Shalabh, Toutenburg, Heumann:

Risk Performance Of Stein-Rule Estimators Over The Least Squares Estimators Of Regression Coefficients Under Quadratic Loss Structures

Sonderforschungsbereich 386, Paper 495 (2006)

Risk Performance Of Stein-Rule Estimators Over The Least Squares Estimators Of Regression Coefficients Under Quadratic Loss Structures

Shalabh[*], H. Toutenburg[†] and C. Heumann[‡]

Discussion Paper 493 SFB386, University of Munich, Munich

November 3, 2006

Abstract

This paper presents a general loss function under quadratic loss structure and discusses the comparison of risk functions associated with the unbiased least squares and biased Stein-rule estimators of the coefficients in a linear regression model.

[*]Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur 208016, India. E-mail: shalab@iitk.ac.in, shalabh1@yahoo.com (Corresponding Author)

[†]Department of Statistics, Ludwig–Maximilians–University Munich, Ludwigstr. 33, 80539 Munich, Germany. E-mail: helge.toutenburg@stat.uni-muenchen.de

[‡]Department of Statistics, Ludwig–Maximilians–University Munich, Ludwigstr. 33, 80539 Munich, Germany. E-mail: chris@stat.uni-muenchen.de

1 Introduction

Performance of any estimation procedure for the parameters in a model is generally evaluated by either the goodness of fitted model or the concentration of the estimates around the true parameter values. In practice, it may often be desirable to employ both the criteria simultaneously; see, for instance, Toutenburg and Shalabh (1996), Shalabh (1995, 2000) and Zellner (1994) for some illustrative examples. Accordingly, Zellner (1994) has introduced the balanced loss function. As the goodness of fitted model can be interpreted as the goodness of the predictions for the actual values of the study variable within the sample while the concentration of estimates may be measured by the goodness of predictions for the average values of the study variable within the sample, Shalabh (1995) has presented the predictive loss function.

In this paper, we present a general loss function of which the loss functions considered by Shalabh (1995) and Zellner (1994) are particular cases, and expose the unbiased least squares and biased Stein-rule estimators of the regression coefficients. In Section 2 we describe the linear regression model and present a general loss function under quadratic loss structure. Section 3 gives a comparison of the risk functions associated with the least squares and Stein-rule estimators, and a condition on the characterizing scalar for the superiority of the Stein-rule estimators over the least squares estimator is obtained. Several particular cases are also considered. Some concluding remarks are then placed in Section 4. Lastly, the Appendix gives the proof of Theorem.

2 Linear Regression Model And The Loss Function

Let us consider the following linear model:

$$y = X\beta + \sigma\epsilon \tag{2.1}$$

where y is a $n \times 1$ vector of n observations on the study variable, X is a $n \times p$ full column rank matrix of n observations on p explanatory variables, σ is an unknown positive scalar and ϵ is a $n \times 1$ vector of disturbances.

It is assumed that the elements of ϵ are independently and identically distributed following a distribution with mean 0, variance 1 and third moment γ_1 measuring skewness.

If $\tilde{\beta}$ denotes any estimator of β, the goodness of the fitted model is reflected in the residual vector $(X\tilde{\beta} - y)$. Similarly, the pivotal quantity for measuring the concentration of estimates around the true parameter values is the estimation error $(\tilde{\beta} - \beta)$. Accordingly, the quadratic loss function for the goodness of fit of the model is

$$(X\tilde{\beta} - y)'(X\tilde{\beta} - y) \tag{2.2}$$

while the commonly employed loss function for the precision of estimation are

$$(\tilde{\beta} - \beta)'(\tilde{\beta} - \beta) \tag{2.3}$$

or

$$(\tilde{\beta} - \beta)'X'X(\tilde{\beta} - \beta) \tag{2.4}$$

Taking both the criteria of the goodness of fit and precision of estimation together, Zellner (1994) has proposed the following balanced loss function:

$$BL(\tilde{\beta}) = \omega(X\tilde{\beta} - y)'(X\tilde{\beta} - y) + (1 - \omega)(\tilde{\beta} - \beta)'X'X(\tilde{\beta} - \beta) \tag{2.5}$$

where ω is a scalar lying between 0 and 1.

From the viewpoint of the prediction of the values of the study variable within the sample, the loss functions (2.2) and (2.4) can be regarded as arising from the prediction of actual values y by $X\tilde{\beta}$ and the prediction of the average values $E(y) = X\beta$ by $X\tilde{\beta}$ respectively. Accordingly, Shalabh (1995) has defined a target function

$$T = \omega y + (1-\omega)E(y) \tag{2.6}$$

and has presented the following predictive loss function

$$\begin{aligned} PL(\tilde{\beta}) &= (X\tilde{\beta} - T)'(X\tilde{\beta} - T) \\ &= \omega^2(X\tilde{\beta} - y)'(X\tilde{\beta} - y) \\ &\quad +(1-\omega)^2(\tilde{\beta} - \beta)'X'X(\tilde{\beta} - \beta) \\ &\quad +2\omega(1-\omega)(X\tilde{\beta} - T)'X(X\tilde{\beta} - T) \end{aligned} \tag{2.7}$$

where ω is a scalar between 0 and 1. Note that $\omega = 0$ and $\omega = 1$ in (2.6) provides predictions for average and actual values of y. Any other value $0 < \omega < 1$ provides the weight assigned to the actual value prediction and provides simultaneous prediction of actual and average values of y.

Looking at the functional forms of the balanced loss function and the predictive loss function, we propose the following weighted loss function:

$$\begin{aligned} WL(\tilde{\beta}) &= \lambda_1(X\tilde{\beta} - y)'(X\tilde{\beta} - y) \\ &\quad +\lambda_2(\tilde{\beta} - \beta)'X'X(\tilde{\beta} - \beta) \\ &\quad +(1-\lambda_1-\lambda_2)(X\tilde{\beta} - y)'X(\tilde{\beta} - \beta) \end{aligned} \tag{2.8}$$

where λ_1 and λ_2 are scalars characterizing the loss functions.

Clearly, the function (2.8) encompasses the loss functions (2.2), (2.4), (2.5) and (2.7) as particular cases. Thus it is fairly general and sufficiently flexible.

3 Comparisons Of Least Squares And Stein-Rule Estimators

The least squares estimator of β is given by

$$b = (X'X)^{-1}X'y \tag{3.1}$$

which is well known for its optimality in the class of linear and unbiased estimators.

If we drop the linearity and unbiasedness, there exist estimators with better performance than the least squares estimator under the risk criterion. One such interesting family of nonlinear and biased estimators of β, popularly known as Stein-rule estimators is defined by

$$\hat{\beta} = \left[1 - \left(\frac{k}{n-p+2}\right)\frac{y'\left(I-H\right)y}{y'Hy}\right] b \tag{3.2}$$

where

$$H = X(X'X)^{-1}X' \tag{3.3}$$

and k is a positive nonstochastic scalar; see, e.g. Judge and Bock (1978) and Saleh (2006).

For comparing the estimators, let us take the criterion as risk, i.e., the expected value of the weighted loss function (2.8).

The exact expressions for the risk functions can be derived bur their nature would be sufficiently intricate. We therefore consider their large sample asymptotic approximations. For this purpose, we assume that the explanatory variables are asymptotically cooperative, i.e., the limiting form of the matrix $n^{-1}X'X$ is finite and nonsingular, as n tends to infinity.

The large sample asymptotic expressions for the risk functions are derived in the Appendix and are presented below.

Theorem 1:

The risk function of b and $\hat{\beta}$ to order $O(n^{-1})$ are given by

$$\begin{aligned} R(b) &= E\left[WL(b)\right] \\ &= \sigma^2\lambda_1 n - \sigma^2 p(\lambda_1-\lambda_2) \end{aligned} \tag{3.4}$$

$$\begin{aligned} R(\hat{\beta}) &= E\left[WL(\hat{\beta})\right] \\ &= \sigma^2\lambda_1 n - \sigma^2 p(\lambda_1-\lambda_2) \\ &\quad -\frac{\sigma^4 k}{n\beta' S\beta}\left[(1-\lambda_1+\lambda_2)\left(\frac{\gamma_1}{\sigma}\bar{X}'\beta+p-2\right)-k\right] \end{aligned} \tag{3.5}$$

where $S=\frac{1}{n}X'X$ and $\bar{X}$ is a $p\times 1$ vector of the means of the observations on the p explanatory variables.

Comparing (3.4) and (3.5), it is observed that the Stein-rule estimator has smaller risk to the order of our approximations, in comparison to the least squares estimator when

$$k<(1-\lambda_1+\lambda_2)\left(\frac{\gamma_1}{\sigma}\bar{X}'\beta+p-2\right) \tag{3.6}$$

provided that

$$(\lambda_1-\lambda_2)<1 \text{ and } \left(\frac{\gamma_1}{\sigma}\bar{X}'\beta+p-2\right)>0 \tag{3.7}$$

or

$$(\lambda_1-\lambda_2)>1 \text{ and } \left(\frac{\gamma_1}{\sigma}\bar{X}'\beta+p-2\right)<0. \tag{3.8}$$

When the distribution of disturbances is symmetric and/or $\bar{X}$ is a null vector, i.e., the observations on the explanatory variables are taken as deviations from their corresponding means, then the condition (3.6) becomes free from unknown parameters β and is satisfied when either of the following two conditions holds true:

$$k < (1 - \lambda_1 + \lambda_2)(p - 2) \text{ and } (\lambda_1 - \lambda_2) < 1 \text{ if } p > 2 \tag{3.9}$$

$$k < (\lambda_1 - \lambda_2 - 1)(2 - p) \text{ and } (\lambda_1 - \lambda_2) > 1 \text{ if } p = 1, 2. \tag{3.10}$$

Now let us examine the performance of estimators under some interesting loss functions.

3.1 Loss Function: $(X\tilde{\beta} - y)'(X\tilde{\beta} - y)$

This loss function is a particular case of (2.8) with $\lambda_1 = 1$ and $\lambda_2 = 0$. It is indeed the residual sum of squares and is the sum of squares of prediction errors when the aim is to predict the actual values of the study of the study variable within the sample.

In this case, it is interesting to observe from (3.4) and (3.5) that the least squares estimator remains unbeaten by all the Stein-rule estimators irrespective of the nature of the observations on the explanatory variables and the distribution of disturbances. This matches with the result obtained by Srivastava and Shalabh (1996, p.143) on the basis of exact risk expressions.

3.2 Loss Function: $(\tilde{\beta} - \beta)'X'X(\tilde{\beta} - \beta)$

If we set $\lambda_1 = 0$ and $\lambda_2 = 1$ in (2.8), we get this loss function. It is essentially the weighted sum of squares of the estimation errors and is the sum of squares of the prediction errors when the aim is to predict the average values of the study variable within the sample.

In this case, the Stein-rule estimators are better than the least squares estimator when

$$k < 2\left(\frac{\gamma_1}{\sigma}\bar{X}'\beta + p - 2\right) \tag{3.11}$$

with the rider that the quantity on the right hand side is positive; see also Vinod and Srivastatave (1995).

This condition reduces to

$$k < 2(p-2);\ p > 2 \tag{3.12}$$

when the distribution of disturbances is symmetric irrespective of the nature of data on the explanatory variables or $\bar{X}$ is a null vector whether the distributions of disturbances is symmetric or asymmetric.

Similarly, the condition (3.11) is satisfied as long as (3.12) holds true provided that γ_1 and $\bar{X}'\beta$ have the same sign, i.e., $\bar{X}'\beta$ is positive for positively skewed distributions of disturbances and is negative for negatively skewed distributions of disturbances. In fact, it is possible to find Stein-rule estimators with better performance than the least squares estimator even for $p = 1$ and $p = 2$ when

$$\gamma_1\bar{X}'\beta > 2\sigma. \tag{3.13}$$

It may be noticed (3.12) is a well-known condition for the superiority of Stein-rule estimators on the basis of exact risk under the normality of disturbances; see, e.g., Judge and Bock (1978) and Saleh (2006).

3.3 Loss Function: $(X\tilde{\beta})'X(\tilde{\beta} - \beta)$

This loss function is obtained from (2.8) by putting $\lambda_1 = \lambda_2 = 0$ and can be regarded as measuring the covariability between the residuals and the estimation errors. From the viewpoint of prediction within the sample, it is the sum of cross products of errors arising from the prediction of the actual and average values of the study variable by $X\tilde{\beta}$. This loss function is, however, not interesting because the exact risk (3.4) of the least squares estimator turns out to be zero which is the risk of Stein-rule estimators to order $O(n^{-1})$ which may be negative.

3.4 Loss Function: $\omega(X\tilde{\beta} - y)'(X\tilde{\beta} - y) + (1 - \omega)(\tilde{\beta} - \beta)'X'X(\tilde{\beta} - \beta)$

If we put $\lambda_1 = \omega$ and $\lambda_2 = (1 - \omega)$ in (2.8), we get the balanced loss function proposed by Zellner (1994). This is indeed a convex combination of the sum of squares of the residuals and the weighted sum of squares of the estimation errors. It can also be interpreted as a convex combination of the two sums of squares of the prediction errors arising from the prediction of the actual and average values of the study variable within the sample.
For $0 \le \omega < 1$, the Stein-rule estimators perform better than the least squares estimator when

$$k < 2(1 - \omega)\left(\frac{\gamma_1}{\sigma}\bar{X}'\beta + p - 2\right). \tag{3.14}$$

When γ_1 is zero and/or $\bar{X}$ is a null vector, the condition (3.14) assumes a simple form:

$$k < 2(1 - \omega)(p - 2);\ p > 2. \tag{3.15}$$

This serves as a sufficient condition for the superiority of the Stein-rule estimators over the least squares estimator in case of the asymmetric distributions of the disturbances provided that the skewness coefficient γ_1 has the same sign as $\bar{X}'\beta$. Further, if (3.14) holds true, one can find Stein-rule estimators better than the least squares estimator even when there is simply one or two explanatory variables in the model.

It may be observed that the condition (3.14) has been derived by Giles, Giles and Ohtani (1996) by considering the exact risk under the normality of disturbances; see also Ohtani (1998).

3.5 Loss Function: $\omega^2(X\tilde{\beta} - y)'(X\tilde{\beta} - y) + (1-\omega)^2(\tilde{\beta} - \beta)'X'X(\tilde{\beta} - \beta) + 2\omega(1-\omega)(X\tilde{\beta} - y)'X(\tilde{\beta} - \beta)$

This loss function is a particular case of (2.8) with $\lambda_1 = \omega^2$ and $\lambda_2 = (1-\omega)^2$. It is a combination of the sum of squares of the residuals, the weighted sum of squares of the estimation errors and the weighted sum of cross products of the residuals and the estimation errors. This is also equal to the sum of squares of prediction errors when $X\tilde{\beta}$ is employed for the prediction of a convex combination of the actual and average values of the study variable; see Shalabh (1995).

From (3.4) and (3.5), it is seen that the Stein-rule estimators have smaller risk in comparison to the least squares estimator when

$$k < 2(1-\omega)\left(\frac{\gamma_1}{\sigma}\bar{X}'\beta + p - 2\right) \tag{3.16}$$

which is precisely the same as (3.14) obtained from risk comparison under the balanced loss function. The condition of (3.16) with $\gamma_1 = 0$ matches with the condition of Shalabh (1995) on the basis of exact risk; see also Shalabh (1999).

4 Some Remarks

Appreciating the simultaneous use of the two performance criteria, viz, the goodness of fitted model and the concentration of estimates around the true parameter values, for judging the efficiency of any estimation procedure for the coefficients in a linear regression model, we have presented a general loss function using the quadratic loss structure. Several popular loss functions are found to be the particular cases of it, and thus the properties of loss function is fairly general and sufficiently flexible.

For the regression coefficient vector, we have considered the unbiased least squares and biased Stein-rule estimators. We have compared their performance according to the risk criterion under the proposed loss function and have obtained a condition on the characterizing scalar for the superiority of the Stein-rule estimates over the least squares estimator.

We have not assumed any functional form for the distribution of the disturbances; we have simply supposed the finiteness of the first three moments. Accordingly, the large sample approximations for risk functions are used for the purpose of comparison. An interesting observation emerging from our investigations is that the condition on the characterizing scalar for the superiority of the Stein-rule estimators over the least squares estimator deduced from the exact risks under the normality of disturbances remains valid for a variety of symmetric and asymmetric distributions. Further, the Stein-rule estimators are found to be quite robust with respect to the choice of loss functions.

A Appendix

From (2.1) and (2.8), we observe that

$$WL(\tilde{\beta}) = \sigma^2\lambda_1\epsilon'\epsilon - \sigma(1+\lambda_1-\lambda_2)\epsilon' X(\tilde{\beta}-\beta) + (\tilde{\beta}-\beta)'X'X(\tilde{\beta}-\beta). \quad (A.1)$$

Setting $\tilde{\beta} = b$, it is easy to see that

$$R(b) = E[WL(b)] = \sigma^2\lambda_1 n - \sigma^2 p(\lambda_1 - \lambda_2) \quad (A.2)$$

which is the result (3.4) of the Theorem.

Now, if we write

$$u = n^{\frac{1}{2}} X'\epsilon$$

and

$$v = n^{\frac{1}{2}}\left(\frac{\epsilon'\epsilon}{n} - 1\right),$$

we have

$$b - \beta = \frac{\sigma}{n^{\frac{1}{2}}} S^{-1}u. \quad (A.3)$$

Next, consider the quantity

$$\begin{aligned} \frac{y'(I-H)y}{(n-p+2)y'Hy} &= \frac{\sigma^2(n + n^{\frac{1}{2}}v - u'S^{-1}u)}{(n-p+2)(n\beta'S\beta + 2n^{\frac{1}{2}}\sigma\beta'u + \sigma^2 u'S^{-1}u)} \\ &= \frac{\sigma^2}{n\beta'S\beta}\left(1 + \frac{v}{n^{\frac{1}{2}}} - \frac{1}{n}u'S^{-1}u\right)\left(1 - \frac{p-2}{n}\right)^{-1} \cdot \\ &\quad \left(1 + \frac{2\sigma\beta'u}{\gamma^{\frac{1}{2}}\beta'S\beta} + \frac{\sigma^2 u'S^{-1}u}{n\beta'S\beta}\right)^{-1}. \end{aligned}$$

Expanding and retaining terms to order $O(n^{-\frac{3}{2}})$, we find

$$\frac{y'(I-H)y}{(n-p+2)y'Hy} = \frac{\sigma^2}{n\beta'S\beta} + \frac{\sigma^2}{n^{\frac{3}{2}}\beta'S\beta}\left(v - \frac{2\sigma\beta'u}{\beta'S\beta}\right) + Op(n^{-2}). \quad (A.4)$$

Substituting (A.3) and (A.4) in (3.2), we find

$$\begin{aligned}(\hat{\beta}-\beta) &= \frac{\sigma}{n^{\frac{1}{2}}}S^{-1}u - \frac{\sigma^2 k}{n\beta' S\beta}\beta \\ &\quad -\frac{\sigma^2 k}{n^{\frac{3}{2}}\beta' S\beta}\left[v\beta + \sigma\left(S^{-1} - \frac{2}{\beta' S\beta}\beta\beta'\right)u\right] + Op(n^{-2}). \quad (A.5)\end{aligned}$$

Using the distributional properties of ϵ, it is easy to verify that

$$\begin{aligned}E(\epsilon' A\epsilon) &= tr(A) \\ E(\epsilon' A\epsilon\cdot\epsilon) &= \gamma_1(I * A)e\end{aligned}$$

where A is any $n\times n$ symmetric matrix with nonstochastic elements, e is a $n\times 1$ vector with all elements unity and * denotes the Hadamard product operator of matrices.

Making use of these results and neglecting terms of higher order of smallness than $O(n^{-1})$, we see from (A.5) that

$$E[\epsilon' X(\hat{\beta}-\beta)] = \sigma p - \frac{\sigma^2 k}{n\beta' S\beta}\left[\frac{\gamma_1}{n}e' X\beta + \sigma(p-2)\right] \quad (A.6)$$

$$E[(\hat{\beta}-\beta)' X' X(\hat{\beta}-\beta)] = \sigma^2 p - \frac{\sigma^3 k}{n\beta' S\beta}\left[\frac{2\gamma_1}{n}e' X\beta + 2\sigma(p-2) - \sigma k\right]. \quad (A.7)$$

Setting $\tilde{\beta} = \hat{\beta}$ in (A.1), utilizing the above results and retaining terms to order $O(n^{-1})$, we find

$$\begin{aligned}R(\hat{\beta}) &= E[WL(\hat{\beta})] \\ &= \sigma^2\lambda_1 n - \sigma^2 p(\lambda_1-\lambda_2) - \frac{\sigma^4 k}{n\beta' S\beta}\left[(1-\lambda_1+\lambda_2)\left(\frac{\gamma_1}{n\sigma}e' X\beta + p - 2 - k\right)\right]\end{aligned}$$

which is the result (3.5) of the Theorem.

Acknowledgement

The first author gratefully acknowledges the support from Alexander von Humboldt Foundation, Germany in the form of Humboldt Fellowship.

References

Giles, J.A., Giles, D.E.A. and Ohtani K. (1996): 'The exact risks of some pre-test and Stein-type regression estimates under balanced loss', Communications in Statistics - Theory and Methods, 25, p. 2901-2924.

Judge, G.G. and Bock, M.E. (1978): '**The Statistical Implication of Pre-Test and Stein-Rule Estimators in Econometrics**', North Holland.

Ohtani K. (1998): 'The exact risk of a weighted average estimator of the OLS and Stein-rule estimators in regression under balanced loss', Statistics and Decisions, 16, p. 35-45.

Saleh, A.K.E. (2006): '**Theory of Preliminary Test and Stein-type Estimation with Application**', John Wiley.

Shalabh (1995): 'Performance of Stein-rule procedure for simultaneous prediction of actual and average values of study variable in linear regression model', Jan Tinbergen Award Paper, Bulletin of the 50th Session of the International Statistical Institute, p. 1375-1390.

Shalabh (1999): 'Improving the predictions in linear regression models', Journal of Statistical Research, 33, p. 33-39.

Shalabh (2000): 'Prediction of values of variables in linear measurement error model', Journal of Applied Statistics, 27, p. 475-482.

Srivastava, A.K. and Shalabh (1996): 'Efficiency properties of least squares and Stein-rule predictions in linear regression model', Journal of Applied Statistical Science, 4, p. 141-145.

Toutenburg, H. and Shalabh (1996): 'Predictive performance of the methods of restricted and mixed regression estimators', Biometrical Journal, 38, p. 951-959.

Vinod, H.D. and Srivastava, V.K. (1995): 'Large sample asymptotic properties of the double k-class estimators in linear regression models', Econometric Reviews, 14, p. 75-100.