



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Toutenburg, Shalabh:

Improved Predictions in Linear Regression Models with Stochastic Linear Constraints

Sonderforschungsbereich 386, Paper 124 (1998)

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

Projektpartner



Improved Predictions in Linear Regression Models with Stochastic Linear Constraints

H. Toutenburg * Shalabh **

August 26, 1998

Abstract

In this article, we have considered two families of predictors for the simultaneous prediction of actual and average values of study variable in a linear regression model when a set of stochastic linear constraints binding the regression coefficients is available. These families arise from the method of mixed regression estimation. Performance properties of these families are analyzed when the objective is to predict values outside the sample and within the sample.

1 Introduction

Quite often we come across situations demanding simultaneous prediction of the actual and average values of study variable. As an illustration, consider a new drug for increasing the duration of sleep in patients suffering from high blood pressure. When a medical practitioner is told to prescribe some new drug, he/she would like to enquire, “What will be the increase, on the average, in the duration of sleep when a specific dose of this drug is administered?” On the other strand, a patient would like to know, “What will be the actual increase in the duration of my sleep if I take the prescribed dose of this drug?”. Thus the medical practitioner is more interested in the prediction of average value in comparison to the prediction of actual value. The opposite is true in case of the patient who is more concerned with the prediction of actual value. It is therefore imperative to consider the simultaneous prediction of actual and average values in such a manner that actual and average values are assigned possibly unequal weightage. Such situations occur not only in medical sciences but in other disciplines too like Economics (see Shalabh, 1995; Zellner, 1994).

Toutenburg and Shalabh (1996) have considered the problem of simultaneous prediction of actual and average values of study variable in a linear regression model assuming the availability of some prior information in the form of few linear constraints binding the regression coefficients. When these constraints are stochastic in nature, they have studied the performance properties of predictors arising from pure and mixed regression methods of estimation for the coefficients. For forecasting values (actual or average or a combination of both)

*Institute of Statistics, LMU München, Ludwigstr. 33, 80539 München, Germany, toutenb@stat.uni-muenchen.de

**Department of Statistics, University of Jammu, Jammu 180004, India

of study variable outside the sample such as future values, their investigations have revealed that the predictions based on mixed regression estimation are superior, at least asymptotically with respect to the criterion of predictive dispersion matrix, to the predictions based on pure regression estimation provided that the degrees of freedom, i.e., excess of the numbers of observations over the numbers of unknown coefficients, are three or more. This is, however, not true when the aim is to predict values within the sample. Here, mixed regression based predictions are superior only when the predictions of average values is assigned a higher weight in comparison to the prediction of actual values provided that the degrees of freedom are at least three.

In view of the above findings reported by Toutenburg and Shalabh (1996), a natural question is as follows. Can we further improve the performance of mixed regression based predictions? In other words, is it possible to construct predictors having superior performance properties at least in these situations where mixed regression based predictions are known to be better than the pure regression based predictions? This article presents a simple effort in this direction.

Our presentation is as follows. In Section 2, we describe the model and present estimators of regression coefficients. Section 3 considers the prediction of values outside the sample and analyzes the properties of some predictors. Similarly, Section 4 deals with the prediction of values within the sample. Some remarks are then placed in Section 5. Finally, the Appendix presents derivation of results.

2 Model Specification and the Estimators

Consider the following linear regression model:

$$y = X\beta + \sigma u \quad (2.1)$$

where y is a $n \times 1$ vector of observations on the study variable, X is $n \times k$ full column rank matrix of n observations on k explanatory variables, β is the vector of regression coefficient, σ is a scalar and u is a vector of disturbances following a multivariate normal distribution with mean vector 0 and variance covariance matrix I .

In addition to (2.1), let us be given some prior information in the form of a set of g stochastic linear constraints binding the regression coefficients as follows:

$$r = R\beta + v \quad (2.2)$$

where r is $g \times 1$ vector with known elements, R is a $g \times k$ matrix with known elements and v is a $g \times 1$ random vector with mean vector 0 and variance covariance matrix Ψ assumed to be known and positive definite.

It is assumed that u and v are stochastically independent and $(n-k)$ exceeds 2.

The pure regression estimator of β is the least squares estimator of β in (2.1):

$$b = (X'X)^{-1}X'y. \quad (2.3)$$

If we apply Stein-rule method to (2.1), we obtain the following family of estimators for β :

$$b_s = \left[1 - \left(\frac{a}{n-k+2} \right) \left(\frac{Q}{b'X'Xb} \right) \right] b \quad (2.4)$$

where $Q = (y - Xb)'(y - Xb)$ is the residual sum of squares and a is the nonstochastic positive scalar characterizing the estimator; see, e.g., Judge and Bock (1978) for an interesting exposition.

The mixed regression estimator proposed by Theil and Goldberger (1961) is given by

$$b_m = \left(X'X + \frac{Q}{n-k} R' \Psi^{-1} R \right)^{-1} \left(X'y + \frac{Q}{n-k} R' \Psi^{-1} r \right) \quad (2.5)$$

see, e.g., Rao and Toutenburg (1995) for an interesting account of its properties.

Synthesizing the mixed regression and Stein-rule estimation procedures, Srivastava and Srivastava (1983) have presented the following families of estimators for β :

$$b_{ms} = \left[1 - \left(\frac{a}{n-k+2} \right) \left(\frac{Q_m}{b'_m X' X b_m} \right) \right] b_m \quad (2.6)$$

$$b_{sm} = \left(X'X + \frac{Q_s}{n-k} R' \Psi^{-1} R \right)^{-1} \left(X'y + \frac{Q_s}{n-k} R' \Psi^{-1} r \right) \quad (2.7)$$

where

$$Q_m = (y - Xb_m)'(y - Xb_m) \quad (2.8)$$

$$Q_s = (y - Xb_s)'(y - Xb_s) \quad (2.9)$$

are the residual sum of squares based on mixed regression and Stein-rule estimators.

3 Prediction Outside The Sample

When the aim is to predict the values of study variable outside the sample such as future values, we assume the validity of model for such values also. Thus we can write

$$y_f = X_f \beta + \sigma u_f \quad (3.1)$$

where y_f is a $n_f \times 1$ vector of values of study variable, X_f is a $n_f \times k$ matrix of given values of explanatory variables and u_f is the vector of disturbances following a normal distribution with mean vector 0 and variance covariance matrix I .

We further assume that the elements of u_f are stochastically independent of the elements of u and v .

For predicting the n_f values of study variable, let us consider the following predictors:

$$\begin{aligned} \hat{T}_f &= X_f b \\ \hat{T}_{fm} &= X_f b_m \\ \hat{T}_{fms} &= X_f b_{ms} \\ \hat{T}_{fsm} &= X_f b_{sm} \end{aligned}$$

which can be used for predicting the actual values y_f as well as the average values $E(y_f) = X_f\beta$.

For the simultaneous prediction of actual and average values, we define the target function as

$$T_f = \lambda y_f + (1 - \lambda) E(y_f) \quad (3.2)$$

where $0 \leq \lambda \leq 1$ is a nonstochastic scalar reflecting the weightage to be given to actual values in relation to average values for the purpose of prediction (see Shalabh, 1995; Toutenburg and Shalabh, 1996).

Toutenburg and Shalabh (1996) have derived an asymptotic approximation for the predictive dispersion matrix of \hat{T}_{fm} employing the small disturbance asymptotic theory and have compared the performance of \hat{T}_f and \hat{T}_{fm} . It is seen that both are weakly unbiased in the sense that $E(\hat{T}_f - T_f)$ and $E(\hat{T}_{fm} - T_f)$ are null vectors. Further, with respect to the criterion of predictive dispersion matrix to order $O(\sigma^4)$, \hat{T}_{fm} is better than \hat{T}_f whether the aim is to predict actual values or average values or both together (see Toutenburg and Shalabh, 1996, Sec. 4.2). Notice that we have assumed $(n - k)$ to be three or more.

Now let us examine whether the predictors \hat{T}_{fms} and \hat{T}_{fsm} have better performance than \hat{T}_{fm} . For this purpose, we consider the small disturbance asymptotic approximations.

Theorem 3.1: The bias vectors of \hat{T}_{fms} and \hat{T}_{fsm} to order $O(\sigma^2)$ are given by

$$\begin{aligned} B(\hat{T}_{fms}) &= E(\hat{T}_{fms} - T_f) \\ &= -\sigma^2 \left(\frac{n-k}{n-k+2} \right) \left(\frac{a}{\beta' X' X \beta} \right) X_f \beta \end{aligned} \quad (3.3)$$

$$\begin{aligned} B(\hat{T}_{fsm}) &= E(\hat{T}_{fsm} - T_f) \\ &= 0 \end{aligned} \quad (3.4)$$

while the differences in dispersion matrices to order $O(\sigma^4)$ are

$$\begin{aligned} D(\hat{T}_{fm}; \hat{T}_{fms}) &= E(\hat{T}_{fm} - T_f)(\hat{T}_{fm} - T_f)' - E(\hat{T}_{fms} - T_f)(\hat{T}_{fms} - T_f)' \\ &= \sigma^4 \left(\frac{n-k}{n-k+2} \right) \left(\frac{a}{\beta' X' X \beta} \right) \\ &\quad X_f \left[2(X'X)^{-1} - \left(\frac{a+4}{\beta' X' X \beta} \right) \beta \beta' \right] X_f' \end{aligned} \quad (3.5)$$

$$\begin{aligned} D(\hat{T}_{fm}; \hat{T}_{fsm}) &= E(\hat{T}_{fm} - T_f)(\hat{T}_{fm} - T_f)' - E(\hat{T}_{fsm} - T_f)(\hat{T}_{fsm} - T_f)' \\ &= 0 \end{aligned} \quad (3.6)$$

These results are derived in Appendix.

It is thus seen from (3.3) that \hat{T}_{fms} is not weakly unbiased like \hat{T}_m at least to the order of our approximation. Next, we observe from (3.5) that the difference matrix is indefinite so that nothing definite can be said about the superiority of one predictor over the other. However, if we take the trace of the difference matrix, we find that \hat{T}_{fms} is better than \hat{T}_m when

$$0 < a < 2 \left[\frac{\beta' X' X \beta}{\beta' X_f' X_f \beta} \text{tr}(X'X)^{-1} X_f' X_f - 2 \right] \quad (3.7)$$

provided that the quantity on the extreme right is positive.

The condition (3.7) is not attractive owing to involvement of β . A sufficient version of it is

$$0 < a < 2 \left(\sum_{i=2}^k \frac{c_i}{c_1} - 1 \right); \quad \sum_{i=2}^k c_i > c_1 \quad (3.8)$$

where $c_1 \geq c_2 \geq \dots$ denote the characteristic roots of the matrix $(X'X)^{-1}X'X_f$.

Next, we observe from (3.4) and (3.6) that \hat{T}_{fsm} is weakly unbiased like \hat{T}_{fm} at least to order $O(\sigma^2)$. Further, its predictive dispersion matrix is same as that of \hat{T}_{fm} to the order of our approximation. Thus we need to consider higher order approximations. These are derived in Appendix and are presented below:

Theorem 3.2: Retaining terms up to order $O(\sigma^6)$, we have

$$\begin{aligned} \mathbf{B}(\hat{T}_{fsm}) &= \mathbf{E}(\hat{T}_{fsm} - T_f) \\ &= \sigma^6 \left[\frac{2a^2}{(n-k+2)(\beta'X'X\beta)^2} \right] X_f(X'X)^{-1}R'\Psi^{-1}R\beta \quad (3.9) \\ \mathbf{D}(\hat{T}_{fm}; \hat{T}_{fsm}) &= \mathbf{E}(\hat{T}_{fm} - T_f)(\hat{T}_{fm} - T_f)' - \mathbf{E}(\hat{T}_{fsm} - T_f)(\hat{T}_{fsm} - T_f)' \\ &= -\sigma^6 \left(\frac{8a^2}{(n-k)(n-k+2)\beta'X'X\beta} \right) \\ &\quad X_f(X'X)^{-1}R'\Psi^{-1}R(X'X)^{-1}X_f' \quad (3.10) \end{aligned}$$

These results indicate that \hat{T}_{fsm} is not only weakly biased but also not efficient in comparison to \hat{T}_{fm} .

We thus find that \hat{T}_{fms} emerges out to be the best predictor among \hat{T}_f , \hat{T}_{fm} , \hat{T}_{fms} and \hat{T}_{fsm} under a mild constraint (3.8) on the characterizing scalar a if the criterion is asymptotic predictive dispersion. This is of course weakly biased. Next best is the weakly unbiased predictor \hat{T}_{fm} . The predictor \hat{T}_{fsm} seems to have no merit.

4 Prediction Within The Sample

When the aim is to predict the values of study variable within the sample, we formulate our target function as follows:

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

with λ a nonstochastic scalar between 0 and 1.

Now consider the following predictors

$$\begin{aligned} \hat{T} &= Xb \\ \hat{T}_m &= Xb_m \\ \hat{T}_{ms} &= Xb_{ms} \\ \hat{T}_{sm} &= Xb_{sm}. \end{aligned}$$

Toutenburg and Shalabh (1996) have observed that the predictors \hat{T} and \hat{T}_m are weakly unbiased. Further, according to the criterion of predictive dispersion matrix to order $O(\sigma^4)$, \hat{T}_m is better than \hat{T} when

$$\lambda < \left(\frac{1}{2} - \frac{1}{n-k} \right) \quad (4.2)$$

provided that $(n-k)$ exceeds two. An interesting implication of it is that it is worthwhile to use mixed regression estimation procedure in preference to the pure regression estimation procedure when prediction of average values is given more weightage in comparison to prediction of actual values.

Let us now examine the performance properties of the predictors \hat{T}_{ms} and \hat{T}_{sm} using small disturbance asymptotic theory.

Theorem 4.1: The bias vector of \hat{T}_{ms} and \hat{T}_{sm} up to order $O(\sigma^2)$ are

$$\begin{aligned} \mathbf{B}(\hat{T}_{ms}) &= \mathbf{E}(\hat{T}_{ms} - T) \\ &= -\sigma^2 \left(\frac{n-k}{n-k+2} \right) \left(\frac{a}{\beta'X'X\beta} \right) X\beta \end{aligned} \quad (4.3)$$

$$\begin{aligned} \mathbf{B}(\hat{T}_{sm}) &= \mathbf{E}(\hat{T}_{sm} - T) \\ &= 0 \end{aligned} \quad (4.4)$$

while the differences in dispersion matrices to order $O(\sigma^4)$ are given by

$$\begin{aligned} \mathbf{D}(\hat{T}_m; \hat{T}_{ms}) &= \mathbf{E}(\hat{T}_m - T)(\hat{T}_m - T)' - \mathbf{E}(\hat{T}_{ms} - T)(\hat{T}_{ms} - T)' \\ &= \sigma^4 \left(\frac{n-k}{n-k+2} \right) \left(\frac{a}{\beta'X'X\beta} \right) \\ &\quad X \left[2(1-\lambda)(X'X)^{-1} - \frac{a+4(1-\lambda)}{\beta'X'X\beta} \beta\beta' \right] X' \end{aligned} \quad (4.5)$$

$$\begin{aligned} \mathbf{D}(\hat{T}_m; \hat{T}_{sm}) &= \mathbf{E}(\hat{T}_m - T)(\hat{T}_m - T)' - \mathbf{E}(\hat{T}_{sm} - T)(\hat{T}_{sm} - T)' \\ &= 0 \end{aligned} \quad (4.6)$$

The derivation of these results is presented in Appendix.

From (4.3), we observe that the predictor \hat{T}_{ms} is weakly biased while \hat{T}_m is unbiased. Comparing them with respect to the criterion of predictive dispersion matrix to order $O(\sigma^4)$, we notice that neither dominates the other. However, if we take the trace so that the criterion is total predictive dispersion, we find that \hat{T}_{ms} is superior to \hat{T}_m when

$$0 < a < 2(1-\lambda)(k-2); \quad k > 2. \quad (4.7)$$

This is an interesting condition as it is easy to use in any given application for finding improved predictions.

So far as the predictor \hat{T}_{sm} is concerned it follows from (4.4) and (4.6) that it is asymptotically equivalent to \hat{T}_m in the sense that both are weakly unbiased to order $O(\sigma^4)$. The difference, however, precipitates when higher order approximations are considered.

Theorem 4.2: If we consider approximations to order $O(\sigma^6)$, we have

$$\begin{aligned}
\mathbf{B}(\hat{T}_{sm}) &= \mathbf{E}(\hat{T}_{sm} - t) \\
&= \sigma^6 \left[\frac{2a^2}{(n-k+2)(\beta'X'X\beta)^2} \right] X(X'X)^{-1}R'\Psi^{-1}R\beta \quad (4.8) \\
\mathbf{D}(\hat{T}_m; \hat{T}_{sm}) &= \mathbf{E}(\hat{T}_m - T)(\hat{T}_m - T)' - \mathbf{E}(\hat{T}_{sm} - T)(\hat{T}_{sm} - T)' \\
&= \sigma^6 \left[\left(1 + \frac{2}{n-k}\right) \left\{ 2 \left(\lambda + \frac{4}{n-k} \right) - 1 \right\} X(X'X)^{-1}R'\Psi^{-1} \right. \\
&\quad \left. R(X'X)^{-1}R'\Psi^{-1}R(X'X)^{-1}X' - \frac{2a^2}{(n-k+2)\beta'X'X\beta} \right. \\
&\quad \left. \left(\lambda + \frac{4}{n-k} \right) X(X'X)^{-1}R'\Psi^{-1}R(X'X)^{-1}X' \right] \quad (4.9)
\end{aligned}$$

From (4.8), we observe that \hat{T}_{sm} is not, while \hat{T}_m is always, weakly unbiased to order $O(\sigma^6)$. From (4.9), it is difficult to determine the nature of matrix and consequently superiority condition of one prediction over the other cannot be found. Even if we consider the trace and obtain a condition on the characterizing scalar a , it is found to depend upon unknown regression coefficients and thus may not serve any useful purpose in practice.

5 Some Remarks

The present investigations have been motivated by the finding reported by Toutenburg and Shalabh (1996) regarding the superiority of mixed regression estimation procedure over the pure regression estimation for the purpose of prediction of values of study variable. When the aim is to predict the values outside the sample, mixed regression based predictions are found to be superior than the pure regression based predictions. Observing that mixed regression based predictions are weakly unbiased for the simultaneous prediction of actual and average values of study variable, we have considered two families of weakly biased predictions stemming from the method of mixed regression estimation. One of them turns out to be useless as it may yield predictions that are inferior to the conventional weakly unbiased predictions. The other, however, provides superior predictions under a mild constraint on the scalar characterizing the family. In other words, one can construct predictors which are biased but more efficient than the predictions arising from the conventional pure and mixed regression methods.

Next, suppose that we wish to predict the values of study variable within the sample, for example, to assess the success of underlying estimation procedure for the unknown parameters of the model. When the prediction of average values is assigned higher weightage in comparison to the prediction of actual values, it is found that mixed regression based predictions are better than the pure regression based predictions provided that the number of observations exceeds the number of unknown coefficients by two. As these predictions are weakly unbiased, we have considered two families of weakly biased predictors arising from mixed regression estimation. One family is found to give predictors that are asymptotically as good as the conventional weakly unbiased predictor. The other family yields predictors that are superior to the weakly unbiased predictor

with respect to the criterion of total predictive dispersion to the order of our approximation.

Appendix

From Toutenburg and Shalabh (1996, p. 957), we can express

$$\begin{aligned}
(b_m - \beta) &= \sigma(X'X)^{-1}X'u + \sigma^2 \left(\frac{u'Mu}{n-k} \right) R\Psi^{-1}v \\
&\quad - \sigma^3 \left(\frac{u'Mu}{n-k} \right) (X'X)^{-1}R'\Psi^{-1}R(X'X)^{-1}X'u \\
&\quad - \sigma^4 \left(\frac{u'Mu}{n-k} \right)^2 (X'X)^{-1}R'\Psi^{-1}R(X'X)^{-1}R'\Psi^{-1}v + O_p(\sigma^5)
\end{aligned} \tag{A.1}$$

where $M = I - X(X'X)^{-1}X'$. Similarly, we can write

$$\begin{aligned}
Q_s &= \sigma^2 u'Mu + \sigma^4 \left(\frac{a u'MU}{n-k+2} \right)^2 \left(\frac{1}{\beta'X'X\beta} \right) \\
&\quad - 2\sigma^5 \left(\frac{a u'Mu}{n-k+2} \right)^2 \frac{\beta'X'u}{(\beta'X'X\beta)^2} + O_p(\sigma^6).
\end{aligned} \tag{A.2}$$

Now we consider the vector

$$\begin{aligned}
E(\hat{T}_{fsm} - T_f) &= E \left[X_f \left(X'X + \frac{Q_s}{n-k} R'\Psi^{-1}R \right)^{-1} \right. \\
&\quad \left. \left(\sigma X'u + \frac{Q_s}{n-k} R'\Psi^{-1}v \right) \right] - \sigma\lambda E(u_f).
\end{aligned}$$

By virtue of stochastic independence of u , u_f and v , we have

$$\begin{aligned}
E(\hat{T}_{fsm} - T_f) &= \sigma E \left[X_f \left(X'X + \frac{Q_s}{n-k} R'\Psi^{-1}R \right)^{-1} X'u \right] \\
&= \sigma X_f (X'X)^{-1} E \left[\left\{ I + \frac{Q_s}{n-k} R'\Psi^{-1}R(X'X)^{-1} \right\}^{-1} X'u \right]
\end{aligned}$$

Substituting (A.2), then expanding and retaining terms upto order $O(\sigma^6)$ only, we get

$$E(\hat{T}_{fsm} - T_f) = X_f (X'X)^{-1} E(\sigma e_1 - \sigma^3 e_3 - \sigma^5 e_5 + \sigma^6 e_6) \tag{A.3}$$

where

$$\begin{aligned}
e_1 &= X'u \\
e_3 &= \left(\frac{u'Mu}{n-k} \right) R'\Psi^{-1}R(X'X)^{-1}X'u
\end{aligned}$$

$$\begin{aligned}
e_5 &= \frac{a^2 (u' M u)^2}{(n-k)(n-k+2)^2 \beta' X' X \beta} R' \Psi^{-1} R (X' X)^{-1} X' u \\
&\quad - \left(\frac{u' M u}{n-k} \right)^2 R' \Psi^{-1} R (X' X)^{-1} R' \Psi^{-1} R (X' X)^{-1} X' u \\
e_6 &= \left(\frac{2}{n-k} \right) \left(\frac{a u' M u}{(n-k+2) \beta' X' X \beta} \right)^2 R' \Psi^{-1} R (X' X)^{-1} X' u u' X \beta
\end{aligned}$$

It is easy to see that

$$\begin{aligned}
\mathbb{E}(e_1) &= \mathbb{E}(e_3) = \mathbb{E}(e_5) = 0 \\
\mathbb{E}(e_6) &= \frac{2a^2}{(n-k+2)(\beta' X' X \beta)^2} R' \Psi^{-1} R \beta
\end{aligned}$$

where use has been made of the independence of $X' u$ and $u' M u$ due to normality of disturbances.

Substituting these results we see that the bias vector to order $O(\sigma^2)$ is null providing the statement (3.4) of Theorem 3.1. If we consider terms upto order $O(\sigma^6)$, we get the result (3.9) of Theorem 3.2.

Next, if we write $N = R' \Psi^{-1} R (X' X)^{-1}$, we observe that

$$\begin{aligned}
&\mathbb{E}(\hat{T}_{fm} - T_f)(\hat{T}_{fm} - T_f)' \\
&= \sigma^2 \lambda^2 I + \mathbb{E} \left[X_f \left(X' X + \frac{Q}{n-k} R' \Psi^{-1} R \right)^{-1} \right. \\
&\quad \left. \left(\sigma^2 X' X + \left(\frac{Q}{n-k} \right)^2 R' \Psi^{-1} R \right) \left(X' X + \frac{Q}{n-k} R' \Psi^{-1} R \right)^{-1} X_f' \right] \\
&= \sigma^2 \lambda^2 I + \sigma^2 X_f (X' X)^{-1} \mathbb{E} \left[I + \sigma^2 \left(\frac{u' M u}{n-k} \right) N \right]^{-1} \\
&\quad \left[I + \sigma^2 \left(\frac{u' M u}{n-k} \right)^2 N \right] \left[I + \sigma^2 \left(\frac{u' M u}{n-k} \right) N \right]^{-1} X_f' \\
&= \sigma^2 \lambda^2 I + \sigma^2 X_f (X' X)^{-1} \mathbb{E} \left[I + \sigma^2 \left(\frac{u' M u}{n-k} \right) N \right]^{-1} X_f' \\
&\quad - \sigma^4 X_f (X' X)^{-1} \mathbb{E} \left(\frac{u' M u}{n-k} \right) \left(1 - \frac{u' M u}{n-k} \right) \\
&\quad \left[I + \sigma^2 \left(\frac{u' M u}{n-k} \right) N \right]^{-1} N \left[I + \sigma^2 \left(\frac{u' M u}{n-k} \right) N \right]^{-1} X_f' \\
&= \sigma^2 [\lambda^2 I + X_f (X' X)^{-1} X_f'] \\
&\quad - \sigma^4 \mathbb{E} \left(\frac{u' M u}{n-k} \right) \left(2 - \frac{u' M u}{n-k} \right) X_f (X' X)^{-1} N X_f' \\
&\quad + \sigma^6 \mathbb{E} \left(\frac{u' M u}{n-k} \right)^2 \left(3 - 2 \frac{u' M u}{n-k} \right) X_f (X' X)^{-1} N^2 X_f' + O(\sigma^8) \\
&= \sigma^2 [\lambda^2 I + X_f (X' X)^{-1} X_f'] - \sigma^4 \left(1 - \frac{2}{n-k} \right) X_f (X' X)^{-1} N X_f' \\
&\quad + \sigma^6 \left(1 + \frac{2}{n-k} \right) \left(1 - \frac{8}{n-k} \right) X_f (X' X)^{-1} N^2 X_f' + O(\sigma^8) \quad (\text{A.4})
\end{aligned}$$

Similarly, we observe that

$$\begin{aligned}
& \mathbb{E}(\hat{T}_{fsm} - T_f)(\hat{T}_{fsm} - T_f)' \\
&= \sigma^2 \lambda^2 I + \mathbb{E} \left[X_f \left(X'X + \frac{Q_s}{n-k} R' \Psi^{-1} R \right)^{-1} \right. \\
&\quad \left. \left(\sigma^2 X' u u' X + \left(\frac{Q_s}{n-k} \right)^2 R' \Psi^{-1} R \right) \left(X'X + \frac{Q_s}{n-k} R' \Psi^{-1} R \right)^{-1} X_f' \right] \\
&= \sigma^2 \lambda^2 I + \mathbb{E} X_f (X'X)^{-1} \left(I + \frac{Q_s}{n-k} N \right)^{-1} \\
&\quad \left[\sigma^2 X' u u' X (X'X)^{-1} + \left(\frac{Q_s}{n-k} \right)^2 N \right] \left[I + \frac{Q_s}{n-k} N \right]^{-1} \\
&= \sigma^2 \left[\lambda^2 I + X_f (X'X)^{-1} X' \mathbb{E}(u u') X (X'X)^{-1} X_f' \right] + \sigma^4 X_f (X'X)^{-1} \\
&\quad \left[E \left(\frac{u' M u}{n-k} \right)^2 N - 2 E \left(\frac{u' M u}{n-k} \right) X' u u' X (X'X)^{-1} N \right] X_f' \\
&\quad + \sigma^6 X_f (X'X)^{-1} \left[\frac{2a^2 \mathbb{E}(u' M u)^3}{(n-k)^2 (n-k+2)^2 \beta' X' X \beta} N - 2 \frac{\mathbb{E}(u' M u)^3}{(n-k)^3} N^2 \right. \\
&\quad \left. - \frac{2a^2}{(n-k)(n-k+2)^2 \beta' X' X \beta} \mathbb{E}(u' M u)^2 X' u u' X (X'X)^{-1} N \right. \\
&\quad \left. + \frac{3}{(n-k)^2} \mathbb{E}(u' M u)^2 X' u u' X (X'X)^{-1} N^2 \right] X_f' + O(\sigma^8) \\
&= \sigma^2 \left[\lambda^2 I + X_f (X'X)^{-1} X_f' \right] - \left(1 - \frac{2}{n-k} \right) X_f (X'X)^{-1} N X_f' \\
&\quad + \sigma^6 \left[\left(1 + \frac{2}{n-k} \right) \left(1 - \frac{8}{n-k} \right) X_f (X'X)^{-1} N^2 X_f' \right. \\
&\quad \left. + \frac{8a^2}{(n-k)(n-k+2) \beta' X' X \beta} X_f (X'X)^{-1} N X_f' \right] + O(\sigma^8).
\end{aligned}$$

Taking the difference, we get

$$D(\hat{T}_{fm}; \hat{T}_{fsm}) = - \frac{8a^2 \sigma^6}{(n-k)(n-k+2) \beta' X' X \beta} X_f (X'X)^{-1} N X_f'$$

which implies the result (3.6) of Theorem 3.1 and provides the result (3.10) of Theorem 3.2.

Next, consider the predictor \hat{T}_{fms} . Using (A.1), we observe that

$$\begin{aligned}
\frac{Q_m}{b_m' X' X b_m} &= \frac{(y - X b_m)' (y - X b_m)}{b_m' X' X b_m} \\
&= \frac{\sigma^2 u' M u + O_p(\sigma^4)}{\beta' X' X \beta + 2\sigma \beta' X' u + O_p(\sigma^2)} \\
&= \sigma^2 \frac{u' M u}{\beta' X' X \beta} - 2\sigma^3 \frac{u' M u \cdot \beta' X' u}{(\beta' X' X \beta)^2} + O_p(\sigma^4)
\end{aligned}$$

whence we see that

$$\begin{aligned}
\mathbb{E}(\hat{T}_{fms} - T_f) &= \mathbb{E}(\hat{T}_{fm} - T) - \left(\frac{a}{n-k+2} \right) \mathbb{E} \left(\frac{Q_m}{b'_m X' X b_m} \hat{T}_{fm} \right) \\
&= -\sigma^2 \left(\frac{a}{n-k+2} \right) \frac{\mathbb{E}(u' M u)}{\beta' X' X \beta} X_f \beta \\
&= -\sigma^2 \frac{a(n-k)}{(n-k+2)\beta' X' X \beta} X_f \beta \tag{A.5}
\end{aligned}$$

upto order $O(\sigma^2)$. This yields the result (3.3). Similarly, the difference matrix upto order $O(\sigma^4)$ is

$$\begin{aligned}
D(\hat{T}_{fm}; \hat{T}_{fms}) &= \mathbb{E}(\hat{T}_{fm} - T_f)(\hat{T}_{fm} - T_f)' - \mathbb{E}(\hat{T}_{fms} - T_f)(\hat{T}_{fms} - T_f)' \\
&= \left(\frac{a}{n-k+2} \right) \mathbb{E} \left(\frac{Q_m}{b'_m X' X b_m} \right) \left[\hat{T}_{fm}(\hat{T}_{fm} - T_f)' + (\hat{T}_{fm} - T_f)\hat{T}'_{fm} \right] \\
&\quad - \left(\frac{a}{n-k+2} \right)^2 \mathbb{E} \left(\frac{Q_m}{b'_m X' X b_m} \right)^2 (\hat{T}_{fm} \hat{T}'_{fm}) \\
&= \left(\frac{a}{(n-k+2)\beta' X' X \beta} \right) [\sigma^3 \mathbb{E}(G_3) + \sigma^4 \mathbb{E}(u' M u G_4)]
\end{aligned}$$

where

$$\begin{aligned}
G_3 &= (u' M u) X_f \beta [u' X (X' X)^{-1} X'_f - \lambda u'_f] \\
&\quad + (u' M u) [X_f (X' X)^{-1} X' u - \lambda u_f] \beta' X'_f \\
G_4 &= X_f \left[(X' X)^{-1} - \frac{2}{\beta' X' X \beta} \beta \beta' \right] X' u [u' X (X' X)^{-1} X'_f - \lambda u'_f] \\
&\quad + [X'_f (X' X)^{-1} X' u - \lambda u_f] u' X \left[(X' X)^{-1} - \frac{2}{\beta' X' X \beta} \beta \beta' \right] X'_f \\
&\quad + \left(\frac{u' M u}{n-k} \right) X_f (\beta v' \Psi^{-1} R + R' \Psi^{-1} v \beta') X'_f \\
&\quad - \left(\frac{a}{n-k+2} \right) \left(\frac{u' M u}{\beta' X' X \beta} \right) X_f \beta \beta' X'_f
\end{aligned}$$

By virtue of normality of u , it is easy to see that

$$\begin{aligned}
\mathbb{E}(G_3) &= 0 \\
\mathbb{E}(u' M u G_4) &= (n-k) X_f \left[2(X' X)^{-1} - \frac{a+4}{\beta' X' X \beta} \beta \beta' \right] X'_f.
\end{aligned}$$

Substituting these results, we get (3.5) of Theorem 3.1.

For the results in Theorem 4.1 and Theorem 4.2, we first see that

$$\mathbb{E}(\hat{T}_{sm} - T) = \sigma \mathbb{E} \left[X \left(X' X + \frac{Q_s}{n-k} R' \Psi^{-1} R \right)^{-1} X' u \right]$$

Proceeding in the same way as in (A.3) we obtain the results (4.4) of Theorem 4.1 and (4.8) of Theorem 4.2.

Observing that

$$(\hat{T}_m - T) = X(X'X)^{-1} \left(I + \frac{Q}{n-k}N \right)^{-1} \left(\sigma X'u + \frac{Q}{n-k}R'\Psi^{-1}v \right) - \sigma\lambda u$$

we have

$$\begin{aligned} & \mathbb{E}(\hat{T}_m - T)(\hat{T}_m - T)' \\ &= \sigma^2 X(X'X)^{-1} \mathbb{E} \left[I + \sigma^2 \left(\frac{u'Mu}{n-k} \right) N \right]^{-1} \left[I + \sigma^2 \left(\frac{u'Mu}{n-k} \right)^2 N \right] \\ & \quad \left[I + \sigma^2 \left(\frac{u'Mu}{n-k} \right) N \right]^{-1} X' \\ & \quad - \sigma^2 \lambda X(X'X)^{-1} \mathbb{E} \left\{ \left[I + \sigma^2 \left(\frac{u'Mu}{n-k} \right) N \right]^{-1} X'uu' \right\} \\ & \quad - \sigma^2 \lambda \mathbb{E} \left\{ uu'X \left[I + \sigma^2 \left(\frac{u'Mu}{n-k} \right) N' \right]^{-1} \right\} (X'X)^{-1} X' \\ & \quad + \sigma^2 \lambda^2 \mathbb{E}(uu') \\ &= \sigma^2 [(1-\lambda)^2 I - (1-2\lambda)M] - \sigma^4 \left[1 - 2 \left(\lambda + \frac{1}{n-k} \right) \right] X(X'X)^{-1}NX' \\ & \quad + \sigma^6 \left(1 + \frac{2}{n-k} \right) \left[1 - 2 \left(\lambda + \frac{4}{n-k} \right) \right] X(X'X)^{-1}N^2X' + O(\sigma^8) \end{aligned}$$

proceeding in the same manner as indicated for (A.4). Similarly, we find

$$\begin{aligned} & \mathbb{E}(\hat{T}_{sm} - T)(\hat{T}_{sm} - T)' \\ &= X(X'X)^{-1} \mathbb{E} \left(I + \frac{Q_s}{n-k}N \right)^{-1} \left[\sigma^2 X'uu'X(X'X)^{-1} + \left(\frac{Q_s}{n-k} \right)^2 N \right] \\ & \quad \left(I + \frac{Q_s}{n-k}N \right)^{-1} X' - \sigma^2 \lambda X(X'X)^{-1} \mathbb{E} \left[\left(I + \frac{Q_s}{n-k}N \right)^{-1} X'uu' \right] \\ & \quad - \sigma^2 \lambda \mathbb{E} \left[uu'X \left(I + \frac{Q_s}{n-k}N' \right)^{-1} \right] (X'X)^{-1} X' + \sigma^2 \lambda^2 \mathbb{E}(uu') \\ &= \sigma^2 [(1-\lambda)^2 I - (1-2\lambda)M] - \sigma^4 \left[1 - 2 \left(\lambda + \frac{1}{n-k} \right) \right] X(X'X)^{-1}NX' \\ & \quad + \sigma^6 \left[\left(1 + \frac{2}{n-k} \right) \left\{ 1 - 2 \left(\lambda + \frac{4}{n-k} \right) \right\} X(X'X)^{-1}N^2X' \right. \\ & \quad \left. + \frac{2a^2}{(n-k+2)\beta'X'X\beta} \left(\lambda + \frac{4}{n-k} \right) X(X'X)^{-1}NX' \right] + O(\sigma^8). \end{aligned}$$

Taking the difference, we find (4.6) of Theorem 4.1 and (4.9) of Theorem 4.2.

Finally, following (A.5), we observe that

$$\begin{aligned} \mathbb{E}(\hat{T}_{ms} - T) &= \mathbb{E}(\hat{T}_m - T) - \left(\frac{a}{n-k+2} \right) \mathbb{E} \left(\frac{Q_m}{b'_m X'X b_m} X b_m \right) \\ &= -\sigma^2 \frac{a(n-k)}{(n-k+2)\beta'X'X\beta} X\beta \end{aligned}$$

upto order $O(\sigma^2)$. In a similar manner, it is easy to see that

$$\begin{aligned}
& D(\hat{T}_m; \hat{T}_{ms}) \\
&= E(\hat{T}_m - T)(\hat{T}_m - T)' - E(\hat{T}_{ms} - T)(\hat{T}_{ms} - T)' \\
&= \left(\frac{a}{n-k+2} \right) E \left(\frac{Q_m}{b'_m X' X b_m} \right) \left[\hat{T}_m (\hat{T}_m - T)' + (\hat{T}_m - T) \hat{T}'_m \right] \\
&\quad - \left(\frac{a}{n-k+2} \right) E \left(\frac{Q_m}{b'_m X' X b_m} \right)^2 \hat{T}_m \hat{T}'_m \\
&= \sigma^4 \frac{a(n-k)}{(n-k+2)\beta' X' X \beta} \left[2(1-\lambda)X(X'X)^{-1}X' - \frac{a+4(1-\lambda)}{\beta' X' X \beta} X\beta\beta' X' \right]
\end{aligned}$$

upto order $O(\sigma^4)$, leading to the result (4.5) stated in Theorem 4.1.

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. and Toutenburg, H. (1995). *Linear Models: Least Squares and Alternatives (corrected second printing, 1997)*, Springer, New York.
- Shalabh (1995). Performance of Stein-rule procedure for simultaneous prediction of actual and average values of study variable in linear regression model, *Bulletin of the International Statistical Institute* **56**: 1375–1390.
- Srivastava, V. K. and Srivastava, A. K. (1983). Improved estimation of coefficients in regression models with incomplete prior information, *Biometrical Journal* **25**: 775–782.
- Theil, H. and Goldberger, A. S. (1961). On pure and mixed estimation in econometrics, *International Economic Review* **2**: 65–78.
- Toutenburg, H. and Shalabh (1996). Predictive performance of the methods of restricted and mixed regression estimators, *Biometrical Journal* **38**: 951–959.
- Zellner, A. (1994). Bayesian and non-bayesian estimation using balanced loss functions, in S. S. Gupta and J. O. Berger (eds), *Statistical Decision Theory and Related Topics V*, Springer, Berlin, pp. 377–390.