



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

INSTITUT FÜR STATISTIK  
SONDERFORSCHUNGSBEREICH 386



Toutenburg, Shalabh:

Use of prior information in the form of interval constraints for the improved estimation of linear regression models with some missing responses

Sonderforschungsbereich 386, Paper 240 (2001)

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

Projektpartner



# Use of Prior Information in the Form of Interval Constraints for the Improved Estimation of Linear Regression Models with Some Missing Responses

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

Shalabh  
Department of Statistics, Panjab University  
Chandigarh, India

May 9, 2001

## Abstract

We have considered the estimation of coefficients in a linear regression model when some responses on the study variable are missing and some prior information in the form of lower and upper bounds for the average values of missing responses is available. Employing the mixed regression framework, we have presented five estimators for the vector of regression coefficients. Their exact as well as asymptotic properties are discussed and superiority of one estimator over the other is examined.

## 1 Introduction

One of the frequently encountered features of the data collection process is the missingness of some responses due to a variety of reasons. For example, in biological and medical sciences, some of the subjects( animal or human being ) may die before the completion of experiment. Similarly, in agricultural studies, some plants may be destroyed during the investigation. Likewise, in socio-economic surveys, some persons may not be available to respond. Or some persons may knowingly or unknowingly provide incorrect or unbelievable responses so that they have to be discarded. In such situations, the standard statistical procedures developed for data with no missing values cannot be immediately and straightforwardly applied for deducing inferences. Consequently various strategies have been evolved to find imputed values for the data set accordingly; see, e. g., Little and Rubin (1987) and Rubin (1987) for an excellent exposition. The thus repaired data resembles the data set without any missing value and permits the application of standard statistical procedures for analysis. This

proposition, however, generally disturbs the known optimal properties of the standard statistical procedures.

Sometimes some prior information about the missing responses may be available in the form of lower and upper bounds of their values. For example, if we are collecting data on various characteristics of households for a socio-economic study, some households in the lower stratum of the society may not disclose their true income. Some of them may refuse outrightly while others may supply figures which are false or unbelievable. Consequently, such values are considered inappropriate by the data collector and are therefore treated as missing observations. However, depending upon the observations on other characteristics of such households and matching them with complete observations or otherwise, it may be possible to specify an interval in each case such that it is expected to contain the income of that particular household. Such intervals specifying the lower and upper bounds may be derived from various sources like the data base, past experience of data collectors and users, long association with similar investigations, stability in repetitive studies, relationships with other correlates and / or some other extraneous considerations. In such circumstances, it may be instructive to utilize this kind of information in finding the imputed values for missing responses, estimating the regression parameters, constructing confidence regions and conducting tests of hypotheses.

When the prior information relates to regression coefficients in the form of interval constraints, the coefficients are estimated by method of interval restricted estimation using least squares procedures. The resulting estimators usually have no closed form except in some particular cases and therefore further analytical investigations are difficult to carry out; see, e. g. , Escobar and Skarpness (1986), Escobar and Skarpness (1987), Judge and Yancy (1986), Klemn and Sposito (1980), Ohtani (1987), Ohtani (1991), Srivastava and Ohtani (1995) and Wan (1996). Free from such a limitation and easy to implement are two alternative procedures forwarded by Toutenburg and Roeder (1978); see also Toutenburg (1982). One procedure involves formulating an ellipsoid enclosing the cuboid defined by the interval constraints and then using the method of minimax linear estimation for the regression coefficients. The other procedure consists of deriving a set of linear stochastic restrictions by treating the lower and upper limits of the intervals as confidence limits and then applying the method of mixed regression estimation. Toutenburg and Srivastava (1996) have presented a critical appraisal of the two procedures for the estimation of regression coefficients and have discussed their asymptotic properties when disturbances are small. A general conclusion emerging from their investigations is that the mixed regression formulation yields more efficient estimators, at least asymptotically.

In view of the above observations, we propose to employ the mixed regression framework for the estimation of coefficients in a linear regression model when some responses on the study variable are missing but prior information in the form of lower and upper bounds for the average values of the missing responses is available. Such a framework is described in Section 2. The estimation of regression coefficients is discussed in Section 3 and five estimators are presented. Exact expressions for the bias vectors and mean squared error matrices for the three estimators are provided and compared in Section 4. Similar expressions in case of the remaining two estimators can be derived but they will be sufficiently

intricate and will not permit us to draw any clear conclusion. In Section 5, we present large sample asymptotic approximations for the bias vectors and mean squared matrices of all the five estimators and carry out efficiency comparisons. The underlying asymptotic theory assumes that only the number of complete observations grows large; the number of incomplete observations stays fixed. Such a specification is relaxed in Section 6 and both are assumed to grow large. Under this supposition, the asymptotic approximations for the bias vectors and mean squared error matrices are given and a comparative study is made. Section 7 presents some concluding remarks. Lastly, the derivation of results stated in Theorems is given in Appendix.

## 2 Specification of Model

Consider the following linear regression model with some missing responses

$$y_c = X_c \beta + \epsilon_c \quad (2.1)$$

$$y_* = X_* \beta + \epsilon_* \quad (2.2)$$

where  $y_c$  is a  $m_c \times 1$  vector of observed responses,  $X_c$  is a  $m_c \times K$  matrix of  $m_c$  observations on  $K$  explanatory variables and is assumed to be of full column rank,  $\beta$  is a column vector of  $K$  regression coefficients and  $\epsilon_c$  is a column vector of  $m_c$  disturbances. Similarly,  $y_*$  denotes a  $m_* \times 1$  vector of missing responses,  $X_*$  is a  $m_* \times K$  matrix (not necessarily of full column rank) of  $m_*$  observations on  $K$  explanatory variables without any missing value and  $\epsilon_*$  is a column vector of  $m_*$  disturbances.

The elements of disturbance vectors  $\epsilon_c$  and  $\epsilon_*$  are assumed to be independently and identically distributed following a normal distribution with mean 0 and variance  $\sigma^2$ .

Suppose that there is some prior information available regarding the average values of the missing responses in the form of lower and upper bounds:

$$L_i \leq E(y_{*i}) \leq U_i \quad (i = 1, 2, \dots, m_*) \quad (2.3)$$

where  $L_i$  and  $U_i$  denote the lower and upper bounds respectively for the expected value of the  $i$ th element of  $y_*$ .

In order to utilize these constraints in the estimation of parameters, let us consider the framework of mixed regression estimation. Following Toutenburg and Srivastava (1996), let us treat them as  $p$  sigma limits so that the value of  $p$  is determined by considerations like credibility and truthfulness of restrictions (2.3). For instance, if we choose  $p = 3$ , the confidence coefficient is more than 88%.

Thus if we write

$$a_i = \frac{1}{2}(L_i - U_i) \quad (2.4)$$

$$\psi_i = \frac{1}{4p^2}(L_i - U_i)^2 \quad (2.5)$$

we can express

$$a_i = E(y_{*i}) + \epsilon_i \quad (i = 1, 2, \dots, m_*) \quad (2.6)$$

where  $\epsilon_i$  is a random variable with mean vector 0 and variance  $\psi_i$ .

Writing compactly, we find

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{m_*} \end{bmatrix} = \begin{bmatrix} E(y_{*1}) \\ E(y_{*2}) \\ \vdots \\ E(y_{*m_*}) \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{m_*} \end{bmatrix} \quad (2.7)$$

or

$$\begin{aligned} a &= E(y_*) + \epsilon \\ &= X_*\beta + \epsilon \end{aligned} \quad (2.8)$$

so that  $\epsilon$  is a  $m_* \times 1$  random vector with mean vector 0 and variance matrix  $\Psi$  where  $\Psi$  is a diagonal matrix with diagonal elements as  $\psi_1, \psi_2, \dots, \psi_{m_*}$ .

It may be observed that (2.1) and (2.7) together mimics the mixed regression framework because the vector  $a$  here is truly not a random vector. The equation (2.7) serves as a kind of approximate representation of the restrictions (2.3); see Toutenburg and Srivastava (1996) for details and its implications on the efficiency properties of the estimators for regression coefficients.

### 3 Estimation of Regression Coefficients

When only the complete observations are used, an application of least squares method to (2.1) provides the following estimator of  $\beta$ :

$$b_c = (X'_c X_c)^{-1} X'_c y_c . \quad (3.1)$$

Obviously, such an estimator discards the remaining incomplete observations in the available data set. We may therefore replace the elements of  $y_*$  in (2.2) by their predicted values as specified by  $X_* b_c$ . Now applying least squares method to the thus repaired model, we find the estimator of  $\beta$  as follows:

$$\begin{aligned} b &= (X'_c X_c + X'_* X_*)^{-1} (X'_c y_c + X'_* X_* b_c) \\ &= b_c. \end{aligned} \quad (3.2)$$

This implies that use of unbiased predicted values does not bring any improvement; see Yates (1933) and Rao and Toutenburg (1999, Chap. 8). Besides it, the estimator (3.2) does not utilize the available prior information.

If we consider the mixed regression framework specified by (2.1) and (2.2), an application of least squares procedure gives the following estimator of  $\beta$ :

$$b_{LS} = (X'_c X_c + X'_* X_*)^{-1} (X'_c y_c + X'_* a) \quad (3.3)$$

while an application of the feasible generalized least squares procedure provides another estimator of  $\beta$ :

$$b_{FGLS} = (X'_c X_c + s^2 X'_* \Psi^{-1} X_*)^{-1} (X'_c y_c + s^2 X'_* \Psi^{-1} a) \quad (3.4)$$

where

$$s^2 = \frac{1}{m_c - K} (y_c - X_c b_c)' (y_c - X_c b_c) \quad (3.5)$$

is an estimator of  $\sigma^2$ . One can take  $(m_c + m_* - K)$  instead of  $(m_c - K)$  in (3.5); see Rao and Toutenburg (1999, Section 8.2).

Now we can take  $X_* b_{LS}$  or  $X_* b_{FGLS}$  as imputed values for the missing responses in (2.2); see Toutenburg and Shalabh (1996), Toutenburg and Shalabh (2000) for the properties and some additional results. Using these to replace  $y_*$  in (2.2) and then applying least squares method to the thus obtained equation and (2.1), we find the following two estimators of  $\beta$ :

$$\hat{\beta}_1 = (X'_c X_c + X'_* X_*)^{-1} (X'_c y_c + X'_* X_* b_{LS}) \quad (3.6)$$

$$\hat{\beta}_2 = (X'_c X_c + X'_* X_*)^{-1} (X'_c y_c + X'_* X_* b_{FGLS}) \quad (3.7)$$

which do not reduce to  $b_c$  like (3.2).

## 4 Comparison of Estimators: Exact Results

It is easy to see that  $b_c$  is an unbiased estimator of  $\beta$  while the remaining four estimators  $b_{LS}$ ,  $b_{FGLS}$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are generally biased.

The bias vectors of  $b_{LS}$  and  $\hat{\beta}_1$  are given by

$$\begin{aligned} B(b_{LS}) &= E(b_{LS} - \beta) \\ &= (X'_c X_c + X'_* X_*)^{-1} X'_* (a - X_* \beta) \\ &= \delta \quad (\text{say}) \end{aligned} \quad (4.1)$$

$$\begin{aligned} B(\hat{\beta}_1) &= E(\hat{\beta}_1 - \beta) \\ &= (X'_c X_c + X'_* X_*)^{-1} X'_* X_* (X'_c X_c + X'_* X_*)^{-1} X'_* (a - X_* \beta) \\ &= (X'_c X_c + X'_* X_*)^{-1} X'_* X_* \delta . \end{aligned} \quad (4.2)$$

Similar expressions for the bias vectors of  $b_{FGLS}$  and  $\hat{\beta}_2$  can be derived but their functional forms will be sufficiently intricate and it will not be possible to draw any clear inference.

From (4.1) and (4.2), we observe that

$$\begin{aligned} [B(\hat{\beta}_1)]'[B(\hat{\beta}_1)] &= \delta' X_*' X_* (X_c' X_c + X_*' X_*)^{-2} X_*' X_* \delta \\ &\leq \delta' \delta = [B(b_{LS})]'[B(b_{LS})] \end{aligned} \quad (4.3)$$

implying that the estimator  $\hat{\beta}_1$  is superior to the estimator  $b_{LS}$  according to the criterion of the length of the bias vector.

The variance covariance matrix of  $b_c$  is given by

$$\begin{aligned} V(b_c) &= E(b_c - \beta)(b_c - \beta)' \\ &= \sigma^2 (X_c' X_c)^{-1} \end{aligned} \quad (4.4)$$

while the mean squared error matrices of  $b_{LS}$  and  $\hat{\beta}_1$  are

$$M(b_{LS}) = E(b_{LS} - \beta)(b_{LS} - \beta)' \quad (4.5)$$

$$= \sigma^2 [(X_c' X_c)^{-1} - \Delta] + \delta' \delta' \quad (4.6)$$

$$\begin{aligned} M(\hat{\beta}_1) &= E(\hat{\beta}_1 - \beta)(\hat{\beta}_1 - \beta)' \\ &= \sigma^2 (I_K + W)' [(X_c' X_c)^{-1} - \Delta] (I_K + W) + W' \delta \delta' W \end{aligned}$$

where

$$W = X_*' X_* (X_c' X_c + X_*' X_*)^{-1} \quad (4.7)$$

$$\Delta = (X_c' X_c)^{-1} - (X_c' X_c + X_*' X_*)^{-1} X_*' X_* (X_c' X_c + X_*' X_*)^{-1} . \quad (4.8)$$

As mentioned earlier, one can derive the exact expressions for the mean squared error matrices of the estimators  $b_{FGLS}$  and  $\hat{\beta}_2$  but they will be considerably complex in comparison to (4.5) and (4.7).

As the matrix

$$(X_c' X_c + X_*' X_*) (X_c' X_c)^{-1} (X_c' X_c + X_*' X_*) - (X_c' X_c)$$

is positive definite, the matrix  $\Delta$  is also so. Consequently, from (4.4) and (4.5), we have

$$D(b_c; b_{LS}) = V(b_c) - M(b_{LS}) = \sigma^2 \Delta - \delta \delta' \quad (4.9)$$

which is a nonnegative definite matrix by virtue of Rao and Toutenburg (1999, Theorem A.57, p. 370) if and only if

$$\delta' \Delta^{-1} \delta \leq \sigma^2 . \quad (4.10)$$

Similarly, the estimator  $\hat{\beta}_1$  can be compared with  $b_c$ . It is, however, difficult to deduce any neat condition from (4.4) and (4.7) for the superiority of one estimator over the other.

Finally, if we compare the expressions (4.5) and (4.7), it may be well appreciated that no clear inference regarding the superiority of  $b_{LS}$  over  $\hat{\beta}_1$  or vice-versa can be drawn.

## 5 Comparison of Estimators: Asymptotic Approximations When $m_c$ Is Large But $m_*$ Stays Fixed

In this Section, we utilize the large sample theory when the number of complete observations grows large while the number of incomplete observations stays fixed. For this purpose, we assume the asymptotic cooperativeness of the explanatory variables in the model so that the matrix  $m_c^{-1}X'_cX_c$  tends to a finite and nonsingular matrix  $Q$  as  $m_c$  grows large.

Utilizing the central limit theorem, it can be easily seen that all the five estimators  $b_c$ ,  $b_{LS}$ ,  $b_{FGLS}$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are consistent for  $\beta$ . Further, if we consider the asymptotic distribution of  $m_c^{1/2}$  times the estimation error, all the five estimators are found to have the same asymptotic distribution which is multivariate normal with mean vector 0 and variance covariance matrix  $\sigma^2Q^{-1}$ . We therefore need higher order approximations so as to discriminate among their performance properties.

Let us first introduce the following notation:

$$\begin{aligned} S_c &= \left(\frac{1}{m_c}X'_cX_c\right)^{-1} \\ \theta &= \frac{1}{m_*}X'_*(a - X_*\beta) \\ \theta_\Psi &= \frac{1}{m_*}X'_*\Psi^{-1}(a - X_*\beta) . \end{aligned} \quad (5.1)$$

From (4.1) and (4.2), it is easy to see that the bias vectors of  $b_{LS}$  and  $\hat{\beta}_1$  to order  $O(m_c^{-2})$  are given by

$$B(b_{LS}) = \frac{m_*}{m_c}S_c\theta - \frac{m_*}{m_c^2}S_cX'_cX_*S_c\theta \quad (5.2)$$

$$B(\hat{\beta}_1) = \frac{m_*}{m_c^2}S_cX'_cX_*S_c\theta . \quad (5.3)$$

Similar asymptotic results for the estimators  $b_{FGLS}$  and  $\hat{\beta}_2$  are derived in Appendix. These are stated below.

**Theorem I:** The bias vector of  $b_{FGLS}$  to order  $O(m_c^{-2})$  is

$$B(b_{FGLS}) = \frac{m_*\sigma^2}{m_c}S_c\theta_\Psi - \frac{m_*}{m_c^2}\sigma^4S_cX'_c\Psi^{-1}X_*S_c\theta_\Psi \quad (5.4)$$

while the estimator  $\hat{\beta}_2$  is unbiased to this order of approximation. However, if we consider bias vector of order  $O(m_c^{-2})$ , it is given by

$$B(\hat{\beta}_2) = \frac{m_*\sigma^2}{m_c^2}S_cX'_cX_*S_c\theta_\Psi . \quad (5.5)$$



Comparing the estimators with respect to the bias vector to order  $O(m_c^{-1})$ , it is seen that the estimators  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are unbiased while  $b_{LS}$  and  $b_{FGLS}$  are generally biased. Comparing  $b_{LS}$  and  $b_{FGLS}$  with respect to the criterion of bias vector length to order  $O(m_c^{-2})$ , we observe from (5.2) and (5.4) that  $b_{FGLS}$  is preferable in comparison to  $b_{LS}$  when  $\Psi_i$  exceeds  $\sigma^2$  for all  $i$ . Such a condition is satisfied so long as  $\sigma^2$  is less than the minimum value of  $\Psi_i$ , i. e. ,

$$\min_i (U_i - L_i)^2 > 4p^2\sigma^2 . \quad (5.6)$$

The opposite is true, i. e. , the estimator  $b_{LS}$  is better than  $b_{FGLS}$  with respect to the criterion of the bias vector length when  $\sigma^2$  exceeds  $\Psi_i$  for all  $i$  which is satisfied as long as

$$\max_i (U_i - L_i)^2 < 4p^2\sigma^2 . \quad (5.7)$$

Similarly, we see from (5.3) and (5.5) that the estimator  $\hat{\beta}_2$  is superior to  $\hat{\beta}_1$  according to the criterion of the length of the bias vector to order  $O(m_c^{-2})$  when  $\Psi_i$  exceeds  $\sigma^2$  for all  $i$  which holds true so long as (5.6) is satisfied. The reverse is true when (5.7) holds good.

Summarizing the results, we conclude that the estimator  $b_c$  is unbiased while the estimators  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are almost unbiased; the remaining two estimators  $b_{LS}$  and  $b_{FGLS}$  are biased.

Next, let us compare the estimators with respect to the criterion of the asymptotic mean squared error matrix.

From (4.5) and (4.7), we observe that

$$M(b_{LS}) = \frac{\sigma^2}{m_c} S_c - \frac{1}{m_c^2} S_c (2\sigma^2 X_*' X_* - m_*^2 \theta \theta') S_c + O(m_c^{-3}) \quad (5.8)$$

$$M(\hat{\beta}_1) = \frac{\sigma^2}{m_c} S_c - \frac{2\sigma^2}{m_c^3} S_c X_*' X_* S_c X_*' X_* S_c + O(m_c^{-4}) . \quad (5.9)$$

Similar results for the estimators  $b_{FGLS}$  and  $\hat{\beta}_2$  are obtained in Appendix.

**Theorem II:** The mean squared error matrices of the estimators  $b_{FGLS}$  and  $\hat{\beta}_2$  are given by

$$M(b_{FGLS}) = \frac{\sigma^2}{m_c} S_c - \frac{\sigma^4}{m_c^2} S_c (2X_*' \Psi^{-1} X_* - m_*^2 \theta_\Psi \theta_\Psi') S_c + O(m_c^{-3}) \quad (5.10)$$

$$M(\hat{\beta}_2) = \frac{\sigma^2}{m_c} S_c - \frac{\sigma^4}{m_c^3} S_c [X_*' X_* S_c X_*' \Psi^{-1} X_* + X_*' \Psi^{-1} X_* S_c X_*' X_*] S_c + O(m_c^{-4}) . \quad (5.11)$$

Observing that the variance covariance matrix of  $b_c$  is

$$V(b_c) = \frac{\sigma^2}{m_c} S_c \quad (5.12)$$

we find from (5.8) that

$$D(b_c; b_{LS}) = \frac{1}{m_c^2} S_c X_*' [2\sigma^2 I_K - (a - X_*\beta)(a - X_*\beta)'] X_* S_c \quad (5.13)$$

which is nonnegative definite if and only if

$$(a - X_*\beta)'(a - X_*\beta) \leq 2\sigma^2$$

or

$$\sum_{i=1}^{m_*} [E(y_{*i}) - \frac{1}{2}(L_i + U_i)]^2 \leq 2\sigma^2 \quad (5.14)$$

where use has been made of Rao and Toutenburg (1999, Theorem A.57, p.370).

Thus the inequality (5.14) provides a necessary and sufficient condition for the superiority of  $b_{LS}$  over  $b_c$ .

Similarly, we observe from (5.10) and (5.12) that the matrix

$$D(b_c; b_{FGLS}) = \frac{\sigma^4}{m_c^2} S_c X_*' \Psi^{-1} [2\Psi - (a - X_*\beta)'(a - X_*\beta)] \Psi^{-1} X_* S_c \quad (5.15)$$

is nonnegative definite if and only if

$$(a - X_*\beta)' \Psi^{-1} (a - X_*\beta) \leq 2 \quad (5.16)$$

or

$$\sum_{i=1}^{m_*} \left[ \frac{2E(y_{*i}) - (L_i + U_i)}{(U_i - L_i)} \right]^2 \leq \frac{2}{p^2} \quad (5.17)$$

which is the condition for the superiority of  $b_{FGLS}$  over  $b_c$ .

Comparing  $b_c$  with  $\hat{\beta}_1$ , it is obvious from (5.8) that the estimator  $\hat{\beta}_1$  is better than  $b_c$ . Similarly, it is again clear from (5.11) that  $\hat{\beta}_2$  is superior to  $b_c$ . Thus both the estimators  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are found to be more efficient than  $b_c$ .

Next, let us compare the estimators  $b_{LS}$ ,  $b_{FGLS}$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  with respect to the criterion of mean squared error matrix.

Taking the performance criterion as mean squared error matrix to order  $O(m_c^{-2})$ , it can be easily seen through a perusal of expressions (5.8)–(5.11) that both the estimators  $b_{LS}$  and  $b_{FGLS}$  are superior to  $\hat{\beta}_1$  and  $\hat{\beta}_2$  both so long as they are more efficient than  $b_c$ . On the other hand, when  $b_{LS}$  and  $b_{FGLS}$  are no better than  $b_c$ , i. e., the conditions (5.14) and (5.17) do not hold true, the estimators  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are not only nearly unbiased but more efficient too in comparison to the biased estimators  $b_{LS}$  and  $b_{FGLS}$ .

Comparing  $\hat{\beta}_1$  and  $\hat{\beta}_2$ , it is observed from (5.9) and (5.11) that

$$\begin{aligned} D(\hat{\beta}_1; \hat{\beta}_2) &= \frac{\sigma^2}{m_c^3} S_c [X_*' X_* S_c X_*' (\sigma^2 \Psi^{-1} - I) X_* \\ &\quad + X_*' (\sigma^2 \Psi^{-1} - I) X_* S_c X_*' X_*] S_c \end{aligned} \quad (5.18)$$

which is nonnegative definite implying the superiority of  $\hat{\beta}_2$  over  $\hat{\beta}_1$ , when  $\sigma^2$  is greater than  $\Psi_i$  for all  $i$ . This condition is satisfied as long as (5.7) holds good.

Similarly, it is seen from (5.17) that  $\hat{\beta}_1$  is superior to  $\hat{\beta}_2$  when  $\sigma^2$  is less than  $\Psi_i$  for all  $i$ . This condition is satisfied as long as (5.6) holds true.

## 6 Comparison of estimators: Asymptotic Approximations When Both $m_c$ and $m_*$ are Large

Now let us analyze the properties of estimators when both the number of complete observations and the number of incomplete observations grow large. Such a specification has, however, little appeal from the viewpoint of situations encountered in practice. It is merely for academic pursuit and the results may interest practitioners to appreciate the consequences when  $m_c$  and  $m_*$  grow large. Accordingly, it is assumed that both the matrices  $m_c^{-1}X'_cX_c$  and  $m_*^{-1}X'_*X_*$  tend to finite and nonsingular matrices as  $m_c$  and  $m_*$  tend to infinity.

We first introduce the following notation:

$$\begin{aligned} S_c &= \left(\frac{1}{m_c}X'_cX_c\right)^{-1} \\ G &= \left(\frac{1}{m_*}X'_*X_*\right) \\ f &= \frac{m_*}{m_c} \\ G_\Psi &= \left(\frac{1}{m_*}X'_*\Psi^{-1}X_*\right) \end{aligned} \quad (6.1)$$

so that

$$(X'_cX_c + X'_*X_*)^{-1} = \frac{1}{m_c}(I_K + fS_cG)^{-1}S_c = \frac{1}{m_c}H \quad (6.2)$$

$$(X'_cX_c + \sigma^2X'_*\Psi^{-1}X_*)^{-1} = \frac{1}{m_c}(I_K + f\sigma^2S_cG_\Psi)^{-1}S_c = \frac{1}{m_c}H_\Psi. \quad (6.3)$$

From Appendix, we have the following results:

Theorem III: The leading terms in the bias vectors are given by

$$B(b_{LS}) = fH\theta \quad (6.4)$$

$$B(b_{FGLS}) = f\sigma^2H_\Psi\theta_\Psi \quad (6.5)$$

$$B(\hat{\beta}_1) = f^2HGH\theta \quad (6.6)$$

$$B(\hat{\beta}_2) = f^2\sigma^2HGH_\Psi\theta_\Psi \quad (6.7)$$

It may be mentioned that (6.4) and (6.6) are exact results while (6.5) and (6.7) are asymptotics.

These results clearly indicate that all the four estimators, viz.,  $b_{LS}$ ,  $b_{FGLS}$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are inconsistent for  $\beta$ , while  $b_c$  is a known consistent estimator.

If we compare the estimators with respect to the length of bias vector, it is easy to see that

$$\left[ B(\hat{\beta}_1) \right]' \left[ B(\hat{\beta}_1) \right] \leq \left[ B(b_{LS}) \right]' \left[ B(b_{LS}) \right] \quad (6.8)$$

$$\left[ B(\hat{\beta}_2) \right]' \left[ B(\hat{\beta}_2) \right] \leq \left[ B(b_{FGLS}) \right]' \left[ B(b_{FGLS}) \right] \quad (6.9)$$

which implies the superiority of  $\hat{\beta}_1$  over  $b_{LS}$  and  $\hat{\beta}_2$  over  $b_{FGLS}$ .

As the estimators are not consistent, we consider their variance covariance matrices rather than their mean squared error matrices for comparing them.

Theorem IV: The asymptotic approximations for the variance covariance matrix to order  $O(m_c^{-1})$  are given by

$$V(b_{LS}) = \frac{\sigma^2}{m_c} H S_c^{-1} H \quad (6.10)$$

$$V(b_{FGLS}) = \frac{\sigma^2}{m_c} (H_\Psi S_c^{-1} H_\Psi + 2f^2 \sigma^2 H_\Psi S_c^{-1} H_\Psi \theta_\Psi \theta_\Psi' H_\Psi S_c^{-1} H_\Psi) \quad (6.11)$$

$$V(\hat{\beta}_1) = \frac{\sigma^2}{m_c} (I_K + fGH) H S_c^{-1} H (I_K + fGH) \quad (6.12)$$

$$V(\hat{\beta}_2) = \frac{\sigma^2}{m_c} H [(I_K + fGH_\Psi) S_c^{-1} (I_K + fH_\Psi G) + 2f^4 \sigma^2 H_\Psi S_c^{-1} H_\Psi \theta_\Psi \theta_\Psi' H_\Psi S_c^{-1} H_\Psi G] H \quad (6.13)$$

It is rather inappropriate, in our opinion, to make a comparison of the consistent estimator  $b_c$  with the inconsistent estimators  $b_{LS}$ ,  $b_{FGLS}$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$ . We therefore delete it and restrict our attention to inconsistent estimators only.

It is interesting to observe from the expressions (6.10), (6.12) and (6.13) that  $b_{LS}$  is better than  $\hat{\beta}_1$  and  $\hat{\beta}_2$  according to the criterion of asymptotic variance covariance matrix.

Next, we observe that

$$H^{-1} - H_\Psi^{-1} = m_c X_*' (I_{m_*} - \sigma^2 \Psi^{-1}) X_* \quad (6.14)$$

is a positive definite matrix when  $\Psi_1, \Psi_2, \dots, \Psi_{m_*}$  are all greater than  $\sigma^2$  which holds true if  $\Psi_{\min}$  exceeds  $\sigma^2$  where  $\Psi_{\min}$  denotes minimum value in  $\Psi_i$ 's; see (5.6). And then  $(H_\Psi - H)$  is positive definite. The opposite is true, i. e. ,  $(H - H_\Psi)$  is positive definite when  $\Psi_{\max}$  is less than  $\sigma^2$  which holds so long as (5.7) holds good.

It thus follows from (6.10) and (6.11) that  $b_{LS}$  is better than  $b_{FGLS}$  when  $\Psi_{\min}$  is greater than  $\sigma^2$  which is satisfied as long as the condition (5.6) holds true.

Similarly, we have

$$\begin{aligned} D(b_{LS}; b_{FGLS}) &= V(b_{LS}) - V(b_{FGLS}) \\ &= \frac{\sigma^2}{m_c} [H S_c^{-1} H - H_\Psi S_c^{-1} H_\Psi \\ &\quad + 2f^2 \sigma^2 H_\Psi S_c^{-1} H_\Psi \theta_\Psi \theta_\Psi' H_\Psi S_c^{-1} H_\Psi] \end{aligned} \quad (6.15)$$

which is nonnegative definite if and only if

$$\theta'_\Psi H_\Psi S_c^{-1} H_\Psi (H S_c^{-1} H - H_\Psi S_c^{-1} H_\Psi)^{-1} H_\Psi S_c^{-1} H_\Psi \theta_\Psi < 2f^2 \sigma^2 \quad (6.16)$$

provided that  $\Psi_{\max}$  is less than  $\sigma^2$  which is satisfied so long as (5.7) holds true.

Under the condition (5.7) combined with (6.16), the estimator  $b_{FGLS}$  is better than  $b_{LS}$ .

In a similar manner, we can compare the expressions (6.12) and (6.13), and can deduce conditions for the superiority of  $\hat{\beta}_1$  over  $\hat{\beta}_2$  and vice-versa but such conditions are not attractive as they are hard to verify in any given application.

## 7 Some Remarks

We have considered the estimation of the coefficients in a linear regression model when some of the responses are missing but plausible lower and upper bounds for them are available. The mixed regression framework for the utilization of such prior information is presented and five estimators for the vector of regression coefficients are envisaged. One estimator is the traditional least squares estimator which discards the incomplete observations and ignores the prior information too. The remaining four estimators use the incomplete observations as well as the prior restrictions. Efficiency properties of these estimators are analyzed employing some exact as well as asymptotic results and conditions are obtained for the superiority of one estimator over the other. However, some further exploration is required. For instance, it will be interesting to conduct a simulation experiment and to investigate the properties of estimators. Application of these estimators to some real-world data sets may also lead to some interesting findings and help us in identifying the situations where use of prior information may bring substantial gain in efficiency. It will be interesting to carry out a similar investigation for evaluating the performance of estimators arising from the ellipsoidal formulation of restrictions (2.3) and use of minimax linear estimation approach considered by Toutenburg and Wargowske (1978).

## Appendix

For the derivation of the results in Theorem I and Theorem II, let us write

$$\begin{aligned} u &= \frac{1}{m_c^{1/2}} X_c' \epsilon_c \\ w &= \left( \frac{1}{m_c^{1/2}} \epsilon_c' \epsilon_c - m_c^{1/2} \sigma^2 \right) \end{aligned}$$

so that  $u$  and  $w$  are of order  $O_p(1)$ .

Using these, we observe from (2.1), (3.4) and (3.5) that

$$\begin{aligned}
s^2 &= (1 - \frac{K}{m_c})^{-1} (\sigma^2 + \frac{w}{m_c^{1/2}} - \frac{u' S_c u}{m_c}) \quad (\text{A.1}) \\
&= \sigma^2 + \frac{w}{m_c^{1/2}} + \frac{K\sigma^2 - u' S_c u}{m_c} + O_p(m_c^{-3/2})
\end{aligned}$$

$$\begin{aligned}
(b_{FGLS} - \beta) &= (X'_c X_c + s^2 X'_* \Psi^{-1} X_*)^{-1} [X'_c \epsilon_c + s^2 X'_* \Psi^{-1} (a - X_* \beta)] \quad (\text{A.2}) \\
&= \left[ I + \left( \frac{\sigma^2}{m_c} + \frac{w}{m_c^{3/2}} \right) S_c X'_* \Psi^{-1} X_* + O_p(m_c^{-2}) \right]^{-1} \\
&\quad \cdot \left[ \frac{1}{m_c^{1/2}} S_c u + m_* \left( \frac{\sigma^2}{m_c} + \frac{w}{m_c^{3/2}} \right) S_c \theta_\Psi + O_p(m_c^{-2}) \right] \\
&= \frac{1}{m_c^{1/2}} S_c u + \frac{m_* \sigma^2}{m_c} S_c \theta_\Psi + \frac{1}{m_c^{3/2}} S_c (m_* w \theta_\Psi - \sigma^2 X'_* \Psi^{-1} X_* S_c u) \\
&\quad + O_p(m_c^{-2})
\end{aligned}$$

Retaining terms to order  $O(m_c^{-1})$ , we get

$$\begin{aligned}
B_{FGLS} &= E(b_{FGLS} - \beta) \\
&= \frac{m_* \sigma^2}{m_c} S_c \theta_\Psi
\end{aligned}$$

which is the result (5.4) of Theorem I.

Similarly, to order  $O(m_c^{-2})$ , we have

$$\begin{aligned}
M(b_{FGLS}) &= E(b_{FGLS} - \beta) E(b_{FGLS} - \beta)' \\
&= \frac{1}{m_c} S_c E(uu') S_c + \frac{\sigma^2}{m_c} S_c [\theta_\Psi E(u') + E(u) \theta'_\Psi] S_c \\
&\quad + \frac{1}{m_c} S_c [\sigma^4 m_*^2 \theta_\Psi \theta'_\Psi + m_* \theta_\Psi E(wu') + m_* E(wu) \theta'_\Psi \\
&\quad - \sigma^2 E(uu') S_c X'_* \Psi^{-1} X_* - \sigma^2 X'_* \Psi^{-1} X_* S_c E(uu')] S_c \\
&= \frac{1}{m_c} S_c - \frac{\sigma^4}{m_c^2} S_c (2X'_* \Psi^{-1} X_* - m_*^2 \theta_\Psi \theta'_\Psi) S_c
\end{aligned}$$

by virtue of the normality of disturbances.

This provides the result (5.10) of Theorem II.

In a similar manner, using (A.2), we can express

$$\begin{aligned}
(\hat{\beta}_2 - \beta) &= (X'_c X_c + X'_* X_*)^{-1} [X'_c \epsilon_c + X'_* X_* (b_{FGLS} - \beta)] \quad (\text{A.3}) \\
&= (I + \frac{1}{m_c} S_c X'_* X_*)^{-1} \left[ \frac{1}{m_c^{1/2}} S_c u + \frac{1}{m_c^{3/2}} S_c X'_* X_* S_c u \right. \\
&\quad + \frac{m_* \sigma^2}{m_c^2} S_c X'_* X_* S_c \theta_\Psi + \frac{1}{m_c^{5/2}} S_c X'_* X_* S_c (m_* w \theta_\Psi \\
&\quad \left. - \sigma^2 X'_* \Psi^{-1} X_* S_c u) + O_p(m_c^{-3}) \right] \\
&= \frac{1}{m_c^{1/2}} S_c u + \frac{m_* \sigma^2}{m_c^2} S_c X'_* X_* S_c \theta_\Psi \\
&\quad + \frac{1}{m_c^{5/2}} S_c X'_* X_* S_c (m_* w \theta_\Psi - \sigma^2 X'_* \Psi^{-1} X_* S_c u) + O_p(m_c^{-3}) .
\end{aligned}$$

Thus the bias vector to order  $O(m_c^{-2})$  is given by

$$B(\hat{\beta}_2) = \frac{m_* \sigma^2}{m_c^2} S_c X'_* X_* S_c \theta_\Psi$$

while the mean squared error matrix to order  $O(m_c^{-3})$  is

$$\begin{aligned}
M(\hat{\beta}_2) &= \frac{\sigma^2}{m_c} S_c + \frac{1}{m_c^3} S_c [m_* X'_* X_* S_c \theta_\Psi E(wu') + m_* E(wu) \theta'_\Psi S_c X'_* X_* \\
&\quad - \sigma^2 X'_* X_* S_c X'_* \Psi^{-1} X_* S_c E(uu') - \sigma^2 E(uu') S_c X'_* \Psi^{-1} X_* S_c X'_* X_*] S_c \\
&= \frac{\sigma^2}{m_c} S_c - \frac{\sigma^4}{m_c^3} S_c [X'_* X_* S_c X'_* \Psi^{-1} X_* + X'_* \Psi^{-1} X_* S_c X'_* X_*] S_c .
\end{aligned}$$

This establishes the result (5.5) of Theorem I and (5.11) of Theorem II.

Next, let us consider the results stated in Theorem III.

It can be easily seen that

$$\begin{aligned}
(b_{LS} - \beta) &= (X'_c X_c + X'_* X_*)^{-1} [X'_*(a - X_* \beta) + X'_c \epsilon_c] \quad (\text{A.4}) \\
&= (I_k + f S_c G)^{-1} (f S_c \theta + \frac{1}{m_c^{1/2}} S_c u) \\
&= H(f \theta + \frac{1}{m_c^{1/2}} S_c u)
\end{aligned}$$

whence we have

$$\begin{aligned}
(\hat{\beta}_1 - \beta) &= (X'_c X_c + X'_* X_*)^{-1} [X'_* X_* (b_{LS} - \beta) + X'_c \epsilon_c] \quad (\text{A.5}) \\
&= f (I_k + f S_c G)^{-1} S_c G (I_k + f S_c G)^{-1} S_c \theta \\
&\quad + (I_k + f S_c G)^{-1} [I_k + f S_c G (I_k + f S_c G)^{-1}] S_c u \\
&= f^2 H G H \theta + H (I_k + f G H) u .
\end{aligned}$$

Taking expectation, we find the results (6.4) and (6.6).

In a similar manner, using (A.1), we can express

$$\begin{aligned}
(b_{FGLS} - \beta) &= \left[ (I_K + f\sigma^2 S_c G_\Psi) + \frac{fw}{m_c^{1/2}} S_c G_\Psi + \dots \right]^{-1} \quad (\text{A.6}) \\
&\quad \cdot \left[ f\sigma^2 S_c \theta_\Psi + \frac{1}{m_c^{1/2}} S_c (wf\theta_\Psi + u) + \dots \right] \\
&= \left[ H_\Psi - \frac{fw}{m_c^{1/2}} S_c H_\Psi G_\Psi H_\Psi + \dots \right] \\
&\quad \left[ f\sigma^2 S_c \theta_\Psi + \frac{1}{m_c^{1/2}} (wf\theta_\Psi + u) + \dots \right] \\
&= f\sigma^2 H_\Psi \theta_\Psi + \frac{1}{m_c^{1/2}} \xi + O_p(m_c^{-1})
\end{aligned}$$

where

$$\xi = wfH_\Psi S_c^{-1} H_\Psi \theta_\Psi + H_\Psi u .$$

Likewise, utilizing (A.6), we can write

$$(\hat{\beta}_2 - \beta) = f^2 \sigma^2 H G H_\Psi \theta_\Psi + \frac{1}{m_c^{1/2}} H (u + fG\xi) + O_p(m_c^{-1}) . \quad (\text{A.7})$$

Taking expectations on both the sides in (A.6) and (A.7), we find the results (6.5) and (6.7) stated in Theorem III.

Finally, we consider the results of Theorem IV.

From (A.4), we observe that

$$\begin{aligned}
V(b_{LS}) &= E[b_{LS} - E(b_{LS})][b_{LS} - E(b_{LS})]' \\
&= \frac{1}{m_c} H E(uu') H = \frac{\sigma^2}{m_c} H S_c^{-1} H
\end{aligned}$$

which is the result (6.10).

Similarly, from (A.5), we find

$$V(\hat{\beta}_1) = \frac{\sigma^2}{m_c} H (I_K + fGH) S_c^{-1} (I_K + fHG) H$$

which yields the result (6.11).

Next, we observe that

$$\begin{aligned}
E(w^2) &= \frac{1}{m_c} E(\epsilon'_c \epsilon_c)^2 - 2\sigma^2 E(\epsilon'_c \epsilon_c) + m_c \sigma^4 \\
&= 2\sigma^4
\end{aligned}$$



so that

$$\begin{aligned} V(b_{FGLS}) &= \frac{1}{m_c} E(\xi\xi') \\ &= \frac{\sigma^2}{m_c} H_\Psi (S_c^{-1} + 2f^2\sigma^2 H_\Psi \theta_\Psi \theta_\Psi' H_\Psi) H_\Psi \end{aligned}$$

and

$$\begin{aligned} V(\hat{\beta}_2) &= \frac{1}{m_c} HE(u + fG\xi)(u' + f\xi'G)H \\ &= \frac{\sigma^2}{m_c} H[S_c^{-1} + f(GH_\Psi S_c^{-1} + S_c^{-1}H_\Psi G) \\ &\quad + f^2GH_\Psi S_c^{-1}H_\Psi G + 2\sigma^2 f^4 H_\Psi S_c^{-1}H_\Psi \theta_\Psi \theta_\Psi' H_\Psi S_c^{-1}H_\Psi]H \\ &= \frac{\sigma^2}{m_c} H[(I_K + fGH_\Psi)S_c^{-1}(I_K + fH_\Psi G) \\ &\quad + 2\sigma^2 f^4 H_\Psi S_c^{-1}H_\Psi \theta_\Psi \theta_\Psi' H_\Psi S_c^{-1}H_\Psi]H \end{aligned}$$

which provides the result (6.11) and (6.13).

## References

- Escobar, L. A. and Skarpness, B. (1986). The bias of the least squares estimator over interval constraints, *Economics Letter* **20**: 331–335.
- Escobar, L. A. and Skarpness, B. (1987). Mean square error efficiency of the least squares estimator over interval constraints, *Communications in Statistics, Part A—Theory and Methods* **16**: 397–406.
- Judge, G. G. and Yancy, T. (1986). *Improved Methods Of Inference In Econometrics*, North Holland, Amsterdam.
- Klemn, R. J. and Sposito, V. A. (1980). Least squares solutions over interval restrictions, *Communications in Statistics, Part B—Simulation and Computation* **9**: 423–425.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley, New York.
- Ohtani, K. (1987). The mse of a least squares estimator over an interval constraint, *Economics Letter* **25**: 351–351.
- Ohtani, K. (1991). Small sample properties of the interval constrained least squares estimator when error terms have a multivariate t-distribution, *Journal of the Japan Statistical Society* **21**: 197–204.
- Rao, C. R. and Toutenburg, H. (1999). *Linear Models: Least Squares and Alternatives*, 2 edn, Springer, New York.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Sample Surveys*, Wiley, New York.

- Srivastava, V. K. and Ohtani, K. (1995). A comparison of interval constrained least squares and mixed regression estimators, *Communications in Statistics, Part A—Theory and Methods* **24**: 395–413.
- Toutenburg, H. (1982). *Prior Information in Linear Models*, Wiley, New York.
- Toutenburg, H. and Roeder, B. (1978). Minimax-linear and Theil estimator for restrained regression coefficients, *Statistics* **9**: 499–505.
- Toutenburg, H. and Shalabh (1996). Predictive performance of the methods of restricted and mixed regression estimators, *Biometrical Journal* **38**: 951–959.
- Toutenburg, H. and Shalabh (2000). Improved predictions in linear regression models with stochastic linear constraints, *Biometrical Journal* **42**: 71–86.
- Toutenburg, H. and Srivastava, V. K. (1996). Estimation of regression coefficients subject to interval constraints, *Sankhya, Series A* **58**: 273–282.
- Toutenburg, H. and Wargowske, B. (1978). On restricted 2-stage-least-squares (2-SLSE) in a system of structural equations, *Statistics* **9**: 167–177.
- Wan, A. T. K. (1996). On the bias and mean squared error in a regression model with two inequality constraints and multivariate t error terms, *Communications in Statistics, Part A—Theory and Methods* **25**: 2079–2091.
- Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete, *Empire Journal of Experimental Agriculture* **1**: 129–142.