THOMSON REUTERS ENDNOTE

[Badawaire* 5(2): February, 2018] ISSN 2349-4506
Impact Factor: 3.799

Global Journal of Engineering Science and Research Management

A COMPARATIVE STUDY OF SOME ESTIMATORS IN ECONOMETRIC MODEL WITH MULTICOLLINEARITY

Abdulrasheed BelloBadawaire*, S. Y. Jackson, M. G. Bukar
* Department of Mathematics & Statistics, Federal University Wukari, P.M.B. 1020, Wukari, Nigeria
Department of Mathematical Sciences, University of Maiduguri, Maiduguri, Nigeria
Department of Basic Science, MohametLawan College of Agriculture Maiduguri, P.M.B. 1427, Maiduguri, Nigeria

DOI: 10.5281/zenodo.1186509

KEYWORDS: Multicollinearity, OLS, Lasso, Ridge.

ABSTRACT
Mostly, economic data are afflicted with the problems of multicollinearity. This leads to inaccurate parameter estimates in Ordinary Least Squares. Therefore, this paper examined the efficiency of three methods of parameter estimation in regression model (Ordinary Least Squares(OLS), Ridge Regression and Least Absolute Shrinkage and Selection Operator (LASSO)) under multicollinearity. Monte-Carlo experiment of 1000 trials was carried out for four sample sizes (20, 50, 100 and 150), each with three levels of collinearity(Low, Mild and Severe). The findings from this paper showed that when the collinearity level between the predictors is low, irrespective of the sample size, OLS is the most efficient estimator. However, under mild or severe collinearity condition, irrespective of the sample size, Lasso is the most efficient estimator.

INTRODUCTION
Multicollinearity is one of the problems in computing the Ordinary Least Squares estimates in regression analysis. It occurs when the assumption of ‘‘no linear dependencies among the predictor variables’’ is violated. It can as well be defined as a situation whereby two or more of the explanatory variables considered in a model are related (Belsley, 1991; Murray, 2006).

Multicollinearity can be perfect or inperfect. A situation occurs whereby two or more predictors considered in a model move exactly in step with each other and is called a perfect multicollinearity condition (Mrray, 2006). If multicollinearity is perfect, the regression coefficients of the X variables are indeterminate and their standard errors are infinite (Gujarati, 2004). When the multicollinearity in the data is inperfect, the linear combination of the relevant columns, though not zero, is small (Stewart, 1987).

Multicollinearity comes into a model through the data collection method employed, model specification errors, constraints on the model or in the population being sampled, over determination of a model and so on (Gujarati, 2004).

Multicollinearity creates improper specification and inflation of variances of the coefficient estimates (Alin, 2010; Murray, 2006).

When the assumptions of classical linear regression model are met, Least Squares estimator has minimum variance (Stock & Watson, 2007). However, if multicollinearity exists in a data, the OLS estimators, though may still be linear, unbiased, and asymptotically normally distributed, would no longer have the minimum variance among linear unbiased estimators, and as such, would be inefficient relative to other linear unbiased estimators.

In this research, the performances of three estimation methods are investigated under the violation of the assumption of ‘‘no multicollinearity among the predictors being considered in a model’’. At various levels of collinearity and sample sizes, the absolute biases, variances and root mean square errors of parameter estimates from these estimators were examined to select the best estimator.

MATERIALS AND METHODS

In this paper, a linear regression with a dependent and four independent variables is considered. While simulating the data for this paper through Monte-Carlo, multicollinearity was injected through the correlation structure. We then examined the efficiency of three methods of parameter estimations under that condition.

Methods of Parameter Estimation Considered

The three methods of estimating parameters of linear regression model with multicollinearity considered in this paper are:

Ordinary Least Squares (OLS): The OLS is a naïve procedure of estimating the parameters of linear regression model. It is written as

$$y = X\beta + \epsilon \tag{1}$$

Where y is an n by 1 column vector of dependent variable, X is an n by p matrix of independent variables, β is a p by 1 vector of the regression coefficients and ϵ is an n by 1 vector of random errors.
The OLS aims at minimizing

$$\sum_{i=1}^{n} \epsilon_i^2 = \epsilon^T\epsilon = (y - X\beta)^T(y - X\beta)$$
$$= y^Ty - 2\beta^TX^Ty + \beta^TX^TX\beta$$

The Least Squares estimators must satisfy

$$\frac{\partial(\epsilon^T\epsilon)}{\partial\beta} = -2X^Ty + 2X^TX\hat{\beta} = 0$$
$$\Rightarrow X^TX\hat{\beta} = X^Ty$$

Thus, the Least Squares estimator of $\beta$is

$$\hat{\beta}_{OLS} = (X^TX)^{-1}X^Ty \tag{2}$$

Properties:
Bias:

$$\begin{aligned}\text{Bias}(\hat{\beta}_{OLS}) &= E(\hat{\beta}_{OLS}) - \beta \\ &= E[(X^TX)^{-1}X^Ty] - \beta \\ &= E[(X^TX)^{-1}X^T(X\beta + \epsilon)] - \beta \\ &= \beta - \beta \\ &= 0\end{aligned}$$

Thus, Least Squares produce an unbiased estimator of the parameter β in the linear regression model.

Variance:

$$var(\hat{\beta}_{OLS}) = var[(X^TX)^{-1}X^Ty]$$
$$= E[(X^TX)^{-1}X^T\epsilon\epsilon^TX(X^TX)^{-1}]$$
$$= (X^TX)^{-1}X^TE(\epsilon\epsilon^T)X(X^TX)^{-1}$$
$$= (X^TX)^{-1}X^T\sigma^2\text{I}X(X^TX)^{-1}$$
$$\Rightarrow var(\hat{\beta}_{OLS}) = \sigma^2(X^TX)^{-1}.$$

Mean Square Error (MSE):

$$MSE(\hat{\beta}_{OLS}) = E(\hat{\beta}_{OLS} - \beta)^2$$
$$= E\{\hat{\beta}_{OLS} - E(\hat{\beta}_{OLS}) + E(\hat{\beta}_{OLS}) - \beta\}^2$$
$$= E\{[\hat{\beta}_{OLS} - E(\hat{\beta}_{OLS})] + [E(\hat{\beta}_{OLS}) - \beta]\}^2$$
$$= E\{[\hat{\beta}_{OLS} - E(\hat{\beta}_{OLS})]^2 + [E(\hat{\beta}_{OLS}) - \beta]^2\}$$
$$= var(\hat{\beta}_{OLS}) + [\text{Bias}(\hat{\beta}_{OLS})]^2$$

But the bias of OLS is zero, $\Rightarrow MSE(\hat{\beta}_{OLS}) = var(\hat{\beta}_{OLS}) = \sigma^2(X^TX)^{-1}$
OLS estimator provides a baseline for comparison with more complex estimators.

THOMSON REUTERS
ENDNOTE
[Badawaire* 5(2): February, 2018] ISSN 2349-4506
Impact Factor: 3.799
Global Journal of Engineering Science and Research Management

Ridge Regression:

The Ridge Regression (RR) (Hoerl and Kennard, 1970) is an estimation procedure based on the matrix $(X^TX + \lambda I)$, where I denoting the $p * p$ identity matrix and$\lambda = diag(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p)$, with $\lambda_i's$ as the biasing parameters (Wan, 2002). Ridge Regression trade off bias for variance (Alin, 2010; Grewal et al., 2004). The reliability of point estimates of the coefficients increases by the reduction of inflated variances caused by multicollinearity in OLS when appropriate λ is chosen in ridge regression (Li et al., 2010).

The procedure is

$$\hat{\beta}_R = \underbrace{argmin}_{\beta} (Y - X\beta)^T(Y - X\beta) + \beta^2 I \tag{3}$$
$$= \underbrace{argmin}_{\beta} (Y - X\beta)^T(Y - X\beta) + \|\beta\|_2^2$$

Differentiating (3) with respect to β and equating the result to zero, we obtain

$$(X^TX + \lambda I)\hat{\beta}_R = X^TY$$
$$\hat{\beta}_R = (X^TX + \lambda I)^{-1}X^TY \tag{4}$$

Where the ridge parameter (λ>0) is choosing arbitrarily.

Properties:

Bias:

Bias$(\hat{\beta}_R) = E(\hat{\beta}_R) - \beta$

$$E(\hat{\beta}_R) = E\{(X^TX + \lambda I)^{-1}X^TY\}$$
$$= [I + \lambda(X^TX)^{-1}]^{-1}E(\hat{\beta})$$
$$= (X^TX + \lambda I)^{-1}X^TX\beta$$

Since $E(\hat{\beta}_R) \neq \beta$ for any λ>0, then, ridge estimator is biased.

Variance:

$$var(\hat{\beta}_R) = (X^TX + \lambda I)^{-1}var(\beta)[(X^TX + \lambda I)^{-1}X^TX\beta]^T$$
$$= \sigma^2(X^TX + \lambda I)^{-1}X^TX\{(X^TX + \lambda I)^{-1}\}^T$$

Mean Square Error (MSE):

$$MSE(\hat{\beta}_R) = \sigma^2 tr\{(X^TX + \lambda I)^{-2}X^TX\} + \lambda^2\beta^T(X^TX + \lambda I)^{-2}\beta.$$

Least Absolute Shrinkage and Selection Operator (LASSO)

Least absolute shrinkage and selection operator (Lasso) is a sparse model proposed by Tibshirani (1996) whereby most of the coefficients of the irrelevant variables considered in the modelare set to zero while other coefficients are shrunk. Lasso estimator which includes only the best subset of regressors considered in its final model has been used by many researchers to handle the problem of multicollinearity, such as (Fu and Knight, 2000; Zhao and Yu, 2006; Yuan and Lin, 2007;;Lounici, 2008).

Lasso estimator uses the same procedure with ridge regression estimator, but the difference is that the squared ℓ_2 norm ($\|\beta\|_2^2$) in the ridge has been replaced by ℓ_1 norm ($\|\beta\|_1$).

Lasso minimized the sum of squares of residuals subject to the sum of absolute value of the coefficient being less than a constant.

$$\hat{\beta}_{Lasso} = \underbrace{argmin}_{\beta} (Y - X\beta)^T(Y - X\beta) + \lambda\|\beta\|_1 \tag{5}$$

Monte Carlo Experiment

The datasets utilized for this study were simulated from R (www.cran.r-project.org) statistical package. Four set of predictors with sizes 20, 50, 100 and 150 were generated from multivariate normal distribution. Also, the residual term was simulated from the univariate normal distribution with mean 0 and standard deviation σ. The response is simulated with the relationship given by

$$y = 20 + 50x_1 + 80x_2 + 10x_3 + 3x_4 + \varepsilon \tag{6}$$

i.e. $\underline{\beta}' = [\beta_0 = 20, \beta_1 = 50, \beta_2 = 80, \beta_3 = 10, \beta_4 = 3]$

The correlation structures used are;

i. Low Collinearity$(0.2 < \rho \leq 0.45)$

$$\rho *= \left\{\begin{bmatrix} 1 & 0.40 & 0.35 & 0.25 \\ . & 1 & 0.38 & 0.45 \\ . & . & 1 & 0.40 \\ . & . & . & 1 \end{bmatrix}\right\}$$

ii. Mild Collinearity$(0.8 \leq \rho < 0.9)$

$$\rho *= \left\{\begin{bmatrix} 1 & 0.83 & 0.83 & 0.87 \\ . & 1 & 0.84 & 0.84 \\ . & . & 1 & 0.85 \\ . & . & . & 1 \end{bmatrix}\right\}$$

iii. Severe Collinearity$(0.9 \leq \rho < 0.9999)$

$$\rho *= \left\{\begin{bmatrix} 1 & 0.99 & 0.99 & 0.99 \\ . & 1 & 0.999 & 0.998 \\ . & . & 1 & 0.995 \\ . & . & . & 1 \end{bmatrix}\right\}$$

Each of the combinations was iterated 1000 times and the three estimators considered were assessed based on the absolute bias, variance and root mean square error of their parameter estimates.

Averages of estimates of parameters $[(\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4)/5]$ from the three criteria above (absolute bias, variance and RMSE) for the three estimators are computed. After that, the estimators were rankedaccording to their relative performance.

RESULTS AND DISCUSSION

The performances of the three estimators considered in this study under low, mild and severe collinearityconditions and at various sample sizes are presented and discussed here. To make it more generalized, the average values of absolute bias, variance and RMSE of the parameter estimates at the three levels of collinearity (low, mild and severe) and four sample sizes (n=20, n=50, n=100 and n=150) are computed. This allows us to see, on average, foe a given criterion, at each sample size, which estimator is the best in terms of having lowest average value.

Table 1: Average of estimates of parameters from the various criteria

Criteria	Sample Size	Level of Collinearity	Estimates		
			OLS	Ridge	Lasso
Absolute Bias	n=20	Low	1725.079	1525.087	1725.135
		Mild	1576.919	1576.690	1576.523
		Severe	9059.565	7915.997	7518.999
	n=50	Low	1101.369	1101.386	1101.428
		Mild	984.6337	984.4975	984.4868
		Severe	5780.961	4717.114	4518.816
	n=100	Low	761.9487	761.9568	761.9781
		Mild	696.5932	696.5215	696.5071
		Severe	3779.616	2791.450	2661.600
	n=150	Low	634.1306	634.4913	634.5297
		Mild	568.2022	568.1277	568.1156
		Severe	3257.156	2285.693	2174.931
Variance	n=20	Low	23658043	23658306	23658404
		Mild	19170828	19165840	19163438
		Severe	578952121	474056937	440010063
	n=50	Low	9360877	9361090	9361265
		Mild	7335220	7333004	7332956
		Severe	245680187	183263902	172191181
	n=100	Low	4558650	4558738	4559042
		Mild	3746132	3745378	3745302

		Severe	10278835	65479536	60841448
	n=150	Low	3104887	3105005	3105297
		Mild	2445341	2444641	2444581
		Severe	75565712	44194248	40810412
RMSE	n=20	Low	2195.891	2195.900	2195.992
		Mild	1988.571	1988.280	1988.116
		Severe	11242.091	10174.941	9803.312
	n=50	Low	1381.303	1381.324	1381.397
		Mild	1230.100	1229.893	1229.873
		Severe	7318.058	6323.498	6129.644
	n=100	Low	962.1928	962.1981	962.2022
		Mild	878.6355	878.4742	878.4466
		Severe	4736.598	3782.861	3646.823
	n=150	Low	794.1306	794.1492	794.2007
		Mild	710.1862	710.0991	710.0867
		Severe	4060.467	3107.574	2986.482

From table 1 above, it can be observed that the absolute biases, variances and root mean square errors of all the estimators decreases as the sample size increases.

Results in table 1 for the estimators are ranked for each criterion and at each sample size, from the one with lowest average value as 1, to the one with highest average value as 3. The ranks are presented in table 2 below.

Table 2. Ranks of the Estimators Using Various Criteria

Criteria	Sample Size	Level of Collinearity	Estimators		
			OLS	Ridge	Lasso
Absolute Bias	n=20	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
	n=50	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
	n=100	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
	n=150	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
Variance	n=20	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
	n=50	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
	n=100	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
	n=150	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
	n=20	Low	1	2	3
		Mild	3	2	1

RMSE		Severe	3	2	1
	n=50	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
	n=100	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1
	n=150	Low	1	2	3
		Mild	3	2	1
		Severe	3	2	1

From table 2 above it could be deduced that, when the collinearity level between the predictors is low, at all the sample sizes considered, OLS has the smallest ranks in all the three criteria for assessment. It can as well be observed that when the collinearity level between the predictors is mild or severe, at all the sample sizes considered, Lasso estimator consistently has the smallest ranks in all the three criteria.

From table 2 above, the table of preference of these estimators at varying degree of multicollinearity and at different sample size is formed, with the estimator having rank 1 as the most preferred estimator.

Table 3: Preference of the estimators

Criteria	Sample Size	Collinearity Level	Most Preferred Estimator
Absolute Bias	n=20	Low	OLS
		Mild	Lasso
		Severe	Lasso
	n=50	Low	OLS
		Mild	Lasso
		Severe	Lasso
	n=100	Low	OLS
		Mild	Lasso
		Severe	Lasso
	n=150	Low	OLS
		Mild	Lasso
		Severe	Lasso
Variance	n=20	Low	OLS
		Mild	Lasso
		Severe	Lasso
	n=50	Low	OLS
		Mild	Lasso
		Severe	Lasso
	n=100	Low	OLS
		Mild	Lasso
		Severe	Lasso
	n=150	Low	OLS
		Mild	Lasso
		Severe	Lasso
	n=20	Low	OLS
		Mild	Lasso
		Severe	Lasso
	n=50	Low	OLS
		Mild	Lasso

RMSE		Severe	Lasso
	n=100	Low	OLS
		Mild	Lasso
		Severe	Lasso
	n=150	Low	OLS
		Mild	Lasso
		Severe	Lasso

From table 3 above, it can be posited that OLS is the most preferred estimator at all the four sample sizes using both the criteria when the collinearity level is low, but when the collinearity level between the regressors is mild or severe, going by the three criteria and at all the four sample sizes, Lasso is the most preferred estimator.

CONCLUSION

In this study, the efficiency of four methods of parameter estimation when the assumption of ‘’no multicollinearity among the predictors considered in a model’’ is violated is investigated.Results from Monte Carlo experiments have shown that under low collinearity condition, irrespective of the sample size, OLS is the most efficient estimator.

Again when the collinearity condition between the predictors is mild or severe, irrespective of the sample size, Lasso is the most efficient estimator.

REFERENCES

1. Alin, A. (2010), Multicollinearity. Willey Interdisciplinary Reviews: Computational Statistics, 2, 370-374.
2. Belsley, D. A. (1991), Conditioning Diagnostics Collinearity and Week Data in Regression. New York: Wiley- Interscience.
3. Fu, W. & Knight, K. (2000), Asymptotic for Lasso-type estimators. Ann. Stat., 28, 1356-1378.
4. Grewal, R., Cote, J. A., & Baumgartner, H. (2004),Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing. Marketing Science, 23, 519-529.
5. Gujarati, D. N. (2004), Basic Econometrics, (4^{th} Ed.).The McGraw-Hill Companies.NewYork.
6. Hoerl, A. E., & Kennard, R. W. (1970), Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12, 55-67.
7. Li, Y. F., Xie, M., &Goh, T. N. (2010), Adaptive Ridge Regression System for Software Cos Estimation On Multi-Collinear Datasets. Jounal of Systems and Software, 83, 2332-2343.
8. Lounici, K. (2008), Sub-norm convergence rate and sign concentration property of Lasso and Dantzig estimators.Electronic Journal of Statistics, 2.
9. Murray, M. P. (2006), Econometrics A Modern Introduction. Boston: Addison-Wesley.
10. Stewart, G. W. (1987),Collinearity and Least Squares Regression. Statistical Science, 2, 68-84.
11. Stock, J. H., & Watson, M. W. (2007), Introduction toEconometrics. (D. Clinton, Ed., 2^{nd}ed.). International Edition: Pearson Education.
12. Wan, A. T. K. (2002),On Generalized Ridge Regression Estimation Under Collinearity and Balanced Loss. Applied Mathematics and Computation, 129, 455-467.
13. Yuan, M., & Lin, Y. (2007),On the non-negative garrotte estimator. J. Roy. Stat. Soc. B, 69, 143-161.
14. Zhao, P., & Yu, B. (2006),On model selection consistency of Lasso. J. Mac. Learn. Res., 7, 2541-2563.