Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.23
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Full Length Research Article

SOME ROBUST REGRESSION ESTIMATORS TO HANDLE OUTLIER AND HIGH DIMENSIONALITY IN LINEAR REGRESSION MODEL

Okeyode O.A. and *Bello A.H.

Department of Statistics, School of Physical Sciences, Federal University of Technology, Akure

*Corresponding Author Email Address: habello@futa.edu.ng

ABSTRACT
This study focuses on addressing the challenges of high dimensionality and outlier detection in linear regression models. With the proliferation of data collection technologies, big data analytics has become prominent, requiring efficient techniques to handle structured, semi-structured, and unstructured data. Traditional linear regression models are often limited in accommodating anomalies such as high dimensionality and outliers that commonly arise in modern datasets. In this research, the aim is on the performance of some robust estimators to handle both high dimensionality and outliers in the linear regression model. The methodology of the research consists of Robust Ridge Regression, Robust Principal Component Analysis, Robust Elastic Net, Least Absolute Shrinkage Selection Operator (LASSO), and Robust Stepwise Regression. The data was generated by conducting a Monte Carlo simulation experiment on a linear Regression Model. The result of the analysis was evaluated and compared using Root Mean Square Error (RMSE). Through a combination of real-life and simulated data, the research findings suggested that Robust Elastic Net estimators provided the most efficient estimator in terms of handling both high dimensionality and outliers. The results highlighted the superiority of the Robust Elastic Net estimator compared to other existing methods, showcasing their efficiency and effectiveness in mitigating the challenges associated with both outliers and high dimensionality in linear regression models.

Keywords: Outlier, High Dimensionality, Robustness, Monte Carlo, and Simulation

INTRODUCTION

Regression analysis provides a statistical framework for analyzing and modelling the relationship between a dependent variable (the response, output, or outcome of interest) and one or more independent variables (the predictors, inputs, or explanatory variables). The fundamental assumption is that the dependent variable can be explained or predicted, to some extent, by variations in the independent variables. The primary objective of regression analysis is often to estimate the expected value of the dependent variable for a given set of known or fixed independent variable values. This estimation typically involves employing the Ordinary Least Squares (OLS) estimator, known for its desirable properties of efficiency, consistency, and sufficiency, under certain assumptions (Johnston, 1992).

One crucial assumption in regression analysis is the orthogonality of the independent variables and that the number of independent features is less than the sample size. Due to the rapid development of advanced technologies over the last decades, high-dimensional data arise in many scientific fields, with the trend towards radically larger numbers of variables (p) but relatively small number of observations (n), i.e. $p \gg n$. For example, in biomedical studies, huge numbers of magnetic resonance images (MRI) and functional MRI data are collected for each subject with hundreds of subjects involved. Satellite imagery has been used in natural resource discovery and agriculture, collecting thousands of high-resolution images. These kinds of examples are many among fields of science, engineering, and humanities, and new knowledge needs to be discovered by using these massive high-throughput data (Donoho *et al*., 2000; Fan & Li, 2006).

Another problem in linear regression data is outlying observations. These observations may be errors, recorded under exceptional circumstances, or belong to another population. These outlying observations are called outliers, and detecting outliers is important to obtain a coherent analysis. Like the problem of multicollinearity, the OLS estimator is extremely sensitive to multiple outliers in linear regression analysis. A single outlier can easily make it biased because of its low breakdown point (Rousseeuw and Leroy, 2003). The breakdown point is the percentage of outliers allowed in a dataset for an estimator to remain unaffected (Donoho and Huber, 1983). In practice, the problem of high dimensionality alone, outliers alone, and both can be present in a real-life data set. This research work focuses on how these problems can be handled jointly.

Traditional methods of estimation using the ordinary least squares estimator (OLSE) are encountering inefficiencies, attributable to the characteristics of contemporary datasets. Scholarly literature has conducted comparative analyses evaluating the efficacy of emergent estimators equipped to handle intricacies in modern big data. Nonetheless, there persists a discernible necessity for a critical examination of the performance of these estimators, particularly in relation to their behaviour in the presence of multiple challenges within a dataset. This research endeavours to investigate the performance of estimators such as enhanced Ridge, LASSO, Elastic-Net, and Principal Components where there exists outliers and in high dimensional settings. The implications of this study are far-reaching and particularly pertinent to the domain of genomic data analysis, as well as other sectors like signal processing, where datasets frequently exhibit high-dimensional characteristics. The insights gleaned from this research are anticipated to facilitate refined analysis methodologies, thereby fostering greater precision and reliability in the interpretation of complex datasets across various scientific and technological fields, such as genetic studies, climatic science, and neuroscience, which involves analyzing data from various sensors to understand brain function.

Outliers constitute observations that markedly deviate from the

collective pattern of a dataset. These anomalous observations are those that are significantly incongruent with the majority of data points and with established domain-specific knowledge regarding plausible values. Such discrepancies may originate from errors in data collection, unique or exceptional circumstances under which data were recorded, or instances where the data points actually pertain to a different population altogether. Given the potentially deleterious impact of outliers on Ordinary Least Squares (OLS) estimations, it is crucial to identify and appropriately address outliers within the dataset prior to conducting analytical procedures. A prevalent method for managing outliers involves their exclusion from the dataset and subsequent reassessment of the influence on regression coefficients or predicted values.

In contrast to the OLS estimator, robust regression methodologies are designed to yield reliable estimators notwithstanding the presence of numerous outliers. Robust regression techniques attenuate the influence of outlier data by assigning diminished weights to such points during the estimation process (Chatterjee and Hadi, 2006). The scholarly literature delineates a multitude of robust regression estimators. Among these, M-estimation stands as the foundational robust approach, further extended by general M-estimation by Huber (1964) and Hampel (1974). The Least Trimmed Squares (LTS) estimation technique is recognized for its high breakdown point, capable of tolerating a substantial proportion of outlier data while retaining robustness (Rousseeuw, 1984). MM-estimation, by Yohai (1987), is distinguished by its high breakdown point and augmented statistical efficiency. Additional methods include the Least Absolute Deviation (LAD), the Least Median of Squares (LMS), and the S estimator.

Hoerl and Kennard (1970) developed the Ridge regression estimator to handle Multicollinearity and with L_2 norm penalty. Frank and Friedman (1993) developed an estimator used in high dimensional case and called it the Bridge regression estimator, which uses the L_p norm penalty. Also in 1996, Tibshirani developed the Least Absolute Shrinkage Selection Operator (LASSO) to handle multicollinearity and/or high dimensionality using the L_1 norm penalty. Zou and Hastie (2005) examined the strengths and weaknesses of LASSO and came up with a list of scenarios where LASSO may experience breakage. They proposed an estimator to address the weakness of LASSO and named it Elastic Nets, which uses both L_1 and L_2 norm penalties. In 2019, Genç and Özkale developed an estimator they named the GO estimator to address the downsides of Elastic Nets.

There are cases in data analytics when $p \gg n$, here, the data contain large number of variables. To perform the task requires some special skills and systems with high configuration; many methods have been proposed in literatures.

When there are outliers in the dataset, all parametric methods experience breakdowns, hence ways of handling outliers in high dimensional setting have been proposed by several authors among which are Robust Principal component regression (Rob PCR) by Rousseeuw (1984) and was improved for high dimensional situation by Hubert and Verboven (2003) and Liu *et al.* (2013) have much details on it.

There is also Robust Partial Least Squares given by Serneels *et. al.* (2005), Filzmoser *et. al.* (2009), and many others to handle high dimensionality and outliers simultaneously. The robust ridge introduced by Maronna (2011) is also a way of addressing outliers in high-dimensional cases. Other methods are the Robust Least Absolute Shrinkage Selection Operator (LASSO) by Alfons *et. al.* (2013) and Alfons (2016), the robust Elastic Net method given by Kurnaz *et. al.* (2018). All these methods are proposed to address outliers in high-dimensional situations.

MATERIALS AND METHODS

The performances of nine different estimation methods were compared. Their robustness in handling outliers and high dimensionality is also examined individually and collectively. The estimators were as follows:

Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS) is a widely used method for estimating the parameters of a linear regression model. It aims to minimize the sum of squared differences between the observed and predicted values. While OLS is effective in many scenarios, its performance can be challenged in high-dimensional and contaminated datasets.

$$Y = X\beta + \varepsilon \qquad (1)$$

where Y is the n −vector of responses (it is also known as the dependent variable), $X_{n\times p}$ is the model matrix of the independent variables, and ε is the n −vector of the error terms.

The parameter estimate can be obtained by the following expression in equation (2)

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

Ridge Regression Estimator

Ridge Regression is a modification of the Ordinary Least Squares (OLS) method that addresses the challenges posed by high-dimensional datasets. It introduced a regularization term to the OLS objective function, providing a solution to multicollinearity and improving the stability of parameter estimates. Ridge Regression introduced a shrinkage parameter, often denoted as λ, which controls the strength of the regularization.

$$\hat{\beta}_{Ridge} = (X'X + \lambda \mathrm{I})^{-1} X'Y \qquad (3)$$

($X'X$ is the correlation matrix (Independent), $\lambda \mathrm{I}$ is the Ridge Parameter, and Y is dependent variable

Least Absolute Shrinkage Selection Operator (LASSO)

LASSO Regression is a regularization technique that extends Ordinary Least Squares (OLS) by adding a penalty term based on the absolute values of the coefficients. LASSO is particularly effective in high-dimensional datasets and is known for inducing sparsity in the model parameters.

Similar to Ridge Regression, LASSO aims to minimize the sum of squared differences between observed and predicted values. However, it introduced a L_1 regularization term, proportional to the absolute values of the coefficients. This penalty encouraged some coefficients to become exactly zero, effectively performing variable selection and simplifying the model.

LASSO's key strength lies in its ability to handle high-dimensional datasets by automatically selecting a subset of relevant features. LASSO introduces a biasing parameter, often denoted as α which controls the strength of the regularization. A higher α increases the penalty on the absolute values of the coefficients, leading to a sparser model. The choice of α is crucial, and cross-validation is commonly used to find the optimal value that balances the trade-off between fitting the data and keeping the model simple.

Elastic Net

Elastic Net Regression is a regularization technique that combines

Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.23
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

the strengths of LASSO (L_1 regularization) and Ridge (L_2 regularization) regression. It introduces both penalty terms, providing a flexible and adaptive approach to handle high-dimensional datasets with potential collinearity among predictors. Elastic Net aims to minimize the sum of squared differences between observed and predicted values, similar to OLS. However, it adds a linear combination of both L_1 and L_2 regularization terms to the objective function. The elastic net mixing parameter, denoted as α, controls the balance between the Lasso and Ridge penalties, allowing for a wide range of regularization strategies.

Principal Components Regression (PCR)

Principal Components Regression (PCR) is a statistical technique that combines the principles of Principal Component Analysis (PCA) with linear regression. PCR aims to address high-dimensional data by transforming the predictors into a new set of uncorrelated variables, the principal components, and then using these components in a regression model. It is a common dimension reduction technique used in the analysis of big data. In standard linear regression, multicollinearity arises when predictors are highly correlated, leading to unstable estimates. PCR begins by performing PCA on the predictor variables, creating principal components that capture the maximum variance in the data. The first few principal components, which retain the most information, are then used as predictors in a linear regression model. PCR provides a natural approach to dimensionality reduction. By using a reduced set of principal components, PCR simplifies the model while retaining most of the information from the original predictors. This can be particularly beneficial in situations where the number of predictors is large, and there is a desire to reduce complexity and improve interpretability.

Robust Ridge

The ridge regression given by Hoerl and Kennard (1970) augments the loss function with an ℓ_2 norm penalty as given below:

$$\mathcal{F}(\beta) = \|Y - X\beta\|_2^2 + \lambda\|\beta\|_2^2 \quad (4)$$

Minimizing $\mathcal{F}(\beta)$ with respect to β yields

$$\hat{\beta}_{Ridge} = (X'X + \lambda I)^{-1}X'y \quad (5)$$

Where I is $p \times p$ identity matrix, and λ is the shrinkage parameter, which satisfies $\lambda \geq 0$ and controls the level of penalty. Obviously, this method is not robust to outliers; therefore, a need for a highly robust estimator for the case $p \gg n$ which has been introduced in Maronna (2011), and the principle is based on MM regression. The classical Ridge estimator can be written in terms of the normal equations, and this is also possible for the MM-type estimator, Hoerl and Kennard (1970):

$$w'(y - \beta_0 1_n - X\hat{\beta}_1) = 0 \quad (6)$$

and

$$(X'wX + \lambda I_p)\hat{\beta}_1 = X'w(y - \beta_0 1_n) \quad (7)$$

This implies

$$\hat{\beta}_1 = (X'wX + \lambda I)^{-1}X'w(y - \beta_0 1_n) \quad (8)$$

Where, 1_n is a vector of ones of length n, $\hat{\beta}_1$ contains all regression coefficients except the intercept, w is a vector of length n with weights, and $W = diag(w)$.

Robust Least Absolute Shrinkage and Selection Operator (RLASSO)

Similar to Ridge regression, LASSO regression minimizes a penalized residual sum-of-squares, but the penalty is not of L_2 type, but of L_1 type. L_1 and L_2 are regularization terms, proportional to the absolute values of the coefficients

$$\mathcal{F}(\beta) = \|Y - X\beta\|_2^2 + \lambda\alpha\|\beta\|_1^1 \quad (9)$$

The L_1 penalty has the effect that some regression coefficients will shrink to exactly zero, which corresponds to a variable selection (Tibshirani, 1996). However, there are fast algorithms in the framework of Least Angle Regression to compute the solution (Efron, *et al.*2004). The (finite-sample) BP of the LASSO estimator is $1/n$. A robust LASSO estimator with BP $(n - h + 1)/n$ is the so-called sparse LTS estimator, defined as:

$$\hat{\beta}_{sparseLTS} = arg\min_{\beta}\{\sum_{i\in\{1:n\}}(y_i - X\beta)^2 + h.\lambda\sum_{j\in\{1:p\}}\alpha|\beta_j|\} \quad (10)$$

This can be implemented in R using the package robustHD (Alfons, 2016). Simulations have demonstrated that this estimator has clear advantages over other robust LASSO estimators (Wang *et. al.*, 2007; Khan *et. al.*, 2007)

Robust Elastic net regression

The Elastic Net estimator is defined as:

$$\hat{\beta}_{E-NET} = arg\min_{\beta}\{\sum_{i\in\{1:n\}}(y_i - X\beta)^2 + \lambda\Psi_\alpha(\beta)\} \quad (11)$$

where $\Psi_\alpha(\beta) = \sum\{\frac{(1-\alpha)}{2}\beta_j^2 + |\beta_j|\}$ (Zou & Hastie, 2005), where j is from 1 to n

This penalty allows us to overcome the limitation of the LASSO estimator to select at most n variables, which is often not enough in settings where p is much larger than n. Moreover, this estimator makes it possible for groups of correlated variables enter the model, and not only single variables from such blocks. A very efficient algorithm for the computation of the Elastic Net estimator is available in the R package glmnet (Friedman *et. al.*, 2016).

The idea of a trimmed version of the Elastic Net estimator has been proposed by Kurnaz *et. al.* (2018). Here, the logistic regression model in the high-dimensional case is treated robustly. The procedures are implemented in the R package enetLTS (Kurnaz *et. al.*, 2018).

Robust Principal Components

With the model matrix, compute robust principal components (PCs) using any of the available methods or packages in R, and take the first rPCs, where the optimal choice of k is usually determined by some rules of thumb, see *Ayinde et. al.,* (2020) for more details. There are many robust PC analysis methods available, typically based on robust covariance estimation (see Maronna *et. al.*, 2019). However, for the case p > n, there are fewer possibilities since, for example, prominent robust covariance estimators like the minimum covariance determinant (MCD) estimator (Rousseeuw & Van, 1999) no longer work in this case because of singularity. There are, however, covariance estimators that work in the high-dimensional situation and methods based on the principle of projection pursuit

Linear Regression Model and Data Generation for Simulation Studies

To examine the proposed and existing estimators, consider a linear regression model of the form:

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots + \beta_p X_{tp} + U_t \quad (12)$$

$t = 1,2, \ldots, n$; $p = 500, 1000$

where $U_t \sim N(0, \sigma^2)$

The model was studied with fixed regressors, X_{ti}, t=1,2,...,n; i=1,2,...,p, such that there exist different levels of multicollinearity among the regressors.

Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.23
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

Procedures for Generating the Error Term

The error term U_t is generated to be normally distributed with mean zero and variance σ^2, $U_t \sim N(0, \sigma^2)$. In this study σ values were 0.5, 1, 5, and 10.

Procedures for Generating the Explanatory Variables

Following the simulation procedure used by Ayinde (2008) and Fayose *et. al.* (2023); the equation to generate the explanatory variables is given as:

$$X_{ti} = (1-\rho^2)^{\frac{1}{2}} Z_{ti} + \rho Z_{tp} \quad (13)$$

t=1, 2, 3,..., n. i=1, 2,...p.

where Z_{ti} is an independent standard normal distribution with mean zero and unit variance, ρ is the correlation between any two explanatory variables, and p is the number of explanatory variables. The values of ρ were taken as 0.8, 0.9, 0.95, 0.99, 0.999, and 0.9999, respectively. Thus, the correlations between the variable is the same. In this study, the number of explanatory variables (p) was taken to be 500 and 1000.

Procedures for Determining the Dependent Variable

β_0 was taken to be identically zero. When P=500, the values of β were chosen to be: β_1=0.8, β_2=0.1, β_3=0.6, and the rest are all zeroes, making only 3 variables to have a contribution to the model. When p=1000, the values of β were chosen to be: β_1=0.4, β_2=0.1, β_3=0.6, β_4=0.2, β_5=0.25, β_6=0.3, β_7=0.53, and others are zeroes. The parameter values were chosen such that $\beta'\beta$=1, which is a common restriction in simulation studies of this type. Ayinde (2008). Sample sizes were varied between 10, 20, 30, 40, and 50. Three different values of σ: 0.5, 1, and 5 were also used. At a specified value of *n, p, σ,* and *U*, the fixed *X*s are first generated, followed by the *U*, and the values of *Y* are then obtained using the regression model. Also, the dependent variable would be computed such that there are 10%, 20% outliers.

$$y_i = y_{\max} \times M + y_i \quad (14)$$

Where M is is the magnitude of the outlier, which would be specified as 5, 10, and 50.

Criterion for Investigation and Performance of Ridge Parameters

Comparisons of the Root Mean Square Errors of various methods would be done and replicated 10000 times. The synthetic simulated dataset would be divided into test (25%) and train (75%) sets. After the training of the models on the data, the validation would be performed using the 25% test dataset, and the RMSE is obtained as:

$$RMSE_i = \sqrt{\frac{1}{N}\sum_j \sum_i (y_{ij} - \hat{y}_{ij})^2} \quad \text{i and j from 1 to n} \quad (15)$$

The Root mean square errors (RMSE) of the existing and the proposed estimators are compared with cross-validation. The RMSE of each estimator is ranked in ascending order, the biasing parameter that produced estimates with MSE ranked less than or equal to five $(Rank_{MSE} \leq 5)$ are counted over six levels of multicollinearity, and four levels of error variances.

Real Life Application (Flu Dataset)

The methods of estimation discussed were applied to the Flu Dataset. This dataset is a benchmark dataset found on the UCI machine learning benchmark dataset repository. It contains data on blood samples from patients who have just received/haven't been placed on antibiotics before 'flu vaccines'.

RESULTS AND DISCUSSION

Performance of penalized estimators in a low-dimensional setting

Table 1: RMSE of Estimators in Low Dimensional Simulation scenario when P=3

P	N	Sig	Percentage	OLS	Ridge	LASSO	E-NET	PC	Rob-Ridge	Rob-LASSO	Rob-ENET	Rob-PC
3	10	0.5	0.1	1.92	1.86	1.84	1.43	1.43	1.32	1.28	1.21	1.15
			0.2	3.88	3.76	3.51	3.04	3.04	2.91	2.82	2.70	2.53
		1	0.1	13.23	9.46	11.90	4.21	4.21	2.90	5.23	5.57	9.16
			0.2	24.58	17.66	17.10	7.71	7.71	5.96	8.58	11.15	15.15
		5	0.1	48.62	29.44	43.38	13.10	13.09	8.08	17.71	28.69	39.32
			0.2	89.56	54.36	59.86	23.85	23.85	17.39	29.67	55.51	69.29
	20	0.5	0.1	2.27	2.26	2.28	2.13	2.13	2.09	2.14	1.83	1.82
			0.2	2.77	2.73	2.67	2.42	2.42	2.34	2.36	2.15	2.12
		1	0.1	10.57	9.29	11.11	4.97	4.97	3.00	3.24	4.76	6.63
			0.2	18.08	15.49	15.73	7.84	7.79	5.46	4.72	8.32	10.89
		5	0.1	36.56	28.51	38.73	14.64	14.63	7.33	8.92	22.38	28.18
			0.2	66.02	50.27	56.64	25.63	25.60	16.48	14.30	41.60	49.74
	30	0.5	0.1	2.84	2.84	2.84	2.80	2.80	2.79	2.81	2.67	2.67

Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.23
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

			0.2	2.70	2.69	2.70	2.62	2.62	2.61	2.64	2.52	2.52
		1	0.1	5.15	5.05	6.60	3.54	3.53	2.41	2.53	2.88	2.92
			0.2	7.00	6.85	9.13	4.13	4.13	2.86	3.40	4.76	4.78
		5	0.1	12.44	11.86	18.16	6.12	6.10	3.56	5.46	9.18	9.73
			0.2	20.61	19.74	29.18	9.14	9.13	4.77	7.85	16.24	17.04
	50	0.5	0.1	2.48	2.48	2.48	2.46	2.45	2.44	2.43	2.35	2.35
			0.2	2.75	2.75	2.75	2.71	2.71	2.69	2.68	2.60	2.60
		1	0.1	3.64	3.62	4.45	2.88	2.88	1.79	1.74	1.99	2.05
			0.2	4.89	4.85	6.25	3.57	3.56	2.39	2.34	2.77	2.79
		5	0.1	7.27	7.11	10.45	4.34	4.33	2.51	3.40	5.55	5.76
			0.2	11.58	11.34	16.90	6.37	6.35	3.34	4.50	8.68	9.02

From Table 1, the RMSE of estimators in low dimensionality when p=3 shows that at a sample size of 10 and at 0.05 significance, Rob-PC performed as the best estimator among others. As the significant increase to 1 and 5, the best estimator is Rob-ridge. As the sample size increases, Rob-ridge, Rob-Lasso, and E-Net compete favorably well as the best estimators.

Table 2: RMSE of Estimators in Low Dimensional Simulation scenario when P=7

P	N	Sig	Percent								
age	OLS	Ridge	LASSO	E-NET	PC	Rob-Ridge	Rob-LASSO		Rob-ENET	Rob-PC	
3	10	0.5	0.1	5.83	5.64	5.58	4.34	4.34	4.01	3.88	3.67
	3.49										
			0.2	11.77	11.41	10.65	9.22	9.22	8.83	8.56	8.19
	7.68										
		1	0.1	40.15	28.71	36.11	12.77	12.77	8.80	15.87	16.90
	27.80										
			0.2	74.59	53.59	51.89	23.40	23.40	18.09	26.04	33.83
	45.97										
		5	0.1	147.53	89.33	131.63	39.75	39.72	24.52	53.74	87.06
	119.31										
			0.2	271.76	164.95	181.64	72.37	72.37	52.77	90.03	168.44
	210.26										
	20	0.5	0.1	6.89	6.86	6.92	6.46	6.46	6.34	6.49	5.55
	5.52										
			0.2	8.41	8.28	8.10	7.34	7.34	7.10	7.16	6.52
	6.43										
		1	0.1	32.07	28.19	33.71	15.08	15.08	9.10	9.83	14.44
	20.12										
			0.2	54.86	47.00	47.73	23.79	23.64	16.57	14.32	25.25
	33.04										
		5	0.1	110.94	86.51	117.52	44.42	44.39	22.24	27.07	67.91
	85.51										
			0.2	200.33	152.54	171.87	77.77	77.68	50.01	43.39	126.23
	150.93										
	30	0.5	0.1	8.62	8.62	8.62	8.50	8.50	8.47	8.53	8.10
	8.10										
			0.2	8.19	8.16	8.19	7.95	7.95	7.92	8.01	7.65
	7.65										
		1	0.1	15.63	15.32	20.03	10.74	10.71	7.31	7.68	8.74
	8.86										
			0.2	21.24	20.79	27.70	12.53	12.53	8.68	10.32	14.44

Science World Journal Vol. 21(No 1) 2026 https://dx.doi.org/10.4314/swj.v21i1.23
www.scienceworldjournal.org
ISSN: 1597-6343 (Online), ISSN: 2756-391X (Print)
Published by Faculty of Science, Kaduna State University

14.50										
	5	0.1	37.75	35.99	55.11	18.57	18.51	10.80	16.57	27.86
29.53										
		0.2	62.54	59.90	88.54	27.73	27.70	14.47	23.82	49.28
51.71										
50	0.5	0.1	7.53	7.53	7.53	7.46	7.43	7.40	7.37	7.13
7.13										
		0.2	8.34	8.34	8.34	8.22	8.22	8.16	8.13	7.89
7.89										
	1	0.1	11.05	10.98	13.50	8.74	8.74	5.43	5.28	6.04
6.22										
		0.2	14.84	14.72	18.97	10.83	10.80	7.25	7.10	8.41
8.47										
	5	0.1	22.06	21.57	31.71	13.17	13.14	7.62	10.32	16.84
17.48										
		0.2	35.14	34.41	51.28	19.33	19.27	10.14	13.65	26.34
27.37										

Performance of penalized estimators in Low Dimensional Setting

This section considers samples of results from a simulation study of penalized regression estimators in low dimensional setting.

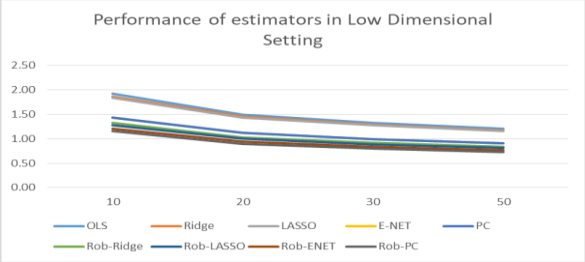

Figure 1: Performance of the estimator when the percentage of outliers in 10%, and the error standard deviation is 0.5 over Sample sizes.

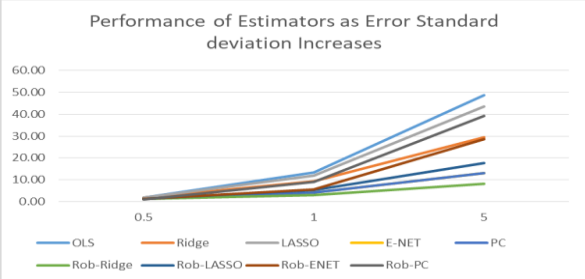

Figure 2: Performance of the estimator when the percentage of outliers in 10%, and the Sample size is 10; As the Error standard deviation increases.

From figures 1 and 2, it can be seen that the RMSE reduces with respect to an increase in sample size; this corroborates the conclusion drawn by Kibria and Lukman (2020) and Ayinde *et al*. (2020). Also, as the error standard deviation increases, the Root Mean Squared Error increases. In the low-dimensional setting, Robust Ridge tends to perform most efficiently among the penalized regression estimators. This is because of the joint capabilities of a robust estimator and the L_2 penalty that handles both outliers and non-orthogonality simultaneously. The OLS estimator would not be included in the high-dimensional setting. This is due to the inability to handle the low rank of the model matrix X.

Performance of penalized estimators in a high-dimensional setting

The results of the performances in terms of RMSE for the penalized and robust penalized estimators in high-dimensional datasets. The RMSE of eight (8) estimators was obtained for all the specifications: From the results, it can be seen that LASSO, E-NET, and PC outperformed the robust penalized estimators. Here, the Elastic Net exhibits the most strength since Multicollinearity was incorporated into the simulated dataset. This corroborates the fact that E-Net combines both the L_1 and L_2 penalties to handle both high dimensionality and multicollinearity. Also, the increase in MSE is steeper as the correlation approaches 1, which indicates that the estimators are becoming less accurate at a faster rate as features become almost perfectly correlated. The sample size has some effect on the MSE. With a larger number of features($P = 1000$), the MSE values are generally higher than with a smaller dimension ($P = 500$), indicating that having more data tends to result in better performance for these estimators

The entire results are summarized in Tables 4 and 5. The data suggests that the rob_enet estimator demonstrated the highest frequency of efficiency across all sample sizes, with a total of 213 occurrences out of 432. This indicated a robust performance by the robust version of Elastic Net in scenarios characterized by outliers and high dimensionality. Its efficiency is most pronounced at the sample size of 500, where it occurred 57 times, suggesting that rob_enet benefited from larger sample sizes under the given conditions.

The PC estimator, a model using Principal Component analysis, showed a moderate frequency of efficiency, with a total of 70 occurrences. It exhibited a steady increase in efficiency frequency as the sample size grew, which indicated its relative suitability in high-dimensional settings, with improved performance as more data became available. Ridge regression displayed a similar trend to 'PC', with a total frequency of 57. It showed a modest increase in frequency as the sample size increased from 50 to 100, but does

not demonstrate a consistent increase with larger sample sizes, indicating a possible plateau in efficiency gains beyond a certain data quantity. The rob_PC estimator showed 43 occurrences of efficiency, with no clear trend in relation to sample size, suggesting that its efficiency may be less sensitive to the number of samples under these conditions.

The rob_lasso and rob_ridge estimators have the lowest frequencies of efficiency, with 25 and 22 occurrences, respectively. This outcome suggested that the robust versions of LASSO and Ridge regression are less frequently efficient in the specified conditions, or they may require more specific conditions or parameter tuning to achieve efficiency. The standard ENET and LASSO estimators each have only one occurrence of efficiency, indicating that in the presence of outliers and high dimensionality, the non-robust versions of these estimators are rarely the most efficient choice. This implied that the robust Elastic Net estimator performed best for high dimensional dataset in the presence of outliers.

Table 3: Frequencies of Efficient Estimators when P=500

Estimator	Sample Sizes				
	50	100	300	500	Total
rob_enet	55	54	47	57	213
PC	15	17	18	20	70
Ridge	14	16	13	14	57
rob_PC	11	9	12	11	43
rob_lasso	6	8	7	4	25
rob_ridge	7	2	11	2	22
enet	0	1	0	0	1
lasso	0	1	0	0	1
Total	108	108	108	108	432

Real-life Application

To establish the facts drawn from the simulation studies, the methods of estimation discussed were applied to the Flu Dataset. This dataset is a benchmark dataset found on the UCI machine learning benchmark dataset repository. It is a high-dimensional dataset with 2436 columns and 108 rows. Table 4 shows the result of the analysis of the dataset and performance of the estimators using the Root Mean Square Error (RMSE) as the criterion for comparison:

Table 4: Real-life performances of the Estimators

Name	RMSE	Active set
Rob_ENET	19.19075	27
PC	22.78247	2435
Rob_Ridge	24.41012	2435
Rob_LASSO	28.80687	25
ENET	29.89856	17
Rob_PC	62.9365	2435
Ridge	349.628	2435
LASSO	379.8551	23

Conclusions

Based on the findings outlined above, it is evident that the robust penalized estimators showcased superior performance when compared to conventional estimators in the presence of observations, regardless of whether the dataset was of high or low dimensionality.

Specifically, when examining real-life datasets and assessing the Root Mean Square Error (RMSE) values, the estimators ranked in ascending order based on their RMSE performance were: Rob_Enet, PC, Rob_Ridge, Rob_LASSO, ENET, Rob_PC, Ridge, and LASSO. This ranking elucidates the comparative effectiveness of these estimators in handling real-world data, with robust versions consistently demonstrating more resilience against outliers and high-dimensional complexities. These collective observations reinforce the robustness of the penalized estimators in handling outlier presence, demonstrating their adaptability across various scenarios, while also highlighting the impact of sample size, multicollinearity, and dataset dimensionality on the accuracy of estimation techniques in linear regression models.

REFERENCES

Alfons, A. (2016). robustHD: Robust methods for high-dimensional data (R package version 0.4.0) [Computer software manual]. Vienna, Austria. Retrieved from http://CRAN.R-project.org/package= robustHD.

Ayinde, K. (2008). Performances of Some Estimators of Linear Model when Stochastic Regressors are correlated with Autocorrelated Error Terms. European Journal of Scientific Research, 20(3), 558-571

Ayinde, K., Lukman, A. F., Alabi, O. O., & Bello, A. H. (2020). A new approach of principal component regression estimator with applications to collinear data. International Journal of Engineering Research and Technology, 13(7), 1616–1622.

Chatterjee, S. and Hadi, A.S. (2006). Regression analysis by example. New Jersey: John Wiley & Sons.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. The Annals of Statistics, 32(2), 407–499.

Fayose Taiwo Stephen, Kayode Ayinde, Olatayo Olusegun Alabi and Abimbola Hamidu Bello (2023). Robust weighted ridge regression based on S – estimator. Nigeria Society of Physical Sciences. African Scientific Reports 2 (2023) 126

Fan, J., & Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. arXiv preprint math/0602133.

Filzmoser, P., Serneels, S., Maronna, R., & Van Espen, P. (2009). Robust multivariate methods in chemometrics. In S. Brown, R. Tauler, & B. Walczak (Eds.), Comprehensive chemometrics: Chemical and biochemical data analysis (pp. 681–722). Amsterdam, Netherlands: Elsevier.

Frank, I. E & Friedman, J. H.(1993). A Statistical View of Some Chemometrics Regression Tools. Technometrics, Vol. 35, No. 2, pp. 109-135

Hampel, F. (1974). The influence curve and its role in robust estimation. Journal of the American Statistical Association, 69, 383–393.

Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12, 55-67.

Huber, P. (1964). Robust estimation of a location parameter. Annals of Mathematical Statistics, 35, 73–101.

Kurnaz, F. S., Hoffmann, I., & Filzmoser, P. (2018). Robust and sparse estimation methods for high-dimensional linear and logistic regression. Chemometrics and Intelligent Laboratory Systems, 172, 211-222.

Liu, J., Wang, C., Gao, J., & Han, J. (2013). Multi-view clustering via joint nonnegative matrix factorization. In Proceedings of the 2013 SIAM international conference on data mining (pp. 252-260). Society for Industrial and Applied Mathematics.

Maronna, R. A. (2011). Robust ridge regression for high-dimensional data. Technometrics, 53(1), 44-53.

Rousseeuw, P. (1984). Least median of squares regression. Journal of the American Statistical Association, 79, 871–880.

Rousseeuw, P., & van Zomeren, B. (1990). Unmasking multivariate outliers and leverage points. Journal of the American Statistical Association, 85(411), 633–639.

Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the lasso". Journal of the Royal Statistical Society. Series B (methodological). Wiley. 58 (1): 267–88. JSTOR 2346178.

Wang, H., Li, G., & Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-Lasso. Journal of Business & Economic Statistics, 25(3), 347–355.

Yohai, V. (1987). High breakdown-point and high efficiency robust estimates for regression. The Annals of Statistics, 15, 642–656.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418-1429.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B, 67(2), 301–32