Toutenburg, Srivastava, Fieger:

Shrinkage Estimation of Incomplete Regression Models by Yates Procedure

Sonderforschungsbereich 386, Paper 69 (1997)

Shrinkage Estimation Of Incomplete Regression Models By Yates Procedure

H. Toutenburg* V.K. Srivastava** A. Fieger*

May 7, 1997

Abstract

The problem of estimating the coefficients in a linear regression model is considered when some of the response values are missing. The conventional Yates procedure employing least squares predictions for missing values does not lead to any improvement over the least squares estimator using complete observations only. However, if we use Stein-rule predictions, it is demonstrated that some improvement can be achieved. An unbiased estimator of the mean squared error matrix of the proposed estimator of coefficient vector is also presented. Some work on the application of the proposed estimation procedure to real-world data sets involving some discrete variables in the set of explanatory variables is under way and will be reported in future.

1 Introduction

More than six decades ago, Yates (1933) has presented a procedure for the estimation of parameters when an incomplete data set due to some missing observations is available for the purpose of statistical analysis. The procedure essentially involves first estimating the parameters of model with the help of complete observations alone and obtaining the predicted values for the missing observations. These predicted values are then substituted in order to get a repaired or completed data set which is finally used for the estimation of parameters; see, e.g., Little and Rubin (1987, Chap. 2) and Rao and Toutenburg (1995, Chap. 8) for an interesting exposition. This strategy is adopted for the estimation of parameters in a linear regression model with some missing observations on the study variable. For the following it is assumed that missingness of the study variable y is independent of the value y itself and independent of the explanatory variables, so that MCAR holds. We begin with an application of the least squares method to estimate the parameters of model with complete observations for finding the predicted values to be substituted in place of missing observations. To the thus obtained repaired model, the least squares method is once again applied for estimating the regression coefficients. The resulting

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

estimators interestingly turn out to be the same as those found in the initial step, i.e., the least squares estimators gotten from complete observations alone. In other words, Yates procedure does not lead to any gain in efficiency so far as the least squares estimation of regression coefficients is concerned. This observation has prompted us to explore the possibility of any modification so that some improvement can be achieved in the performance properties of estimators. An attempt in this direction is the subject matter of present paper.

Shrinkage estimators arising from the traditional unbiased estimation of parameters are generally found to have superior performance properties under some mild constraints when the available data set has no missing observations. For example, the Stein-rule estimation of regression coefficients provides more efficient estimators than least squares under some modest constraint on the characterizing scalar; see, e.g., Judge and Bock (1978). Let us employ such an estimation procedure for the formulation of predicted values to be used for replacement of missing observartions. Using it to complete the data set, we now apply the least squares method for the estimation of regression parameters. It is proposed to study the efficiency properties of such a modified estimation procedure.

The plan of our presentation is as follows. Section 2 describes the model and the modified procedure for the estimation of regression coefficients. Asymptotic properties of the modified estimator are discussed in Section 3 under two specifications, viz., when the number of complete observations grows large but the number of missing observations stays fixed and when both the numbers grow large. Sufficient conditions for the superiority of modified procedure are deduced. These are simple and easy to check in any given application. They may also help practitioners in choosing an appropriate estimator. Section 4 presents a formula for the unbiased estimation of exact gain in efficiency of the modified procedure in comparison to the traditional procedure. Some concluding remarks are then placed in Section 5. Finally, the derivations of results stated in Theorems are outlined in Appendix.

2 Model Specification And The Estimation Procedures

Let us postulate the following linear regression model with some missing observations on the study variable:

$$\mathbf{y}_c = \mathbf{X}_c\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}_c \tag{2.1}$$

$$\mathbf{y}_{mis} = \mathbf{X}_*\boldsymbol{\beta} + \sigma\boldsymbol{\epsilon}_* \tag{2.2}$$

where $\mathbf{y}_c$ and $\mathbf{y}_{mis}$ denote the column vectors of m_c complete and m_* missing observations respectively on the study variable, $\mathbf{X}_c$ is a $m_c \times p$ full column rank matrix of m observations on p explanatory variables, $\mathbf{X}_*$ is similarly a $m_* \times p$ matrix of m_* observations on explanatory variables corresponding to missing values of study variable, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients, $\boldsymbol{\epsilon}_c$ and $\boldsymbol{\epsilon}_*$ are $m_c \times 1$ and $m_* \times 1$ vectors of disturbances and σ is an unknown scalar.

It is assumed that the elements of vectors $\boldsymbol{\epsilon}_c$ and $\boldsymbol{\epsilon}_*$ are independently and identically distributed following a normal probability law with zero mean and unit variance.

Writing (2.1) and (2.2) compactly, we get

$$\begin{pmatrix} \mathbf{y}_c \\ \mathbf{y}_{mis} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_c \\ \mathbf{X}_* \end{pmatrix} \boldsymbol{\beta} + \sigma \begin{pmatrix} \boldsymbol{\epsilon}_c \\ \boldsymbol{\epsilon}_* \end{pmatrix} . \tag{2.3}$$

Applying the least squares method to (2.3) for the estimation of $\boldsymbol{\beta}$, we find an estimator which cannot be used in practice due to involvement of m_* missing observations on the study variable. Now if we ignore such a set of m_* observations in the data and employ only the complete observations, application of least squares method yields the following estimator of $\boldsymbol{\beta}$:

$$\mathbf{b}_c = (\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y}_c . \tag{2.4}$$

As the estimator does not utilize the available set of m_* observations on the explanatory variables, Yates has suggested to repair the model by using predicted values for the missing observations on the study variable and then to apply least squares method employing the repaired or completed data set.

Now there are two popular ways for obtaining the predicted values of study variable. One is the least squares method which gives the following predictions for the missing observations on study variable:

$$\check{\mathbf{y}}_{mis} = \mathbf{X}_*\mathbf{b}_c \tag{2.5}$$

while the other is the Stein-rule method providing the following predictions:

$$\begin{aligned} \hat{\mathbf{y}}_{mis} &= \left[1 - \frac{kR_c}{(m_c - p + 2)\mathbf{b}_c'\mathbf{X}_c'\mathbf{X}_c\mathbf{b}_c}\right]\mathbf{X}_*\mathbf{b}_c \\ &= \left[1 - \frac{k\hat{\boldsymbol{\epsilon}}_c'\hat{\boldsymbol{\epsilon}}_c}{(m_c - p + 2)\hat{\mathbf{y}}_c'\hat{\mathbf{y}}_c}\right]\mathbf{X}_*\mathbf{b}_c \end{aligned} \tag{2.6}$$

where

$$\hat{\mathbf{y}} = \mathbf{X}_c\mathbf{b}_c \quad \text{and} \quad R_c = (\mathbf{y}_c - \mathbf{X}_c\mathbf{b}_c)'(\mathbf{y}_c - \mathbf{X}_c\mathbf{b}_c) = \hat{\epsilon}_c'\hat{\epsilon}_c \tag{2.7}$$

is the residual sum of squares and k is a positive nonstochastic scalar.

Ali and Abusalih (1988) have adopted a different approach for obtaining the predictions. They start with $f\check{\mathbf{y}}_{mis}$ as the vector of predictions for $\mathbf{y}_{mis}$ where f is a nonstochastic scalar. Assuming that the matrix $\mathbf{X}_*(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_*'$ is positive definite, the scalar f is chosen such that the quantity

$$\mathrm{E}(f\check{\mathbf{y}}_{mis} - \mathbf{y}_{mis})'[\mathbf{X}_*(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_*']^{-1}(f\check{\mathbf{y}}_{mis} - \mathbf{y}_{mis})$$

is minimum. Such an optimum value of f, however, turns out to involve some unknown parameters and does not serve any useful purpose. Replacing such unknowns by their estimators based on complete observations, they have presented the following predicitons for $\mathbf{y}_{mis}$ in the spirit of Stein-rule method:

$$\tilde{\mathbf{y}}_{mis} = \left[1 - \frac{hR_c}{(m_c - p + 2)\mathbf{b}_c'\mathbf{X}_*'[\mathbf{X}_*(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_*']^{-1}\mathbf{X}_*\mathbf{b}_c}\right]\mathbf{X}_*\mathbf{b}_c \tag{2.8}$$

where h is any positive nonstochastic scalar.

Notice that in case of $m_* > p$ the inverse $[\mathbf{X}_*(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_*']^{-1}$ does not exist and may be replaced by its g-inverse. In the following we assume $m_* \leq p$ and $\mathbf{X}_*$ of full row rank m_*, so that $\tilde{y}_{mis}$ (2.8) is well defined.

If we replace $\mathbf{y}_{mis}$ in (2.3) by $\check{\mathbf{y}}_{mis}$ and then apply the least squares method for estimating $\boldsymbol{\beta}$, we get the following estimator:

$$\check{\boldsymbol{\beta}} = (\mathbf{X}_c'\mathbf{X}_c + \mathbf{X}_*'\mathbf{X}_*)^{-1}(\mathbf{X}_c'\mathbf{y}_c + \mathbf{X}_*'\check{\mathbf{y}}_{mis}) \equiv \mathbf{b}_c \tag{2.9}$$

which means that least squares predictions do not bring any improvement at all in the estimation of regression coefficients.

If we replace $\mathbf{y}_{mis}$ in (2.3) by $\hat{\mathbf{y}}_{mis}$ and $\tilde{\mathbf{y}}_{mis}$ and then apply the least squares method, the following estimators of $\boldsymbol{\beta}$ are obtained:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}_c'\mathbf{X}_c + \mathbf{X}_*'\mathbf{X}_*)^{-1}(\mathbf{X}_c'\mathbf{y}_c + \mathbf{X}_*'\hat{\mathbf{y}}_{mis}) && (2.10)\\
&= \mathbf{b}_c - \frac{kR_c}{(m_c - p + 2)\mathbf{b}_c'\mathbf{X}_c'\mathbf{X}_c\mathbf{b}_c}\left[\mathbf{I} + (\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_c'\mathbf{X}_c\right]^{-1}\mathbf{b}_c\,. \\
\tilde{\boldsymbol{\beta}} &= (\mathbf{X}_c'\mathbf{X}_c + \mathbf{X}_*'\mathbf{X}_*)^{-1}(\mathbf{X}_c'\mathbf{y}_c + \mathbf{X}_*'\tilde{\mathbf{y}}_{mis}) && (2.11)\\
&= \mathbf{b}_c - \frac{hR_c}{(m_c - p + 2)\mathbf{b}_c'\mathbf{X}_*'\left[\mathbf{X}_*(\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_*'\right]^{-1}\mathbf{X}_*\mathbf{b}_c} \\
&\quad \times \left[\mathbf{I} + (\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{X}_c'\mathbf{X}_c\right]^{-1}\mathbf{b}_c
\end{aligned}$$

3 Efficiency Properties

It is well known that the estimator $\mathbf{b}_c$ is consistent and unbiased with variance covariance matrix as

$$V(\mathbf{b}_c) = \sigma^2(\mathbf{X}_c'\mathbf{X}_c)^{-1} = \frac{\sigma^2}{m_c}\mathbf{S}_c \quad \text{say}\,. \tag{3.1}$$

Similarly, it can be easily seen that $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ are consistent but biased. The exact expressions for their bias vectors and mean squared error matrices can be straightforwardly obtained, for example, from Judge and Bock (1978). However, they turn out to be intricate enough and may not lead to some clear inferences regarding the gain/loss in efficiency of $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ with respect to $\mathbf{b}_c$. We therefore consider their asymptotic approximations.

Application of large sample asymptotic theory requires $(m_c + m_*)$ to be large which can happen in three ways. First is that m_c stays fixed but m_* increases. Second is its converse, i.e., m_c increases but m_* stays fixed. Third is that both m_c and m_* increase. Out of these, first way is somewhat less appealing from the viewpoint of a practitioner. We therefore drop it and pursue the remaining two alternatives.

3.1 Specification: m_c Increases But m_* Stays Fixed

In order to analyze the asymptotic properties of $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ when m_c increases but m_* stays fixed, we assume that $\mathbf{S}_c$ as specified in (3.1) tends to a finite nonsingular matrix as m_c tends to infinity.

Theorem 1: *The asymptotic approximations for the bias vector of $\hat{\boldsymbol{\beta}}$ up to order $O(m_c^{-1})$ and the mean squared error matrix up to order $O(m_c^{-2})$ are given by*

$$
\begin{aligned}
B(\hat{\boldsymbol{\beta}}) &= -\frac{\sigma^2 k}{m_c \boldsymbol{\beta}' \mathbf{S}_c^{-1} \boldsymbol{\beta}} \boldsymbol{\beta} && (3.2)\\
M(\hat{\boldsymbol{\beta}}) &= \frac{\sigma^2}{m_c} \mathbf{S}_c - \frac{2\sigma^4 k}{m_c^2 \boldsymbol{\beta}' \mathbf{S}_c^{-1} \boldsymbol{\beta}} \left[\mathbf{S}_c - \left(\frac{4+k}{2\boldsymbol{\beta}' \mathbf{S}_c^{-1} \boldsymbol{\beta}} \right) \boldsymbol{\beta}\boldsymbol{\beta}' \right] && (3.3)
\end{aligned}
$$

while similar results for the estimator $\tilde{\boldsymbol{\beta}}$ are

$$
\begin{aligned}
B(\tilde{\boldsymbol{\beta}}) &= -\frac{\sigma^2 h}{m_c \boldsymbol{\beta}' \mathbf{X}_*' (\mathbf{X}_* \mathbf{S}_c \mathbf{X}_*')^{-1} \mathbf{X}_* \boldsymbol{\beta}} \boldsymbol{\beta} && (3.4)\\
M(\tilde{\boldsymbol{\beta}}) &= \frac{\sigma^2}{m_c} \mathbf{S}_c - \frac{2\sigma^4 h}{m_c^2 \boldsymbol{\beta}' \mathbf{X}_*' (\mathbf{X}_* \mathbf{S}_c \mathbf{X}_*')^{-1} \mathbf{X}_* \boldsymbol{\beta}} \Bigg[\mathbf{S}_c - && (3.5)\\
&\quad - \frac{1}{\boldsymbol{\beta}' \mathbf{X}_*' (\mathbf{X}_* \mathbf{S}_c \mathbf{X}_*')^{-1} \mathbf{X}_* \boldsymbol{\beta}} \Big(\boldsymbol{\beta}\boldsymbol{\beta}' \mathbf{X}_*' (\mathbf{X}_* \mathbf{S}_c \mathbf{X}_*')^{-1} \mathbf{X}_* \mathbf{S}_c \\
&\quad + \mathbf{S}_c \mathbf{X}_*' (\mathbf{X}_* \mathbf{S}_c \mathbf{X}_*')^{-1} \mathbf{X}_* \boldsymbol{\beta}\boldsymbol{\beta}' + \frac{h}{2} \boldsymbol{\beta}\boldsymbol{\beta}' \Big) \Bigg]
\end{aligned}
$$

Proof: See Appendix.

From (3.2) and (3.4), we observe that the elements of bias vectors have signs opposite to those of corresponding elements in $\boldsymbol{\beta}$. Further, the magnitude of bias declines as k or h respectively tends to be small and/or m_c grows large.

Observing that $\left[\mathbf{S}_c^{1/2} \mathbf{X}_*' (\mathbf{X}_* \mathbf{S}_c \mathbf{X}_*')^{-1} \mathbf{X}_* \mathbf{S}_c^{1/2} \right]$ is an idempotent matrix, we have

$$
\begin{aligned}
& \boldsymbol{\beta}' \mathbf{X}_*' (\mathbf{X}_* \mathbf{S}_c \mathbf{X}_*')^{-1} \mathbf{X}_* \boldsymbol{\beta} \\
&\quad = \boldsymbol{\beta}' \mathbf{S}_c^{-1/2} \left[\mathbf{S}_c^{1/2} \mathbf{X}_*' (\mathbf{X}_* \mathbf{S}_c \mathbf{X}_*')^{-1} \mathbf{X}_* \mathbf{S}_c^{1/2} \right] \mathbf{S}_c^{-1/2} \boldsymbol{\beta} \\
&\quad \leq \boldsymbol{\beta}' \mathbf{S}_c^{-1} \boldsymbol{\beta} && (3.6)
\end{aligned}
$$

whence it follows from (3.2) and (3.4) that if we take $k = h$, $\hat{\boldsymbol{\beta}}$ has smaller bias in magnitude than $\tilde{\boldsymbol{\beta}}$.

Comparing (3.1) and (3.3), we notice that the expression $[V(\mathbf{b}_c) - M(\hat{\boldsymbol{\beta}})]$ cannot be positive definite for positive values of k. Similarly, the expression $[M(\hat{\boldsymbol{\beta}}) - V(\mathbf{b}_c)]$ cannot be positive definite except in the trivial case of $p = 1$. We thus find that none of the two estimators $\mathbf{b}_c$ and $\hat{\boldsymbol{\beta}}$ dominates over the other with respect to the criterion of mean squared error matrix, at least to the order of our approximation. A similar result holds if we compare $\tilde{\boldsymbol{\beta}}$ with $\mathbf{b}_c$.

Next, let us compare $\mathbf{b}_c$ and $\hat{\boldsymbol{\beta}}$ with respect to the weaker criterion of risk under weighted squared error loss function. If $\mathbf{Q}$ denotes a positive definite matrix of order $O(1)$, the estimator $\hat{\boldsymbol{\beta}}$ is superior to $\mathbf{b}_c$ in the sense that $\operatorname{tr}\mathbf{Q}M(\hat{\boldsymbol{\beta}})$ is smaller than $\operatorname{tr}\mathbf{Q}V(\mathbf{b}_c)$ when

$$k < 2\left(\frac{\boldsymbol{\beta}'\mathbf{S}_c^{-1}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}}\operatorname{tr}\mathbf{Q}\mathbf{S}_c - 2\right) \tag{3.7}$$

provided that the quantity on the right hand side is positive.

The condition (3.7) is certainly satisfied if

$$k < 2(T-1)\ ; \quad T > 1 \tag{3.8}$$

where $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p$ denote the characteristic roots of $\mathbf{Q}\mathbf{S}_c$ and $T = \sum_{i=1}^{p-1}\frac{\lambda_i}{\lambda_p}$.

If we choose the MSE-III criterion, i.e. if we choose $\mathbf{Q} = \mathbf{S}_c^{-1}$, the inequality (3.7) assumes a simple form

$$k < 2(p-2) \tag{3.9}$$

provided that p exceeds 2.

Similarily, if we compare $\mathbf{b}_c$ and $\tilde{\boldsymbol{\beta}}$, it is seen that $\tilde{\boldsymbol{\beta}}$ is superior to $\mathbf{b}_c$ when

$$h < 2\left(\frac{\boldsymbol{\beta}'\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}}\operatorname{tr}\mathbf{Q}\mathbf{S}_c - \frac{2\boldsymbol{\beta}'\mathbf{Q}^{1/2}\mathbf{S}_c^{1/2}\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\mathbf{S}_c^{1/2}\mathbf{Q}^{1/2}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}}\right) \tag{3.10}$$

provided that the quantity on the right hand side is positive.

If $\phi_1 \leq \phi_2 \leq \ldots \leq \phi_p$ are the characteristic roots of the matrix

$$\mathbf{Q}^{1/2}\left[\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\right]^{-1}\mathbf{Q}^{1/2},$$

we have

$$\frac{\boldsymbol{\beta}'\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}} \geq \frac{1}{\phi_p}\,. \tag{3.11}$$

As $\mathbf{S}_c^{1/2}\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\mathbf{S}_c^{1/2}$ is an idempotent matrix, we have

$$\frac{\boldsymbol{\beta}'\mathbf{Q}^{1/2}\mathbf{S}_c^{1/2}\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\mathbf{S}_c^{1/2}\mathbf{Q}^{1/2}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta}} \leq 1 \tag{3.12}$$

Employing (3.11) and (3.12), we observe that the condition (3.10) is satisfied as long as

$$h < 2(\check{T}-2)\ ; \quad \check{T} = \left(\frac{1}{\phi_p}\sum_{i=1}^{p}\lambda_i\right) > 2\,. \tag{3.13}$$

For the choice $\mathbf{Q} = \mathbf{S}_c^{-1}$, the condition (3.10) reduces to the following:

$$h < 2(p-2)\left(\frac{\boldsymbol{\beta}'\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{S}_c^{-1}\boldsymbol{\beta}}\right) < 2(p-2)\ ; \quad p > 2 \tag{3.14}$$

where use has been made of (3.6).

Next, let us compare the estimators $\hat{\boldsymbol\beta}$ and $\tilde{\boldsymbol\beta}$. Supposing $k = h$ and $\mathbf{Q} = \mathbf{S}_c^{-1}$ for the sake of clarity in exposition, we observe from (3.3) and (3.5) that the estimator $\hat{\boldsymbol\beta}$ is superior to $\tilde{\boldsymbol\beta}$ with respect to risk criterion when

$$k > 2(p-2)\left(1 + \frac{\boldsymbol\beta'\mathbf{S}_c^{-1}\boldsymbol\beta}{\boldsymbol\beta'\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\boldsymbol\beta}\right)^{-1}; \quad p > 2 \tag{3.15}$$

which is satisfied as long as

$$k > 2(p-2)\,; \quad p > 2\,. \tag{3.16}$$

The reverse is true, i. e., $\tilde{\boldsymbol\beta}$ is superior to $\hat{\boldsymbol\beta}$ when the inequality for k in (3.15) holds with an opposite sign. This is satisfied as long as

$$k < (p-2)\,; \quad p > 2\,. \tag{3.17}$$

3.2 Specification: Both m_c And m_* Increase

Let us define m as m_c for m_c less than m_* and as m_* for m_c greater than m_* so that $m \to \infty$ is equivalent to $m_c \to \infty$ and $m_* \to \infty$.

Assuming the asymptotic cooperativeness of explanatory variables, i.e., $\mathbf{S}_c = m_c(\mathbf{X}_c'\mathbf{X})^{-1}$ and $\mathbf{S}_* = m_*(\mathbf{X}_*'\mathbf{X}_*)^{-1}$ tend to finite nonsingular matrices as m_c and m_* grow large, we have the following results for $\hat{\boldsymbol\beta}$.

Theorem 2: *For the estimator $\hat{\boldsymbol\beta}$, the asymptotic approximations for the bias vector to order $O(m^{-1})$ and the mean squared error matrix to order $O(m^{-2})$ are given by*

$$B(\hat{\boldsymbol\beta}) = -\frac{\sigma^2 k}{m_c\boldsymbol\beta'\mathbf{S}_c\boldsymbol\beta}\mathbf{G}\boldsymbol\beta \tag{3.18}$$

$$\begin{aligned} M(\hat{\boldsymbol\beta}) &= \frac{\sigma^2}{m_c}\mathbf{S}_c \\ &\quad - \frac{2\sigma^2 k}{m_c^2\boldsymbol\beta'\mathbf{S}_c^{-1}\boldsymbol\beta}\left[\mathbf{G}\mathbf{S}_c - \frac{1}{\boldsymbol\beta'\mathbf{S}_c^{-1}\boldsymbol\beta}(\mathbf{G}\boldsymbol\beta\boldsymbol\beta' + \boldsymbol\beta\boldsymbol\beta'\mathbf{G}' + \frac{k}{2}\mathbf{G}\boldsymbol\beta\boldsymbol\beta'\mathbf{G}')\right] \end{aligned} \tag{3.19}$$

where

$$\mathbf{G} = \mathbf{S}_c\left(\mathbf{S}_c + \frac{m_c}{m_*}\mathbf{S}_*\right)^{-1}. \tag{3.20}$$

Similar results for the estimator $\tilde{\boldsymbol\beta}$ do not exist because of the constraint $m_* \leq p$.

Choosing the preformance criterion to be the risk under weighted squared error loss function specified by weight matrix $\mathbf{Q}$ of order $O(1)$, we find from (3.1) and (3.19) that $\hat{\boldsymbol\beta}$ is superior to $\mathbf{b}_c$ when

$$k < 2\left(\frac{\boldsymbol\beta'\mathbf{S}_c^{-1}\boldsymbol\beta}{\boldsymbol\beta'\mathbf{G}'\mathbf{Q}\mathbf{G}\boldsymbol\beta}\operatorname{tr}\mathbf{Q}\mathbf{G}\mathbf{S}_c - \frac{2\boldsymbol\beta'\mathbf{G}'\mathbf{Q}\boldsymbol\beta}{\boldsymbol\beta'\mathbf{G}'\mathbf{Q}\mathbf{G}\boldsymbol\beta}\right) \tag{3.21}$$

provided that the quantity on the right hand side of the inequality is positive.

Let δ be the largest characteristic root of $\mathbf{QS}_*$ or $\mathbf{Q}^{1/2}\mathbf{S}_*\mathbf{Q}^{1/2}$ in the metric of $\mathbf{QS}_c$ or $\mathbf{Q}^{1/2}\mathbf{S}_c\mathbf{Q}^{1/2}$. Now we observe that

$$\begin{aligned} \operatorname{tr}\mathbf{QGS}_c &\geq \left(\frac{m_*}{m_*+\delta m_c}\right)\operatorname{tr}\mathbf{QS}_c \\ \frac{\boldsymbol{\beta}'\mathbf{G}'\mathbf{QG}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{S}_c^{-1}\boldsymbol{\beta}} &\leq \left(\frac{m_*}{m_*+\delta m_c}\right)^2\lambda_p \\ \frac{\boldsymbol{\beta}'\mathbf{G}'\mathbf{Q}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{G}'\mathbf{QG}\boldsymbol{\beta}} &\leq \left(\frac{m_*}{m_*+\delta m_c}\right) \end{aligned} \tag{3.22}$$

whence we see that the condition (3.21) is satisfied as long as

$$k < 2\left(1+\frac{\delta m_c}{m_*}\right)\left[T-2\left(\frac{m_*}{m_*+\delta m_*}\right)^{-1}\right] \; ; \quad T > \left(\frac{m_*}{m_*+\delta m_*}\right)^2 \tag{3.23}$$

which is easy to check in any given application owing to absence of $\boldsymbol{\beta}$.

Like (3.23), it may be added, one can derive various sufficient versions of the condition (3.21).

4 Estimation Of Gain In Efficiency

In applied work, we need an unbiased estimator of the exact mean squared error matrices of $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ so that standard errors of the estimates can be computed in order to appreciate the actual gain or loss in efficiency of $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ relative to $\mathbf{b}_c$. These standard errors may also help in finding confidence regions and conducting tests of hypotheses.

We now present an unbiased estimator of the expression

$$\begin{aligned} \mathbf{D} &= \mathrm{E}(\mathbf{b}_c-\boldsymbol{\beta})(\mathbf{b}_c-\boldsymbol{\beta})' - \mathrm{E}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})' \\ &= \left(\frac{k}{m_c-p+2}\right)\mathrm{E}\left(\frac{R_c}{\mathbf{b}_c'\mathbf{S}_c^{-1}\mathbf{b}_c}\right)[\mathbf{Gb}_c(\mathbf{b}_c-\boldsymbol{\beta})' + (\mathbf{b}_c-\boldsymbol{\beta})_\mathbf{c}'\mathbf{G}'] \\ &\quad - \left(\frac{k}{m_c-p+2}\right)^2\mathrm{E}\left[\left(\frac{R_c}{\mathbf{b}_c'\mathbf{S}_c^{-1}\mathbf{b}_c}\right)^2\mathbf{Gb}_c\mathbf{b}_c'\mathbf{G}'\right] \end{aligned} \tag{4.1}$$

where $\mathbf{G}$ is given by (3.22).

From Carter, Srivastava, Srivastava and Ullah (1990), we can express

$$\begin{aligned} &\mathrm{E}\left[\left(\frac{R_c}{\mathbf{b}_c'\mathbf{S}_c^{-1}\mathbf{b}_c}\right)\mathbf{b}_c(\mathbf{b}_c-\boldsymbol{\beta})'\right] \\ &= \left(\frac{1}{m_c-p+2}\right)\mathrm{E}\left[\left(\frac{R_c^2}{\mathbf{b}_c'\mathbf{S}_c^{-1}\mathbf{b}_c}\right)\left(\mathbf{S}_c-\frac{2}{\mathbf{b}_c'\mathbf{S}_c^{-1}\mathbf{b}_c}\mathbf{b}_c\mathbf{b}_c'\right)\right] . \end{aligned} \tag{4.2}$$

Using it in (4.1), we get

$$\mathbf{D} = \mathrm{E}(\hat{\mathbf{D}}) \tag{4.3}$$

where

$$\begin{aligned}
\hat{\mathbf{D}} &= \frac{2kR_c^2}{(m_c-p+2)^2\mathbf{b}_c'\mathbf{S}_c^{-1}\mathbf{b}_c}\left[\mathbf{G}\mathbf{S}_c - \frac{1}{\mathbf{b}_c'\mathbf{S}_c^{-1}\mathbf{b}_c}(\mathbf{G}\mathbf{b}_c\mathbf{b}_c' + \mathbf{b}_c\mathbf{b}_c'\mathbf{G}' + \frac{k}{2}\mathbf{G}\mathbf{b}_c\mathbf{b}_c'\mathbf{G}')\right] \\
&= \frac{2kR_c^2}{(m_c-p+2)^2\mathbf{b}_c'\mathbf{S}_c^{-1}\mathbf{b}_c}\left[\mathbf{S}_c(\mathbf{S}_c + \frac{m_c}{m_*}\mathbf{S}_*)^{-1}\mathbf{S}_c\right. \\
&\quad - \frac{1}{\mathbf{b}_c'\mathbf{S}_c^{-1}\mathbf{b}_c}\left(\left(\frac{m_c}{m_*}\mathbf{S}_c^{-1}\mathbf{S}_* + \mathbf{I}\right)^{-1}\mathbf{b}_c\mathbf{b}_c' + \mathbf{b}_c\mathbf{b}_c'\left(\frac{m_c}{m_*}\mathbf{S}_c^{-1}\mathbf{S}_* + \mathbf{I}\right)^{-1}\right. \\
&\quad \left.\left. + \frac{k}{2}\left(\frac{m_c}{m_*}\mathbf{S}_c^{-1}\mathbf{S}_* + \mathbf{I}\right)^{-1}\mathbf{b}_c\mathbf{b}_c'\left(\frac{m_c}{m_*}\mathbf{S}_c^{-1}\mathbf{S}_* + \mathbf{I}\right)\right)\right] . \qquad (4.4)
\end{aligned}$$

Thus $\hat{\mathbf{D}}$ serves as an unbiased estimator of $\mathbf{D}$ which is the exact difference between the mean squared error matrix of biased estimator $\hat{\boldsymbol{\beta}}$ and the variance covariance matrix of unbiased estimator $\mathbf{b}_c$.

Similarly, an unbiased estimator for the difference between the mean squared error matrix of $\tilde{\boldsymbol{\beta}}$ and the variance covariance matrix of $\mathbf{b}_c$ can be easily obtained.

As $\mathbf{b}_c$ is identically equal to the unbiased estimator $\mathbf{b}_c$ based on complete observations, its variance covariance matrix can be estimated in the traditional manner; see, e.g., Rao and Toutenburg (1995, Chap. 8) for the details.

5 Some Concluding Remarks

Recognizing that the use of least squares predictions for some missing values of study variable in a linear regression model does not lead to improved estimation of regression coefficients, we have considered the application of shrinkage method. Employing the Stein-rule predictions, we have presented a class of estimators for coefficients. We have also given another class of estimators arising from shrunken predictions proposed by Ali and Abusalih (1988). However, their predictions assume the positive definiteness of the variance covariance matrix of least squares predictions of missing values which requires that the number of missing values must not be greater than the number of regression coefficients in the model. Our estimation procedure is free from such a limitation.

For the estimation of regression coefficients, the estimator $\mathbf{b}_c$ employing the least squares predictions is unbiased while the estimators $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ arising from shrunken predictions of Ali and Abusalih (1988) and Stein-rule predicitons respectively are biased. Examining the large sample asymptotic approximations for the bias vectors, it is found that $\hat{\boldsymbol{\beta}}$ is superior to $\tilde{\boldsymbol{\beta}}$ with respect to the magnitude of bias. Next, comparing the estimators $\mathbf{b}_c$, $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ according to the criterion of risk under weighted squared error loss function to the order of our approximaiton, we have deduced sufficient conditions for the superiority of $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ over $\mathbf{b}_c$. Similar sufficient conditions are obtained for the superiority of $\tilde{\boldsymbol{\beta}}$ over $\hat{\boldsymbol{\beta}}$ and vice-versa. A distinguishing feature of these conditions is that they are simple and easy to check in practice.

Finally, we have presented an expression for the unbiased estimation of the mean squared error matrix of proposed estimators. Such a formula may help in computing the standard errors of estimates and in appreciating the gain/loss in efficiency in actual practice. It may also help in constructing confidence regions and conducting tests of hypotheses.

It may be remarked that some work dealing with the application of proposed estimation procedure to real-world data sets involving some discrete variables in the set of explanatory variables is under way and will be reported in future.

A Appendix

Proof of Theorem 1

Let us write

$$\begin{aligned} \mathbf{u} &= m_c^{-1/2}\mathbf{X}_c'\boldsymbol{\epsilon}_c\,, \\ \mathbf{v} &= (m_c^{-1/2}\boldsymbol{\epsilon}_c'\boldsymbol{\epsilon}_c - m_c^{1/2})\,, \\ \boldsymbol{\theta} &= \boldsymbol{\beta}'\mathbf{S}_c^{-1}\boldsymbol{\beta} \end{aligned}$$

so that $\mathbf{u}$ and $\mathbf{v}$ are of order $O_p(1)$ while $\boldsymbol{\theta}$ is of order $O(1)$.

From (2.1) and (2.7), we can express

$$\begin{aligned} & \frac{R_c}{(m_c - p + 2)\mathbf{b}_c'\mathbf{X}_c'\mathbf{X}_c\mathbf{b}_c} \\ &\quad = \frac{\sigma^2}{m_c\boldsymbol{\theta}}\left(1 - \frac{p-2}{m_c}\right)^{-1}\left(1 + \frac{\mathbf{v}}{m_c^{1/2}} - \frac{\mathbf{u}'\mathbf{S}_c\mathbf{u}}{m_c}\right) \\ &\quad\quad \times\left(1 + \frac{2\sigma\boldsymbol{\beta}'\mathbf{u}}{m_c^{1/2}\boldsymbol{\theta}} + \frac{\sigma^2\mathbf{u}'\mathbf{S}_c\mathbf{u}}{m_c\boldsymbol{\theta}}\right)^{-1} \\ &\quad = \frac{\sigma^2}{m_c\boldsymbol{\theta}} + \frac{\sigma^2(\boldsymbol{\theta}\mathbf{v} - 2\sigma\boldsymbol{\beta}'\mathbf{u})}{m_c^{3/2}\boldsymbol{\theta}^2} + O_p(\frac{1}{m_c^2}) \end{aligned} \tag{A.1}$$

$$\begin{aligned} & \left[\mathbf{I} + \frac{1}{m_c}(\mathbf{X}_*'\mathbf{X}_*)^{-1}\mathbf{S}_c^{-1}\right]^{-1}\mathbf{b}_c \\ &\quad = \boldsymbol{\beta} + \frac{\sigma}{m_c^{1/2}}\mathbf{S}_c\mathbf{u} + O_p(\frac{1}{m_c})\,. \end{aligned} \tag{A.2}$$

Using (A.1) and (A.2) in (2.10), we obtain

$$\begin{aligned} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \frac{\sigma}{m_c^1/2}\mathbf{S}_c\mathbf{u} - \frac{\sigma^2 k}{m_c\boldsymbol{\theta}}\boldsymbol{\beta} \\ &\quad - \frac{\sigma^2 k}{m_c^{2/3}\boldsymbol{\theta}}\left[\mathbf{v}\boldsymbol{\beta} + \sigma(\mathbf{S}_c - \frac{2}{\boldsymbol{\theta}}\boldsymbol{\beta}\boldsymbol{\beta}')\mathbf{u}\right] + O_p(\frac{1}{m_c^2}) \end{aligned} \tag{A.3}$$

whence the bias vector upto order $O_p(m_c^{-1})$ is given by

$$\begin{aligned} B(\hat{\boldsymbol{\beta}}) &= \mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &= \frac{\sigma}{m_c^{1/2}}\mathbf{S}_c\,\mathrm{E}(\mathbf{u}) - \frac{\sigma^2 k}{m_c\boldsymbol{\theta}}\boldsymbol{\beta} \end{aligned}$$

$$= -\frac{\sigma^2 k}{m_c \boldsymbol{\theta}} \boldsymbol{\beta}$$

which is the result (3.2) of Theorem 1.

Similarly, the mean squared error matrix of $\hat{\boldsymbol{\beta}}$ upto order $O(m_c^{-2})$ is

$$\begin{aligned}
M(\hat{\boldsymbol{\beta}}) &= \mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \\
&= \frac{\sigma^2}{m_c} \mathbf{S}_c \,\mathrm{E}(\mathbf{u}\mathbf{u}')\mathbf{S}_c \\
&\quad - \frac{\sigma^3 k}{m_c^{3/2} \boldsymbol{\theta}} \left[\boldsymbol{\beta}\,\mathrm{E}(\mathbf{u}')\mathbf{S}_c + \mathbf{S}_c\,\mathrm{E}(\mathbf{u})\boldsymbol{\beta}'\right] \\
&\quad - \frac{\sigma^4 k}{m_c^2 \boldsymbol{\theta}} \left[\boldsymbol{\beta}\,\mathrm{E}(\mathbf{v}\mathbf{u}')\mathbf{S}_c + \mathbf{S}_c\,\mathrm{E}(\mathbf{v}\mathbf{u})\boldsymbol{\beta}'\right. \\
&\quad + (\mathbf{S}_c - \frac{2}{\boldsymbol{\theta}}\boldsymbol{\beta}\boldsymbol{\beta}')\,\mathrm{E}(\mathbf{u}\mathbf{u}')\mathbf{S}_c \\
&\quad \left. + \mathbf{S}_c\,\mathrm{E}(\mathbf{u}\mathbf{u}')(\mathbf{S}_c - \frac{2}{\boldsymbol{\theta}}\boldsymbol{\beta}\boldsymbol{\beta}') - \frac{k}{\boldsymbol{\theta}}\boldsymbol{\beta}\boldsymbol{\beta}'\right] \\
&= \frac{\sigma^2}{m_c}\mathbf{S}_c - \frac{\sigma^4 k}{m_c^2 \boldsymbol{\theta}}\left[2\mathbf{S}_c - \left(\frac{4+k}{\boldsymbol{\theta}}\right)\boldsymbol{\beta}\boldsymbol{\beta}'\right]
\end{aligned}$$

which is the result (3.3) of Theorem 1.

In a similar manner, the results stated in Theorem 2 can be derived.

Equation 3.6

$$\begin{aligned}
& \boldsymbol{\beta}'\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\boldsymbol{\beta} \;\le\; \boldsymbol{\beta}'\mathbf{S}_c^{-1}\boldsymbol{\beta} \\
\Leftrightarrow\;& \underbrace{\boldsymbol{\beta}'\mathbf{S}_c^{-1/2}}_{\mathbf{a}}\underbrace{\mathbf{S}_c^{1/2}\mathbf{X}_*'(\mathbf{X}_*\mathbf{S}_c\mathbf{X}_*')^{-1}\mathbf{X}_*\mathbf{S}_c^{1/2}}_{\mathbf{A}}\underbrace{\mathbf{S}_c^{-1/2}\boldsymbol{\beta}}_{\mathbf{a}} \;\le\; \boldsymbol{\beta}'\mathbf{S}_c^{-1/2}\mathbf{S}_c^{-1/2}\boldsymbol{\beta} \\
\Leftrightarrow\;& \mathbf{a}'\mathbf{A}\mathbf{a} \;\le\; \mathbf{a}'\mathbf{I}\mathbf{a} \\
\Leftrightarrow\;& \mathbf{a}'(\mathbf{I}-\mathbf{A})\mathbf{a} \;\ge\; \mathbf{0}
\end{aligned}$$

which is true, as $\mathbf{I} - \mathbf{A}$ is idempotent.

References

Ali, M. A. and Abusalih, M. S. (1988). On estimation of missing observations in linear regression models, *Sankhya, Series B* **50**: 404–411.

Carter, R. A. L., Srivastava, M. S., Srivastava, V. K. and Ullah, A. (1990). Unbiased estimation of the mse-matrix of stein-rule-estimators, confidence ellipsoids and hypotheses testing, *Econometric Theory* **6**: 63–74.

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. 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*, Springer, New York.

Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete, *Empire Journal of Experimental Agriculture* **1**: 129–142.