kibria, B. M. G. (2009). On Some Ridge Regression Estimators: An Empirical Comparisons. Communications in Statistics - Simulation and Computation, 38(3), 621–630. https://doi.org/10.1080/03610910802592838

[47] Muniz, G., Kibria, B. M. G., Mansson, K., and Shukur, G. (2012). On developing ridge regression parameters: A graphical investigation. SORT, 36(2), 115–138.

[48] Nayem, H. M., and Kibria, B. M. G. (2024). A simulation study on some confidence Intervals for Estimating the Population Mean Under Asymmetric and Symmetric Distribution Conditions. Journal of Statistics Applications & Probability Letters, 11(3), 123–144.

[49] Park, H., and Konishi, S. (2016). Robust logistic regression modelling via the elastic net-type regularization and tuning parameter selection. Journal of Statistical Computation and Simulation, 86(7), 1450–1461. https://doi.org/10.1080/00949655.2015.1073290

[50] Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[51] Trenkler, G., and Toutenburg, H. (1990). Mean squared error matrix comparisons between biased estimators — An overview of recent results. Statistical Papers, 31(1), 165–179. https://doi.org/10.1007/BF02924687

[52] Walker, E., and Birch, J. B. (1988). Influence Measures in Ridge Regression. Technometrics, 30(2), 221–227. https://doi.org/10.1080/00401706.1988.10488370

[53] Yang, S.-P., and Emura, T. (2017). A Bayesian approach with generalized ridge estimation for high-dimensional regression and testing. Communications in Statistics - Simulation and Computation, 46(8), 6083–6105. https://doi.org/10.1080/03610918.2016.1193195

[54] Yousif, H., Eledum, A., and Alkhalifa, A. A. (2012). Generalized Two Stages Ridge regression Estimator GTR for Multicollinearity and Autocorrelated Errors. Canadian Journal on Science and Engineering Mathematics, 3(3), 79–84.

[55] Yuan, M., and Lin, Y. (2007). On the Non-Negative Garrotte Estimator. Journal of the Royal Statistical Society Series B: Statistical Methodology, 69(2), 143–161. https://doi.org/10.1111/j.1467-9868.2007.00581.x

[56] Yüzbaşı, B., Arashi, M., and Ejaz Ahmed, S. (2020). Shrinkage Estimation Strategies in Generalised Ridge Regression Models: Low/High-Dimension Regime. International Statistical Review, 88(1), 229–251. https://doi.org/10.1111/insr.12351

[57] Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 67(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x