Far East Journal of Theoretical Statistics
Published Online: October 2013
Available online at http://pphmj.com/journals/fjts.htm
Volume 44, Number 2, 2013, Pages 135-144
© 2013 Pushpa Publishing House, Allahabad, India

ON SHRUNKEN ESTIMATORS IN LINEAR MODELS

Ghadban Khalaf

Department of Mathematics
Faculty of Science
King Khalid University
Saudi Arabia

Abstract

It is well-known that one of the major consequences of multicollinearity on the ordinary least (OLS) estimator is that the estimator produces large sampling variances, which in turn might inappropriately lead to exclusion of otherwise significant coefficients from the model. To circumvent this problem, two accepted estimation procedures which are often suggested are the James-Stein method and the ridge regression method. Both of the two methods ensure a smaller mean square error (MSE) value for the estimator in the presence of multicollinearity. In this paper, we have proposed a new estimator and it has been shown, view simulation, that this estimator is superior to both the James-Stein as well as the ordinary ridge regression estimators by the criterion of MSE of the estimator of the regression coefficients.

1. Introduction

The problem of multicollinearity and its statistical consequences on a linear regression model are very well-known in statistics. It is, for instance, known that one of the major consequences of multicollinearity on the OLS

Received: August 21, 2013; Accepted: September 21, 2013

2010 Mathematics Subject Classification: Primary 62J07; Secondary 62J05.

Keywords and phrases: OLS estimator, multicollinearity, ridge regression, James-Stein estimator, simulation.

method of estimation is that it produces large sampling variances for the estimation regression coefficients, which in turn gives rise to the possibility that otherwise significant coefficients may be excluded from the model improperly. Several suggestions have been made by researchers to contain this inflation of sampling variances so that meaningful inferences can be drawn for linear regression models having multicollinearity among two or more regressors.

In this attempt towards improving the precision of the OLS estimator, two very standard and well-established procedures are the James-Stein method of estimation and the method of ridge regression.

The James-Stein [4] estimator is a biased estimator but it can be shown that this estimator dominates the OLS approach, i.e., it have lower variances. An earlier version of this estimator was developed by Stein in 1956 and it is sometimes referred to as Stein’s estimator.

As regards the method of ridge regression, it is one of the most widely-used “ad-hoc” solutions to the problem of multicollinearity although the detailed sampling properties of this estimator are largely unknown, Hoerl and Kennard [2]. The principle result concerning the ordinary ridge regression is that; while it is always a biased estimator, it is superior to the OLS estimator by the criterion of sampling variance. However, the sampling variance being not the proper criterion for evaluating the performance of a biased estimator, Hoerl and Kennard [2] and Vinod [13] examined the performance of the ordinary ridge regression estimator by the MSE criterion and showed that there always exists an ordinary ridge regression estimator having smaller MSE value than the OLS estimator (see, for example, Judge et al. [5], Vinod and Ullah [14], Mason and Perreault [11], Grapentine [1], Mardikyan and Cetin [10], Khalaf and Shukur [8], Khalaf [6] and Khalaf et al. [7]).

A consequence of the above discussion is the following counterintuitive result: when three or more related parameters are measured, their total MSE can be reduced by using a combined estimator such as the James-Stein and the ridge regression estimators.

In this paper, we have suggested a new estimator by combining in a particular way the two approaches followed in obtaining the James-Stein and the ordinary ridge regression estimators. In other words, we have here suggested a new estimator by grafting the ridge regression philosophy into the James-Stein estimator. Further, we have established the MSE superiority of the proposed estimator over both the James-Stein and the ordinary ridge regression estimators.

The arrangements in the paper are as follows: the model as well as the James-Stein and the ordinary ridge regression estimators are described in Section 2. The proposed estimator is given in Section 3. The comparative performance of this estimator vis-à-vis the James-Stein and the ordinary ridge regression estimators are studied in Section 4. The paper ends with some concluding remarks in Section 5.

2. The Model, James-Stein Estimator and the Ridge Regression

Consider the standard *p*-variable linear regression model:

$$Y = X\beta + e, \tag{1}$$

where Y is an $(n \times 1)$ vector of observations on the dependent variable, X is an $(n \times p)$ matrix of observations on p nonstochastic independent variables, β is the $(p \times 1)$ vector of parameters associated with the p regressors and e is an $(n \times 1)$ vector of disturbances having mean zero and variance-covariance matrix $\sigma^2 I_n$. We assume that two or more regressors in X are closely related so that the model suffers from the problem of multicollinearity.

The usual OLS estimator of β denoted by $\hat{\beta}$, which is obtained by minimizing $(Y - X\beta)'(Y - X\beta)$, can be written as:

$$\hat{\beta} = (X'X)^{-1} X'Y. \tag{2}$$

It is found that among all linear, unbiased estimates, the OLS estimator produces the smallest variances. But in the presence of multicollinearity, both the least squares estimator and its variance are inflated. That is why we relaxed the condition of unbiasedness and considered MSE as criterion for comparing estimators by introducing a new kind of estimate that acts by shrinking the OLS coefficients and studied its MSE.

James and Stein [4] showed that the estimator:

$$\hat{\beta}_{JS} = \left(1 - \frac{(p-2)\sigma^2}{\hat{\beta}' X'X \hat{\beta}}\right)\hat{\beta} \tag{3}$$

dominates the OLS estimator $\hat{\beta}$ when $X'X \neq I$, for any $p > 2$, meaning that the James-Stein estimator always achieves lower MSE than the OLS. By definition, this makes the least square estimator inadmissible when $p > 2$. We will henceforth refer to this estimator as JS.

The family of ordinary ridge regression estimators is defined by:

$$\hat{\beta}(k) = (X'X + kI_p)^{-1} X'Y, \quad k > 0. \tag{4}$$

Obviously, the OLS estimator refers to the case where $k = 0$. It may be noted here that the ordinary ridge regression estimator can be obtained by minimizing $\beta'\beta$ subject to $(\beta - \hat{\beta})' X'X(\beta - \hat{\beta}) = Q$, where Q is fixed, Hoerl and Kennard [2]. It is easily seen from (4) that $\hat{\beta}(k)$ is always a biased estimator of β, and the bias equals $-k(X'X + kI_p)^{-1}$. Since the ridge regression method always yields a biased estimator, Hoerl and Kennard [2] studied its MSE property and showed that there always exists a $k > 0$ such that $\hat{\beta}(k)$ has a smaller MSE value than $\hat{\beta}$, and that a sufficient condition for this to hold is that

$$k < \frac{\sigma^2}{\alpha_{\max}^2}, \tag{5}$$

where $\alpha_{\max}^2$ is the squared value of the largest element of the vector

$\alpha = Q'\beta$, Q being the matrix of orthogonal characteristic vectors of $X'X$. In principle, with the right choice of k, we can get an estimator with a better MSE. The estimate is not unbiased, but what we pay for in bias, we make up for in variance, since the MSE of $\hat{\beta}(k)$ is given by:

$$\begin{aligned} MSE(\hat{\beta}(k)) &= E[(\hat{\beta}(k) - \beta)'(\hat{\beta}(k) - \beta)] \\ &= \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \sum_{i=1}^{p} \frac{\beta_i^2}{(\lambda_i + k)^2} \\ &= Var(\hat{\beta}(k)) + Bias^2(\hat{\beta}(k)), \end{aligned} \tag{6}$$

where λ_i is the eigenvalue of the matrix $X'X$. Hoerl et al. [3] argued that a reasonable choice of k is:

$$k = \frac{p\sigma^2}{\alpha'\alpha}, \tag{7}$$

if these quantities were known. They suggested using;

$$\hat{k}_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}, \tag{8}$$

as an estimate of k in (7). This estimator will be denoted by HKB.

3. The Proposed Estimator

In this section, we suggest a new estimator by combining in a particular way the two approaches underlying the James-Stein and the ordinary ridge regression estimators and study its performance using simulation techniques. Specifically, we propose modifying the James-Stein estimator in the ordinary ridge regression philosophy so as to obtain a new estimator for β, which may be designated as modified ridge-Stein estimator, denoted by $\beta^*(k)$, and given by:

$$\beta^*(k) = c\hat{\beta}(k), \qquad k > 0, \tag{9}$$

where $\hat{\beta}(k)$ is the ordinary ridge regression estimator, given by (4), and

$$c = (I_p - k\beta^*_{JS})^{-1}. \tag{10}$$

Obviously, $\beta^*(k) = \hat{\beta}(k)$ when $k = 0$, where β^*_{JS} is a modified estimator of James and Stein [4] suggested by Khalaf and Iguernane [13], given by:

$$\beta^*_{JS} = \hat{c}\hat{\beta}, \tag{11}$$

and

$$\hat{c} = \left[\frac{\lambda_{\max} + \lambda_{\min}}{p}\right]\left(1 - \frac{(p-2)\hat{\sigma}^2}{\hat{\beta}' X' X \hat{\beta}}\right), \tag{12}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and the smallest eigenvalues of the matrix $X'X$, p is the number of the explanatory variables and $\hat{\sigma}^2$ is the unbiased estimator of σ^2, defined by:

$$\hat{\sigma}^2 = \frac{RSS}{n - p - 2}, \tag{13}$$

and RSS is the residual sum of squares. For our proposed estimator, defined by (9), we use the acronym GK.

4. The Performance of the Proposed Estimator

Since the proposed estimator is always a biased estimator of β unless $k = 0$, and hence the appropriate criterion for gauging the performance of our suggested estimator is the MSE of the estimator of the coefficients by using the simulation technique.

4.1. Simulations and results

A simulation study was conducted in order to draw conclusions about the performance of our suggested estimator relative to HKB, JS and the OLS estimators depending on the MSE for individual estimators and the predictive accuracy of model based on the coefficient of determination R^2 for each method. In this paper, we simulate a set of data using SAS package, where the correlation coefficients between the predictor variables (the X's) are large. The data consist of 50 observations on six regressors and a response. This sample of size 50 observations is simulated for 10000 iterations. The corresponding summary with $X'X$ in correlation form based on a set of the simulated data is reproduced in Table 1.

Table 1. Correlation coefficients for the 6-variables, $n = 50$

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1					
X_2	0.965	1				
X_3	0.978	0.969	1			
X_4	0.970	0.994	0.967	1		
X_5	0.993	0.980	0.994	0.984	1	
X_6	0.957	0.982	0.958	0.945	0.990	1

The eigenvalues of the correlation matrix, ordered by their magnitudes, were: 6.6081, 0.0798, 0.0608, 0.0257, 0.0143, 0.0105. It is clear that the last five eigenvalues are fairly small. Thus, for our 6-factor data, the data are multicollinear because the condition number is large, $C.N. = \frac{\lambda_{\max}}{\lambda_{\min}} = \frac{\lambda_1}{\lambda_6}$ $= 629$.

From the previous indicators, it is obvious that there is a serious multicollinearity problem because there is more than one of the eigenvalues close to zero and the condition number is more than 5.

By using the OLS and the methods of HKB, JS and our proposed method to analyze the simulated data set, we get the following results, shown in Table 2.

Table 2. The estimated regression coefficients and its MSE

OLS		HKB		JS		GK	
Coeff.	MSE	Coeff.	MSE	Coeff.	MSE	Coeff.	MSE
1.1115	0.1761	1.0021	0.1744	1.0522	0.1831	0.3232	0.0060
-0.3806	0.2061	-0.3147	0.1753	0.3304	0.1841	0.4087	0.1359
-0.1087	0.1872	-0.0321	0.1573	-0.0337	0.1652	0.6746	0.1804
0.4831	0.1736	0.4326	0.1462	0.4542	0.1535	0.2762	0.1377
-0.6747	0.3411	-0.5226	0.2309	-0.5487	0.2424	0.7772	0.3240
0.2642	0.3052	0.1539	0.1736	0.1616	0.1823	0.4172	0.2659

Table 3 shows the computation of the coefficient of determination (R^2) for each model.

Table 3. R^2 for the OLS estimator and the other shrunken estimators

OLS	HKB	JS	GK
0.763	0.802	0.773	0.931

From the previous results, it is noticed that the shrunken regression models have smaller MSE and large R^2 than the OLS estimator when multicollinearity problem exists in the data.

5. Conclusions

The estimates of the OLS are great, but we can improve on them. We then considered shrinkage estimates and showed that if we were willing to give up a little in terms of bias, we could do better in terms of MSE.

In this paper, we have combined the criteria underlying the James-Stein estimator and the ordinary ridge regression estimator to obtain a new estimator for the regression coefficients of a linear regression model which suffers from the problem of multicollinearity. The performance of the proposed estimator, designated as modified ridge-Stein estimator, as compared to those of the James-Stein and ordinary ridge regression estimators has been studied by the criterion of MSE of the estimator of the regression coefficients. We have established that our new estimator has a small MSE value than the ordinary ridge regression estimator, and that there exists a value for $k > 0$ such that the new estimator is dominated over the James-Stein estimator.

References

[1] A. Grapentine, Managing multicollinearity, J. Market. Res. 9 (1997), 11-21.

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

[3] A. E. Hoerl, R. W. Kennard and K. F. Baldwin, Ridge regression: some simulation, Comm. Statist. Theory Methods 4 (1975), 105-124.

[4] J. F. James and C. M. Stein, Estimation with quadratic loss, Proceedings of the Fourth Berkeley Symposium on Math. Statist. Prob., Vol. 1, 1961, pp. 361-379.

[5] G. G. Judge, W. E. Griffiths, R. C. Hill and T. C. Lee, The Theory and Practice of Econometrics, John Wiley, New York, 1980.

[6] G. Khalaf, Suggested ridge regression estimators under multicollinearity, J. Natural and Applied Sciences 15 (2011), 170 -193.

[7] G. Khalaf, K. Mansson, G. Shukur and P. Sjolander, A Tobit ridge regression estimator, Comm. Statist. Theory Methods, 2012 (to appear).

[8] G. Khalaf and G. Shukur, Choosing ridge parameter for regression problems, Comm. Statist. Theory Methods 34 (2005), 1177-1182.

[9] G. Khalaf and M. Iguernan, Shrunken estimators and the problem of multicollinearity, Far East J. Theor. Stat. 43(2) (2013), 121-131.

[10] S. Mardikyan and E. Cetin, Efficient choice of biasing constant for ridge regression, Int. J. Contemp. Math. Sci. 3 (2008), 527-536.

[11] C. H. Mason and W. D. Perreault, Collinearity, power and interpretation regression analysis, J. Market. Res. 28 (1991), 268-280.

[12] C. M. Stein, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proceedings of the Third Berkeley Symposium on Math. Statist. Prob., University of California Press, Berkeley, Vol. 1, 1956, pp. 197-206.

[13] H. D. Vinod, A ridge estimator whose MSE dominates OLS, International Econ. Rev. 19 (1978), 727-737.

[14] H. D. Vinod and A. Ullah, Recent Advances in Regression Model, Marcel Dekker Inc., New York, 1981.