African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 9, Issue 1, 2026 (pp. 108-123)
www.abjournals.org

ENHANCING PREDICTIVE ACCURACY IN SEEMINGLY UNRELATED REGRESSION MODELS: A COMPARATIVE STUDY OF SHRINKAGE AND ENSEMBLE ESTIMATORS

Afeez Abolaji Lawal and Oluwayemisi Oyeronke Alaba

Department of Statistics, Faculty of Sciences, University of Ibadan.

Corresponding Author's Email: heyyzeed@gmail.com

Cite this article:

A. A., Lawal, O. O., Alaba (2026), Enhancing Predictive Accuracy in Seemingly Unrelated Regression Models: A Comparative Study of Shrinkage and Ensemble Estimators. African Journal of Mathematics and Statistics Studies 9(1), 108-123. DOI: 10.52589/AJMSS-VI2TUSHU

Manuscript History

Received: 15 Feb 2026

Accepted: 17 Mar 2026

Published: 8 Apr 2026

ABSTRACT: *In econometric models, a system of equations is designed to increase the efficiency of estimation. However, with an increase in data, even in large datasets, multicollinearity compromises the estimation efficiency of the models, particularly Seemingly Unrelated Regression (SUR) models. Shrinkage estimators are effective in overcoming the problem of multicollinearity but pose challenges of overshrinkage bias in large datasets, whereas ensemble learning improves prediction by aggregating multiple models. However, there is a dearth of study in the combined application of shrinkage estimators and ensemble learning methods. The objective of the study is to assess the performance of shrinkage and ensemble shrinkage-based estimators in the sense of alleviating the multicollinearity problem in SUR models under different sample size scenarios. The Monte Carlo method simulates the covariance matrix with collinearity level of 0.5, 0.6, 0.7, 0.8 and 0.9 among explanatory variables, the error variance-covariance matrix with non-diagonal elements of 0.8. The estimators used in the study are Ridge, Lasso, Elastic Net, Boosting Ridge, Boosting Lasso, Boosting Elastic Net, Stacking Ridge, Stacking Lasso, and Stacking Elastic Net. Regularization parameters of 0.01 and 0.8, along with sample sizes of 30, 100, 500, 1000, 5000, 10,000, 50,000, 100,000 and 200,000 were considered. Based on the results obtained in relation to the findings, it is clear that an improvement in SUR coefficients can be achieved by using ensemble shrinkage-based estimators instead of traditional shrinkage when dealing with a large sample size under high levels of multicollinearity. Of all the proposed estimators, Stacking Elastic Net performed better under high multicollinearity in large datasets.*

KEYWORDS: Ensemble Learning Methods, Large Datasets, Multicollinearity, System of Equations.

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 9, Issue 1, 2026 (pp. 108-123)

www.abjournals.org

INTRODUCTION

Regression analysis is frequently used in statistics, econometrics and applied sciences but its efficiency can be compromised by multicollinearity. High correlation among predictors inflate variance, which result to inefficient parameter estimate (Hastie, Tibshirani and Friedman, 2009). Shrinkage estimators such as ridge regression, Lasso and Elastic Net addresses the challenges posed by multicollinearity (Zou and Hastie, 2005). Studies have shown that high level of multicollinearity is associated with lower power rates substantially increase type II error rates under certain conditions (Alabi et al., 2008).

In large datasets, regression analysis presents huge challenges majorly in precision and interpretation of results due to the presence of multicollinearity that causes standard errors of regression coefficient to increase resulting in inefficiency for the test of hypothesis (Herawati et al., 2018). Seemingly Unrelated Regression (SUR) model developed to simultaneously estimate multiple related equations which are applied extensively in economic modeling for their ability to model correlations between multiple equations (Zellner, 1962). The SUR models are often plagued by multicollinearity issues that can affect the efficiency of parameter estimation.

According to Chan et al., (2022), shrinkage estimators maintain efficiency of regression coefficient by reducing associated with high standard errors. Conventional statistical analysis has been the backbone of econometric analysis, however, recently the emergence of machine learning has brought in methods that are efficient when dealing with complex datasets. In handling the shortcomings of individual models, ensemble learning methods have shown to be an efficient model in statistical modelling to give a more reliable result (Jadama et al., 2024).

This paper leverage on the robustness of ensemble learning method by combining shrinkage estimators with ensemble learning methods. Despite these advances, evidence is limited on the comparative performance of traditional shrinkage estimators and ensemble-based shrinkage estimators under varying sample size and multicollinearity. This study investigates their predictive performance through simulation, providing guide on when ensemble shrinkage methods outperform traditional methods.

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 9, Issue 1, 2026 (pp. 108-123)
www.abjournals.org

METHODOLOGY

Model Specification

Considering a system of m linear regression equations with several response variables each containing different number of observations. Each of the equations in this system is assumed to satisfy the classical assumption of regression model. The system can be represented by;

$$y_i = X_i\beta_i + \varepsilon_i \qquad i = 1,\ 2,\ \cdots,\ m \tag{2.1}$$

The system in matrix form is;

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_m & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix} \tag{2.2}$$

Stacking equation (2.2) becomes:

$$y = X\beta + \varepsilon, \qquad \varepsilon \sim (0, \Sigma \otimes I_n) \tag{2.3}$$

where, y is an $(mn \times 1)$ stacked response vector, X is an $\left(mn \times \sum_{i=1}^{m} p_i\right)$ block diagonal design matrix of explanatory variables, β is an $\left(\sum_{i=1}^{m} p_i \times 1\right)$ vector of unknown parameter and ε is an $(mn \times 1)$ vector of disturbance terms. The disturbance term satisfies;

$E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \Omega$

where;

$$\Omega = \begin{bmatrix} \sigma_{11}I_n & \sigma_{12}I_n & \cdots & \sigma_{1m}I_n \\ \sigma_{21}I_n & \sigma_{22}I_n & \cdots & \sigma_{2m}I_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}I_n & \sigma_{m2}I_n & \cdots & \sigma_{mn}I_n \end{bmatrix} = \Sigma \otimes I_n \tag{2.4}$$

Multicollinearity is SUR

Given the diagonal matrix of response variables $X = diag(X_1, X_2, \cdots, X_m)$, multicollinearity may arise from linear dependence among the response variables in the system. Suppose that for equation (2.3) two regressors X_{ij} and X_{ik} are highly correlated such that;

$$corr(X_{ij}, X_{ik}) \simeq \pm 1 \tag{2.5}$$

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 9, Issue 1, 2026 (pp. 108-123)

www.abjournals.org

This indicates the presence of severe multicollinearity within i^{th} equation. Given that multicollinearity is present within the component matrices X_i, then this cause instability in $X'\Omega^{-1}X$, thereby affecting the efficiency of SUR estimator.

Shrinkage Estimators within the SUR Framework

Ridge Estimator

Given equation (2.3), the ridge regression estimator is obtained as:

$$\hat{\beta}_R = \left(X'\Omega^{-1}X + \lambda I\right)^{-1} X'\Omega^{-1}y \tag{2.6}$$

Where $\lambda \succ 0$ is the biasing (regularization) parameter

Lasso Estimator

The lasso estimator is obtained as;

$$\hat{\beta}_L = \arg\min_{\beta}\left\{(y - X\beta)'\Omega^{-1}(y - X\beta) + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{2.7}$$

Elastic Net Estimator

The elastic net estimator combines ridge and Lasso penalties which is obtained as;

$$\hat{\beta}_{EN} = \arg\min_{\beta}\left\{\left(y - X\beta\right)'\Omega^{-1}\left(y - X\beta\right) + \lambda\left[\alpha\sum_{j=1}^{p}|\beta_j| + \left(1-\alpha\right)\sum_{j=1}^{p}\beta_j^{\,2}\right]\right\} \tag{2.8}$$

where $0 \le \alpha \le 1$

Ensemble Learning Method

This method is considered essential in regression based on their capability to enhance to enhance predictive accuracy and efficiency by merging multiple models. This study considered the boosting and stacking ensemble methods integrated with shrinkage estimators

Boosting Based Shrinkage SUR

Boosting within SUR framework is implemented by iteratively fitting weak learners based on shrinkage estimators (Ridge, Lasso and Elastic Net), while explicitly accounting for cross equation error correlation through covariance matrix Ω. Given equation (2.3), boosting proceeds by sequentially fitting base learners to the residuals using a Generalized Least Squares (GLS) loss function.

At m iteration, the residual is expressed as:

$$r^{(m)} = y - f^{(m-1)}\left(X\right) \tag{2.9}$$

The base learner $h_m\left(X\right)$ is obtained by minimize the SUR weighted loss given as;

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 9, Issue 1, 2026 (pp. 108-123)

www.abjournals.org

$$h_m = \arg\min_h \left\{ \left(r^{(m)} \right)' \Omega^{-1} \left(r^{(m)} \right) + penalty(h) \right\} \tag{2.10}$$

Then the model is updated as:

$$f^{(m)}(X) = f^{(m-1)}(X) + v h_m(X) \tag{2.11}$$

Where v is the learning rate $\{v \in (0,1)\}$ and $h_m(X)$ is the base learner at step m.

a) Boosting Ridge SUR

When Ridge regularization is incorporated into boosting, the loss function under SUR is modified to include $L2$ penalty given as:

$$l_{BR} = (y - X\beta)' \Omega^{-1} (y - X\beta) + \lambda \sum_{j=1}^{p} \beta_j^{\,2} \tag{2.12}$$

at m boosting ridge iteration, the base learner is expressed as:

$$h_m^{(Ridge)} = \arg\min_{\beta} \left\{ \left(r^{(m)} \right)' \Omega^{-1} \left(r^{(m)} \right) + \lambda \sum_{j=1}^{p} \beta_j^{\,2} \right\} \tag{2.13}$$

b) Boosting Lasso SUR

The boosting lasso regularization is incorporated into boosting combined with $L1$ penalties given as:

$$l_{BL} = (y - X\beta)' \Omega^{-1} (y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j| \tag{2.14}$$

at each iteration, the base learner is given as:

$$h_m^{(Lasso)} = \arg\min_{\beta} \left\{ \left(r^{(m)} \right)' \Omega^{-1} \left(r^{(m)} \right) + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{2.15}$$

c) Boosting Elastic Net SUR

The Elastic Net penalties is incorporated within the SUR boosting framework given the resulting loss function as:

$$l_{BEN} = (y - X\beta)' \Omega^{-1} (y - X\beta) + \lambda \left[\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^{\,2} \right] \tag{2.16}$$

at each iteration, the base learner is given as:

$$h_m^{(EN)} = \arg\min_{\beta} \left\{ \left(r^{(m)} \right)' \Omega^{-1} \left(r^{(m)} \right) + \lambda \left[\alpha \sum_{j=1}^{p} |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \beta_j^{\,2} \right] \right\} \tag{2.17}$$

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 9, Issue 1, 2026 (pp. 108-123)

www.abjournals.org

Stacking Based Shrinkage Seemingly Unrelated Regression (SUR)

Stacking within the SUR framework combines multiple base learners by constructing a meta estimator that optionally weights their predictions, while accounting for cross equation error correlation. Let the base learners be $f^{(1)}, f^{(2)}, \cdots, f^{(k)}$ each base learner produces $\hat{y}(k)$ the final prediction is given as;

$$\hat{y} = g\left(\hat{y}^{(1)}, \hat{y}^{(2)}, \cdots, \hat{y}^{(k)}\right) \tag{2.18}$$

The final stacked prediction is given as:

$$\hat{y} = \sum_{k=1}^{K} w_k \hat{y}^{(k)} \tag{2.19}$$

where w_k are the stacking weights.

Minimizing the GLS loss function, the stacking weights are estimated as:

$$w = \arg\min_{w} \left\{ \left(y - \sum_{k=1}^{K} w_k \hat{y}^{(k)} \right)' \Omega^{-1} \left(y - \sum_{k=1}^{K} w_k \hat{y}^{(k)} \right) + penalty(w) \right\} \tag{2.20}$$

a) Stacking Ridge SUR

The stacking weight of ridge is expressed as:

$$w = \arg\min_{w} \left\{ \left(y - \sum_{k=1}^{K} w_k \hat{y}_R{}^{(k)} \right)' \Omega^{-1} \left(y - \sum_{k=1}^{K} w_k \hat{y}_R{}^{(k)} \right) + \lambda \sum_{k=1}^{K} w_k^2 \right\} \tag{2.21}$$

b) Stacking Lasso SUR

The stacking weight of Lasso is expressed as:

$$w = \arg\min_{w} \left\{ \left(y - \sum_{k=1}^{K} w_k \hat{y}_L{}^{(k)} \right)' \Omega^{-1} \left(y - \sum_{k=1}^{K} w_k \hat{y}_L{}^{(k)} \right) + \lambda \sum_{k=1}^{K} |w| \right\} \tag{2.22}$$

d) Stacking Elastic Net SUR

The stacking weight of the Elastic Net is expressed as:

$$w = \arg\min_{w} \left\{ \left(y - \sum_{k=1}^{K} w_k \hat{y}_{EN}{}^{(k)} \right)' \Omega^{-1} \left(y - \sum_{k=1}^{K} w_k \hat{y}_{EN}{}^{(k)} \right) + \lambda \left[\alpha \sum_{k=1}^{K} |w_k| + (1-\alpha) \sum_{k=1}^{K} w_k^2 \right] \right\} \tag{2.23}$$

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 9, Issue 1, 2026 (pp. 108-123)

www.abjournals.org

Choice of Biasing Parameter

Small regularization parameters have been adopted by different researchers in shrinkage regression to improve numerical efficiency under multicollinearity without inducing substantial bias (Zou and Hastie, 2005). The regularization parameter adopted was fixed at $\lambda = 0.01$ and 0.8 for the simulation scenarios.

Simulation Design

The simulation study is detailed as follows;

- We generated the i-th explanatory matrix x_i from $MVN_{p_i}(0,\Sigma_x)$ where diag $(\Sigma_x = 1)$ and off diag $(\Sigma_x = \rho_x)$.
- The ρ_x regulate the strength of collinearity among explanatory variables per equation. This study considered $\rho_x = 0.5, 0.6, 0.7, 0.8,$ and 0.9
- The variance covariance matrix of errors is defined by diag $\Sigma_\varepsilon = 1$ and off diag $\Sigma_\varepsilon = 0.8$ and generated from $MVN_m(0,\Sigma_\varepsilon)$.
- Data for a two equation Seemingly Unrelated Regression (SUR) model were generated.
- Sample sizes of 30, 100, 500, 1000, 5,000, 10,000, 50,000, 100,000 and 200,000 observations were considered
- Data set was split into training 70% and testing 30% sets
- The performance of the estimators was evaluated using the Root Mean Squared Error (RMSE) criterion.

$$RMSE = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\beta}^{(r)} - \beta\right)^2} \qquad (2.24)$$

the lower the RMSE value implies better estimation accuracy.

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 9, Issue 1, 2026 (pp. 108-123)
www.abjournals.org

RESULTS AND INTERPRETATION

Table 4.1 presents the Root Mean Squared Errors (RMSE) of traditional shrinkage estimators (Ridge, Lasso, Elastic Net), as well as boosting and stacking shrinkage-based estimators, under varying levels of multicollinearity (ρ_x = 0.5, 0.6, 0.7, 0.8, 0.9), sample sizes (n = 30, 100 and 500) and regularization parameter λ = 0.01.

For small sample size (n = 30), traditional Lasso and Elastic Net consistently exhibit the lowest RMSE, under severe multicollinearity scenario (ρ_x = 0.9). This is as a result of Lasso type penalties imposing strong shrinkage that effectively reduce variance at the cost of introducing bias. However, boosting shrinkage-based estimators shows higher RMSE values and this is due to the iterative nature of boosting which tends to accumulate estimation noise when the sample size is small leading to increasing variance rather than reducing it. Stacking shrinkage-based estimators on the other hand perform slightly better than boosting estimators but fails to outperform traditional shrinkage estimators as the meta learning step introduces additional estimation uncertainty in small samples.

At moderate sample size (n = 100), Lasso estimator continues to perform better under high multicollinearity (ρ_x = 0.8, 0.9) showing its effectiveness in handling sparsity and correlated predictors. Under moderate multicollinearity (ρ_x = 0.5 - 0.7), stacking Elastic Net estimator outperform traditional shrinkage estimators. This implies that as the sample size increases, the variance of the stacking weights decreases allowing ensemble methods to effectively approximate the optimal combination of base learners.

At large sample (n = 500), stacking shrinkage-based estimator (stacking Lasso) exhibits the lowest RMSE under high multicollinearity (ρ_x = 0.8, 0.9). The results show that shrinkage methods dominate in small samples due to variance reduction, however, ensemble shrinkage methods become asymptotically efficient as sample size increases and estimation uncertainty diminishes.

Table 1: RMSE with Varying Degree of Multicollinearity λ = 0.01 when sample sizes n = 30, 100, and 500

n	ρ_x	Ridge	Lasso	Elastic Net	B_Ridge	B_Lasso	B_Elastic Net	S_Ridge	S_Lasso	S_Elastic Net
30	0.9	0.84119	0.76476	0.79622	1.15845	1.10238	1.12499	1.00871	0.78598	0.82167
	0.8	1.60946	1.51859	1.54810	1.36704	1.32911	1.34237	1.60308	1.49506	1.54782
	0.7	1.32315	1.29568	1.31165	1.71716	1.69738	1.70585	1.58692	1.68147	1.57468
	0.6	1.03279	0.98695	1.00904	1.08459	1.07704	1.08065	1.00093	0.89551	1.02235
	0.5	1.42433	1.35468	1.38975	1.39778	1.37917	1.38910	1.39271	1.22724	1.33634
100	0.9	0.86843	0.85531	0.86066	1.20368	1.20652	1.20513	0.88943	0.86739	0.88777
	0.8	0.92915	0.92373	0.92585	1.23604	1.24611	1.24094	0.93647	0.92908	0.92786
	0.7	1.11824	1.10044	1.10894	1.48970	1.48670	1.48813	1.11744	1.11102	1.08601
	0.6	0.88758	0.88434	0.88510	1.10517	1.10824	1.10651	0.89268	0.89634	0.89143
	0.5	1.13545	1.12498	1.13010	1.32510	1.32589	1.32546	1.10096	1.10367	1.08688
500	0.9	1.08022	1.07069	1.07447	1.40845	1.40775	1.40750	1.08153	1.06901	1.07391
	0.8	1.04406	1.04151	1.04286	1.33655	1.33878	1.33771	1.04395	1.04294	1.04217
	0.7	0.96516	0.96415	0.96449	1.31429	1.31595	1.31506	0.96604	0.96408	0.96523
	0.6	1.04612	1.04127	1.04337	1.25921	1.25885	1.25889	1.04728	1.04031	1.04272

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

	0.5	1.01121	1.00928	1.00995	1.28828	1.29036	1.28926	1.01018	1.00971	1.01272

Table 2 presents the RMSE of Ridge, Lasso, Elastic Net, as well as boosting and stacking shrinkage-based estimators across different levels of multicollinearity (ρ_x = 0.5, 0.6, 0.7, 0.8, and 0.9) and large sample sizes (n = 1000, 5000, and 10,000) with λ = 0.01.

At n = 1000 and high multicollinearity (ρ_x = 0.9), Lasso and stacking Lasso exhibit lowest RMSE showing the ability of L1 based regularization to effectively control variance inflation induced. When the ρ_x = 0.8, stacking shrinkage-based estimators outperformed traditional shrinkage estimators, but boosting shrinkage based continue to exhibit substantially higher RMSE, indicating that the sequential fitting process introduces additional variance that is not sufficiently offset by shrinkage.

As the sample size increases to n = 5000 under high multicollinearity level (ρ_x = 0.9), stacking Lasso exhibit the lowest RMSE value. At ρ_x = 0.8, stacking Lasso and stacking Elastic Net have a better performance. This suggest that as n increases, the variance reduction advantage of stacking becomes more pronounced in the presence of severe multicollinearity. When the sample size n = 10,000 under high multicollinearity (ρ_x = 0.9), stacking shrinkage-based estimators (stacking Lasso) exhibit lowest RMSE. Also, ridge estimator performs better when the sample size is very large. This indicates that as n tend to infinity, the variance of the estimator diminishes and the bias introduced by ridge regularization becomes less consequential. However, at ρ_x = 0.8, the performance of stacking shrinkage-based estimator and traditional shrinkage estimators (Lasso and Elastic Net) seems to be similar.

The results indicates that traditional shrinkage estimators effectively control variance in finite samples, while stacking shrinkage-based estimators exhibit a better performance by optimally combining multiple regularized estimators.

Table 2: RMSE with Varying Degree of Multicollinearity λ = 0.01 when sample sizes n = 1000, 5000, and 10000

n	ρ_x	Ridge	Lasso	Elastic Net	B_Ridge	B_Lasso	B_Elastic Net	S_Ridge	S_Lasso	S_Elastic Net
1000	0.9	0.99747	0.98712	0.99136	1.25556	1.25412	1.25475	0.99776	0.98708	0.99163
	0.8	0.98631	0.98694	0.98655	1.28193	1.28541	1.28378	0.98675	0.98655	0.98674
	0.7	0.95474	0.95323	0.95380	1.23532	1.23720	1.23622	0.95584	0.95292	0.95347
	0.6	1.06506	1.06343	1.06413	1.28701	1.28836	1.28766	1.06552	1.06321	1.06390
	0.5	1.04317	1.04132	1.04190	1.31232	1.31416	1.31315	1.04357	1.04094	1.04203
5000	0.9	1.00479	1.00474	1.00486	1.23648	1.23869	1.23764	1.00481	1.00492	1.00501
	0.8	1.00112	1.00037	1.00065	1.25524	1.25744	1.25636	1.00092	0.99965	1.00012
	0.7	0.97613	0.97625	0.97641	1.19687	1.19951	1.19827	0.97619	0.97629	0.97643
	0.6	0.99931	0.99868	0.99887	1.19039	1.19260	1.19142	0.99938	0.99879	0.99894
	0.5	0.99539	0.99477	0.99497	1.18514	1.18750	1.18634	0.99546	0.99477	0.99492
10000	0.9	1.01268	1.01265	1.01262	1.25529	1.25750	1.25637	1.01268	1.01262	1.01261
	0.8	1.01308	1.01234	1.01251	1.22271	1.22485	1.22372	1.01314	1.01253	1.01264
	0.7	0.99809	0.99780	0.99779	1.20551	1.20782	1.20659	0.99810	0.99783	0.99780
	0.6	1.00213	1.00202	1.00205	1.18395	1.18625	1.18507	1.00222	1.00225	1.00221
	0.5	1.00344	1.00314	1.00318	1.19662	1.19895	1.19776	1.00337	1.00296	1.00308

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 9, Issue 1, 2026 (pp. 108-123)

www.abjournals.org

Table 3 presents the RMSE values of traditional shrinkage estimators (Ridge, Lasso, Elastic Net), boosting and stacking shrinkage-based estimators under varying level of multicollinearity (ρ_x = 0.5, 0.6, 0.7, 0.8 and 0.9) for very large sample sizes n = 50,000, 100,000, and 200,000.

At n = 50,000 under severe multicollinearity (ρ_x = 0.9), the Elastic Net exhibit the lowest RMSE, while stacking shrinkage-based estimators yield nearly identical but marginally higher values. This shows the advantage of Elastic Net which promote stability and sparsity making it effective when predictors are highly correlated. At ρ_x = 0.8, the stacking shrinkage-based estimators dominate the traditional estimator(s). For n = 100,000 under Severe Multicollinearity (ρ_x = 0.9), Elastic Net retained the minimum RMSE, showing that traditional shrinkage estimators remain highly competitive even when the dataset is extremely large. However, under moderate multicollinearity (ρ_x = 0.6 and ρ_x = 0.8), stacking Lasso consistently achieves the lowest RMSE. This suggests that, as sample size increases, the estimation error of stacking weights diminishes, allowing the meta learner to approximate an optimal combination of base estimators to achieve better variance reduction.

At n = 200,000, stacking shrinkage-based estimators frequently yield the lowest RMSE, particularly under moderate multicollinearity (ρ_x = 0.5 - 0.8). This confirms the efficiency of stacking as both base learners and combination weights are estimated with high precision. Under severe multicollinearity (ρ_x = 0.9), stacking estimators also outperform traditional shrinkage estimators, suggesting that ensemble methods offer efficiency gains beyond what single penalty estimator can achieve.

The results suggested that while traditional shrinkage estimator remain efficient under extreme multicollinearity, stacking shrinkage-based estimators become more asymptotically efficient as sample size increase due to their ability to approximate an optimal GLS weighted combination of biased estimators and further reduce estimation variance.

Table 3: RMSE with Varying Degree of Multicollinearity λ = 0.01 when sample sizes n = 50,000, 100,000, and 200,000

n	ρ_x	Ridge	Lasso	Elastic Net	B_Ridge	B_Lasso	B_Elastic Net	S_Ridge	S_Lasso	S_Elastic Net
50000	0.9	1.00829	1.00822	1.00819	1.24265	1.24487	1.24375	1.00830	1.00821	1.00820
	0.8	1.00034	1.00039	1.00035	1.22689	1.22921	1.22805	1.00033	1.00031	1.00031
	0.7	0.99320	0.99320	0.99316	1.19877	1.20112	1.19995	0.99320	0.99313	0.99313
	0.6	0.99610	0.99613	0.99610	1.19095	1.19336	1.19216	0.99608	0.99604	0.99606
	0.5	1.00011	1.00006	1.00007	1.17492	1.17721	1.17607	1.00013	1.00008	1.00010
100000	0.9	1.00162	1.00162	1.00158	1.24258	1.24479	1.24366	1.00160	1.00153	1.00153
	0.8	1.00169	1.00172	1.00168	1.22188	1.22413	1.22299	1.00169	1.00164	1.00165
	0.7	1.00444	1.00448	1.00442	1.21004	1.21235	1.21118	1.00441	1.00433	1.00434
	0.6	0.99464	0.99467	0.99464	1.18568	1.18806	1.18688	0.99464	0.99459	0.99460
	0.5	0.99836	0.99838	0.99835	1.17556	1.17795	1.17676	0.99837	0.99835	0.99836
200000	0.9	0.99821	0.99822	0.99818	1.24031	1.24255	1.24141	0.99820	0.99813	0.99813
	0.8	1.00226	1.00230	1.00226	1.22354	1.22579	1.22465	1.00226	1.00223	1.00223
	0.7	0.99921	0.99927	0.99922	1.20654	1.20884	1.20768	0.99919	0.99914	0.99915

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

	0.6	0.99513	0.99517	0.99511	1.18955	1.19191	1.19072	0.99510	0.99504	0.99504
	0.5	1.00004	1.00014	1.00006	1.18063	1.18303	1.18183	1.00001	0.99999	0.99999

Table 4 presents the RMSE performance of traditional shrinkage, boosting and stacking shrinkage-based estimators, under varying degrees of multicollinearity $(\rho_x = 0.5,0.6,0.7,0.8,0.9)$ for sample sizes n = 30, 100, and 500 with a relatively large regularization parameter λ = 0.8.

At small sample size (n = 30), estimator performance is highly sensitive to multicollinearity. Under this condition, stacking shrinkage-based estimators (stacking Elastic Net and stacking Lasso) consistently exhibit the lowest RMSE across most values of ρ_x. For severe multicollinearity (ρ_x = 0.9), stacking Elastic Net and stacking Lasso outperform traditional shrinkage estimators. This is attributed to the combined effect of strong shrinkage and variance reduction through ensembling which stabilizes estimation even when individual models are highly biased. Also, boosting shrinkage-based estimators exhibit inefficiency at moderate to high multicollinearity levels. This is attributed to the iterative nature of boosting which tends to propagate estimation noise in small samples and the high shrinkage parameter does not sufficiently offset this variance accumulation. Moreso, traditional Lasso performs relatively poor as the multicollinearity decreases, suggesting that under strong penalization, bias dominates when the variance inflation problem is less severe.

As the sample size increases to n = 100, RMSE values decrease across all estimators, confirming an improvement in the estimation efficiency as the variance decreases. Stacking Elastic Net performs better by recording the lowest RMSE values across all levels of multicollinearity, indicating that combination of L1 and L2 penalties within an ensemble framework provides an optimal balance between bias and variance under strong regularization.

This implies that when shrinkage parameter is large, variance reduction dominates the bias variance trade off and stacking based estimators become effective by combining multiple bias but stable estimators, thereby achieving lower RMSE within the SUR framework.

Table 4: RMSE with Varying Degree of Multicollinearity λ = 0.8 when sample sizes n = 30, 100, and 500

n	rho_x	Ridge	Lasso	Elastic Net	B_Ridge	B_Lasso	B_Elastic Net	S_Ridge	S_Lasso	S_Elastic Net
30	0.9	1.21347	1.30176	1.15865	1.21803	1.28825	1.22174	1.16989	1.06615	1.05493
	0.8	1.31169	1.23977	1.18425	1.97390	1.86912	1.90698	1.32414	1.31564	1.22629
	0.7	1.33954	1.94690	1.55286	1.69737	1.98940	1.83604	1.26819	1.52018	1.41465
	0.6	1.20505	1.67882	1.20872	1.22935	1.42335	1.25079	1.20902	1.39012	1.05213
	0.5	1.57178	2.38976	1.93607	2.21409	2.56998	2.37833	1.42734	3.41757	1.39030
100	0.9	1.09846	1.35124	1.17221	1.20681	1.28558	1.22821	1.07759	1.05348	1.04295
	0.8	0.98910	1.19927	1.01012	1.12596	1.27379	1.18223	0.98707	0.92999	0.89108
	0.7	1.47895	1.67501	1.51587	1.61911	1.71288	1.64320	1.46147	1.48264	1.41523
	0.6	1.10156	1.40404	1.17940	1.43008	1.53793	1.45531	1.11600	1.17149	1.10150
	0.5	0.99535	1.35318	1.12043	1.32366	1.45030	1.37119	0.98919	1.05987	1.00021

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

500	0.9	1.11097	1.29694	1.15792	1.16285	1.20117	1.16866	1.10040	1.06058	1.06492
	0.8	1.04686	1.39368	1.17346	1.13591	1.25355	1.18075	1.00875	1.01220	0.99698
	0.7	1.08020	1.40320	1.18814	1.21354	1.30478	1.24468	1.05599	1.04256	1.04057
	0.6	1.01836	1.40974	1.15633	1.16864	1.30088	1.21497	0.98219	1.02485	0.97451
	0.5	1.02529	1.45568	1.17033	1.14377	1.29600	1.19625	0.98576	1.05100	0.98152

Table 5 presents the RMSE performance of traditional shrinkage estimator, boosting and stacking shrinkage-based estimators under varying levels of multicollinearity (ρ_x = 0.5, 0.6, 0.7, 0.8 and 0.9) for large sample sizes n = 1000, 5000, and 10,000.

At n = 1000, stacking Elastic Net consistently exhibit the lowest RMSE across all levels of multicollinearity. At ρ_x = 0.9, the stacking Elastic Net shows the lowest RMSE, however traditional ridge estimator remains competitive due to its strong variance reduction properties. As the sample size increases, the variance component of the estimators reduces and the performance differences are increasingly driven by how effectively each methods handles multicollinearity. The result shows that stacking Elastic Net is more efficient as it combines the strengths of multiple shrinkage methods while maintaining stability through regularization. The dominance of stacking Elastic Net reflects it ability to approximate an optimal GLS-weighted combination of estimators within the SUR context, thereby achieving variance without excessively increasing bias, especially when the regressor matrix is highly correlated.

Table 5: RMSE with Varying Degree of Multicollinearity λ = 0.8 when sample sizes n = 1000, 5000, and 10,000

n	rho_x	Ridge	Lasso	Elastic Net	B_Ridge	B_Lasso	B_Elastic Net	S_Ridge	S_Lasso	S_Elastic Net
1000	0.9	0.99262	1.28106	1.08243	1.08794	1.16626	1.11145	0.97828	0.96260	0.95475
	0.8	1.03065	1.38299	1.15383	1.13754	1.25029	1.17780	1.00049	1.00166	0.98226
	0.7	1.05650	1.50719	1.22109	1.20450	1.36506	1.26533	1.00770	1.04110	0.99840
	0.6	1.03871	1.43899	1.18139	1.16694	1.28371	1.20636	0.99216	1.02471	0.98541
	0.5	1.09214	1.51967	1.25021	1.19630	1.34535	1.25346	1.03773	1.06965	1.03381
5000	0.9	1.02677	1.30331	1.11104	1.08659	1.15797	1.10652	1.01685	1.01198	0.99738
	0.8	1.06385	1.40563	1.18624	1.14651	1.24321	1.18070	1.03397	1.05128	1.02577
	0.7	1.05294	1.40396	1.17196	1.14310	1.23677	1.17258	1.02630	1.04995	1.01815
	0.6	1.04338	1.43994	1.18444	1.14056	1.24792	1.17645	1.00441	1.01317	0.99298
	0.5	1.06658	1.52613	1.23594	1.15487	1.29067	1.20245	1.01453	1.05602	1.01618
10000	0.9	1.04116	1.32153	1.13111	1.09728	1.17328	1.12065	1.02507	1.02065	1.00586
	0.8	1.02810	1.34123	1.13104	1.10772	1.18637	1.13094	1.01120	1.02083	0.99913
	0.7	1.04525	1.42915	1.18451	1.11503	1.22555	1.15355	1.00718	1.02618	1.00002
	0.6	1.03577	1.45882	1.19053	1.11292	1.23463	1.15570	0.99492	1.03136	0.99526
	0.5	1.04269	1.49136	1.20682	1.11176	1.23616	1.15430	0.99782	1.03420	1.00033

Table 6 presents the RMSE performance of traditional shrinkage estimators, boosting and stacking shrinkage-based estimators under varying degrees of multicollinearity (ρ_x = 0.5, 0.6, 0.7, 0.8 and 0.9) for large sample sizes (n = 50,000, 100,000, and 200,000). Stacking

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

shrinkage estimators consistently out-perform the traditional shrinkage estimator and boosting shrinkage-based estimators, with stacking Elastic Ner achieving the lowest RMSE.

Stacking Elastic Net exhibit, the lowest RMSE consistently outperform traditional shrinkage estimators and boosting shrinkage-based estimators. Traditional ridge and Elastic Net remain stable but are dominated by stacking shrinkage-based estimators. But boosting shrinkage-based estimators show a marginal improvement over traditional shrinkage estimators and consistently underperform to stacking estimators. This indicates that in large samples, the nature of boosting does not sufficiently reduce variance beyond what is achieved through direct regularization, most especially when compare to more efficient aggregation mechanism of stacking.

The result shows that in very large samples, estimator performance is driven less by variance reduction and by the ability to optimally combine biased estimators.

Table 6: RMSE with Varying Degree of Multicollinearity λ = 0.8 when sample sizes n = 50,000, 100,000, and 200,000

n	rho_x	Ridge	Lasso	Elastic Net	B_Ridge	B_Lasso	B_Elastic Net	S_Ridge	S_Lasso	S_Elastic Net
50000	0.9	1.03540	1.31419	1.12394	1.05002	1.12580	1.07286	1.02174	1.01199	1.00233
	0.8	1.03637	1.36636	1.15041	1.05766	1.14800	1.08693	1.01288	1.02152	1.00155
	0.7	1.03661	1.40144	1.16622	1.05566	1.15791	1.08994	1.00681	1.03010	1.00236
	0.6	1.04557	1.45185	1.19122	1.06844	1.18221	1.10634	1.00894	1.04307	1.00726
	0.5	1.04640	1.49542	1.21093	1.06390	1.19313	1.10797	0.99795	1.03697	0.99941
100000	0.9	1.04102	1.32157	1.13025	1.03881	1.11546	1.06164	1.02567	1.01496	1.00471
	0.8	1.03643	1.35912	1.14661	1.03930	1.12831	1.06781	1.01485	1.02446	1.00368
	0.7	1.04978	1.40976	1.17631	1.05354	1.15412	1.08656	1.02076	1.04008	1.01444
	0.6	1.04543	1.45925	1.19633	1.04665	1.16605	1.08751	1.00487	1.04035	1.00458
	0.5	1.05444	1.50698	1.22053	1.05373	1.18458	1.09821	1.00656	1.04811	1.00863
200000	0.9	1.03698	1.31357	1.12466	1.02470	1.10120	1.04755	1.02185	1.00799	0.99976
	0.8	1.04285	1.36304	1.15188	1.03603	1.12551	1.06460	1.02048	1.02729	1.00746
	0.7	1.03930	1.40326	1.16723	1.03088	1.13388	1.06459	1.00929	1.03105	1.00246
	0.6	1.04225	1.44621	1.18741	1.03340	1.14836	1.07173	1.00550	1.03706	1.00356
	0.5	1.05417	1.50760	1.22110	1.04054	1.17361	1.08611	1.00432	1.04613	1.00666

DISCUSSION OF FINDINGS

The results findings show that the predictive efficiency of shrinkage estimators is critically dependent upon sample size and multicollinearity level, which is also confirmed by previously developed regularization and ensemble learning theory. As expected, in small sample sizes, traditional shrinkage like Lasso and Elastic Net have lower values of RMSE, particularly in cases of strong multicollinearity. Simulation and methodological studies consistently show that penalized regression remains reliable in low-sample, high-correlation scenario because shrinkage stabilizes coefficient estimation through an improved bias–variance trade-off (Zou & Hastie, 2005; Hastie et al., 2009). With increased sample size, stacking shrinkage-based estimators tend to gain an escalating advantage over traditional shrinkage estimators, especially in conditions of moderate to high multicollinearity. This advantage reflects the nature of stacking in optimally combining multiple estimators with

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

varied strengths in a way that minimizes variance, hence reducing prediction error. Both theoretical and empirical evidence reported that stacking and model-averaging methods outperform single regularized models when there is enough information in the sample about the models (Zhang & Yang, 2021).

Boosting shrinkage-based estimators consistently realized higher RMSE values across most scenarios, thereby raising concerns about their suitability for regression problems dominated by highly correlated predictors. Similar conclusions have been reported in several applied and simulation studies where boosting performs poorly to regularization and averaging-based ensemble methods in linear regression contexts (Bühlmann & Yu, 2003; Friedman, 2001). This performance is due to the behavior that lies in the sequential nature of boosting. When there is severe multicollinearity, the design matrix of the response variables becomes highly ill-conditioned which result to instability in the estimation of early-stage learners. These initial learners tend to overfit noise associated with highly correlated regressors, and since boosting updates the model recursively, this early overfitting is propagated and compounded across iterations, resulting in variance amplification rather than reduction and consequently higher RMSE values. This shows why boosting fails to achieve efficiency gains in the presence of multicollinearity.

In contrast, stacking shrinkage-based estimators perform best across all levels of multicollinearity in large and very large sample sizes. The strong performance of stacking Elastic Net and stacking Lasso under high correlation further supports the efficiency gains associated with averaging based estimators in high-dimensional datasets (Hansen, 2007; Hansen and Racine, 2012). Even if Elastic Net remains competitive under extreme multicollinearity due to its grouped-variable penalty structure, the consistent lower RMSE values obtained from stacking indicate efficiency gains from optimally combining multiple biased estimators within the SUR framework.

CONCLUSION

This study has shown that the performance of shrinkage estimators within Seemingly Unrelated Regression (SUR) model is highly sensitive to sample size and the level of multicollinearity involved. The result shows that traditional shrinkage estimators such as ridge, Lasso and Elastic Net are more efficient when the sample size is small but were consistently outperformed by stacking shrinkage-based estimators as the sample size tends to be large under moderate and severe multicollinearity. However, boosting shrinkage-based estimators exhibit persistent poor performance across the degrees of multicollinearity and this is due to variance amplification arising from their sequential learning structure.

Implication for Practice

This study provides guidance to researchers estimating SUR models under highly correlated regressors with large datasets should prioritize stacking shrinkage-based estimators, especially stacking Elastic Net over the traditional estimators. In small datasets, traditional shrinkage estimators still remain preferable due to their efficiency and lower estimation variability.

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

Limitation and Direction for Future Research

This study is subjected to several limitations inherent in simulation-based analysis. The data generation process employed assumed a specific structure for the explanatory variables and error terms which may not capture the complexity of real-World data. This analysis is based on assumptions of homoscedastic and well-behaved error distribution across equations.

Future studies could extend this by examining the performance of these estimators under more general conditions like heteroskedastic disturbance, non-normal distributions, and also time dependent SUR systems. Also, exploring high dimensional settings where the number of predictors exceeds the sample size, as well as data driven selection of regularization would provide more information into the efficiency of stacking shrinkage-based methods.

REFERENCES

Alabi, O., Ayinde, K., & Olatayo, T. (2008). Effect of multicollinearity on power rates of the ordinary least squares estimators. Journal of Mathematics and Statistics, 4(2), 75–80.

Breiman, L. (1996). Stacked regressions. Machine Learning, 24(1), 49–64. https://doi.org/10.1007/BF00117832

Bühlmann, P., & Yu, B. (2003). Boosting with the L2 loss: Regression and classification. Journal of the American Statistical Association, 98(462), 324–339. https://doi.org/10.1198/016214503000125

Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the multicollinearity problem and its machine learning approach: A review. Mathematics, 10(8), Article 1283. https://doi.org/10.3390/math10081283

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Hansen, B. E. (2007). Least squares model averaging. Econometrica, 75(4), 1175–1189. https://doi.org/10.1111/j.1468-0262.2007.00785.x

Hansen, B. E., & Racine, J. S. (2012). Jackknife model averaging. Journal of Econometrics, 167(1), 38–46. https://doi.org/10.1016/j.jeconom.2011.06.019

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Herawati, N., Nisa, K., Setiawan, E., & Nusyirwan, T. (2018). Regularized multiple regression methods to deal with severe multicollinearity. International Journal of Statistics and Applications, 8(4), 167–172.

Jadama, A. F., Jobarteh, B., Islam, M. M., & Toray, M. K. (2024). Ensemble learning: Methods, techniques, application [Preprint]. ResearchGate. https://doi:10.13140/RG.2.2.28017.08802

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Zhang, Y., & Yang, Y. (2021). Model selection and model averaging: A survey. Statistical Science, 36(2), 200–222. https://doi.org/10.1214/20-STS774

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, 57(298), 348–368. https://doi.org/10.2307/2281644

Zou, H., & 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

Article DOI: 10.52589/AJMSS-VI2TUSHU
DOI URL: https://doi.org/10.52589/AJMSS-VI2TUSHU