Table 4.2. VIF after the Application of Ridge Regression (k=1)

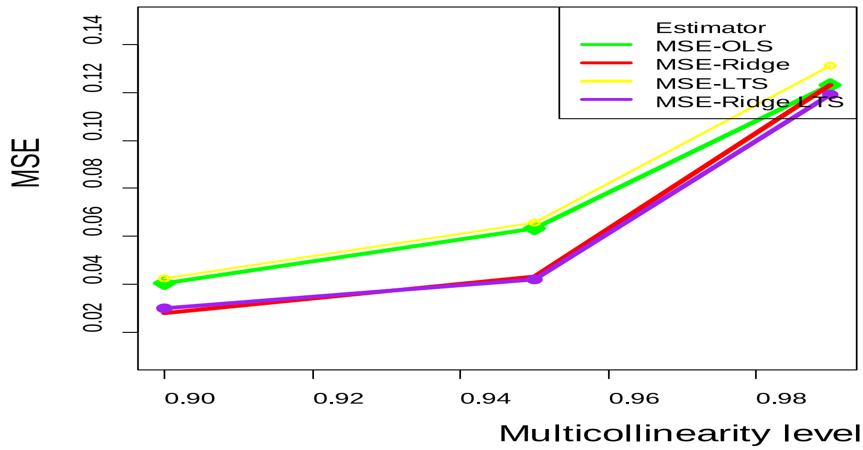
Variable	Sample Size (n)	Variance Inflation Factor (VIF)		
		r = 0.90	r= 0.95	r= 0.99
×1	25	0.1623	0.0880	0.0707
	50	0.1464	0.1103	0.0960
	100	0.1655	0.0781	0.0804
x2	25	0.1384	0.0793	0.0661
	50	0.1053	0.0853	0.0737
	100	0.1654	0.0821	0.0693
x3	25	0.1507	0.0805	0.0641
	50	0.1264	0.0993	0.1067
	100	0.2492	0.1029	0.0664

The VIF is reexamined after the ridge regression is performed to the data to see if the multicollinearity problem has been overcome. The results reveal that when ridge regression is used, the VIF values drop dramatically approaching one. It shows that ridge regression is quite good at dealing with multicollinearity.

Table 4.3 Estimated MSE For OLS, Ridge, LTS, Ridge.LTS Estimators for Different Sample Sizes and levels of Multicollinearity Without Outliers

Sample Size (n)	Estimator	Mean Square Error (MSE)		
		r=0.90	r=0.95	r=0.99
25	OLS	0.0405	0.0632	0.1233
	Ridge	0.0280	0.0430	0.1230
	LTS	0.0423	0.0658	0.1310
	Ridge.LTS	0.0300	0.0420	0.1190
50	OLS	0.0144	0.0170	0.0244
	Ridge	0.0070	0.0090	0.0190
	LTS	0.0125	0.0148	0.0213
	Ridge.LTS	0.0060	0.0080	0.0160
100	OLS	0.0073	0.0122	0.0160
	Ridge	0.0070	0.0080	0.0160
	LTS	0.0079	0.0157	0.0172
	Ridge.LTS	0.0070	0.0070	0.0150

When data is simulated using sample sizes of 25, 50, and 100, Table 4.3 shows the relative performance of the estimators in the presence of three different levels of multicollinearity.



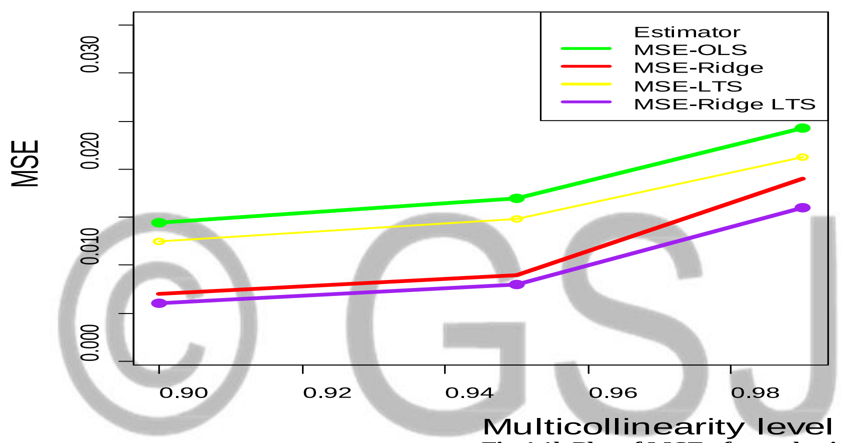
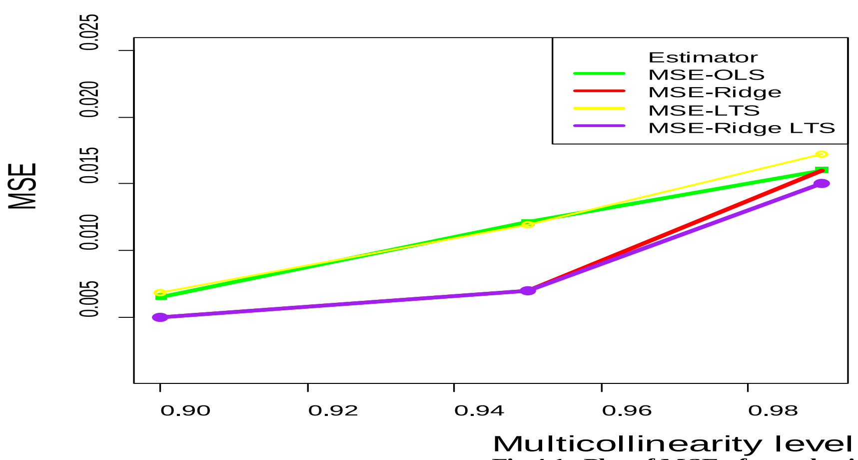


Fig.4.1b Plot of MSE of sample size n=



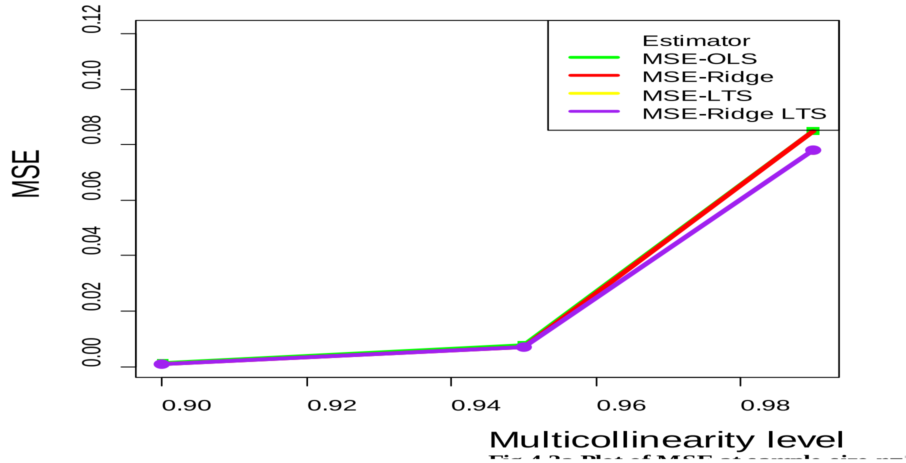
The MSE of the Ridge and Ridge LTS are shown in Table 4.1. When the errors are normally distributed and multicollinearity is present at a correlation value of r = 0.90, Ridge LTS is smaller than the other estimators. The result in Table 4.2 favors Ridge and Ridge LTS at r = 0.95. LTS is used for normal error distributions when collinearity is present in the data. Ridge is better than OLS, LTS, and its performance is virtually as excellent as Ridge, according to the MSE in Table 4.2, which corroborated the conclusion from Table 4.3. At r = 0.99, LTS has a high collinearity level. Ridge LTS is superior in every other way. The simulation results for bigger samples, n=100, are, nonetheless, consistent with the results for smaller samples. Because the MSE values are smaller, the results suggest that estimates for bigger samples are more efficient than those for smaller samples. For further clarity, the MSE values were recorded in tables 4.1, 4.2, and 4.3, and the data were presented in Figs. 4.1a, 4.1b, and 4.1c, respectively. Ridge LTS estimators were found to be the best estimators since they have the lowest values of the criterion considered in the assessment. Furthermore, as the level of multicollinearity was raised, the estimators' performance deteriorated.

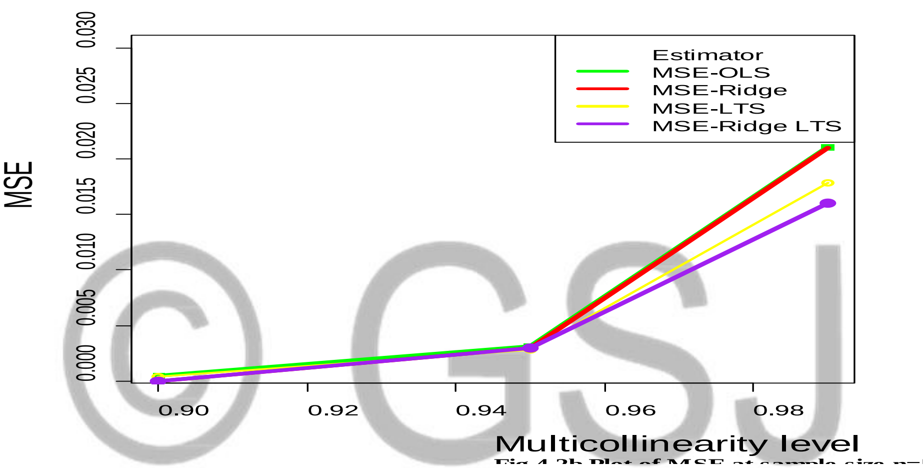
4.4 Effect of Multicollinearity and Outliers on the Estimators.

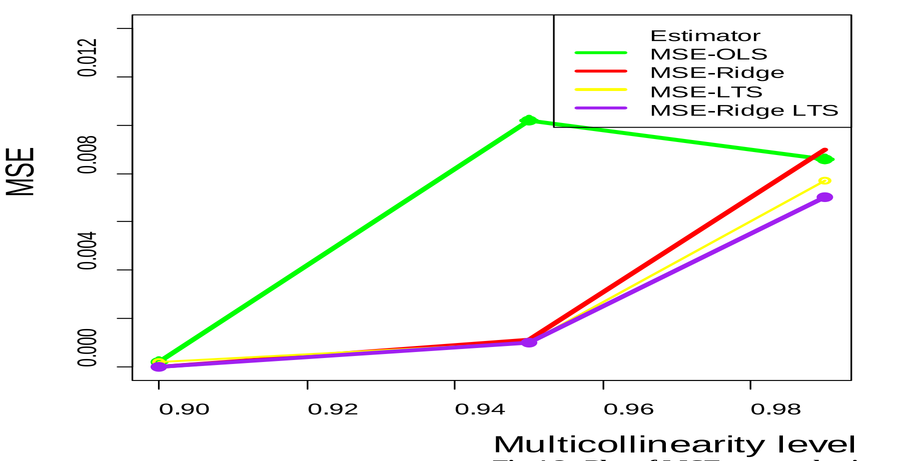
When data are simulated using sample sizes of 25, 50, and 100 with 20% outliers, Table 4.4 shows the relative performance of the estimators in the presence of three distinct levels of multicollinearity.

Table 4.4 Estimated MSE For OLS, Ridge, LTS, Ridge.LTS Estimators for Different Sample Sizes, levels of Multicollinearity and 20% Outliers

		Mean Square Error (MSE)		
Sample Size (n)	Estimator	r=0.90	r=0.95	r=0.99
25	OLS	0.0012	0.0078	0.0849
	Ridge	0.0010	0.0070	0.0850
	LTS	0.0011	0.0071	0.0779
	Ridge.LTS	0.0010	0.0070	0.0780
50	OLS	0.0004	0.0031	0.0211
	Ridge	0.0000	0.0030	0.0210
	LTS	0.0004	0.0028	0.0179
	Ridge.LTS	0.0000	0.0030	0.0170
100	OLS	0.0001	0.0011	0.0086
	Ridge	0.0000	0.0010	0.0090
	LTS	0.0002	0.0010	0.0077
	Ridge.LTS	0.0000	0.0010	0.0070







For data with a non-normal error distribution and multicollinearity, and for each sample size and number of outliers, Tables 4.4, 4.5, and 4.6 indicate that Ridge LTS produces the least MSE value, followed by LTS and OLS. In any number of sample sizes, Ridge LTS handles multicollinearity and the number of outliers substantially better than Ridge, OLS, and LTS. For further clarity, the MSE values were recorded in table

4.4, 4.5, and 4.6, and the data were presented in Fig. 4.2a, 4.2b, and 4.2c, respectively. Ridge LTS estimators were found to be the best estimators since they have the lowest values of the criterion considered in the assessment.

Furthermore, as the level of multicollinearity was raised, the estimators' performance deteriorated. The results of simulations for bigger samples are like those of smaller samples. The results also show that the estimator for bigger samples is more efficient than for smaller samples, as evidenced by the lower MSE values. As a result, when the errors are uniformly distributed in the presence of multicollinearity and outliers, the MSE of the Ridge LTS is smaller than the other estimators. When multicollinearity and outliers are present, Ridge LTS is more efficient than LTS and Ridge, and certainly much more efficient than OLS.

5.0 Conclusion

The MSE derived via Ridge LTS is the lowest, as can be shown. When both multicollinearity and outliers are present, simulation tests clearly reveal that the Ridge LTS estimate is the most practical option over other estimators. The MSE value for a large sample size is much lower than for a small sample size, implying that a larger sample size produces better and more reliable results. As the sample size grows, the case results of the estimation methods become more stable. It can be concluded that Ridge LTS is a better method for handling multicollinearity and outliers than OLS, Ridge, and LTS for small and large sample sizes.

References

Adegoke, A. S., Adewuyi, E., Ayinde, K. and Lukman, A. F. (2016). A Comparative Study of Some Robust Ridge and Liu Estimators. *Science World Journal.* **11** (4). 16-20 *www.scienceworldjournal.org*

Fitrianto, A. and Yik, L.C. (2014). Performance of ridge regression estimator methods on small sample size by varying correlation coefficients: a simulation study. *Journal of Mathematics and Statistics*. 10 (1): 25-29, doi:10.3844/jmssp.2014.25.29

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

Hoerl, A. E, Kennard, R. W., and Baldwin, K. F. (1975). Ridge regression: Some simulations Communications in Statistical Theory, **4** (2), 105-123.

Liu, K. (1993). A new class of biased estimate in linear regression. *Communication in Statistics.* **22** (2): 393-402.

Montgomery, D.C and Peck, E.A Introduction to Linear Regression Analysis, New York; Wiley and Sons,Inc.,(1982).

Roozbeh, A. M., M, Hamzah, N. A. and Gasparini, M (2021) Ridge regression and its applications in genetic studies.PLoS ONE 16(4):e0245376. doi.org/10.1371/journal.pone.0245376

Rousseeuw, P.J. and K. Van Driessen, (1998). "Computing LTS regression for large data sets",Technical Report, University of Antwer.

Yahya, W. B. and Olaif, J. B. (2014). A note on ridge regression modeling techniques. *Electronic Journal of Applied Statistical Analysis,* **07** (02), 343-361 DOI: 10.1285/i20705948v7n2p343

Zhang, J. and Mahmud Ibrahim (2021). A simulation study on SPSS ridge regression and ordinary least squares regression procedures for multicollinearity data. 571-588 doi.org//10.1080/02664760500078946.