



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Toutenburg, Shalabh:

Prediction of Response Values in Linear Regression Models from Replicated Experiments

Sonderforschungsbereich 386, Paper 112 (1998)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



Prediction of Response Values in Linear Regression Models from Replicated Experiments

H. Toutenburg
Institut für Statistik, Universität München
80799 München, Germany

Shalabh
Department of Statistics
Jammu University
Jammu 180004, India

June 2, 1998

Abstract: This paper considers the problem of prediction in a linear regression model when data sets are available from replicated experiments. Pooling the data sets for the estimation of regression parameters, we present three predictors—one arising from the least squares method and two stemming from Stein-rule method. Efficiency properties of these predictors are discussed when they are used to predict actual and average values of response variable within/outside the sample.

Key words: least squares estimator, prediction, Stein-type estimator

1 Introduction

Linear regression analysis provides an interesting framework for relating the responses to a set of explanatory variables particularly in planned experimentation. When observations are available from replicated experiments performed under the same protocol, the pooled data set provides a typical blend of information contained in individual data sets. It is then desirable to use combined evidence for deducing inferences about model parameters. Accordingly, Srivastava and Toutenburg (1994) and Rao, Srivastava and Toutenburg (1997) have considered the estimation of regression coefficients and have analyzed the properties of one unbiased and two biased estimators with respect to criteria of bias, mean squared error and Pitman nearness. This article studies the problem of prediction and analyzes the performance of some predictors for the actual and average values of response variable.

The organization of our presentation is as follows. Section 2 describes the model and presents three estimators for the vector of regression coefficients. These are then utilized for the prediction of values of response variable outside the sample in Section 3 and within the sample in Section 4. In both the cases, a

comparative study of the performances of predictors is reported. Some concluding remarks are then placed in Section 5. Finally, proof of a result is outlined in the Appendix.

2 Model and the Estimators of Regression Coefficients

Following Srivastava and Toutenburg (1994), we postulate the following framework for modelling the data obtained from two similar experiments under the same protocol:

$$\mathbf{y}_1 = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_1 \quad (2.1)$$

$$\mathbf{y}_2 = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_2 \quad (2.2)$$

where \mathbf{y}_1 and \mathbf{y}_2 are $n \times 1$ vectors of responses in the two experiments, \mathbf{X} is an $n \times p$ full column rank matrix of n values of p explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of coefficients associated with them and \mathbf{u}_1 and \mathbf{u}_2 are $n \times 1$ vectors of disturbances.

It is assumed that \mathbf{u}_1 and \mathbf{u}_2 follow multivariate normal distributions with same mean vector $\mathbf{0}$ but possibly different variance covariance matrices, viz., $\sigma^2\mathbf{I}_n$ and $\lambda\sigma^2\mathbf{I}_n$ where σ^2 and λ both are unknown. Further, we assume that \mathbf{u}_1 and \mathbf{u}_2 are stochastically independent.

Applying least squares method to (2.1) and (2.2) for the estimation of coefficient vector $\boldsymbol{\beta}$, we get the following two estimators:

$$\mathbf{b}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_1, \mathbf{b}_2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_2 \quad (2.3)$$

which lead to the combined estimator

$$\mathbf{b} = (1 - \alpha)\mathbf{b}_1 + \alpha\mathbf{b}_2 \quad (2.4)$$

where α is the combining (scalar) parameter lying between 0 and 1; see Srivastava and Toutenburg (1994).

Similarly, we can apply Stein estimation procedure to (2.1) and (2.2) in order to get two estimators of $\boldsymbol{\beta}$; see Judge and Bock (1978) for details. These estimators when combined as in (2.4) yield the following combined estimator:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_1 &= (1 - \alpha) \left[1 - \left(\frac{k}{n} \right) \frac{\mathbf{y}'_1 \mathbf{M} \mathbf{y}_1}{\mathbf{b}'_1 \mathbf{X}' \mathbf{X} \mathbf{b}_1} \right] \mathbf{b}_1 + \alpha \left[1 - \left(\frac{k}{n} \right) \frac{\mathbf{y}'_2 \mathbf{M} \mathbf{y}_2}{\mathbf{b}'_2 \mathbf{X}' \mathbf{X} \mathbf{b}_2} \right] \mathbf{b}_2 \\ &= \mathbf{b} - \left(\frac{k}{n} \right) \left[(1 - \alpha) \frac{\mathbf{y}'_1 \mathbf{M} \mathbf{y}_1}{\mathbf{b}'_1 \mathbf{X}' \mathbf{X} \mathbf{b}_1} \mathbf{b}_1 + \alpha \frac{\mathbf{y}'_2 \mathbf{M} \mathbf{y}_2}{\mathbf{b}'_2 \mathbf{X}' \mathbf{X} \mathbf{b}_2} \mathbf{b}_2 \right] \end{aligned} \quad (2.5)$$

where $\mathbf{M} = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$ and k is a nonstochastic positive scalar.

Alternatively, we can apply Stein procedure to the combined estimator \mathbf{b} so as to get another estimator of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_2 = \mathbf{b} - \left(\frac{k}{n} \right) \frac{\mathbf{y}'_1 \mathbf{M} \mathbf{y}_1 + \mathbf{y}'_2 \mathbf{M} \mathbf{y}_2}{\mathbf{b}' \mathbf{X}' \mathbf{X} \mathbf{b}} \mathbf{b}. \quad (2.6)$$

We shall employ these estimators for the formulation of predictors of response values.

3 Prediction of Responses outside the Sample

For the prediction of m values of the response variable corresponding to a set of given values of explanatory variables, let us assume that

$$\mathbf{y}_f = \mathbf{X}_f \boldsymbol{\beta} + \mathbf{u}_f \quad (3.1)$$

where \mathbf{y}_f denotes a column vector of m unobserved responses outside the sample such as future values, \mathbf{X}_f is an $m \times p$ matrix of m given values of p explanatory variables and \mathbf{u}_f is a $m \times 1$ vector of disturbances following a multivariate normal distribution with mean vector $\mathbf{0}$ and variance covariance matrix $\sigma_f^2 \mathbf{I}_m$. Further, \mathbf{u}_f is independent of \mathbf{u}_1 and \mathbf{u}_2 .

Employing the estimators (2.3), (2.4) and (2.5), we formulate the following three predictors:

$$\mathbf{F} = \mathbf{X}_f \mathbf{b}, \mathbf{F}_1 = \mathbf{X}_f \hat{\boldsymbol{\beta}}_1, \mathbf{F}_2 = \mathbf{X}_f \hat{\boldsymbol{\beta}}_2 \quad (3.2)$$

which can be utilized for predicting the actual values (\mathbf{y}_f) or average values ($\mathbf{X}_f \boldsymbol{\beta}$) of the response variable. We therefore define the following composite target function

$$\mathbf{T}_f = \phi \mathbf{y}_f + (1 - \phi) \mathbf{X}_f \boldsymbol{\beta} \quad (3.3)$$

where ϕ is a nonstochastic scalar lying between 0 and 1; see, e.g., Shalabh (1995).

Setting $\phi = 0$ and $\phi = 1$, we observe that \mathbf{T}_f reduces to the vector of average and actual response values, respectively. The composite target function thus provides us a kind of unified framework for handling the prediction of actual and average responses.

Let us now analyze the performance properties of the predictors (3.2) when they are used for \mathbf{T}_f .

It is easy to see that \mathbf{F} is weakly unbiased for \mathbf{T}_f in the sense that

$$\mathbf{E}(\mathbf{F} - \mathbf{T}_f) = \mathbf{0}. \quad (3.4)$$

Next we have

$$\mathbf{E}(\mathbf{F}_i - \mathbf{T}_f) = \mathbf{X}_f \mathbf{E}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}); i = 1, 2 \quad (3.5)$$

which is generally a non-null vector implying the biased nature of Stein-type predictors \mathbf{F}_1 and \mathbf{F}_2 . Further, we observe that bias remains unaltered whether the predictor is used for actual values or for average values. Such is, however, not the case with the dispersion or MSE-matrices given by

$$\mathbf{V}(\mathbf{F}) = \sigma_f^2 \phi^2 \mathbf{I}_m + \mathbf{X}_f \mathbf{E}(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})' \mathbf{X}_f' \quad (3.6)$$

$$\mathbf{M}(\mathbf{F}_i) = \sigma_f^2 \phi^2 \mathbf{I}_m + \mathbf{X}_f \mathbf{E}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta})' \mathbf{X}_f' \quad (i = 1, 2) \quad (3.7)$$

It is seen from the above expressions that all the three predictors exhibit larger variability when they are used for actual values ($\phi = 1$) in comparison to the case when they are used for average values ($\phi = 0$). In other words, they provide more efficient predictions for average values rather than actual values.

If we consider matrix differences like $[\mathbf{M}(\mathbf{F}_i) - \mathbf{V}(\mathbf{F})]$ and $[\mathbf{M}(\mathbf{F}_1) - \mathbf{M}(\mathbf{F}_2)]$, we observe that the gain/loss in efficiency of one predictor over the other remains unchanged whether they are employed for actual values or for average

values; see also Trenkler and Toutenburg (1992). The large sample asymptotic approximations for these matrix differences can be straightforwardly found from Srivastava and Toutenburg (1994). It is observed that there exist no conditions for the superiority of one predictor over the other except in the trivial case $p = 1$. Such is not the case if we change the performance criterion and take it as trace of mean squared error matrix (i.e., predictive mean squared error) instead of the matrix itself.

In order to study the relative performance of predictors with respect to the criterion of predictive mean squared error using large sample asymptotic theory, we assume the asymptotic cooperativeness of explanatory variables meaning thereby that the limiting form of $\mathbf{V} = n(\mathbf{X}'\mathbf{X})^{-1}$ as n tends to infinity is a nonsingular matrix with finite elements. Further, we introduce the following notation:

$$\left. \begin{aligned} f_1 &= (1 - \alpha) + \lambda\alpha, f_2 = 1 + \lambda \\ g &= (1 - \alpha)^2 + \lambda\alpha^2, \theta = \boldsymbol{\beta}'\mathbf{V}^{-1}\boldsymbol{\beta}. \end{aligned} \right\} \quad (3.8)$$

Using Srivastava and Toutenburg (1994), the difference in the predictive mean squared errors to order $O(n^{-2})$ is given by

$$\begin{aligned} \mathbf{D}(\mathbf{F}; \mathbf{F}_i) &= \mathbf{E}(\mathbf{F} - \mathbf{T}_f)'(\mathbf{F} - \mathbf{T}_f) - \mathbf{E}(\mathbf{F}_i - \mathbf{T}_f)'(\mathbf{F}_i - \mathbf{T}_f) \\ &= \frac{\sigma^4 f_i^2 \boldsymbol{\beta}' \mathbf{X}'_f \mathbf{X}_f \boldsymbol{\beta}}{n^2 \theta^2} (2d_i - k)k \end{aligned} \quad (3.9)$$

where

$$d_i = \frac{g}{f_i} \left(\frac{\theta \text{tr} \mathbf{V} \mathbf{X}'_f \mathbf{X}_f}{\boldsymbol{\beta}' \mathbf{X}'_f \mathbf{X}_f \boldsymbol{\beta}} - 2 \right); i = 1, 2. \quad (3.10)$$

The expression (3.9) is negative implying the superiority of \mathbf{F} over Stein-type predictor \mathbf{F}_i when

$$k > 2d_i; d_i > 0. \quad (3.11)$$

On the other hand, the predictor \mathbf{F}_i is superior to \mathbf{F} when

$$0 < k < 2d_i; d_i > 0. \quad (3.12)$$

As d_i involves unknown $\boldsymbol{\beta}$ and λ , the conditions (3.11) and (3.12) are hard to check in any given application. For this purpose, let us consider their sufficient versions.

If δ_* and δ^* denote the minimum and maximum characteristic roots of the matrix $(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}} \mathbf{X}'_f \mathbf{X}_f (\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}$ and S is the sum of all the characteristic roots, we have

$$\delta_* \leq \frac{\boldsymbol{\beta}' \mathbf{X}'_f \mathbf{X}_f \boldsymbol{\beta}}{\boldsymbol{\beta}' \mathbf{X}' \mathbf{X} \boldsymbol{\beta}} \leq \delta^*. \quad (3.13)$$

Similarly, if we write

$$\begin{aligned} \nu_{1*} &= \alpha, \nu_1^* = 1 - \alpha, \nu_{2*} = \alpha^2, \nu_2^* = (1 - \alpha)^2 & \text{if } \alpha \leq \frac{1}{2} \\ \nu_{1*} &= 1 - \alpha, \nu_1^* = \alpha, \nu_{2*} = (1 - \alpha)^2, \nu_2^* = \alpha^2 & \text{if } \alpha > \frac{1}{2} \end{aligned} \quad (3.14)$$

it is easy to see that

$$\nu_{i*} \leq \left(\frac{g}{f_i} \right) \leq \nu_i^*. \quad (3.15)$$

Using (3.13) and (3.15), we observe that

$$d_{i*} = \nu_{i*} \left(\frac{S}{\delta_*} - 2 \right) \leq d_i \leq d_i^* = \nu_i^* \left(\frac{S}{\delta_*} - 2 \right) \quad (3.16)$$

from which it follows that the inequality (3.11) holds true as long as

$$k > 2d_i^*; d_i^* > 0 \quad (3.17)$$

while the inequality (3.12) holds true at least so long as

$$0 < k < 2d_{i*}; d_{i*} > 0. \quad (3.18)$$

According to the criterion of predictive mean squared error to the given order of approximation, we thus observe that the predictor \mathbf{F} is not only unbiased but has smaller predictive variance in comparison to the predictive mean squared error of the biased predictor \mathbf{F}_i under the condition (3.17). The opposite is true, i.e., \mathbf{F}_i is superior to \mathbf{F} when the condition (3.18) is satisfied. Both the conditions (3.17) and (3.18), it may be noticed, are easy to check in practice.

Next, let us make a similar comparison of the two Stein-type predictors \mathbf{F}_1 and \mathbf{F}_2 .

If we take as criterion the mean squared error matrix up to order $O(n^{-2})$ and utilize Srivastava and Toutenburg (1994), no predictor is found to be superior to the other for p exceeding one. If we consider the trace of mean squared error matrix, we observe from (3.9) that

$$\begin{aligned} \mathbf{D}(\mathbf{F}_1; \mathbf{F}_2) &= \mathbf{E}(\mathbf{F}_1 - \mathbf{T}_f)'(\mathbf{F}_1 - \mathbf{T}_f) - \mathbf{E}(\mathbf{F}_2 - \mathbf{T}_f)'(\mathbf{F}_2 - \mathbf{T}_f) \\ &= \frac{\sigma^4(f_2^2 - f_1^2)\boldsymbol{\beta}'\mathbf{X}'_f\mathbf{X}_f\boldsymbol{\beta}}{n^2\theta^2}(2d - k)k \end{aligned} \quad (3.19)$$

where

$$d = \left(\frac{g}{f_1 + f_2} \right) \left(\frac{\theta \text{tr} \mathbf{V} \mathbf{X}'_f \mathbf{X}_f}{\boldsymbol{\beta}' \mathbf{X}'_f \mathbf{X}_f \boldsymbol{\beta}} - 2 \right). \quad (3.20)$$

It is thus seen from (3.19) that \mathbf{F}_1 is superior to \mathbf{F}_2 when

$$k > 2d; d > 0 \quad (3.21)$$

while the converse is true when

$$0 < k < 2d; d > 0. \quad (3.22)$$

Using (3.15) we find

$$\nu_* \leq \left(\frac{g}{f_1 + f_2} \right) \leq \nu^*. \quad (3.23)$$

where

$$\begin{aligned} \nu^* &= \begin{cases} \frac{(1-\alpha)^2}{2-\alpha} & \text{if } \alpha \leq \frac{1}{2} \\ \frac{\alpha^2}{1+\alpha} & \text{if } \alpha > \frac{1}{2} \end{cases} \\ \nu_* &= \begin{cases} \frac{\alpha^2}{1+\alpha} & \text{if } \alpha \leq \frac{1}{2} \\ \frac{(1-\alpha)^2}{2-\alpha} & \text{if } \alpha > \frac{1}{2}. \end{cases} \end{aligned} \quad (3.24)$$

From (3.13) and (3.23), we obtain

$$d_* = \nu_* \left(\frac{S}{\delta_*} - 2 \right) \leq d \leq d^* = \nu^* \left(\frac{S}{\delta_*} - 2 \right). \quad (3.25)$$

Utilizing it, we observe from (3.21) that the predictor \mathbf{F}_1 is superior to \mathbf{F}_2 as long as

$$k > 2d^*; d^* > 0. \quad (3.26)$$

Similarly, it follows from (3.22) that the reverse is true, i. e., \mathbf{F}_2 is superior to \mathbf{F}_1 at least so long as

$$0 < k < 2d_*; d_* > 0. \quad (3.27)$$

The conditions (3.26) and (3.27) are free from any unknown quantity and therefore can be easily verified in practice.

It may be remarked that the conditions (3.17), (3.18), (3.26) and (3.27) for the superiority of one predictor over the other require the bound (lower or upper as the case may be) of the characterizing scalar k to be positive. This constraint is fairly mild and will be tenable at least so long as the maximum characteristic root of the matrix $(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}\mathbf{X}'_f\mathbf{X}_f(\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}$ is less than half of the sum of all the roots.

4 Prediction of Responses within the Sample

Prediction of values of the response variable within the sample may shed some light on the suitability of fitted model. Let us therefore consider the prediction of responses in equation (2.1) without any loss of generality as a similar investigation for the other equation (2.2) can be easily carried out.

Defining the composite target function as

$$\mathbf{T} = \phi \mathbf{y}_1 + (1 - \phi) \mathbf{E}(\mathbf{y}_1) \quad (4.1)$$

in the spirit of (3.3) with ϕ as a nonstochastic scalar between 0 and 1, we consider the following three predictors:

$$\mathbf{P} = \mathbf{X}\mathbf{b}, \mathbf{P}_1 = \mathbf{X}\hat{\beta}_1, \mathbf{P}_2 = \mathbf{X}\hat{\beta}_2. \quad (4.2)$$

It is easy to see that

$$\mathbf{E}(\mathbf{P} - \mathbf{T}) = \mathbf{0} \quad (4.3)$$

so that \mathbf{P} is weakly unbiased for \mathbf{T} , and

$$\begin{aligned} \mathbf{V}(\mathbf{P}) &= \mathbf{E}(\mathbf{P} - \mathbf{T})(\mathbf{P} - \mathbf{T})' \\ &= \sigma^2 \phi^2 \mathbf{I}_n + \sigma^2 [(1 - \alpha)(1 - \alpha - 2\phi) + \lambda \alpha^2] \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \\ &= \sigma^2 \phi^2 \mathbf{I}_n + \sigma^2 [g - 2\phi(1 - \alpha)] \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'. \end{aligned} \quad (4.4)$$

Thus \mathbf{P} provides unbiased predictions for both the actual and average responses and in fact for any convex combination of actual and average responses. However, increased variability in predictions may be observed when \mathbf{P} is used for actual values in comparison to average values provided that α exceeds 0.5. The converse is true if α is less than 0.5.

Similarly, \mathbf{P}_1 and \mathbf{P}_2 are found to be biased. The bias vectors up to order $0(n^{-1})$ can be easily obtained from Srivastava and Toutenburg (1994). The resulting expressions reveal that \mathbf{P}_1 is superior to \mathbf{P}_2 with respect to the criterion of magnitude of bias to the order of our approximation. Further, we observe that the predictor \mathbf{P}_i has same bias whether it is used for predicting the actual values or average values or any convex linear combination of these.

If we look at the mean squared error matrices of \mathbf{P}_1 and \mathbf{P}_2 up to order $0(n^{-1})$ only, the resulting expressions are identical and equal to the exact variance covariance matrix (4.4) of \mathbf{P} . Thus all the three predictors have same performance with respect to mean squared error matrix criterion to order $0(n^{-1})$. This is not true if we take the criterion as trace of mean squared error matrix to the same order of approximation.

Result: The asymptotic approximation for the difference between the predictive mean squared errors (i. e., traces of mean squared error matrices) of \mathbf{P} and \mathbf{P}_i up to order $0(n^{-1})$ is given by

$$\begin{aligned} \mathbf{D}_\phi(\mathbf{P}; \mathbf{P}_i) &= \mathbf{E}(\mathbf{P} - \mathbf{T})'(\mathbf{P} - \mathbf{T}) - \mathbf{E}(\mathbf{P}_i - \mathbf{T})'(\mathbf{P}_i - \mathbf{T}) \\ &= \frac{\sigma^4 f_i^2}{n\theta} \left[2 \left(\frac{p-2}{f_i} \right) (g - \phi + \phi\alpha) - k \right] k. \end{aligned} \quad (4.5)$$

This result is derived in Appendix.

When the aim is to predict the actual values, this difference is

$$\mathbf{D}_1(\mathbf{P}; \mathbf{P}_i) = \frac{\sigma^4 f_i^2}{n\theta} \left[2\alpha \left(\frac{p-2}{f_i} \right) (\lambda\alpha + \alpha - 1) - k \right] k \quad (4.6)$$

which is negative when one of the following is true for all choices of k :

- (i) $p = 2$
- (ii) $\lambda = \left(\frac{1-\alpha}{\alpha} \right)$
- (iii) $p < 2$ and $\lambda > \left(\frac{1-\alpha}{\alpha} \right)$
- (iv) $p > 2$ and $\lambda < \left(\frac{1-\alpha}{\alpha} \right)$.

An additional fifth case that restricts the choice of k is specified as follows:

- (v) $k > 2\alpha \left(\frac{p-2}{f_i} \right) (\lambda\alpha + \alpha - 1)$

provided that $(p-2)$ and $(\lambda\alpha + \alpha - 1)$ have same sign.

In all these cases, the unbiased predictor \mathbf{P} is superior to the biased Stein-type predictor \mathbf{P}_i .

On the other hand, the predictor \mathbf{P}_i is superior to \mathbf{P} when

$$0 < k < 2\alpha \left(\frac{p-2}{f_i} \right) (\lambda\alpha + \alpha - 1)$$

provided that either of the following is true

- (i) $p = 1$ and $\lambda < \left(\frac{1-\alpha}{\alpha}\right)$
(ii) $p > 2$ and $\lambda > \left(\frac{1-\alpha}{\alpha}\right)$.

When the aim is to predict the average values of response variable, the difference (4.5) reduces to the following:

$$\mathbf{D}_0(\mathbf{P}; \mathbf{P}_i] = \frac{\sigma^4 f_i^2}{n\theta} \left[2(p-2) \left(\frac{g}{f_i}\right) - k \right] k. \quad (4.7)$$

If p is one or two, this difference is negative irrespective of the value of k . If $p > 2$, it is so when

$$k > 2(p-2) \left(\frac{g}{f_i}\right). \quad (4.8)$$

Under the above circumstances, Stein-type predictor \mathbf{P}_i is no better than the unbiased predictor \mathbf{P} .

Conversely, the predictor \mathbf{P}_i is superior to \mathbf{P} when

$$0 < k < 2(p-2) \left(\frac{g}{f_i}\right); p > 2. \quad (4.9)$$

The conditions (4.8) and (4.9) are not very attractive as they are difficult to check due to involvement of unknown λ . This limitation can be overcome by using (3.15). Thus the inequality (4.8) holds true as long as

$$k > 2(p-2)\nu_i^*; p > 2 \quad (4.10)$$

while the condition (4.9) is satisfied as long as

$$0 < k < 2(p-2)\nu_{i*}; p > 2. \quad (4.11)$$

The sufficient conditions (4.10) and (4.11) are simple and easy to check.

Next, let us compare the two Stein-type predictors.

It is seen from (4.5) that

$$\begin{aligned} \mathbf{D}_\phi(\mathbf{P}_1; \mathbf{P}_2) &= \mathbf{E}(\mathbf{P}_1 - \mathbf{T})'(\mathbf{P}_1 - \mathbf{T}) - \mathbf{E}(\mathbf{P}_2 - \mathbf{T})'(\mathbf{P}_2 - \mathbf{T}) \\ &= \frac{\sigma^4(f_2^2 - f_1^2)}{n\theta} \left[2\left(\frac{p-2}{f_1 + f_2}\right) (g - \phi + \phi\alpha) - k \right] k. \end{aligned} \quad (4.12)$$

When the aim is to predict the actual values, the difference (4.12) becomes

$$\mathbf{D}_1(\mathbf{P}_1; \mathbf{P}_2) = \frac{\sigma^4(f_2^2 - f_1^2)}{n\theta} \left[2\alpha \left(\frac{p-2}{f_1 + f_2}\right) (\lambda\alpha + \alpha - 1) - k \right] k. \quad (4.13)$$

This difference is negative under any one of the four cases cited for the negativity of (4.5). In addition to these, it is also negative when

$$k > 2\alpha \left(\frac{p-2}{f_1 + f_2}\right) (\lambda\alpha + \alpha - 1)$$

provided that $(p-2)$ and $(\lambda\alpha + \alpha - 1)$ have same sign, i.e., λ exceeds $(1-\alpha)/\alpha$ for $p > 2$ but it is less than $(1-\alpha)/\alpha$ for $p = 1$.

The difference (4.13) is positive implying the superiority of \mathbf{P}_2 over \mathbf{P}_1 when

$$0 < k < 2\alpha \left(\frac{p-2}{f_1+f_2} \right) (\lambda\alpha + \alpha - 1)$$

provided that $(p-2)$ and $(\lambda\alpha + \alpha - 1)$ have same sign.

When the aim is to predict the average values, we get the following expression from (4.12):

$$\mathbf{D}_0(\mathbf{P}_1; \mathbf{P}_2) = \frac{\sigma^4(f_2^2 - f_1^2)}{n\theta} \left[2(p-2) \left(\frac{g}{f_1+f_2} \right) - k \right] \quad (4.14)$$

whence it follows that \mathbf{P}_1 is superior to \mathbf{P}_2 for all values of k if p is either one or two. If p exceeds two, this result continues to remain true when

$$k > 2(p-2) \left(\frac{g}{f_1+f_2} \right). \quad (4.15)$$

The opposite is true, i.e., \mathbf{P}_2 is superior to \mathbf{P}_1 when

$$0 < k < 2(p-2) \left(\frac{g}{f_1+f_2} \right); p > 2. \quad (4.16)$$

Utilizing (3.23), we find that the condition (4.15) is satisfied as long as

$$k > 2(p-2)\nu^*; p > 2 \quad (4.17)$$

while the condition (4.16) is satisfied at least so long as

$$0 < k < 2(p-2)\nu_*; p > 2. \quad (4.18)$$

It may be observed that a user can easily check the conditions (4.17) and (4.18) for determining the superiority of one predictor over the other with respect to the criterion of predictive mean squared error to the order of our approximation.

5 Some concluding remarks

We have considered the problem of predicting the values of response variable when the available data set consists of observations from two similar experiments conducted independently. Pooling the two data sets and employing the combined evidence, three estimators of the regression coefficients are presented following Srivastava and Toutenburg (1994). Out of these, one is based on the least squares procedure while the other two emerge from Stein procedure. These three estimators are utilized to form three predictors for response values.

It is observed that the least squares predictor is unbiased while the Stein-type predictors are not unbiased whether the response values are other than the sample observations such as some future values or they are a part of the given sample values and whether we use them for predicting the actual values or average values or any weighted (convex) combination of these. Examining the bias vectors up to order $O(n^{-1})$ only, each Stein-type predictor is found to have the same bias irrespective of its use for actual or average values meaning

thereby that it is immaterial whether the predictor is used for actual values or average values or both. Comparing the two Stein-type predictors, it is seen that the first predictor is better than the second predictor with respect to the criterion of magnitude of bias to the order of our approximation.

When the three predictors are compared according to the mean squared error matrix criterion to order $0(n^{-2})$, there exist no conditions for the superiority of one predictor over the other except in a trivial case. Such a situation takes an interesting turn when we take the criterion as trace of mean squared error matrix (predictive mean squared error) to the order of our approximation. And we are able to identify situations where one predictor will have superior performance than the other. A salient feature of these comparisons is that we often get conditions which involve some unknown quantities and consequently cannot be used in practice. We have been able to overcome this unattractive aspect in many cases, and have succeeded in deducing sufficient conditions on the choice of scalar characterizing the predictor. These conditions are easy to check in actual practice.

Some of the sufficient conditions which we have stated are such that they provide a lower bound for the choice of the characterizing scalar. This lower bound in some cases may have a sufficiently large value and then a choice of k may alter the signs of Stein-type predictions which is obviously an undesirable feature. The user should be cautious about it.

Finally, our investigations have revealed that relative gain or loss in efficiency arising from the use of one predictor over the other remains same whether they are employed for predicting actual values or average values outside the sample. If the aim is to predict values within the sample, the gain or loss in efficiency vary and depends upon whether we use the predictors for actual values or average values.

A Appendix

Let us write

$$\begin{aligned}\boldsymbol{\epsilon} &= n^{-\frac{1}{2}} [(1-\alpha)\mathbf{X}'\mathbf{u}_1 + \alpha\mathbf{X}'\mathbf{u}_2] \\ \omega_1 &= n^{-\frac{1}{2}} \left[(1-\alpha) \left(\frac{\mathbf{u}_1'\mathbf{u}_1}{n} - \sigma^2 \right) + \alpha \left(\frac{\mathbf{u}_2'\mathbf{u}_2}{n} - \lambda\sigma^2 \right) \right] \\ \omega_2 &= n^{-\frac{1}{2}} \left[\left(\frac{\mathbf{u}_1'\mathbf{u}_1}{n} - \sigma^2 \right) + \left(\frac{\mathbf{u}_2'\mathbf{u}_2}{n} - \lambda\sigma^2 \right) \right].\end{aligned}$$

From Srivastava and Toutenburg (1994), we can express

$$(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}) = \frac{1}{n^{\frac{1}{2}}}\mathbf{V}\boldsymbol{\epsilon} - \frac{\sigma^2 k f_i}{n\theta}\boldsymbol{\beta} - \frac{k}{n^{\frac{3}{2}}\theta} \left[\omega_i\boldsymbol{\beta} + \sigma^2 f_i(\mathbf{V} - \frac{2}{\theta}\boldsymbol{\beta}\boldsymbol{\beta}')\boldsymbol{\epsilon} \right] + 0_p(n^{-2}).$$

Using it, we observe that

$$(\mathbf{P} - \mathbf{T})'(\mathbf{P} - \mathbf{T}) - (\mathbf{P}_i - \mathbf{T})'(\mathbf{P}_i - \mathbf{T}) = 2\xi_{-\frac{1}{2}} + \xi_{-1} + 0_p(n^{-\frac{3}{2}})$$

where

$$\begin{aligned}\xi_{-\frac{1}{2}} &= \frac{\sigma^2 k f_i}{n\theta} [(1-\alpha-\phi)\mathbf{u}_1'\mathbf{X}\boldsymbol{\beta} + \alpha\mathbf{u}_2'\mathbf{X}\boldsymbol{\beta}] \\ \xi_{-1} &= \frac{2k}{n^{\frac{3}{2}}\theta} [(1-\alpha-\phi)\mathbf{u}_1'\mathbf{X} + \alpha\mathbf{u}_2'\mathbf{X}] \left[\omega_i\boldsymbol{\beta} + \sigma^2 f_i(\mathbf{V} - \frac{2}{\theta}\boldsymbol{\beta}\boldsymbol{\beta}')\boldsymbol{\epsilon} \right] - \frac{\sigma^4 k^2 f_i^2}{n\theta}.\end{aligned}$$

By virtue of the distributional properties of \mathbf{u}_1 and \mathbf{u}_2 , it is easy to see that

$$\begin{aligned} E(\xi_{-\frac{1}{2}}) &= 0 \\ E(\xi_{-1}) &= \frac{\sigma^4 k f_i}{n\theta} (p-2) [(1-\alpha)(1-\alpha-\phi) + \lambda\alpha^2] - \frac{\sigma^4 k^2 f_i^2}{n\theta}. \end{aligned}$$

Using these, we obtain the result (4.5).

References

- Judge, G. G. and Bock, M. E. (1978). *The statistical implications of pre-test and Stein-rule estimators in econometrics*, North Holland, Amsterdam.
- Rao, C. R., Srivastava, V. K. and Toutenburg, H. (1997). Pitman nearness comparisons of stein-type estimators for regression coefficients in replicated experiments, *to be published in Statistical Papers*.
- Shalabh (1995). Performance of Stein-type procedure for simultaneous prediction of actual and average values of study variable in linear regression model, *Proceedings of the Fiftieth Session of the International Statistical Institute*, pp. 1375–1390.
- Srivastava, V. K. and Toutenburg, H. (1994). Application of stein-type estimation in combining regression estimates from 'replicated experiments', *Statistical Papers* **35**: 101–112.
- Trenkler, G. and Toutenburg, H. (1992). Pre-test procedures and forecasting in the regression model under restrictions, *Journal of Statistical Planning and Inference* **30**: 249–256.