Bayesian Estimation of Regression Coefficients Under Extended Balanced Loss Function

ANOOP CHATURVEDI\textsuperscript{1} AND SHALABH\textsuperscript{2}

\textsuperscript{1}Department of Statistics, Allahabad University, Allahabad, India
\textsuperscript{2}Department of Mathematics & Statistics, Indian Institute of Technology, Kanpur, India

Appreciating the desirability of simultaneously using both the criteria of goodness of fitted model and clustering of estimates around true parameter values, an extended version of the balanced loss function is presented and the Bayesian estimation of regression coefficients is discussed. The thus obtained optimal estimator is then compared with the least squares estimator and posterior mean vector with respect to the criteria like posterior expected loss, Bayes risk, bias vector, mean squared error matrix and risk function.

**Keywords** Balanced loss function; Bayesian estimation; Bayes risk; Linear regression model; Posterior expected loss; Predictive loss function.

**Mathematics Subject Classification** 62J05.

1. Introduction

The estimation of parameters and prediction of values of variable are inseparable parts of any regression modeling. For the estimation of parameters in a linear regression model, the efficiency of any estimation procedure is generally examined under a loss function based on either the goodness-of-fit model or the clustering of estimates around the true parameter values. The goodness of an estimator depends on its variability around its mean as well as how good it is in predicting the new values. Generally, the criteria of fitting of model and its predictive capability are separately considered to judge the performance of estimator. In fact, both the criteria should be considered together in many practical situations in judging the performance of any estimator. For example, consider the use and calibration of a thermometer for reading the temperature of human body. The thermometer has some markings on it which read out the temperature of the body when inserted inside the mouth. Such readings are based on the height achieved by the expansion of mercury inside the glass body. While constructing the thermometer, some of these markings are created
based on some known temperature and other markings are created using the statistical
calibration technique. Now a person with unknown temperature inserts the thermometer
inside his mouth and the mercury rises to certain height. The main issue is now to predict
the corresponding value of the temperature. Note that the calibration part involved the
use of only a good estimator in the sense of having small variability so that the markings
on the thermometer do not make very large variation in determining the correct value of
temperature. The other part of the use of thermometer involves the prediction of value
which is again dependent on the estimated parameter. So, in a nut shell, the goodness of
thermometer will depend on the success of estimating the parameter which is related to the
variability around the true value of the parameter as well as using it for prediction. So both
the criteria, viz, the goodness-of-fit model and the clustering of estimates are important in
such a case. Zellner (1994) advocated that situations may occur where both the criteria need
to be incorporated simultaneously in the formulation of loss function. Accordingly, he has
propounded the balanced loss function employing the quadratic structure and has discussed
the Bayesian as well as non Bayesian estimation of parameters like the population means
and regression coefficients. His pioneering work has motivated a number of applications of
balanced loss functions; see, e.g., Giles et al. (1996), Ohtani (1998), Rodrigues and Zellner
(1994), and Wan (1994).

In this article, we present an extended version of the balanced loss function and discuss
the Bayesian estimation of regression coefficients. Section 2 develops an extended version
of balanced loss function providing its motivation. Its application to the estimation of re-
gression coefficients in a linear regression model under a Bayesian framework is considered
in Sec. 3 and the optimal estimator is deduced. Section 4 analyzes the posterior expected
losses and Bayes risk associated with the Bayesian estimator, least squares estimator and
posterior mean. In Sec. 5, we compare the bias vectors, dispersion matrices, and risk
functions in a non-Bayesian framework. Finally, some concluding remarks are placed in
Sec. 6.

2. An Extended Balanced Loss Function

Let us postulate the following linear regression model:

\[ y = X\beta + u, \]  

(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, \( \beta \) is a \( p \times 1 \) vector of regression
coefficients and \( u \) is a \( n \times 1 \) vector with mean vector 0 and variance covariance matrix \( \sigma^2 I \).

If \( \hat{\beta} \) is any estimator of \( \beta \), the goodness-of-fit model can be measured by the residual
vector \( (X\hat{\beta} - y) \) while the closeness of estimates to the true parameter values can be
measured by \( (\hat{\beta} - \beta) \) or \( X(\hat{\beta} - \beta) \). Employing the framework of quadratic loss structure,
Zellner (1994) proposed the following balanced loss function:

\[ BL(\hat{\beta}) = w(X\hat{\beta} - y)'(X\hat{\beta} - y) + (1 - w)(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta), \]  

(2.2)

where \( w \) is a non stochastic scalar lying between 0 and 1 reflecting the weightage being
given to the goodness-of-fit model in comparison to the precision of estimation.

Given the estimator \( \hat{\beta} \), it is natural to obtain the predictions for the values of study
variable within the sample from \( X\hat{\beta} \). If these predictions are used for the actual values
\( y \) of the study variable, the quality of predictions is judged by \( (X\hat{\beta} - y) \). On the other
hand, if these predictions are used for the average values \( E(y) = X\beta \) of the study variable,
the quality of predictions is judged by \((X\tilde{\beta} - X\beta)\). Thus, the balanced loss function (2.2) measures the quality of predictions when they are used simultaneously for the actual values and average values of the study variable within the sample; see Zellner (1994).

The problem of simultaneous prediction of actual and average values of the study variable is considered by Shalabh (1995). He formulated a target function

\[ T = \lambda y + (1 - \lambda)E(y), \quad (2.3) \]

where \(0 \leq \lambda \leq 1\) as a non stochastic scalar depicting the weight being assigned to the prediction of actual values in comparison to the prediction of average values, and considered the following predictive loss function:

\[ PL(\hat{\beta}) = (X\tilde{\beta} - T)'(X\tilde{\beta} - T) \quad (2.4) \]

\[ = \lambda^2 (X\tilde{\beta} - y)'(X\tilde{\beta} - y) + (1 - \lambda)^2 (\beta - \beta)'X'X(\beta - \beta) \]

\[ + 2\lambda(1 - \lambda)(X\tilde{\beta} - y)'X(\tilde{\beta} - \beta). \]

If we define a vector

\[ d = \begin{pmatrix} X\tilde{\beta} - y \\ X\tilde{\beta} - X\beta \end{pmatrix} \quad (2.5) \]

then it can be regarded as a vector quantity reflecting the quality of predictions or measuring the goodness-of-fit model and closeness of estimates to true parameter values. Based on it, let us consider the following weighted loss function:

\[ WL(\tilde{\beta}; Q) = d'Qd, \quad (2.6) \]

where \(Q\) is any \(2n \times 2n\) matrix with non stochastic elements.

This formulation implies an interesting class of loss functions by substituting \(Q = (A \otimes I_p)\) with \(A\) any \(2 \times 2\) matrix. This yields the following loss function

\[ WL(\tilde{\beta}; A) = d'(A \otimes I_p) d. \quad (2.7) \]

If we choose

\[ A = \begin{pmatrix} w & 0 \\ 0 & 1 - w \end{pmatrix} \quad (2.8) \]

we get the balanced loss function proposed by Zellner (1994).

Similarly, if we set

\[ A = \begin{pmatrix} \lambda \\ 1 - \lambda \end{pmatrix} \begin{pmatrix} \lambda & 1 - \lambda \end{pmatrix} \quad (2.9) \]

we obtain the predictive loss function considered by Shalabh (1995).

If \(\alpha_1\) and \(\alpha_2\) are the diagonal elements of \(A\) and \((1 - \alpha_1 - \alpha_2)/2\) is the off-diagonal element along with the constraint that \(\alpha_1\) and \(\alpha_2\) lie between 0 and 1, the loss function (2.7) reduces to the following:

\[ L(\tilde{\beta}) = \alpha_1 (X\tilde{\beta} - y)'(X\tilde{\beta} - y) + \alpha_2 (\beta - \beta)'X'X(\beta - \beta) \]

\[ + (1 - \alpha_1 - \alpha_2)(X\tilde{\beta} - y)'X(\tilde{\beta} - \beta) \quad (2.10) \]
which can be regarded as a simple extension of the loss functions (2.2) and (2.4), and can be termed as extended balanced loss function in the honor of Zellner. The main advantage of this extended balanced loss function is that it is more flexible in comparison to (2.2) and (2.4). The balanced loss function in (2.2) takes care only either the precision of estimation or the goodness of fit whereas (2.4) extends it by considering the interaction or covariation between the precision of estimation and goodness of fit. The weights assigned in (2.4) for precision of estimation, goodness of fit and their covariation depends only on one factor \( \lambda \). The balanced loss function in (2.10) gives more freedom and one can choose the weights for precision of estimation as well as goodness of fit. Referring to the example cited in Section 1, one can expect that the two aspects of precision of estimation and goodness of fit are not independent in the sense that they are simply different functions of \( y, X \) and \( \beta \). So their covariation may exist in data and ignoring it in the formulation of loss function may lead to wrong statistical inferences. The loss function (2.4) takes care of such a covariation which is between the precision of estimation and the goodness of fit. In this sense, this provides an improved alternative to (2.2) but it has limitation of choosing the weight only for one of the criterion. The proposed extended balanced loss function (2.10) provides a more flexibility option of choosing different weights for precision of estimation, goodness of fit and covariation terms and, in the presence of covariation, will obviously lead to improved statistical inferences.

3. Bayesian Estimation of Regression Coefficients

If \( \hat{\beta} \) is any estimator of \( \beta \), the posterior expected loss under (2.10) is given by

\[
E_{\beta}[L(\hat{\beta})] = \alpha_1(X\hat{\beta} - y)'(X\hat{\beta} - y) + \alpha_2E_{\beta}[(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)] + (1 - \alpha_1 - \alpha_2)E_{\beta}[(X\hat{\beta} - y)'X(\hat{\beta} - \beta)]
\]

\[
= \alpha_1(X\hat{\beta} - y)'(X\hat{\beta} - y) + \alpha_2(\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta}) + \alpha_2E_{\beta}[(\beta - \tilde{\beta})'X'X(\beta - \tilde{\beta})] + (1 - \alpha_1 - \alpha_2)(X\hat{\beta} - y)'X(\tilde{\beta} - \beta),
\]

where \( \tilde{\beta} \) denotes the posterior mean vector.

If we write

\[
b = (X'X)^{-1}X'y
\]

\[
\hat{\beta}^* = \frac{1}{2}[(1 + \alpha_1 - \alpha_2)b + (1 - \alpha_1 + \alpha_2)\tilde{\beta}]
\]

it is easy to verify that

\[
E_{\beta}[L(\hat{\beta})] = (\tilde{\beta} - \hat{\beta}^*)'X'X(\tilde{\beta} - \hat{\beta}^* + [4\alpha_1\alpha_2 - (1 - \alpha_1 - \alpha_2)^2](b - \tilde{\beta})'X'X(b - \tilde{\beta}) + \alpha_1(Xb - y)'(Xb - y) + \alpha_2E_{\beta}[(\beta - \tilde{\beta})'X'X(\beta - \tilde{\beta})].
\]

Minimizing it with respect to \( \hat{\beta} \), we find that the Bayesian estimator conditional on the sample and prior information is \( \hat{\beta}^* \).

It is interesting to observe that the Bayesian estimator \( \hat{\beta}^* \) reduces to the least squares estimator in case \( (\alpha_1 - \alpha_2) = 1 \) while it is equal to the posterior mean vector \( \hat{\beta} \) in case of \( (\alpha_1 - \alpha_2) = -1 \) or when \( \hat{\beta} = b \). For all other values of \( (\alpha_1 - \alpha_2) \) between -1 and 1, the estimator \( \hat{\beta}^* \) is a weighted combination of \( b \) and \( \tilde{\beta} \).
Observing that
\[
\tilde{\beta}^* = b - \frac{1}{2}(1 - \alpha_1 + \alpha_2)(b - \bar{\beta})
\]
(3.4)
\[
= \bar{\beta} - \frac{1}{2}(1 + \alpha_1 - \alpha_2)(\bar{\beta} - b)
\]
the Bayesian estimator can be interpreted as a shrinkage estimator, shrinking \(b\) or \(\bar{\beta}\).

Now let us be given a prior distribution of \(\bar{\beta}\), conditional on \(\sigma^2\), such that it has mean vector \(\beta_0\) and variance covariance matrix \(\sigma^2 Q\).

When, for \(\beta\), a \(g\)-prior distribution \(N(\beta_0, \sigma^2(gX'X)^{-1})\) is utilized, it is observed in Zellner (1986) that the posterior mean vector is expressible as a scalar weighted combination of the least squares estimator \(b\) and the prior mean vector \(\beta_0\) as follows:
\[
\tilde{\beta} = \theta b + (1 - \theta)\beta_0 \quad ; \quad 0 < \theta < 1
\]
(3.5)
so that
\[
\tilde{\beta}^* = hb + (1 - h)\beta_0
\]
(3.6)
implying that the Bayesian estimator is a scalar weighted combination of \(b\) and \(\beta_0\) where
\[
h = \frac{1}{2}[1 + \theta + (\alpha_1 - \alpha_2)(1 - \theta)], \quad \theta = \frac{g}{1 + g}.
\]
(3.7)

It is interesting to note that if we use the \(g\)-prior with dispersion matrix \(\sigma^2(g^*X'X)^{-1}\) with
\[
g^* = \frac{1 + 2g + \alpha_1 - \alpha_2}{1 - \alpha_1 + \alpha_2}
\]
then \(\tilde{\beta}^*\) is the prior mean, i.e., Bayes estimator under the quadratic loss. This leads to existence of duality between the specification of loss function and specification of prior distribution in the sense that there exists a prior distribution such that prior mean equals to minimum expected loss estimator. Hence, it is possible to compare \(\tilde{\beta}^*\) and the estimators \(\tilde{\beta}\) and \(b\) by comparing two posterior means based on different prior distributions.

Similarly, if we assume a natural conjugate prior for \(\beta\) so that
\[
\tilde{\beta} = (X'X + Q^{-1})^{-1}(X'Xb + Q^{-1}\beta_0)
\]
(3.8)
we have
\[
\tilde{\beta}^* = Hb + (I - H)\beta_0
\]
(3.9)
which is a matrix weighted average of \(b\) and \(\beta_0\) with
\[
H = I - \frac{1}{2}(1 - \alpha_1 + \alpha_2)(I + QX'X)^{-1}.
\]
(3.10)

Comparing (3.6) with (3.5) and (3.9) with (3.8), it is seen that the Bayesian estimator has the same form as the posterior mean vector.
Similar observations can be made if we consider the problem of predicting the values of study variable within the sample. Then the Bayesian predictions are specified by

$$X \tilde{\beta}^* = \frac{1}{2} \left[ (1 + \alpha_1 - \alpha_2) Xb + (1 - \alpha_1 + \alpha_2) X \bar{\beta} \right],$$

(3.11)

where $Xb$ and $X \bar{\beta}$ denote the predictions arising from least squares and posterior distribution respectively.

If we employ the balanced loss function (2.2), we observe that

$$X \tilde{\beta}^* = w Xb + (1 - w) X \bar{\beta}$$

(3.12)

while if we use the predictive loss function (2.4), we have

$$X \tilde{\beta}^* = \lambda Xb + (1 - \lambda) X \bar{\beta}$$

(3.13)

which has got the same form as the target function (2.3).

4. The Posterior Expected Loss And Bayes Risk

It is easy to see from (3.3) that the posterior expected loss function associated with the Bayesian estimator $\tilde{\beta}^*$ of $\beta$ is given by

$$E_\beta[L(\tilde{\beta}^*)] = [4\alpha_1\alpha_2 - (1 - \alpha_1 - \alpha_2)^2](b - \bar{\beta})'X'X(b - \bar{\beta})$$

$$+ \alpha_1(Xb - y)'(Xb - y) + \alpha_2 E_\beta[(\beta - \bar{\beta})'X'X(\beta - \bar{\beta})].$$

(4.1)

Thus, the posterior expected loss is influenced by the difference between the least squares estimator and posterior mean, the residual sum of squares based on least squares method, the posterior variability of $\beta$ and the scalars characterizing the extended balanced loss function. This matches the observation made by Zellner (1994, Sec. 3).

Treating the two components $b$ and $\bar{\beta}$ of $\tilde{\beta}^*$ as estimators of $\beta$, we observe that the differences in the posterior expected losses are given by

$$D(b; \tilde{\beta}^*) = E_\beta[L(b) - L(\tilde{\beta}^*)]$$

$$= \frac{1}{4}(1 - \alpha_1 + \alpha_2)^2(b - \bar{\beta})'X'X(b - \bar{\beta})$$

(4.2)

$$D(\bar{\beta}; \tilde{\beta}^*) = E_\beta[L(\bar{\beta}) - L(\tilde{\beta}^*)]$$

$$= \frac{1}{4}(1 + \alpha_1 - \alpha_2)^2(b - \bar{\beta})'X'X(b - \bar{\beta})$$

(4.3)

when it follows that the increase in the posterior expected loss depends upon the characterizing scalars of loss functions and the difference between the least squares estimator and the posterior mean vector or equivalently the difference between predictions based on least squares and posterior mean.

It is seen from (4.2) and (4.3) that the magnitudes of the elements of the difference vector between $Xb$ and $X \bar{\beta}$ play an important role. Substantial increase in the posterior expected loss may arise when $Xb$ and $X \bar{\beta}$ differ markedly.

When $\tilde{\beta}$ is expressible as the weighted combination of the least squares estimator $b$ and the prior mean vector $\beta_0$, then the crucial role is determined by the nearness of $b$ to $\beta_0$.
Bayesian Regression Coefficients Balanced Loss 4259

As

\[(b - \bar{\beta})'X'(b - \bar{\beta}) = \begin{cases} 
(1 - \theta)(b - \beta_0)'X'X(b - \beta_0) & \text{for } \bar{\beta} \text{ given by (3.5)} \\
(b - \beta_0)'(I + X'Q)^{-1}X'X(I + QX'X)^{-1}(b - \beta_0) & \text{for } \bar{\beta} \text{ given by (3.8)}
\end{cases} \] (4.4)

Next, let us consider the Bayes risk under the extended loss function. For this purpose, following Zellner (1994), we assume that the disturbances are normally distributed and the prior distribution of \(\beta\) given \(\sigma^2\) is specified in such a manner that the posterior mean vector \(\bar{\beta}\) has the form (3.5). In particular, let us assume it to be a \(g\)-prior density introduced by Zellner (1986). That is, the prior distribution of \(\beta\) given \(\sigma^2\) is multivariate normal with mean vector \(\beta_0\) and variance covariance matrix \((\sigma^2/g)(X'X)^{-1}\) where \(g\) is a positive known number. In this case, the posterior mean vector is given by

\[\bar{\beta} = \left(\frac{1}{1+g}\right)(b + g\beta_0)\] (4.5)

and therefore the Bayesian estimator is

\[\tilde{\beta}^* = \frac{1}{2(1+g)}[2 + g(1 + \alpha_1 - \alpha_2)]b + g(1 - \alpha_1 + \alpha_2)\beta_0].\] (4.6)

Using these expressions along with (4.2) and (4.3), it is easy to obtain the differences in Bayes risks under the extended balanced loss function:

\[D(b; \tilde{\beta}^*) = BR(b) - BR(\tilde{\beta}^*) = \frac{(1 - \alpha_1 + \alpha_2)g\sigma^2}{4(1+g)^2}\] (4.7)

\[D(\bar{\beta}; \tilde{\beta}^*) = BR(\bar{\beta}) - BR(\tilde{\beta}^*) = \frac{(1 + \alpha_1 - \alpha_2)g\sigma^2}{4(1+g)^2}\] (4.8)

which are clearly non-negative implying the superiority of \(\tilde{\beta}^*\) over \(b\) and \(\bar{\beta}\).

In order to examine the role of characterizing scalars \(\alpha_1\) and \(\alpha_2\) on the magnitude of increase in Bayes risk, let us consider the quantities:

\[f_b = (1 - \alpha_1 + \alpha_2)^2\] (4.9)

\[f_{\beta} = (1 + \alpha_1 - \alpha_2)^2.\] (4.10)

It is easy to see that \(f_b = 4(1 - \lambda)\) and \(f_{\beta} = 4\lambda^2\) for the balanced loss function (2.2) while \(f_b = 4(1 - \lambda)^2\) and \(f_{\beta} = 4\lambda^2\) for the predictive loss function (2.4). In Table 1, we have presented the values of (4.9) and (4.10) for some selected values of \(\alpha_1\) and \(\alpha_2\).

5. Performance Properties

Let us compare the performance properties of three estimators, viz., least squares estimator of \(b\), posterior mean vector \(\bar{\beta}\), and Bayesian estimator \(\tilde{\beta}^*\) under a non Bayesian framework.

For the sake of simplicity in exposition, it is assumed that the prior distribution of regression parameters is such that \(\bar{\beta}\) and \(\tilde{\beta}^*\) are given by (3.5) and (3.6), respectively.
Table 1

Values of $f_b$ and $f_{\bar{\beta}}$ throwing light on the increase in posterior expected loss

<table>
<thead>
<tr>
<th>$\alpha_2$</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.5</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>Values of $f_b$</td>
<td>Values of $f_{\bar{\beta}}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>1.00</td>
<td>1.44</td>
<td>1.96</td>
<td>2.56</td>
<td>3.24</td>
<td>4.00</td>
</tr>
<tr>
<td>0.2</td>
<td>0.64</td>
<td>1.00</td>
<td>1.44</td>
<td>1.96</td>
<td>2.56</td>
<td>3.24</td>
</tr>
<tr>
<td>0.4</td>
<td>0.36</td>
<td>0.64</td>
<td>1.00</td>
<td>1.44</td>
<td>1.96</td>
<td>2.56</td>
</tr>
<tr>
<td>0.6</td>
<td>0.16</td>
<td>0.36</td>
<td>0.64</td>
<td>1.00</td>
<td>1.44</td>
<td>1.96</td>
</tr>
<tr>
<td>0.8</td>
<td>0.04</td>
<td>0.16</td>
<td>0.36</td>
<td>0.64</td>
<td>1.00</td>
<td>1.44</td>
</tr>
<tr>
<td>1.0</td>
<td>0.00</td>
<td>0.04</td>
<td>0.16</td>
<td>0.36</td>
<td>0.64</td>
<td>1.00</td>
</tr>
</tbody>
</table>

5.1. The Bias Vector

It is easy to see from (3.5) and (3.6) that $b$ is unbiased while $\bar{\beta}$ and $\tilde{\beta}^*$ are not. Further, we have

$$E(\bar{\beta} - \beta) = (1 - \theta)(\beta_0 - \beta) \quad (5.1)$$
$$E(\tilde{\beta}^* - \beta) = (1 - h)(\beta_0 - \beta) \quad (5.2)$$

$$= \frac{1}{2}(1 - \alpha_1 + \alpha_2)(1 - \theta)(\beta_0 - \beta)$$

from which it follows that the elements of bias vector have the same sign as the corresponding elements of the difference vector $(\beta_0 - \beta)$. So far as the magnitude of bias is concerned, the Bayesian estimator $\tilde{\beta}^*$ is superior to $\bar{\beta}$.

5.2. The Dispersion Matrix

First, we state the following results which are used later.

If $a$ is a $p \times 1$ vector and $A$ is a $p \times p$ positive definite matrix, we have the following results.

**Results I.** The matrix $(A - aa')$ is positive definite if and only if $a'A^{-1}a$ is less than 1; see Farebrother (1976).

**Results II.** The matrix $(aa' - A)$ is never nonnegative definite for $p$ greater than 1; see Guilky and Price (1981).
The variance covariance matrix of $b$ is given by

$$V(b) = \sigma^2(X'X)^{-1}$$  \hspace{1cm} (5.3)

while the mean squared error matrices of $\bar{\beta}$ and $\tilde{\beta}^*$ are

$$M(\bar{\beta}) = \theta^2 \sigma^2(X'X)^{-1} + (1 - \theta)^2(\beta_0 - \beta)(\beta_0 - \beta)'$$  \hspace{1cm} (5.4)

$$M(\tilde{\beta}^*) = h^2 \sigma^2(X'X)^{-1} + (1 - h)^2(\beta_0 - \beta)(\beta_0 - \beta)'.$$  \hspace{1cm} (5.5)

Comparing (5.3) and (5.4), we observe that

$$V(b) - M(\bar{\beta}) = (1 - \theta)[(1 + \theta)^2(X'X)^{-1} - (1 - \theta)(\beta_0 - \beta)(\beta_0 - \beta)']$$  \hspace{1cm} (5.6)

which is positive definite if and only if

$$q = \frac{1}{\sigma^2}(\beta_0 - \beta)'X'X(\beta_0 - \beta) < \frac{(1 + \theta)}{(1 - \theta)}$$  \hspace{1cm} (5.7)

where use has been made of the result I.

Similarly, it is observed from (5.3) and (5.6) that the variance covariance matrix of $b$ exceeds the mean squared error matrix of $\tilde{\beta}^*$ by a positive definite matrix if and only if

$$q < \frac{(1 + h)}{(1 - h)} = \frac{3 + \theta + (\alpha_1 - \alpha_2)(1 - \theta)}{(1 + \alpha_1 - \alpha_2)(1 - \theta)}$$  \hspace{1cm} (5.8)

provided that $(1 + \alpha_1 - \alpha_2)$ is not zero.

As

$$\frac{(1 + h)}{(1 - h)} > \frac{(1 + \theta)}{(1 - \theta)}$$  \hspace{1cm} (5.9)

it is observed from (5.7) and (5.8) that superiority of $\bar{\beta}$ over $b$ with respect to the criterion of mean squared error matrix implies the superiority of $\tilde{\beta}^*$ over $b$ while the converse may not be necessarily true. In other words, the estimator $\tilde{\beta}^*$ dominates $b$ over a wider range of parametric space in comparison to the range over which the estimator $\bar{\beta}$ dominates $b$.

Next, let us consider the differences

$$M(\bar{\beta}) - V(b) = (1 - \theta)[(1 + \theta)(\beta_0 - \beta)(\beta_0 - \beta)']$$

$$- (1 + \theta)\sigma^2(X'X)^{-1}$$

$$M(\tilde{\beta}^*) - V(b) = (1 - h)[(1 - h)(\beta_0 - \beta)(\beta_0 - \beta)']$$

$$- (1 + h)\sigma^2(X'X)^{-1}.$$  \hspace{1cm} (5.11)

Applying the Result II, it is seen that $b$ cannot dominate the estimators $\bar{\beta}$ and $\tilde{\beta}^*$ with respect to the criterion of mean squared error matrix for $p$ exceeding 1.

Finally, let us compare $\bar{\beta}$ and $\tilde{\beta}^*$:

$$M(\bar{\beta}) - M(\tilde{\beta}^*) = (\theta - h)[(\theta + h)\sigma^2(X'X)^{-1}$$

$$- (2 - \theta - h)(\beta_0 - \beta)(\beta_0 - \beta)']$$

$$= \frac{1}{4}(1 - \theta)^2(1 + \alpha_1 - \alpha_2)(3 - \alpha_1 + \alpha_2)$$  \hspace{1cm} (5.12)
which cannot be non negative definite for \( p > 1 \) by virtue of the Result II. This implies that \( \tilde{\beta}^{\ast} \) does not uniformly dominate \( \bar{\beta} \).

Similarly, using the Result I, we find that \( \bar{\beta} \) dominates \( \tilde{\beta}^{\ast} \) if and only if

\[
q < \frac{1}{4} (1 - \theta)(3 - \alpha_1 - \alpha_2).
\]

5.3. The Risk Function

Using the extended balanced loss function (2.4), it is easy to see that the risk functions associated with the three estimators are given by

\[
R(b) = E[L(b)] = \sigma^2[(n - p)\alpha_1 + p\alpha_2] (5.14)
\]

\[
R(\bar{\beta}) = E[L(\bar{\beta})] = \sigma^2[n\alpha_1 + \theta p(\theta - 1 - \alpha_1 + \alpha_2) + (1 - \theta)^2 q] (5.15)
\]

\[
R(\tilde{\beta}^{\ast}) = E[L(\tilde{\beta}^{\ast})] = \sigma^2[n\alpha_1 + hp(h - 1 - \alpha_1 + \alpha_2) + (1 - h)^2 q]. (5.16)
\]

It is thus seen that \( \bar{\beta} \) has smaller risk than \( b \) when

\[
q < \frac{p(\theta - \alpha_1 + \alpha_2)}{(1 - \theta)}. (5.17)
\]

Similarly, \( \tilde{\beta}^{\ast} \) has smaller risk than \( b \) when

\[
q < \frac{p(h - \alpha_1 + \alpha_2)}{(1 - h)} = \frac{p(1 + \theta)}{(1 - \theta)}. (5.18)
\]

Comparing (5.15) and (5.16), we observe that

\[
R(\bar{\beta}) - R(\tilde{\beta}^{\ast}) = \sigma^2(\theta - h)[p(h + \theta - 1 - \alpha_1 + \alpha_2) - q(2 - h - \theta)] (5.19)
\]

\[
= \sigma^2 \frac{4}{4} (1 - \theta)^2 (1 + \alpha_1 - \alpha_2)(3 - \alpha_1 + \alpha_2)
\]

\[
\left[ q - p \left( \frac{4}{(1 - \theta)(3 - \alpha_1 + \alpha_2)} - 1 \right) \right].
\]

Thus, \( \tilde{\beta}^{\ast} \) has smaller risk than \( \bar{\beta} \) when

\[
q < p \left( \frac{4}{(1 - \theta)(3 - \alpha_1 + \alpha_2)} - 1 \right) (5.20)
\]

while the opposite is true, i.e., \( \bar{\beta} \) is better than \( \tilde{\beta}^{\ast} \) when the condition (5.20) with the reverse inequality sign holds true.
From (5.18) and (5.20), we observe that if the prior distribution is properly specified and $\beta_0$ is close to $\beta$ or the prior precision $g$ is large, the parametric range in which $\tilde{\beta}^*$ dominates both $b$ and $\hat{\beta}$ also increases. Further, it can be easily verified that, $\tilde{\beta}^*$ dominates both $b$ and $\hat{\beta}$ at least for

$$\frac{(\beta_0 - \beta)X'X(\beta_0 - \beta)}{p\sigma^2} < 1$$

irrespective of the values of $\alpha_1$ and $\alpha_2$ in the range 0 and 1.

6. Some Remarks

Appreciating the desirability of simultaneously using both the criteria of fitted model and clustering of estimates around true parameter values for judging the quality of any estimator for the coefficients in a linear regression model, we presented a loss function which can be regarded as an extended version of the balanced loss function introduced by Zellner (1994). Utilizing a Bayesian framework, we derived the optimal Bayesian estimator for the regression coefficient vector. This optimal estimator turns out to be a weighted combination of the least squares estimator and the posterior mean vector.

Comparing the optimal estimator with its two components, viz., the least squares estimator and posterior mean vector, it is found that the optimal estimator is uniformly better than the two estimators with respect to the criterion of posterior expected loss and Bayes risk.

Next, it is observed that the least squares estimator is unbiased while the posterior mean vector and the optimal estimator are biased. Further, the optimal estimator is found to have smaller bias in magnitude in comparison to the posterior mean.

Taking the performance criterion as mean squared error matrix, no uniform dominance of one estimator over the other is observed. However, the two biased estimators (posterior mean and optimal estimator) dominate the unbiased estimator under certain conditions specified by (5.7) and (5.8). Comparing the biased estimators, we have observed that the optimal estimator does not dominate the posterior mean as long as there are two or more explanatory variables. On the other hand, the posterior mean is found to be better than the optimal estimator under the condition (5.13).

If the performance criterion is the risk function under the extended loss function, it is again seen that no estimator dominates uniformly the other estimator. The conditions under which one estimator is better than the other are then deduced for each pair of estimators.

It may be remarked that the proposed loss function can be generalized in several directions. For instance, one may consider more general quadratic structures for measuring the goodness-of-fit model and the clustering of estimates around true parameter values; see Shinozaki (1980) and Zellner (1994). Another interesting direction is to consider other measures for the goodness-of-fit model and the clustering of estimates such as Pitman nearness. Yet another direction is to use matrix measures in place of scalar measures for the goodness-of model and the clustering of estimates.

References


