Toutenburg, Srivastava:

Improving the Estimation of Coefficients in Linear Regression Models with Some Missing Observations on Some Explanatory Variables

Sonderforschungsbereich 386, Paper 129 (1998)

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

Improving the Estimation of Coefficients in Linear Regression Models with Some Missing Observations on Some Explanatory Variables

H. Toutenburg* V. K. Srivastava**

September 3, 1998

1 Introduction

Considerable attention has been paid to various kinds of statistical issues in linear regression models when some observations on some of the explanatory variables are missing; see, e. g., Afifi and Elashoff (1966), Afifi and Elashoff (1967), Afifi and Elashoff (1969), Hartley and Hocking (1971) and Little (1992) for interesting reviews of literature. Among them, an important issue relates to the estimation of regression coefficients when the missing values in the available data set are replaced by some kind of imputed values and the model is thus repaired; see, e. g., Little (1992), Little and Rubin (1987) and Rao and Toutenburg (1995) for an interesting exposition of various imputation procedures.

If least squares method is used for the estimation of regression coefficients employing only the complete observations, the resulting estimators are unbiased. Contending that outright discard of the remaining incomplete data set may not necessarily be a good strategy, one may employ some kind of imputation procedure to find substitutes for missing observations. If the imputation procedure provides nonstochastic values for the replacement of missing observations and the least squares method is applied to the thus repaired model, the resulting estimators of regression coefficients are biased except in a trivial case where imputed values and true values of missing observations are identical. Performance properties of such estimators have been analyzed by Toutenburg, Heumann, Fieger and Park (1995); see also Rao and Toutenburg (1995, Chap. 8).

In view of the biased nature of least squares estimators, a question arises whether we can find other biased estimators having better efficiency properties. An effort in this direction is reported in this article. It essentially consists of applying the Stein-rule estimation method to the repaired model; see, e. g., Judge and Bock (1978) for a detail account of Stein-rule estimation method.

The plan of this article is as follows. Section 2 describes the linear regression model with missing observations and presents the estimators for regression coefficients. Section 3 presents the large sample asymptotic approximations for

*Institut für Statistik, Universität München, Akademiestr. 1, 80799 München, Germany
**Department of Statistics, Lucknow University, Lucknow 226007, India

the bias vector and the mean squared error matrix of the estimators. A comparison of efficiency properties is made in Section 4. Finally, some remarks are presented in Section 5.

2 Model Specification and Estimators

Let us consider the following linear regression model with missing observations:

$$y_c = X_c\beta + \epsilon_c \tag{2.1}$$
$$y_* = X_*\beta + \epsilon_* \tag{2.2}$$

where y_c and y_* are column vectors of T and m observations respectively on the study variable, X_c and X_* are matrices of T and m observations respectively on k explanatory variables, ϵ_c and ϵ_* are column vectors of T and m disturbances respectively and β is a column vector of k regression coefficients.

It is assumed that the matrix X_c has full column rank and is completely known while the matrix X_* may not necessarily have full column rank and is partially known in the sense that each row vector of X_* contains at least one value missing.

Finally, we assume that the elements of ϵ_c and ϵ_* are independently and identically distributed following a normal distribution with zero mean and unit variance.

We thus observe that equations (2.1) and (2.2) describe the linear regression model with complete and incomplete observations respectively.

If we apply the least squares method to (2.1) and (2.2) together, we get the following estimator of β:

$$b = (X_c'X_c + X_*'X_*)^{-1}(X_c'y_c + X_*'y_*) \tag{2.3}$$

which is the optimal estimator in the class of linear and unbiased estimators. This estimator has, however, no practical use as some elements in the matrix X_* are missing.

If we restrict our attention to complete data only and accordingly apply least squares to (2.1), we find the estimator of β as

$$b_c = (X_c'X_c)^{-1}X_c'y_c \tag{2.4}$$

which is unbiased.

If we wish to utilize incomplete data set also for the estimation of β, we need to employ some kind of imputation procedure so as to find substitutes for the missing elements in X_*. An interesting description of various imputation procedures is available in Little (1992, Sec. 4 and Sec. 7) and Rao and Toutenburg (1995, Sec. 8.3). Accordingly, let X_R denote a $m \times k$ matrix such that it is same as X_* except that missing values are replaced by nonstochastic quantities obtained from some imputation procedure.

Substituting X_R in place of X_* in (2.3), we get an operational version of least squares estimator:

$$b_R = (X_c'X_c + X_R'X_R)^{-1}(X_c'y_c + X_R'y_*) \tag{2.5}$$

which is a biased estimator of β.

As b_R is biased, it is tempting to consider other biased estimators which may have better performance properties than b_R. There could possibly be many ways to do it, but we propose to consider a shrunken estimator based on b_R. In particular, we choose to apply the method of Stein-rule estimation. This yields the following family of estimators for β:

$$\hat{\beta_R} = \left[1 - h^* \frac{(y_c - X_c b_R)'(y_c - X_c b_R) + (y_* - X_R b_R)'(y_* - X_R b_R)}{b_R'(X_c'X_c + X_R'X_R)b_R}\right] b_R \quad (2.6)$$

where $h^* = \left(\frac{h}{T+m-k+2}\right)$ and h is any positive nonstochastic scalar characterizing the estimator; see, e.g., Judge and Bock (1978).

3 Asymptotic Properties

Toutenburg et al. (1995) have presented exact expressions for the bias vector and mean squared error matrix of the estimator b_R and have examined its efficiency with respect to the estimator b_c. Similar expressions for the Stein-rule estimators can be derived following Judge and Bock (1978) but they will be sufficiently intricate and will not be helpful in deducing some clear inferences regarding the superiority of one estimator over the other. We therefore consider the large sample asymptotic approximations. For this purpose, we assume that the number (m) of incomplete observations stays fixed and only the number (T) grows large. Further, we assume the asymptotic cooperativeness of the explanatory variables in the model so that the limiting form of the matrix $(T^{-1}X_c'X_c)$ as T tends to infinity is finite and nonsingular.

First of all, we notice that the bias vector of b_c is null and its variance covariance matrix is given by

$$V(b_c) = \sigma^2(X_c'X_c)^{-1} = \frac{\sigma^2}{T} S \text{ (say)}. \quad (3.1)$$

Further, the distribution of $T^{\frac{1}{2}}(b_c - \beta)$ is multivariate normal with mean vector 0 and variance covariance matrix $\sigma^2 S$.

Similarly, if we consider the random vectors $T^{\frac{1}{2}}(b_R - \beta)$ and $T^{\frac{1}{2}}(\hat{\beta_R} - \beta)$, it can be easily verified by applying central limit theorem that both the quantities have identical asymptotic distributions, and this asymptotic distribution is same as the distribution of $T^{\frac{1}{2}}(b_c - \beta)$. Thus, on the basis of asymptotic distribution, we cannot prefer one estimator over the other. We therefore consider large sample asymptotic approximations for the estimators b_R and $\hat{\beta}_R$.

It is easy to see from (2.1) and (2.2) that

$$\begin{aligned}(b_R - \beta) &= (X_c'X_c + X_R'X_R)^{-1}[X_c'\epsilon_c + X_R'\epsilon_* + X_R'(X_* - X_R)\beta] \\ &= \left(I + \frac{1}{T}SX_R'X_R\right)^{-1}\left[\frac{1}{T^{\frac{1}{2}}}Su + \frac{1}{T}(SX_R'\epsilon_* + \delta)\right] \\ &= \frac{1}{T^{\frac{1}{2}}}Su + \frac{1}{T}(SX_R'\epsilon_* + \delta) - \frac{1}{T^{\frac{3}{2}}}SX_R'X_RSu + 0_p(T^{-2}), \quad (3.2)\end{aligned}$$

where $u = \frac{1}{T^{\frac{1}{2}}}X_c'\epsilon_c$ and $\delta = SX_R'(X_* - X_R)\beta$.

It thus follows from (3.2) that the bias vector of b_R to order $0(T^{-1})$ is

$$B(b_R) = \frac{1}{T}\delta\,. \tag{3.3}$$

Similarly, observing that

$$[b_R - E(b_R)] = S\left(\frac{1}{T^{\frac{1}{2}}}u + \frac{1}{T}X_R'\epsilon_* - \frac{1}{T^{\frac{3}{2}}}X_R'X_RSu\right) + 0_p(T^{-2})\,,$$

the variance covariance matrix to order $0(T^{-2})$ is

$$\begin{aligned} V(b_R) &= \frac{1}{T}S\left[E(uu') + \frac{1}{T^{\frac{1}{2}}}E(X_R'\epsilon_*u' + u\epsilon_*'X_R)\right. \\ &\quad \left. -\frac{1}{T}E(X_R'X_RSuu' + uu'SX_R'X_R - X_R'\epsilon_*\epsilon_*'X_R)\right]S \\ &= \frac{\sigma^2}{T}S - \frac{\sigma^2}{T^2}SX_R'X_RS\,. \end{aligned} \tag{3.4}$$

From (3.3) and (3.4), we obtain the mean squared error matrix of b_R to order $0(T^{-2})$ as follows:

$$M(b_R) = \frac{\sigma^2}{T}S - \frac{1}{T^2}(\sigma^2SX_R'X_RS - \delta\delta')\,. \tag{3.5}$$

It may be remarked that if we consider the exact expressions for the bias vector, variance covariance matrix and mean squared error matrix obtained by Toutenburg et al. (1995) and therefrom deduce large sample asymptotic approximations to the order of our approximation, they are found to match the results (3.3), (3.4) and (3.5).

For obtaining similar results in case of $\hat{\beta}_R$, we observe that

$$\begin{aligned} (y_c - X_cb_R) &= \epsilon_c - X_c(b_R - \beta) \\ (y_* - X_Rb_R) &= \epsilon_* + (X_* - X_R)\beta - X_R(b_R - \beta)\,. \end{aligned}$$

Using these along with (3.2) and writing

$$w = T^{-\frac{1}{2}}\epsilon_c'\epsilon_c - T^{\frac{1}{2}}\sigma^2$$

we can express

$$\begin{aligned} &\frac{(y_c - X_cb_R)'(y_c - X_cb_R) + (y_* - X_Rb_R)'(y_* - X_Rb_R)}{(T-m-k+2)b_R'(X_c'X_c + X_R'X_R)b_R} \qquad (3.6) \\ &= \frac{1}{T\beta'S^{-1}\beta}\left[\sigma^2 + \frac{w}{T^{\frac{1}{2}}} + 0_p(T^{-1})\right]\left[1 + \frac{m-k+2}{T}\right]^{-1} \times \\ &\quad \times\left[1 + 2\frac{\beta'u}{T^{\frac{1}{2}}\beta'S^{-1}\beta} + 0_p(T^{-1})\right]^{-1} \\ &= \frac{1}{T\beta'S^{-1}\beta}\left[\sigma^2 + \frac{w}{T^{\frac{1}{2}}} + 0_p(T^{-1})\right]\left[1 + 0_p(T^{-1})\right] \times \\ &\quad \times\left[1 - 2\frac{\beta'u}{T^{\frac{1}{2}}\beta'S^{-1}\beta} + 0_p(T^{-1})\right] \\ &= \frac{\sigma^2}{T\beta'S^{-1}\beta} + \frac{1}{T^{\frac{3}{2}}\beta'S^{-1}\beta}\left(w - 2\frac{\sigma^2\beta'u}{\beta'S^{-1}\beta}\right) + 0_p(T^{-2}) \end{aligned}$$

whence we find

$$\begin{aligned}(\hat{\beta}_R-\beta) &= (b_R-\beta)-\frac{h}{T\beta'S^{-1}\beta}\left[\sigma^2+\frac{1}{T^{\frac{1}{2}}}\left(w-2\frac{\sigma^2\beta'u}{\beta'S^{-1}\beta}\right)+0_p(T^{-1})\right]\times\\ &\quad\times[\beta+(b_R-\beta)]\\ &= \frac{1}{T^{\frac{1}{2}}}Su+\frac{1}{T}f-\frac{1}{T^{\frac{3}{2}}}g+0_p(T^{-2}) \qquad (3.7)\end{aligned}$$

where

$$\begin{aligned}f &= SX_R'\epsilon_*+\left(\delta-\frac{\sigma^2h}{\beta'S^{-1}\beta}\beta\right)\\ g &= \left[SX_R'X_RS+\frac{\sigma^2h}{\beta'S^{-1}\beta}\left(S-\frac{2}{\beta'S^{-1}\beta}\beta\beta'\right)\right]u+\frac{hw}{\beta'S^{-1}\beta}\beta\,.\end{aligned}$$

Thus the bias vector, to order $0(T^{-1})$, of the estimator $\hat{\beta}_R$ is

$$B(\hat{\beta}_R)=\frac{1}{T}\left(\delta-\frac{\sigma^2h}{\beta'S^{-1}\beta}\beta\right)\,. \qquad (3.8)$$

Similarly, we see that

$$[\hat{\beta}_R-E(\hat{\beta}_R)]=\frac{1}{T^{\frac{1}{2}}}Su+\frac{1}{T}SX_R'\epsilon_*-\frac{1}{T^{\frac{3}{2}}}g+0_p(T^{-2})$$

whence the variance covariance matrix to order $0(T^{-2})$ is

$$\begin{aligned}V(\hat{\beta}_R) &= \frac{1}{T}SE(uu')S+\frac{1}{T^{\frac{3}{2}}}SE(X_R'\epsilon_*u'+u\epsilon_*'X_R)S\\ &\quad-\frac{1}{T^2}E(gu'S+Sug'-SX_R'\epsilon_*\epsilon_*'X_RS)\,.\end{aligned}$$

By virtue of normality and stochastic independence of ϵ_c and ϵ_*, it is easy to verify that

$$V(\hat{\beta}_R)=\frac{\sigma^2}{T}S-\frac{\sigma^2}{T^2}\left[SX_R'X_RS+\frac{2\sigma^2h}{\beta'S^{-1}\beta}\left(S-\frac{2}{\beta'S^{-1}\beta}\beta\beta'\right)\right]\,. \qquad (3.9)$$

It thus follows from (3.8) and (3.9) that the mean squared error matrix of the estimator $\hat{\beta_R}$ is given by

$$\begin{aligned}M(\hat{\beta}_R) &= \frac{\sigma^2}{T}S-\frac{1}{T^2}(\sigma^2SX_R'X_RS-\delta\delta')\\ &\quad-\frac{\sigma^2h}{T^2\beta'S^{-1}\beta}\left[2\sigma^2S+\beta\delta'+\delta\beta'-\frac{\sigma^2(h+4)}{\beta'S^{-1}\beta}\beta\beta'\right] \qquad (3.10)\end{aligned}$$

to order $0(T^{-2})$ of approximation.

4 Efficiency Comparisons

Let us now compare the asymptotic properties of the three estimators b_c, b_R and $\hat{\beta}_R$ of β. Such a comparison may shed light on the usefulness of imputed values for repairing the model so far as the estimation of regression coefficients is concerned.

4.1 Bias

We have observed that the estimators b_R and $\hat{\beta}_R$ obtained from the imputation of missing observations are generally biased while the estimator b_c which ignores the incomplete observations is unbiased.

Comparing the two biased estimators, it is observed from (3.3) and (3.8) that $\hat{\beta}_R$ is better than b_R with respect to the criterion of length of bias vector when the characterizing scalar h satisfies the following constraint:

$$h < \frac{2}{\sigma^2 \beta'\beta} \beta' S^{-1} \beta \beta' S X_R' (X_* - X_R)\beta \tag{4.1}$$

which is not an attractive condition due to presence of unknown quantities.

4.2 Variability Around Mean Vector

First we state two results for any non-null column vector a and any positive definite matrix A of order $k \times k$.

Lemma 1: The matrix $(A - aa')$ is positive definite if and only if $a' A^{-1} a$ is less than 1.

Proof: See Farebrother (1976).

Lemma 2: The matrix $(aa' - A)$ cannot be non-negative definite for k greater than 1.

Proof: See Guilkey and Price (1981).

Now we observe from (3.1) and (3.4) that

$$D(b_c; b_R) = V(b_c) - V(b_R) = \frac{\sigma^2}{T^2} S X_R' X_R S \tag{4.2}$$

which is a non-negative definite matrix. This implies that the imputation procedure leads to a gain in efficiency when the criterion is variance covariance matrix to order $0(T^{-2})$.

Similarly, from (3.1) and (3.8), we have

$$\begin{aligned} D(b_c; \hat{\beta}_R) &= V(b_c) - V(\hat{\beta}_R) \\ &= \frac{\sigma^2}{T^2} \left[S X_R' X_R S + \frac{2\sigma^2 h}{\beta' S^{-1} \beta} \left(S - \frac{2}{\beta' S^{-1} \beta} \beta\beta' \right) \right] \end{aligned} \tag{4.3}$$

which is positive definite, using Lemma 1, if and only if

$$\frac{4\sigma^2 h}{(\beta' S^{-1} \beta)^2} \beta' \left(S X_R' X_R S + \frac{2\sigma^2 h}{\beta' S^{-1} \beta} S \right)^{-1} \beta < 1$$

or

$$\frac{2}{\beta' S^{-1} \beta} \beta' \left[S^{-1} - X_R' \left(X_R S X_R' + \frac{2\sigma^2 h}{\beta' S^{-1} \beta} I \right)^{-1} X_R \right] \beta < 1$$

or

$$\frac{\beta' X_R' \left(X_R S X_R' + \frac{2\sigma^2 h}{\beta' S^{-1} \beta} I \right)^{-1} X_R \beta}{\beta' S^{-1} \beta} > 1\,. \tag{4.4}$$

If $\alpha_{\min}$ and $\alpha_{\max}$ denote the minimum and maximum values among the characteristic roots of $X_R S X_R'$, we observe that the condition (4.4) is satisfied when

$$\frac{\alpha_{\min}}{\alpha_{\max} + \frac{2\sigma^2 h}{\beta' S^{-1} \beta}} > 1$$

which cannot hold true. Consequently, (4.4) can never hold good implying that $\hat{\beta}_R$ cannot be better than b_c.

Next, let us check whether the converse is true. Thus we have

$$\begin{aligned} D(\hat{\beta}_R; b_c) &= V(\hat{\beta}_R) - V(b_c) \\ &= \frac{\sigma^2}{T^2}\left[\frac{4\sigma^2 h}{(\beta' S^{-1}\beta)^2}\beta\beta' - \left(S X_R' X_R S + \frac{2\sigma^2 h}{\beta' S^{-1}\beta} S\right)\right] \end{aligned} \tag{4.5}$$

which cannot be non-negative by virtue of Lemma 2 except in the trivial case $k = 1$. This means that the estimator b_c cannot be better than $\hat{\beta}_R$ except in the special case of $k = 1$.

Finally, from (3.4) and (3.9), we get

$$D(b_R; \hat{\beta}_R) = V(b_R) - V(\hat{\beta}_R) = \frac{2\sigma^4 h}{T^2 \beta' S^{-1}\beta}\left(S - \frac{2}{\beta' S^{-1}\beta}\beta\beta'\right) \tag{4.6}$$

which cannot be positive definite from Lemma 1. Similarly, it follows from Lemma 2 that the matrix difference $D(\hat{\beta}_R; b_R)$ cannot be non-negative definte except when $k = 1$. Thus none of the two estimators b_R and $\hat{\beta}_R$ is generally superior to other.

Now let us compare the estimators with respect to a scalar measure of variability around mean vector. Choosing this weak criterion to be trace of S^{-1} times the variance covariance matrix to order $0(T^{-2})$, we observe from (4.2) that

$$\operatorname{tr} S^{-1} D(b_c; b_R) = \frac{\sigma^2}{T^2} \operatorname{tr} S X_R' X_R \tag{4.7}$$

which is obviously positive implying the superiority of b_R over b_c.

Similarly, from (4.3), we have

$$\operatorname{tr} S^{-1} D(b_c; \hat{\beta}_R) = \frac{\sigma^2}{T^2}\left[\operatorname{tr} S X_R' X_R + \frac{2\sigma^2 h(k-2)}{\beta' S^{-1}\beta}\right] \tag{4.8}$$

which is positive for $k > 1$. If $k = 1$, the condition for its positivity is

$$h > \frac{\beta' S^{-1}\beta}{2\sigma^2} \operatorname{tr} S X_R' X_R\,. \tag{4.9}$$

Comparing b_R and $\hat{\beta}_R$, we see from (4.6) that

$$\operatorname{tr} S^{-1} D(b_R; \hat{\beta}_R) = \frac{2\sigma^4 h(k-2)}{T^2 \beta' S^{-1}\beta} \tag{4.10}$$

whence it follows that $\hat{\beta}_R$ is better than b_R for all positive choices of characterizing scalar h provided that k exceeds 1. When $k = 1$, the estimator b_R is better than $\hat{\beta}_R$ for all positive values of h.

4.3 Variability Around True Parameter Vector

Now let us compare the estimators according to the criterion of mean squared error matrix to the order of our approximation.

From (3.1) and (3.5), we observe that

$$\Delta(b_c; b_R) = V(b_c) - M(b_R) = \frac{1}{T^2}(\sigma^2 S X_R' X_R S - \delta\delta') \tag{4.11}$$

which is positive definite, according to Lemma 1, if and only if

$$\frac{1}{\sigma^2}\delta' S^{-1}(X_R' X_R)^{-1} S^{-1}\delta < 1$$

provided that $(X_R'X_R)$ is invertible. When $(X_R'X_R)$ is not invertible, it is hard to determine the nature of matrix on the right hand side of (4.11).

On the other hand, if we consider the difference $\Delta(b_R; b_c)$, it follows from Lemma 2 that it cannot be positive definite for $k > 1$ provided that $(X_R'X_R)$ is nonsingular. If $(X_R'X_R)$ is singular, nothing definite can be said.

Similarly, from (3.1), (3.5) and (3.10), we find that

$$\begin{aligned}
\Delta(b_c; \hat\beta_R) &= V(b_c) - M(\hat\beta_R) \\
&= \frac{1}{T^2}(\sigma^2 S X_R' X_R S - \delta\delta') \\
&\quad + \frac{\sigma^2 h}{T^2 \beta' S^{-1}\beta}\left[2\sigma^2 S + \beta\delta' + \delta\beta' - \frac{\sigma^2(h+4)}{\beta' S^{-1}\beta}\beta\beta'\right] \qquad (4.12) \\
\Delta(b_R; \hat\beta_R) &= V(b_R) - M(\hat\beta_R) \\
&= \frac{\sigma^2 h}{T^2 \beta' S^{-1}\beta}\left[2\sigma^2 S + \beta\delta' + \delta\beta' - \frac{\sigma^2(h+4)}{\beta' S^{-1}\beta}\beta\beta'\right]. \qquad (4.13)
\end{aligned}$$

It is, however, difficult to draw any clear inference from these expressions regarding the superiority of one estimator over the other.

Next, let us compare the risk functions under a quadratic loss structure with loss matrix as S^{-1}.

Premultiplying (4.11) by S^{-1} and then taking trace, we observe that b_R has smaller risk in comparison to b_c when

$$\operatorname{tr} S X_R' X_R > \frac{1}{\sigma^2}\delta' S^{-1}\delta = \frac{1}{\sigma^2}\beta'(X_* - X_R)' X_R S X_R'(X_* - X_R)\beta \tag{4.14}$$

while the reverse is true when the condition (4.14) holds with an opposite inequality sign.

Similarly, it follows from (4.13) that $\hat\beta_R$ is better than b_R when

$$h < 2\left[(k-2) + \frac{1}{\sigma^2}\beta' X_R'(X_* - X_R)\beta\right] \tag{4.15}$$

provided that the quantity in square brackets on the right hand side is positive.

In a similar manner it is seen from (4.13) that $\hat\beta_R$ is better than b_c when

$$\left[\frac{\beta' S^{-1}\beta}{\sigma^2}\left(\operatorname{tr} S X_R' X_R - \frac{1}{\sigma^2}\delta' S^{-1}\delta\right) + 2h\left(k - 2 + \frac{1}{\sigma^2}\delta' S^{-1}\beta\right) - h^2\right] > 0 \tag{4.16}$$

which holds true so long as (4.14) and (4.15) are satisfied.

If the quantity on the left hand side of inequality (4.16) is negative, b_c is superior to $\hat{\beta}_R$. This is true at least as long as the inequalities (4.14) and (4.15) hold true with a reversed sign. Then b_c turns out to be better than both the estimators b_R and $\hat{\beta}_R$ implying that it is not worthwhile to employ any imputation procedure and it is better to use complete data set only.

5 Some Remarks

We have considered the problem of estimating the coefficients in a linear regression model when some observations on some explanatory variables are missing. For this purpose, we have followed two alternative strategies. One strategy consists of ignoring the incomplete data set and utilizing simply the complete observations. Now an application of least squares method yields unbiased estimators which may not be necessarily efficient. The other strategy consists of finding imputed values for missing observations through some imputation procedure and then employing the thus repaired data set. Now an application of least squares method provides generally biased estimators. Extending a bit further, we have considered the Stein-rule family of biased estimators.

As the unbiased as well as the biased estimators of regression coefficients are found to share the same asymptotic distributional properties, we have obtained large sample asymptotic approximations and have analyzed their performance properties with respect to criteria like the bias, variability around mean vector and variability around true coefficient vector. Such an exercise has helped us in specifying the situations where use of imputation procedure is worthwhile in comparison to the strategy of ignoring the incomplete observations and vice-versa.

Shrinkage techniques like ridge regression and Stein-rule estimation are well documented for their capabilities to handle the problems arising due to presence of harmful multicollinearity. It will therefore be interesting to examine the performance of Stein-rule estimators with respect to varying degree of harmful multicollinearity, for example, on the lines of Hill and Ziemer (1983). Other kinds of shrinkage estimators may be included and their relative performance may be analyzed. It will perhaps be equally interesting to investigate the behaviour of estimators when the distribution of disturbances departs from normality. For studying these issues, we are planning a study based on simulation and bootstrap methodologies, and we hope to come back with some findings in near future.

References

Afifi, A. A. and Elashoff, R. M. (1966). Missing observations in multivariate statistics: Part I: review of the literature, *Journal of the American Statistical Association* **61**: 595–604.

Afifi, A. A. and Elashoff, R. M. (1967). Missing observations in multivariate statistics: Part II: point estimation in simple linear regression, *Journal of the American Statistical Association* **62**: 10–29.

Afifi, A. A. and Elashoff, R. M. (1969). Missing observations in multivariate statistics: Part III: large sample analysis of simple linear regression, *Journal of the American Statistical Association* **64**: 331–358.

Farebrother, R. W. (1976). Further results on the mean square error of ridge regression, *Journal of the Royal Statistical Society, Series B* **38**: 248–250.

Guilkey, D. K. and Price, J. M. (1981). On comparing restricted least squares estimators, *Journal of Econometrics* **15**: 397–404.

Hartley, H. O. and Hocking, R. R. (1971). The analysis of incomplete data, *Biometrics* **27**: 783–808.

Hill, R. C. and Ziemer, R. F. (1983). Missing regressor values under conditions of multicollinearity, *Communications in Statistics, Part A–Theory and Methods* **12**: 2557–2573.

Judge, G. G. and Bock, M. E. (1978). *The statistical implications of pre-test and Stein-rule estimators in econometrics*, North Holland, Amsterdam.

Little, R. J. A. (1992). Regression with missing X's: a review, *Journal of the American Statistical Association* **87**: 1227–1237.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical analysis with missing data*, Wiley, New York.

Rao, C. R. and Toutenburg, H. (1995). *Linear Models: Least Squares and Alternatives (corrected second printing, 1997)*, Springer, New York.

Toutenburg, H., Heumann, C., Fieger, A. and Park, S. H. (1995). Missing values in regression: Mixed and weighted mixed estimation, *in* V. Mammitzsch and H. Schneeweiß (eds), *Gauss Symposium*, de Gruyter, Berlin, pp. 289–301.