Toutenburg, Shalabh:

# Estimation of linear models with missing data: The role of stochastic linear constraints

Sonderforschungsbereich 386, Paper 239 (2001)

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

# Estimation of linear regression models with missing data: The role of stochastic linear constraints

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

Shalabh
Department of Statistics, Panjab University
Chandigarh 160014, India

January 9, 2001

**Abstract**

Assuming the nonavailability of some observations and the availability of some stochastic linear constraints connecting the coefficients in a linear regression, the technique of mixed regression estimation is considered and a set of five unbiased estimators for the vector of coefficeints is presented. They are compared with respect to the criterion of variance covariance matrix and conditions are obtained for the superiority of one estimator over the other.

## 1 Introduction

When the data available for statistical analysis involves some missing observations, a popular strategy is to find the imputed values for the missing responses on the basis of an analysis of the complete observations and then to use them for completing the data set. The thus obtained repaired data set can now be utilized for the statistical analysis. For example, if the aim is to estimate the coefficients in a linear model, one may first apply the least squares procedure to complete observations alone and employ the estimated regression relationship to find predicted values for the missing responses. These predicted values serve as imputed values for the missing observations and their substitution in place of missing responses provides a repaired data set. Now if we estimate the regression coefficients by least squares procedure using all the observations, the resulting estimator turns out to be precisely the same as the estimators which discards the complete observations; see, e. g., Little and Rubin (1987) and Rao and Toutenburg (1999, Chapter 8). In other words, repairing of data in this manner does not bring any improvement at all.

Quite often some stochastic linear restrictions connecting the regression coefficients are available from some sources other than the current sample. These restrictions, for instance, may araise from past investigations and/or repeated experimentation and/or long association with the resembling studies and/or some exogeneous source. When such a prior information is to be utilized in a non-Bayesian framework, the method of mixed regression estimation forwarded by Theil and Goldberger (1961) is available for the estimation of parameters; see, e. g. Srivastava (1980) for an annotated bibliography of earlier work, Judge, Griffiths, Hill, Lütkepohl and Lee (1985) for an interesting review and Rao and Toutenburg (1999) as well as Toutenburg and Shalabh (1996, 2000) for some recent results. Utilization of the mixed regression framework in the presence of some missing observations is the subject matter of this investigation with an aim to examine the role of prior restrictions in the efficient estimation of regression coefficients.

The organization of this paper is as follows. In Section 2, we describe the model in which some observations on the study variable are missing and a set of stochastic linear restrictions binding the regression coefficients is available. In all six estimators for the vector of regression coefficients are considered out of which six estimators are such that they can be readily used in practice. The first estimator discards the incomplete observations as well as the prior restrictions. The second estimator incorporates the prior restrictions but ignores the incomplete observations. The remaining four estimators utilize the prior restrictions as well as the complete and incomplete observations. In Section 3, the properties of these etimators are investigated and conditions for the superiority of one estimator over the other under the strong criterion of the variance covariance matrix are obtained. Section 4 provides some concluding remarks. Lastly, the Appendix gives derivation of main expressions.

## 2 Model specification and the estimators

Let us postulate a linear regression model involving some missing observations on the study variable.

Corresponding to the complete observations, we have

$$Y_c = X_c\beta + \sigma\epsilon_c \tag{2.1}$$

where $Y_c$ is a $n_c \times 1$ vector of $n_c$ observations on the study variable, $X_c$ is a $n_c \times K$ full column rank matrix of $n_c$ observations on $K$ explanatory variables, $\beta$ is a $K \times 1$ vector of regression coefficients, $\sigma$ is a positive scalar and $\epsilon_c$ is a $n_c \times 1$ vector of disturbances.

Similarly, corresponding to the incomplete observations, we have

$$Y_{mis} = X_m\beta + \sigma\epsilon_m \tag{2.2}$$

where $Y_{mis}$ denotes the $n_m \times 1$ vector of missing observations on the study variable, $X_m$ is a $n_m \times K$ matrix of $n_m$ observations on $K$ explanatory variables with no missing value and $\epsilon_m$ is a $n_m \times 1$ vector of disturbances.

Further, we are given a set of $J$ stochastic linear restrictions binding the regression coefficients of the model

$$r = R\beta + \epsilon \tag{2.3}$$

where $r$ is a $J \times 1$ vector of known values, $R$ is a $J \times K$ known matrix and $\epsilon$ is a random vector.

It is assumed that the elements of $\epsilon_c$ and $\epsilon_m$ are independently and identically distributed following a normal distribution with zero mean and unit variance. Similarly, $\epsilon$ has a multivariate normal distribution with mean vector null and known variance covariance matrix $\Sigma^{-1}$. Further, $\epsilon$ is stochastically independent of $\epsilon_c$ and $\epsilon_m$.

The application of least squares method to (2.1) provides the following estimator of $\beta$:

$$b_1 = (X_c'X_c)^{-1}X_c'Y_c \tag{2.4}$$

which discards the prior restrictions as well as the incomplete observations.

Similarly, the application of generalized least squares to (2.1) and (2.3) gives the following estimator:

$$\beta^* = (X_c'X_c + \sigma^2R'\Sigma R)^{-1}(X_c'Y_c + \sigma^2R'\Sigma r) \tag{2.5}$$

which incorporates the prior restrictions but ignores the incomplete observations.

Replacing $\sigma^2$ in (2.5) by its unbiased estimator $s^2 = (n_c - K)^{-1}(Y_c - X_cb_1)'(Y_c - X_cb_1)$, we get the mixed regression estimator as follows:

$$b_2 = (X_c'X_c + s^2R'\Sigma R)^{-1}(X_cY_c + s^2R'\Sigma r)\,. \tag{2.6}$$

Next, let us apply the least squares method to (2.1) and (2.2). This gives the following estimator:

$$\tilde{\beta}^* = (X_c'X_c + X_m'X_m)^{-1}(X_c'Y_c + X_m'Y_{mis}) \tag{2.7}$$

which utilizes all the complete and incomplete observations but does not use the prior restrictions.

This estimator has no utility owing to involvement of the vector $Y_{mis}$. Replacing it by its predicted values $X_m b_1$ and $X_m b_2$ yields the following estimators of $\beta$:

$$\begin{aligned} \tilde{\beta}_1 &= (X_c' X_c + X_m' X_m)^{-1}(X_c Y_c + X_m' X_m b_1) \qquad (2.8) \\ &= b_1 \\ \tilde{\beta}_2 &= (X_c' X_c + X_m' X_m)^{-1}(X_c Y_c + X_m' X_m b_2)\,. \qquad (2.9) \end{aligned}$$

Lastly, if we apply the generalized least squares method to (2.1), (2.2) and (2.3), we obtain the following estimator of $\beta$:

$$\hat{\beta}^* = (X_c' X_c + X_m' X_m + \sigma^2 R' \Sigma R)^{-1}(X_c' Y_c + X_m' Y_{mis} + \sigma^2 R' \Sigma r) \qquad (2.10)$$

which uses all the available information.

Substituting $s^2$ and $X_m b_1$ in place of $\sigma^2$ and $Y_{mis}$ in (2.10), we find a feasible version of $\hat{\beta}^*$ as follows:

$$\hat{\beta}_1 = (X_c' X_c + X_m' X_m + s^2 R' \Sigma R)^{-1}(X_c' Y_c + X_m' X_m b_1 + s^2 R' \Sigma r)\,. \qquad (2.11)$$

Similarly, if we utilize $b_2$ for obtaining a feasible version, we get

$$\hat{\beta}_2 = (X_c' X_c + X_m' X_m + s_{MR}^2 R' \Sigma R)^{-1}(X_c' Y_c + X_m' X_m b_2 + s_{MR}^2 R' \Sigma r)\,. \qquad (2.12)$$

where

$$s_{MR}^2 = \left(\frac{1}{n_c - K}\right)(Y_c - X_c b_2)'(Y_c - X_c b_2)\,. \qquad (2.13)$$

Instead of $b_2$ if we use $\tilde{\beta}_2$ and write

$$\tilde{\sigma}^2 = \left(\frac{1}{n_c - K}\right)(Y_c - X_c \tilde{\beta}_2)'(Y_c - X_c \tilde{\beta}_2)\,, \qquad (2.14)$$

another feasible version of $\hat{\beta}^*$ is as follows:

$$\hat{\beta}_3 = (X_c' X_c + X_m' X_m + \tilde{\sigma}^2 R' \Sigma R)^{-1}(X_c' Y_c + X_m' X_m \tilde{\beta}_2 + \tilde{\sigma}^2 R' \Sigma r)\,. \qquad (2.15)$$

We have thus six distinct estimators $b_1, b_2, \tilde{\beta}_2, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ that can be readily used in practice.

# 3 Efficiency Properties

It is easy to see that the estimator $b_1$ is unbiased with variance covariance matrix as

$$V(b_1) = \sigma^2 (X_c' X_c)^{-1} = \sigma^2 S_c (\text{say})\,. \qquad (3.1)$$

Similarly, using the result in Kakwani (1968), it can be straightforwardly seen that the estimators $b_2, \tilde{\beta}_2, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ are all unbiased for $\beta$. Their variance covariance matrix can be derived following Swamy and Mehta (1969) but the resulting expressions will be sufficiently complex and will not permit us to deduce neat conditions for the superiority of one estimator over the other. We therefore restrict our attention to their asymptotic approximations. For this purpose, we propose to employ the small disturbance asymptotic theory in view of findings reported by Srivastava and Upadhyaha (1975). Using it, the asymptotic variance covariance matrices are derived in Appendix and are stated below:

Theorem I: When disturbances are small, the asymptotic approximations for the variance covariance matrices of five estimators $b_2, \tilde{\beta}_2, \hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$ to order $O(\sigma^4)$ are given by

$$V(b_2) = \sigma^2 S_c - \sigma^4 \left(1 - \frac{2}{n_c - K}\right) S_c R' \Sigma R S_c \tag{3.2}$$

$$V(\tilde{\beta}_2) = \sigma^2 S_c - \sigma^4 [S_c R' \Sigma R S_c - S R' \Sigma R S - \frac{2}{(n_c - K)} (S_c - S) R' \Sigma R (S_c - S)] \tag{3.3}$$

$$V(\hat{\beta}_1) = \sigma^2 S_c - \sigma^4 [S_c R' \Sigma R S_c - (S_c - S) R' \Sigma R (S_c - S) - \frac{2}{(n_c - K)} S R' \Sigma R S] \tag{3.4}$$

$$V(\hat{\beta}_2) = \sigma^2 S_c - \sigma^4 \left(1 - \frac{2}{n_c - K}\right) S_c R' \Sigma R S_c \tag{3.5}$$

$$V(\hat{\beta}_3) = \sigma^2 S_c - \sigma^4 [S_c R \Sigma R S_c - (S - S S_c^{-1} S) R' \Sigma R (S - S S_c^{-1} S) - \frac{2}{(n_c - K)} (S_C - S + S S_c^{-1} S) R' \Sigma R (S_c - S + S S_c^{-1} S)] \tag{3.6}$$

where

$$S = (X_c' X_c + X_m' X_m)^{-1}. \tag{3.7}$$

In order to compare the variance covariance matrices, we first note that $(S_c - S)$ is a nonnegative matrix. Further, all the eigen values of $S$ in the metric of $S_c$ or equivalently the eigen values of $S S_c^{-1}$ and $S_c^{-1} S$ are nonnegative and do not exceed 1. Let the minimum and maximum eigen values be $\lambda$ and $\mu$ respectively.

Comparing (3.1) and (3.2), we observe that the variance covariance matrix of $b_1$ exceeds the variance covariance matrix of $b_2$ by a nonnegative definite matrix implying the superiority of mixed regression estimator $b_2$ over the pure estimator $b_1$ provided that $(n_c - K)$ is greater than 2.

Next, we notice that

$$\begin{aligned} D(b_2,\tilde{\beta}_2) &= V(b_2)-V(\tilde{\beta}_2) \\ &= \sigma^4 S_c[(\frac{2}{n_c-K})\{R'\Sigma R \\ &\quad -(I-S_c^{-1}S)R'\Sigma R(I-SS_c^{-1})\} \\ &\quad -S_c^{-1}SR'\Sigma RSS_c^{-1}]S_c \end{aligned} \tag{3.8}$$

which is nonnegative definite so long as

$$\left[\left(\frac{2}{n_c-K}\right)\{1-(1-\lambda)^2\}-\mu^2\right]>0$$

or

$$2\lambda(2-\lambda)>(n_c-K)\mu^2\,. \tag{3.9}$$

This is a sufficient condition for the superiority of $\tilde{\beta}_2$ over $b_2$, and it can be easily verified in practice.

The converse is true, i.e., the estimator $b_2$ is superior than $\tilde{\beta}_2$ at least as long as

$$2\mu(2-\mu)^2<(n_c-K)\lambda^2\,. \tag{3.10}$$

Similarly, we observe from (3.2), (3.4) and (3.6) that the estimator $\hat{\beta}_1$ is better than $b_2$ as long as

$$2(1-\mu^2)>(n_c-K)(1-\lambda)^2 \tag{3.11}$$

and $b_2$ is better than $\hat{\beta}_1$ so long as

$$2(1-\lambda^2)<(n_c-K)(1-\mu)^2\,. \tag{3.12}$$

Likewise, the estimator $\hat{\beta}_3$ is better than $b_2$ when

$$2(\lambda-\mu)^2(2-\lambda+\mu^2)>(n_c-K)(1-\lambda)\mu \tag{3.13}$$

and $b_2$ is better than $\hat{\beta}_3$ when

$$2(\mu-\lambda^2)(2-\mu+\lambda^2)<(n_c-K)(1-\mu)\lambda\,. \tag{3.14}$$

As $b_2$ and $\hat{\beta}_2$ have identical variance covariance matrices to order $O(\sigma^4)$, we need to consider the higher order terms so that the difference between them preciptates. This yields the following result:

<u>Theorem II:</u> To order $O(\sigma^6)$, we have

$$\begin{aligned} D(b_2,\hat{\beta}_2) &= V(b_2)-V(\hat{\beta}_2) \\ &= 2\sigma^6(\mathrm{tr}\,R'\Sigma RS_c)\left(1+\frac{2}{n_c-K}\right)^2 \\ &\quad (SR'\Sigma RS_c+S_cR'\Sigma RS)\,. \end{aligned} \tag{3.15}$$

This result indicates that $\hat{\beta}_2$ is uniformly superior to $b_2$.

Next, comparing $\tilde{\beta}_2$ with $\hat{\beta}_1$, we observe that

$$\begin{aligned} D(\tilde{\beta}_2, \hat{\beta}_1) &= V(\tilde{\beta}_2) - V(\hat{\beta}_1) &(3.16)\\ &= \sigma^4 \left(1 - \frac{2}{n_c - K}\right) [SR'\Sigma RS \\ &\quad -(S - S_c)R'\Sigma R(S - S_c)] &(3.17)\end{aligned}$$

which is nonnegative definite for $(n_c - K) > 2$. This implies the superiority of $\hat{\beta}_1$ over $\tilde{\beta}_2$ under the mild constraint that the excess of the number of observations over the number of coefficients exceeds two. However, $\tilde{\beta}_2$ is superior than $\hat{\beta}_1$ in the trivial case $(n_c - K) = 1$.

Similarly, $\tilde{\beta}_2$ is found to dominate $\hat{\beta}_2$ under the condition (3.9) and is dominated by $\hat{\beta}_2$ under the condition (3.10).

Next, we observe that

$$\begin{aligned} D(\tilde{\beta}_2, \hat{\beta}_3) &= V(\tilde{\beta}_2) - V(\hat{\beta}_3) &(3.18)\\ &= \sigma^2[SR'\Sigma RS - (S - SS_c^{-1}S)R'\Sigma R(S - SS_c^{-1}S) \\ &\quad + \left(\frac{2}{n_c - K}\right) \{(S_c - S)R'\Sigma R(S_c - S) \\ &\quad -(S_c - S + SS_c^{-1}S)R'\Sigma R(S_c - S + SS_c^{-1}S)\}]\end{aligned}$$

which is nonnegative definite so long as

$$2[\mu(1+\mu) - \lambda][2 - \lambda - \mu(1-\mu)] < (n_c - K)(2 - \lambda)\lambda\,. \tag{3.19}$$

This is a sufficient condition for the superiority of $\hat{\beta}_3$ over $\tilde{\beta}_2$.

A similar sufficient condition for the superiority of $\tilde{\beta}_2$ over $\hat{\beta}_3$ is

$$2[\lambda(1+\lambda) - \mu][2 - \mu - \lambda(1-\lambda)] > (n_c - K)(2 - \mu)\mu\,. \tag{3.20}$$

Likewise, the estimator $\hat{\beta}_1$ is better than $\hat{\beta}_2$ under the condition (3.11) while $\hat{\beta}_2$ is better than $\hat{\beta}_1$ under the constraint (3.12).

In a similar manner, it is seen that $\hat{\beta}_3$ is always better than $\hat{\beta}_1$ as the matrix expression

$$\begin{aligned} D(\hat{\beta}_1, \hat{\beta}_3) &= V(\hat{\beta}_1) - V(\hat{\beta}_3) &(3.21)\\ &= \sigma^4[(S - S_c)R'\Sigma R(S - S_c) \\ &\quad -(S - SS_c^{-1}S)R'\Sigma R(S - SS_c^{-1}S) \\ &\quad + \left(\frac{2}{n_c - K}\right) \{SR'\Sigma RS \\ &\quad -(S - S_c - SS_c^{-1}S)R'\Sigma R(S - S_c - SS_c^{-1}S)\}]\end{aligned}$$

is nonnegative definite by virtue of the nonnegative definiteness of the matrix $(SS_c^{-1}S - S_c)$.

Lastly, it follows from (3.5) and (3.6) that $\hat{\beta}_3$ is superior to $\hat{\beta}_2$ under the condition (3.13) while reverse is the case, i. e., $\hat{\beta}_2$ is superior to $\hat{\beta}_3$ under the condition (3.14).

For quick perusal, we have assembled all the comparison results in a tabular form and have stated sufficient conditions; see the accompanying table.

# 4 Some concluding remarks

For the estimation of coefficients in a linear regression model when some observations on the study variable are not available for some reasons but available are a set of stochastic linear constraints connecting the regression coefficients, we have presented six unbiased estimators and have compared their variance covariance matrices employing small disturbance asymptotic theory. In each case, we have succeeded in deducing a condition for the superiority of one estimator over the other and have presented them in a tabular form for ready reference. An interesting aspect of all such conditions is that they do not involve any unknown quantity and are therefore easy to check in practice. And thus a suitable choice of estimator can be exercised without any difficulty.

Once a choice of estimator is made, we confront the problem of estimating the standard error. A simple solution is to insert a consistent estimator of $\sigma^2$ in the asymptotic approximation for the variance covariance matrix presented in the preceding section. If one wishes to employ the unbiased estimator of the exact variance covariance matrix, it can be easily found following Giles and Srivastava (1991) who have derived such an expression for the traditional mixed regression estimator.

It may be remarked that our investigations can be straightforwardly extended to the situations where sample observations and prior information are required to be given unequal weights. One can then employ the technique of weighted mixed regression estimation developed by Schaffrin and Toutenburg (1990). In fact, if the situation demands that unequal weights should be assigned to complete observations, incomplete observations and prior information, it can be easily accomplished without any conceptual difficulty through at the cost of little analytical complexity.

Table 4.1: Sufficient conditions for the superiority of one estimator over the other

| | |
|---|---|
| Superiority of $b_2$ over | |
| $\tilde{\beta}_2$ | $2\mu(2-\mu)^2 < (n_c - K)\lambda^2$ |
| $\tilde{\beta}_1$ | $2(1-\lambda^2) < (n_c - K)(1-\mu)^2$ |
| $\hat{\beta}_2$ | Never |
| $\hat{\beta}_3$ | $2(\mu-\lambda^2)(2-\mu+\lambda^2) < (n_c - K)(1-\mu)\lambda$ |
| Superiority of $\tilde{\beta}_2$ over | |
| $b_2$ and $\hat{\beta}_2$ | $2\lambda(2-\lambda) > (n_c - K)\mu^2$ |
| $\hat{\beta}_1$ | $(n_c - K) = 1$ |
| $\hat{\beta}_3$ | $2[\lambda(1+\lambda)-\mu][2-\mu-\lambda(1-\lambda)] > (n_c - K)(2-\mu)\mu$ |
| Superiority of $\hat{\beta}_1$ over | |
| $b_2$ and $\hat{\beta}_2$ | $2(1-\mu^2) > (n_c - K)(1-\lambda)^2$ |
| $\tilde{\beta}_2$ | $(n_c - K) > 2$ |
| $\hat{\beta}_3$ | Never |
| Superiority of $\hat{\beta}_2$ over | |
| $b_2$ | Always |
| $\tilde{\beta}_2$ | $2\mu(2-\mu)^2 < (n_c - K)\lambda^2$ |
| $\hat{\beta}_1$ | $2(1-\lambda^2) < (n_c - K)(1-\mu)^2$ |
| $\hat{\beta}_3$ | $2(\mu-\lambda^2)(2-\mu+\lambda^2) < (n_c - K)(1-\mu)\lambda$ |
| Superiority of $\hat{\beta}_3$ over | |
| $b_2$ and $\hat{\beta}_2$ | $2(\lambda-\mu^2)(2-\lambda+\mu^2) > (n_c - K)(1-\lambda)\mu$ |
| $\tilde{\beta}_2$ | $2[\mu(1+\mu)-\lambda][2-\lambda-\mu(1-\mu)] < (n_c - K)(2-\lambda)\lambda$ |
| $\hat{\beta}_1$ | Always |

# Appendix

Derivation of Results

From (2.1) and (2.3), we can write

$$\begin{aligned}
(b_2 - \beta) &= [X_c'X_c + \sigma^2(\frac{\epsilon_c' M\epsilon_c}{n_c - K})R'\Sigma R]^{-1} \qquad \text{(A.1)} \\
&\quad [\sigma X_c'\epsilon_c + \sigma^2 \frac{\epsilon_c' M\epsilon}{n_c - K} R'\Sigma\epsilon] \\
&= [S_c - \sigma^2(\frac{\epsilon_c' M\epsilon_c}{n_c - K})S_c R'\Sigma R S_c + O_p(\sigma^4)] \\
&\quad [\sigma X_c'\epsilon_c + \sigma^2(\frac{\epsilon_c' M\epsilon_c}{n_c - K})R'\Sigma\epsilon] \\
&= \sigma S_c X_c'\epsilon_c + \sigma^2(\frac{\epsilon_c' M\epsilon_c}{n_c - K})S_c R'\Sigma\epsilon - \\
&\quad \sigma^3(\frac{\epsilon_c' M\epsilon_c}{n_c - K})S_c R'\Sigma R S_c X_c'\epsilon_c + O_p(\sigma^4)
\end{aligned}$$

where

$$M = I_{n_c} - X_c(X_c'X_c)^{-1}X_c' .$$

Observing that

$$\begin{aligned}
E[\epsilon_c' M\epsilon_c \cdot \epsilon_c'\epsilon_c] &= (\mathrm{tr}\, M)I_{n_c} + 2M \\
&= (n_c - K + 2)I_{n_c} + 2X_c(X_c'X_c)^{-1}X_c'
\end{aligned}$$

we find

$$\begin{aligned}
V(b_2) &= E(b_2 - \beta)(b_2 - \beta)' \\
&= \sigma^2 S_c X_c' E(\epsilon_c\epsilon_c') X_c S_c \\
&\quad + \frac{\sigma^3}{(n_c - K)} S_c [X_c' E(\epsilon_c' M\epsilon_c \cdot \epsilon_c) E(\epsilon')\Sigma R \\
&\quad + R'\Sigma E(\epsilon) E(\epsilon_c' M\epsilon_c \cdot \epsilon_c') X_c] S_c \\
&\quad + \frac{\sigma^4}{(n_c - K)} S_c [\frac{E(\epsilon_c' M\epsilon_c)^2}{(n_c - K)} R'\Sigma E(\epsilon\epsilon')\Sigma R \\
&\quad - R'\Sigma R S_c X_c' E(\epsilon_c' M\epsilon_c \cdot \epsilon_c\epsilon_c') X_c \\
&\quad - X_c' E(\epsilon_c M\epsilon_c \cdot \epsilon_c\epsilon_c') X_c S_c R'\Sigma R] S_c \\
&= \sigma^2 S_c - \sigma^4(1 - \frac{2}{n_c - K}) S_c R'\Sigma R S_c
\end{aligned}$$

which is the result (3.2) of Theorem I.

From (2.9), we observe that

$$\begin{aligned}
\tilde{\beta}_2 &= (X_c'X_c + X_m'X_m)^{-1}(X_c'X_c b_1 + X_m'X_m b_2) \\
&= b_2 - (X_c'X_c + X_m'X_m)^{-1}X_c'X_c(b_2 - b_1)
\end{aligned}$$

so that using (A.1), we can express

$$
\begin{aligned}
(\tilde{\beta}_2 - \beta) &= \sigma S_c X_c' \epsilon_c + \sigma^2 (\frac{\epsilon_c' M \epsilon_c}{n_c - K})(S_c - S) R' \Sigma \epsilon \\
&\quad - \sigma^3 (\frac{\epsilon_c' M \epsilon_c}{n_c - K})(S_c - S) R' \Sigma R S_c X_c' \epsilon_c + O_p(\sigma^4)
\end{aligned} \tag{A.2}
$$

from which the result (3.3) can be easily obtained.

Similarly, utilizing (A.1), we find

$$
\begin{aligned}
(\hat{\beta}_1 - \beta) &= \sigma S_c X_c' \epsilon_c + \sigma^2 (\frac{\epsilon_c' M \epsilon_c}{n_c - K}) S R' \Sigma \epsilon \\
&\quad - \sigma^3 (\frac{\epsilon_c M' \epsilon_c}{n_c - K}) S R' \Sigma R S_c X_c' \epsilon_c + O_p(\sigma^4)
\end{aligned} \tag{A.3}
$$

whence it is straightforward to deduce the expression (3.4).

Noticing that

$$
\begin{aligned}
s^2_{MR} &= \left(\frac{1}{n_c - K}\right) [\sigma \epsilon_c - X_c(b_2 - \beta)]' [\sigma \epsilon_c - X_c(b_2 - \beta)] \qquad \text{(A.4)} \\
&= \sigma^2 \left(\frac{\epsilon_c' M \epsilon_c}{n_c - K}\right) + \sigma^4 \left(\frac{\epsilon_c' M \epsilon_c}{n_c - K}\right)^2 \epsilon' \Sigma R S_c R' \Sigma \epsilon + O_p(\sigma^6) \\
\tilde{\sigma}^2 &= \left(\frac{1}{n_c - K}\right) [\sigma \epsilon_c - X_c(\tilde{\beta}_2 - \beta)]' [\sigma \epsilon_c - X_c(\tilde{\beta}_2 - \beta)] \qquad \text{(A.5)} \\
&= \sigma^2 \left(\frac{\epsilon_c' M \epsilon_c}{n_c - K}\right) + \sigma^4 \left(\frac{\epsilon_c' M \epsilon_c}{n_c - K}\right)^2 \epsilon' \Sigma R \\
&\quad (S_c - 2S + S S_c^{-1} S) R' \Sigma R \epsilon + O_p(\sigma^6)
\end{aligned}
$$

it can be easily seen that

$$
(\hat{\beta}_2 - \beta) = (b_2 - \beta) + O_p(\sigma^4) \tag{A.6}
$$

$$
\begin{aligned}
(\hat{\beta}_3 - \beta) &= \sigma S_c X_c' \epsilon_c + \sigma^2 (\frac{\epsilon_c' M \epsilon_c}{n_c - K})(S_c - S + S S_c^{-1} S) R' \Sigma R \\
&\quad - \sigma^3 (\frac{\epsilon_c' M \epsilon_c}{n_c - K})(S_c - S + S S_c^{-1} S) R' \Sigma R S_c X_c' \epsilon_c + O_p(\sigma^4 .)
\end{aligned} \tag{A.7}
$$

Using these expressions, we find the result (3.5) and (3.6) of Theorem I.

For the results in Theorem II, we need to retain the terms upto order $O(\sigma^5)$ in

(A.1) and (A.6). Such an exercise yields the following results:

$$\begin{aligned}(b_2-\beta) &= \sigma S_c X_c'\epsilon_c - \sigma^2(\frac{\epsilon_c' M\epsilon_c}{n_c-K})S_c R'\Sigma\epsilon \\ &+\sigma^3(\frac{\epsilon_c' M\epsilon_c}{n_c-K})S_c R'\Sigma R S_c X_c'\epsilon_c - \sigma^4(\frac{\epsilon_c' M\epsilon_c}{n_c-K})^2 S_c R'\Sigma R S_c R'\Sigma\epsilon \\ &+\sigma^5(\frac{\epsilon_c' M\epsilon_c}{n_c-K})^2 S_c R'\Sigma R S_c R'\Sigma R S_c X_c'\epsilon_c + O_p(\sigma^6)\end{aligned} \tag{A.8}$$

$$\begin{aligned}(\hat{\beta}_2-\beta) &= \sigma S_c X_c'\epsilon_c - \sigma^2(\frac{\epsilon_c' M\epsilon_c}{n_c-K})S_c R'\Sigma\epsilon \\ &+\sigma^3(\frac{\epsilon_c' M\epsilon_c}{n_c-K})S_c R'\Sigma R S_c X_c'\epsilon_c \\ &-\sigma^4(\frac{\epsilon_c' M\epsilon_c}{n_c-K})^2[S_c R'\Sigma R S_c R'\Sigma\epsilon - (\epsilon'\Sigma R S_c R'\Sigma\epsilon)S R'\Sigma\epsilon] \\ &+\sigma^5(\frac{\epsilon_c' M\epsilon_c}{n_c-K})^2[S_c R'\Sigma R S_c R'\Sigma R S_c X_c'\epsilon_c \\ &-(\epsilon'\Sigma R S_c R'\Sigma\epsilon) S R'\Sigma R S_c X_c'\epsilon_c] + O_p(\sigma^6)\end{aligned} \tag{A.9}$$

so that

$$\begin{aligned}&E(b_2-\beta)(b_2-\beta)' - E(\hat{\beta}_2-\beta)(\hat{\beta}_2-\beta)' \\ =\;& -\sigma^5 E[(\frac{\epsilon_c' M\epsilon_c}{n_c-K})^2(\epsilon'\Sigma R S_c R'\Sigma\epsilon) \\ &(S R'\Sigma\epsilon\epsilon_c' X_c S_c + S_c X_c'\epsilon_c\epsilon'\Sigma R S)] \\ &+\sigma^6 E[(\epsilon'\Sigma R S_c R'\Sigma\epsilon)\{(\frac{\epsilon_c' M\epsilon_c}{n_c-K})^2(S R'\Sigma R S_c X_c'\epsilon_c\epsilon_c' X_c S_c \\ &+S_c X_c'\epsilon_c\epsilon_c' X_c S_c R'\Sigma R S) + (\frac{\epsilon_c' M\epsilon_c}{n_c-K})^3(S R'\Sigma\epsilon\epsilon'\Sigma R S_c \\ &+S_c R'\Sigma\epsilon\epsilon'\Sigma R S)\}] + O(\sigma^7)\,.\end{aligned} \tag{A.10}$$

From Srivastava and Tiwari (1976), we have

$$\begin{aligned}E(\epsilon'\Sigma R S_c R'\Sigma\epsilon\cdot\epsilon\epsilon') &= (\mathrm{tr}\, R'\Sigma R S_c)\Sigma^{-1} + 2 R S_c R' \\ E[(\epsilon_c' M\epsilon_c)^2\epsilon_c\epsilon_c'] &= (n_c-K+2)[(n_c-K)I_{n_c} + 4M]\,.\end{aligned}$$

Substituting these in (A.10), we find the result stated in Theorem II.

# References

Giles, D. E. A. and Srivastava, V. K. (1991). An unbiased estimator of the covariance matrix of the mixed regression estimator, *Journal of the American Statistical Association* .

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H. and Lee, T.-C. (1985). *The theory and practice of econometrics*, Wiley.

Kakwani, N. C. (1968). Note on the unbiasedness of mixed regression estimation, *Econometrica* .

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, Wiley.

Rao, C. R. and Toutenburg, H. (1999). *Linear Models: Least Squares and Alternatives*, Springer.

Schaffrin, B. and Toutenburg, H. (1990). Weighted mixed regression, *Zeitschrift für Angewandte Mathematik und Mechanik* .

Srivastava, V. K. (1980). Estimation of linear single-equation and simultaneous-equation models under stochastic linear constraints: An annotated bibliography, *International Statistical Review* .

Srivastava, V. K. and Tiwari, R. (1976). Evaluation of expectations of products of stochastic matrices, *Scandinavian Journal of Statistics* .

Srivastava, V. K. and Upadhyaha, S. (1975). Small-disturbance and large sample approximations in mixed regression estimation, *Eastern Economic Journal* .

Swamy, P. A. V. B. and Mehta, J. S. (1969). On Theil's mixed regression estimator, *Journal of the American Statistical Association* .

Theil, H. and Goldberger, A. S. (1961). On pure and mixed estimation in econometrics, *International Economic Review* .

Toutenburg, H. and Shalabh (1996). Predictive performance of the methods of restricted and mixed regression estimators, *Biometrical Journal* .

Toutenburg, H. and Shalabh (2000). Improved Predictions in Linear Regression Models with Stochastic Linear Constraints, *Biometrical Journal* .