Toutenburg, Shalabh:

Use of minimum risk approach in the estimation of regression models with missing observation

Sonderforschungsbereich 386, Paper 118 (1998)

Use of minimum risk approach in the estimation of regression models with missing observations

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

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

July 8, 1998

Abstract

This article considers a linear regression model with some missing observations on the response variable and presents two estimators of regression coefficients employing the approach of minimum risk estimation. Asymptotic properties of these estimators along with the traditional unbiased estimator are analyzed and conditions, that are easy to check in practice, for the superiority of one estimator over the other are derived.

1 Introduction

The standard linear regression analysis assumes the availability of all the observations. Such a specification may often be violated in many practical situations, and it may be hard to collect data on response variable in some cases. For instance, in sample survey, the responses related to some delicate and sensitive questions may not be available due to caprice nature. Or the investigator may feel that some respondents have not deliberately provided correct information and therefore such responses should be discarded. Similarly, in biological experiments, some animals or plants may die due to some reasons which are not related to the treatments being tested. A certain laboratory instrument or measuring device may break down before the completion of experiment.

When some observations on response variable are missing, there are two alternatives. One is to use complete observations alone and to discard the incomplete data set while the other alternative is to substitute estimated values for missing observations and to use the thus repaired data set; see, e.g. Little (1992), Little and Rubin (1987) and Rao and Toutenburg (1995) for an interesting account. The substitutions are generally constructed on the basis of regression analysis of complete observations. If a weakly unbiased substitution is utilized for missing observations and least squares procedure is applied to the model with the repaired data set, the estimators of regression coefficients remain same as those obtained from the use of complete observations alone. Such a finding has prompted us to consider biased substitutions for missing observations and to analyze the properties of least squares estimators of regression coefficients based on repaired data set.

In this article, we describe the model with some missing observations on response variable in Section 2 and utilize the minimum risk approach in order to obtain substitutions for missing observations. Such an approach yields optimal substitutions but they have no utility owing to involvement of unknown quantities. We therefore consider their operational version. The thus obtained operational substitutions are then used to repair the data set. Now we present three estimators of regression coefficients arising from an application of least squares procedure. In Section 3, we analyze the performance properties of these estimators employing the small disturbance asymptotic theory. Sufficient conditions for the superiority of one estimator over the other are also derived. An elegant aspect of these conditions is that they do not involve any unknown quantities and are thus easy to check in any application. Finally, the derivation of main results is presented in Appendix.

2 Specification of model and estimators

Consider the following linear regression model

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

$$Y_{mis} = X_*\beta + \sigma\epsilon_* \tag{2.2}$$

where Y_c and Y_{mis} are the column vectors of m_c available and m_* missing observations on the response variable, X_c is a full column rank matrix of order $m_c \times p$ consisting of m_c observations on p explanatory variables corresponding to available observations on the response variables, X_* is a not necessarily full column rank matrix of order $m_* \times p$ consisting of m_* observations on p explanatory variables corresponding to missing observations on the response variable, β is a column vector of p unknown coefficients, ϵ_c and ϵ_* are column vectors of m_c and m_* disturbances respectively and σ is an unknown scalar.

We assume that the elements of ϵ_c and ϵ_* are independently, identically and normally distributed with zero mean and unit variance.

If we employ simply the m_c complete observations for the estimation of β, the least squares estimator is given by

$$\hat{\beta}_c = (X_c'X_c)^{-1}X_c'Y_c \tag{2.3}$$

On the other hand, if we use all the $(m_c + m_*)$ observations, the least squares estimator of β is

$$\hat{\beta} = (X_c'X_c + X_*'X_*)^{-1}(X_c'Y_c + X_c'Y_{mis}). \tag{2.4}$$

Observing that

$$
\begin{aligned}
&(X_c'X_c + X_*'X_*)^{-1} \\
&\quad = \;\; (X_c'X_c)^{-1} - (X_c'X_c)^{-1}X_*'\left[I + X_*(X_c'X_c)^{-1}X_*'\right]^{-1}X_*(X_c'X_c)^{-1}
\end{aligned}
$$

we can express

$$
\hat{\beta} = \hat{\beta}_c + (X_c'X_c)^{-1}X_*'\left[I + X_*(X_c'X_c)^{-1}X_*'\right]^{-1}(Y_{mis} - X_*\hat{\beta}_c). \tag{2.5}
$$

Obviously, this estimator has no practical utility due to involvement of missing observations. A simple solution is then to replace Y_{mis} by an observable quantity.

If we employ a weakly unbiased substitution $X_*\hat{\beta}_c$ for missing observations in (2.5), i.e.,

$$
\mathrm{E}(X_*\hat{\beta}_c) = \mathrm{E}(Y_{mis}) \tag{2.6}
$$

the resulting operational version of $\hat{\beta}$ is nothing but the estimator $\hat{\beta}_c$ itself. This is a celebrated result due to Yates; see, for instance, Little and Rubin (1987, chap. 2) or Rao and Toutenburg (1995, chap. 8).

Let us next consider biased substitutions for the vector of missing observations. If we consider substitutions of the type A_cY_c with matrix A_c of order $m_* \times m_c$ for the replacement of $X_*\beta$ in (2.5) and choose A_c such that the risk under a general quadratic loss function is minimum, the optimal substitution for $X_*\beta$ is given by

$$
\left[\frac{\beta'X_c'Y_c}{\beta'X_c'X_c\beta + \sigma^2}\right]X_*\beta \tag{2.7}
$$

which is again not operative; see, for example, Rao and Toutenburg (1995, p. 161). A simple operational version of it, however, can be deduced by using unbiased estimators of β and σ^2 based on complete observations. This yields the following operational substitution for Y_{mis}:

$$
\left[\frac{\hat{\beta}_c'X_c'Y_c}{\hat{\beta}_cX_c'X_c\hat{\beta}_c + \hat{\sigma}_c^2}\right]X_*\hat{\beta}_c = \left[1 - \frac{\hat{\sigma}_c^2}{\hat{\beta}_cX_c'X_c\hat{\beta}_c + \hat{\sigma}_c^2}\right]X_*\hat{\beta}_c \tag{2.8}
$$

where

$$
\hat{\sigma}_c^2 = \frac{1}{(m_c - p)}(Y_c - X_c\hat{\beta}_c)'(Y_c - X_c\hat{\beta}_c). \tag{2.9}
$$

Putting it in place of Y_{mis} in (2.5) leads to the following estimator of β:

$$
b_1 = \hat{\beta}_c - \left[\frac{\hat{\sigma}_c^2}{\hat{\beta}_cX_c'X_c\hat{\beta}_c + \hat{\sigma}_c^2}\right](X_c'X_c)^{-1}X_*\left[I + X_*(X_c'X_c)^{-1}X_*'\right]^{-1}X_*\hat{\beta}_c \tag{2.10}
$$

In the same spirit, if we consider A_*Y_{mis} instead of A_cY_c, we get the following optimal substitution for $X_*\beta$:

$$\left[\frac{\beta' X_*' Y_{mis}}{\beta' X_*' X_* \beta + \sigma^2}\right] X_*\beta \tag{2.11}$$

of which an operational version is

$$\left[\frac{\hat{\beta}_c' X_*' X_* \hat{\beta}_c}{\hat{\beta}_c X_*' X_* \hat{\beta}_c + \hat{\sigma}_c^2}\right] X_*\hat{\beta}_c = \left[1 - \frac{\hat{\sigma}_c^2}{\hat{\beta}_c{}' X_*' X_* \hat{\beta}_c + \hat{\sigma}_c^2}\right] X_*\hat{\beta}_c. \tag{2.12}$$

Substituting it in place of Y_{mis} in (2.5) provides another estimator of β:

$$b_2 = \hat{\beta}_c - \left[\frac{\hat{\sigma}_c^2}{\hat{\beta}_c{}' X_*' X_* \hat{\beta}_c + \hat{\sigma}^2}\right] (X_c'X_c)^{-1} X_*' \left[I + X_*(X_c'X_c)^{-1}X_*'\right]^{-1} X_*\hat{\beta}_c. \tag{2.13}$$

We have thus formulated two estimators b_1 and b_2 arising from the minimum risk approach.

3 Properties of estimators

For analyzing the relative performance of the estimators, we first observe that b_c is unbiased for β and its variance covariance matrix is given by

$$\begin{aligned} \mathrm{V}(b_c) &= \mathrm{E}(b_c - \beta)(b_c - \beta)' \\ &= \sigma^2 (X_c'X_c)^{-1}. \end{aligned} \tag{3.1}$$

Next, let us consider the estimators b_1 and b_2 which are biased for β. Expressions for their bias vectors and mean squared error matrices can be derived but they will be sufficiently intricate and will not lead to any clear conclusion. We therefore consider their approximations employing the small disturbance asymptotic theory.

<u>Theorem I:</u> If we write

$$S = X_*' \left[I + X_*(X_c'X_c)^{-1}X_*'\right]^{-1} X_* \tag{3.2}$$

then, according to small disturbance asymptotic theory, the bias vectors of the estimators b_1 and b_2 to order $O(\sigma^2)$ are given by

$$B(b_1) = -\frac{\sigma^2}{\beta' X_c' X_c \beta} (X_c'X_c)^{-1} S\beta \tag{3.3}$$

$$B(b_2) = -\frac{\sigma^2}{\beta' X_*' X_* \beta} (X_c'X_c)^{-1} S\beta. \tag{3.4}$$

These expressions are obtained from the result (A.1) of Appendix by putting $Q = (X_c'X_c)$ and $Q = (X_*'X_*)$ respectively.

It is seen that both the estimators of β are biased which means that the process of operationalization of estimator $\hat{\beta}$ introduces bias. Further, the elements of bias vector have a sign which is opposite to the sign of corresponding elements in $(X_c'X_c)^{-1}S\beta$.

If we compare the estimators with respect to the criterion of magnitude of bias or equivalently the length of bias vector, we find that the estimator b_1 is better than b_2 when

$$\frac{\beta' X_*' X_* \beta}{\beta' X_c' X_c \beta} < 1 \tag{3.5}$$

which is satisfied at least as long as the largest eigen value of $X_*'X_*$ in the metric of $X_c'X_c$ is less than one. The opposite is true, i.e., b_2 is better than b_1 with respect to the criterion of magnitude of bias as length of bias vector so long as the smallest eigen value of $X_*'X_*$ in the metric of $X_c'X_c$ exceeds one.

<u>Theorem II:</u> The asymptotic approximations for the mean squared error matrices of the estimators b_1 and b_2 to order $O(\sigma^4)$ are given by

$$\begin{aligned} M(b_1) &= \sigma^2(X_c'X_c)^{-1} - \frac{\sigma^4}{\beta' X_c' X_c \beta}[2(X_c'X_c)^{-1}S(X_c'X_c)^{-1} \\ &\quad - \frac{m_c - p + 2}{(m_c - p)\beta' X_c' X_c \beta}(X_c'X_c)^{-1}S\beta\beta' S(X_c'X_c)^{-1} \\ &\quad - \frac{2}{\beta' X_c' X_c \beta}\{\beta\beta' S(X_c'X_c)^{-1} + (X_c'X_c)^{-1}S\beta\beta'\}] \end{aligned} \tag{3.6}$$

$$\begin{aligned} M(b_2) &= \sigma^2(X_c'X_c)^{-1} - \frac{\sigma^4}{\beta' X_*' X_* \beta}[2(X_c'X_c)^{-1}S(X_c'X_c)^{-1} \\ &\quad - \frac{m_c - p + 2}{(m_c - p)\beta' X_*' X_* \beta}(X_c'X_c)^{-1}S\beta\beta' S(X_c'X_c)^{-1} \\ &\quad - \frac{2}{\beta' X_*' X_* \beta}(X_c'X_c)^{-1}(X_*'X_*\beta\beta' S + S\beta\beta' X_*'X_*)(X_c'X_c)^{-1}]. \end{aligned} \tag{3.7}$$

These expressions are obtained from the result (A.2) of Appendix.

Let us first compare the estimators b_1 and b_2 with b_c according to the criterion of mean squared error matrix to the order of our approximation. For this purpose, we employ the following two results for any $m \times m$ positive definite matrix G and a column vector g:

1. The matrix $(G - gg')$ is nonnegative definite if and only if $g'G^{-1}g \leq 1$, see, e.g., Yancey, Judge and Bock (1974).
2. The matrix $(gg' - G)$ cannot be nonnegative definite for $m > 1$; see, e.g. Guilkey and Price (1981).

Now from (3.1) and (3.6), we observe that the variance covariance matrix of b_c exceeds the mean squared error matrix of b_1 to order $O(\sigma^4)$ by at least a positive semi-definite matrix when the matrix

$$\begin{aligned}
&2(X_c'X_c)^{-1}S(X_c'X_c)^{-1} - \frac{(m_c - p + 2)}{(m_c - 2)\beta' X_c' X_c \beta}(X_c'X_c)^{-1}S\beta\beta' S(X_c'X_c)^{-1} \\
&- \frac{2}{\beta' X_c' X_c \beta}\{\beta\beta' S(X_c'X_c)^{-1} + (X_c'X_c)^{-1}S\beta\beta'\} \\
&\quad = \quad 2(X_c'X_c)^{-1}S(X_c'X_c)^{-1} - \delta\delta' + \frac{4(m_c - p)}{(m_c - p + 2)\beta' X_c' X_c \beta}\beta\beta' \qquad (3.8)
\end{aligned}$$

is nonnegative definite where

$$\delta = \left[\frac{m_c - p + 2}{(m_c - 2)\beta' X_c' X_c \beta}\right]^{\frac{1}{2}} \left[(X_c'X_c)^{-1}S + \frac{2(m_c - p)}{(m_c - p + 2)}I\right]\beta. \qquad (3.9)$$

As the matrix

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

is nonnegative definite, the matrix expression (3.8) will be nonnegative definite at least as long as

$$2(X_c'X_c)^{-1}S(X_c'X_c)^{-1} - \delta\delta' \qquad (3.10)$$

is nonnegative definite. For this to be true, using the first result, a necessary and sufficient condition is that the quantity

$$\begin{aligned}
&\frac{1}{2}\delta' X_c' X_c S^{-1} X_c' X_c \delta \\
&\quad = \quad 2 + \frac{m_c - p + 2}{2(m_c - p)}\frac{\beta' S \beta}{\beta' X_c' X_c \beta} + \frac{2(m_c - p)}{m_c - p + 2}\frac{\beta' X_c' X_c S^{-1} X_c' X_c \beta}{\beta' X_c' X_c \beta} \qquad (3.11)
\end{aligned}$$

does not exceed 1. Obviously, this condition cannot hold true implying that b_1 cannot be superior to b_c.

Next, if we consider the matrix $[M(b_1) - M(b_c)]$, we observe that it is nonnegative definite when the matrix expression (3.8) with reversed sign is nonnegative definite. This cannot be true by virtue of second result. This means that b_c cannot be superior to b_1.

We thus observe that neither the estimator b_c dominates nor is dominated by the estimator b_1 according to the mean squared error matrix criterion to the order of our approximation. In a similar manner, it can be easily demonstrated that the estimator b_2 is neither dominated by b_c nor dominates b_c.

Similarly, if we analyze the expressions (3.6) and (3.7), it is interesting to find that between the estimators b_1 and b_2, no one is superior to the other with respect to the mean squared error matrix criterion.

Next, let us compare the risk functions, associated with these estimators, under a quadratic loss structure with loss matrix as $(X_c'X_c)$. Now we observe from (3.1) and (3.6) that

$$\begin{aligned}\triangle(b_c;b_1) &= \mathrm{E}(b_c-\beta)'X_c'X_c(b_c-\beta)-\mathrm{E}(b_1-\beta)'X_c'X_c(b_1-\beta)\\ &= \frac{\sigma^4}{\beta'X_c'X_c\beta}\Bigg[2\,\mathrm{tr}\,S(X_c'X_c)^{-1}-4\frac{\beta'S\beta}{\beta'X_c'X_c\beta}\\ &\quad -\left(\frac{m_c-p+2}{m_c-p}\right)\frac{\beta'S(X_c'X_c)^{-1}S\beta}{\beta'X_c'X_c\beta}\Bigg]. \qquad (3.12)\end{aligned}$$

Let $\lambda_1,\lambda_2,\lambda_3,\ldots$ be the eigen values of $X_*(X_c'X_c)^{-1}X_*'$. These are then the eigen values of $X_*'X_*$ in the metric of $X_c'X_c$. Further, let λ_L and λ_S denote the largest and smallest values among the nonzero eigen values. Now we observe that

$$\left(\frac{\lambda_S}{1+\lambda_L}\right)^2\leq\frac{\beta'S(X_c'X_c)^{-1}S\beta}{\beta'X_c'X_c\beta}\leq\left(\frac{\lambda_L}{1+\lambda_S}\right)^2 \qquad (3.13)$$

$$\left(\frac{\lambda_S}{1+\lambda_L}\right)\leq\frac{\beta'S\beta}{\beta'X_c'X_c\beta}\leq\left(\frac{\lambda_L}{1+\lambda_S}\right) \qquad (3.14)$$

whence it follows that the difference (3.12) is positive so long as

$$\sum_i\left(\frac{\lambda_i}{1+\lambda_i}\right)>\left(\frac{\lambda_L}{1+\lambda_S}\right)\left[2+\frac{(m_c-p+2)\lambda_L}{2(m_c-p)(1+\lambda_S)}\right] \qquad (3.15)$$

which is a sufficient condition for the superiority of b_1 over b_c.

Similarly, the difference (3.12) is negative meaning thereby the superiority of b_c over b_1, with respect to the criterion of risk function, at least as long as

$$\sum_i\left(\frac{\lambda_i}{1+\lambda_i}\right)<\left(\frac{\lambda_S}{1+\lambda_L}\right)\left[2+\frac{(m_c-p+2)\lambda_S}{2(m_c-p)(1+\lambda_L)}\right]. \qquad (3.16)$$

In a similar manner, if we compare the risk functions of b_c and b_2 using (3.1) and (3.7), it can be easily seen that b_2 is superior to b_c so long as

$$\sum_i\left(\frac{\lambda_i}{1+\lambda_i}\right)>\left(\frac{\lambda_L}{1+\lambda_S}\right)\left[2+\frac{(m_c-p+2)}{2(m_c-p)(1+\lambda_S)}\right] \qquad (3.17)$$

where use has been made of the following results

$$\frac{\lambda_S}{(1+\lambda_L)^2}\leq\frac{\beta'S(X_c'X_c)^{-1}S\beta}{\beta'X_*'X_*\beta}\leq\frac{\lambda_L}{(1+\lambda_S)^2} \qquad (3.18)$$

$$\left(\frac{\lambda_S}{1+\lambda_L}\right)\leq\frac{\beta'X_*'X_*(X_c'X_c)^{-1}S\beta}{\beta'X_*'X_*\beta}\leq\left(\frac{\lambda_L}{1+\lambda_S}\right). \qquad (3.19)$$

Just the opposite is true, i.e., b_c is superior to b_2 when

$$\sum_i \left(\frac{\lambda_i}{1+\lambda_i}\right) < \left(\frac{\lambda_S}{1+\lambda_L}\right)\left[2+\frac{(m_c-p+2)}{2(m_c-p)(1+\lambda_L)}\right]. \tag{3.20}$$

Finally, let us compare the risk functions of the biased estimators b_1 and b_2. It is easy to see from (3.6) and (3.7) that b_1 is better than b_2 when

$$\begin{aligned}
&\left[\left(\frac{\beta' X_*' X_*\beta}{\beta' X_c' X_c\beta}-1\right)\mathrm{tr}(X_c'X_c)^{-1}S + 2\left(\frac{\beta' S X_*' X_*\beta}{\beta' X_*' X_*\beta}\right)\right] \\
&> \left[2\left(\frac{\beta' S\beta}{\beta' X_c' X_c\beta}\right)\left(\frac{\beta' X_*' X_*\beta}{\beta' X_c' X_c\beta}\right)\right. \\
&\qquad \left. + \frac{m_c-p+2}{2(m_c-p)}\left\{\left(\frac{\beta' X_*' X_*\beta}{\beta' X_c' X_c\beta}\right)^2 - 1\right\}\frac{\beta' S(X_c'X_c)^{-1}\beta}{\beta' X_*' X_*\beta}\right].
\end{aligned} \tag{3.21}$$

For this inequality to hold true, several sufficient conditions can be deduced employing the results (3.13), (3.14), (3.18) and (3.19). For instance, the inequality (3.21) is satisfied so long as

$$\left[(\lambda_S-1)\sum_i\left(\frac{\lambda_i}{1+\lambda_i}\right)+\frac{2\lambda_S}{1+\lambda_L}\right] > \left(\frac{\lambda_L}{1+\lambda_S}\right)\left[2\lambda_L+\frac{(m_c-p+2)(\lambda_L^2-1)}{2(m_c-p)(1+\lambda_S)}\right]. \tag{3.22}$$

Similarly, the opposite is true, i.e., b_2 is better than b_1 when the inequality (3.21) holds with a reversed sign. This happens, for instance, as long as

$$\left[(\lambda_L-1)\sum_i\left(\frac{\lambda_i}{1+\lambda_i}\right)+\frac{2\lambda_L}{1+\lambda_S}\right] > \left(\frac{\lambda_S}{1+\lambda_L}\right)\left[2\lambda_S+\frac{(m_c-p+2)(\lambda_S^2-1)}{2(m_c-p)(1+\lambda_L)}\right]. \tag{3.23}$$

It is interesting to note that the sufficient conditions for the superiority of one estimator over the other are free from unknown quantities and are easy to check in any practical situation.

Appendix

If we define

$$b = \hat\beta_c - \left[\frac{\hat\sigma_c^2}{\hat\beta_c' Q\hat\beta_c + \hat\sigma_c^2}\right](X_c'X_c)^{-1}S\hat\beta_c$$

with Q as a nonstochastic symmetric matrix and S given by (3.2), the bias vector to order $O(\sigma^2)$ is

$$
\begin{aligned}
B(b) &= \mathrm{E}(b-\beta) && \text{(A.1)}\\
&= -\frac{\sigma^2}{\beta' Q\beta}(X_c' X_c)^{-1} S\beta
\end{aligned}
$$

and the mean squarred error matrix to order $O(\sigma^4)$ is given by

$$
\begin{aligned}
M(b) &= \mathrm{E}(b-\beta)(b-\beta)' \qquad \text{(A.2)}\\
&= \sigma^2 (X_c' X_c)^{-1} - \frac{\sigma^4}{\beta' Q\beta}(X_c' X_c)^{-1}\left[2S - \frac{m_c - p + 2}{(m_c - p)\beta' Q\beta} S\beta\beta' S \right.\\
&\quad \left. - \frac{2}{\beta' Q\beta}(Q\beta\beta' S + S\beta\beta' Q)\right](X_c' X_c)^{-1}
\end{aligned}
$$

<u>Proof:</u> Using (2.1), we can write

$$
\begin{aligned}
\hat{\beta}_c &= \beta + \sigma (X_c' X_c)^{-1} X_c' \epsilon_c\\
\hat{\sigma}_c^2 &= \frac{\sigma^2}{m_c - p}\epsilon_c' \left[I - X_c (X_c' X_c)^{-1} X_c'\right]\epsilon_c
\end{aligned}
$$

so that

$$
\begin{aligned}
&\frac{\hat{\sigma}_c^2}{\hat{\beta}_c' Q\beta_c + \hat{\sigma}_c^2}\\
= &\frac{\sigma^2 \epsilon_c' \left[I - X_c (X_c' X_c)^{-1} X_c'\right]\epsilon_c}{(m_c - p)}\left[\beta' Q\beta + 2\sigma\beta' Q (X_c' X_c)^{-1} X_c' \epsilon_c \right.\\
&\left. + \sigma^2 \left\{\epsilon_c' X_c (X_c' X_c)^{-1} Q (X_c' X_c)^{-1} X_c' \epsilon_c + \frac{\epsilon_c' \left[I - X_c (X_c' X_c)^{-1} X_c'\right]\epsilon_c}{(m_c - p)}\right\}\right]^{-1}\\
= &\frac{\sigma^2 \epsilon_c' \left[I - X_c (X_c' X_c)^{-1} X_c'\right]}{(m_c - p)\beta' Q\beta}\left[1 + 2\sigma \frac{\beta' Q (X_c' X_c)^{-1} X_c' \epsilon_c}{\beta' Q\beta} + O_p(\sigma^2)\right]^{-1}
\end{aligned}
$$

Expanding and retaining terms up to order $O(\sigma^3)$ only, we get

$$
\begin{aligned}
\frac{\hat{\sigma}_c^2}{\hat{\beta}_c' Q\hat{\beta}_c + \hat{\sigma}_c^2} &= \frac{\sigma^2 \epsilon_c' \left[I - X_c (X_c' X_c)^{-1} X_c'\right]\epsilon_c}{(m_c - p)\beta' Q\beta}\\
&\quad - 2\sigma^3 \frac{\epsilon_c' \left[I - X_c (X_c' X_c)^{-1} X_c'\right]\epsilon_c \beta' Q (X_c' X_c)^{-1} X_c' \epsilon_c}{(m_c - p)(\beta' Q\beta)^2} + O_p(\sigma^4)
\end{aligned}
$$

Using it we can express

$$
(b-\beta) = \sigma\xi_1 - \sigma^2\xi_2 - \sigma^3\xi_3 + O_p(\sigma^4)
$$

where

$$
\begin{aligned}
\xi_1 &= (X_c'X_c)^{-1}X_c'\epsilon_c \\
\xi_2 &= \frac{\epsilon_c'\left[I - X_c(X_c'X_c)^{-1}X_c'\right]\epsilon_c}{(m_c-\beta)\beta'Q\beta}(X_c'X_c)^{-1}S\beta \\
\xi_3 &= \frac{\epsilon_c'\left[I - X_c(X_c'X_c)^{-1}X_c'\right]\epsilon_c}{(m_c-\beta)\beta'Q\beta}(X_c'X_c)^{-1}S \\
&\quad \left(I - \frac{2}{\beta'Q\beta}\beta\beta'Q\right)(X_c'X_c)^{-1}X_c'\epsilon_c.
\end{aligned}
$$

Thus the bias vector to order $O(\sigma^2)$ is

$$
\begin{aligned}
B(b) &= \sigma\,\mathrm{E}(\xi_1) - \sigma^2\,\mathrm{E}(\xi_2) \\
&= -\sigma^2\xi_2
\end{aligned}
$$

which is the result (A.1).

In a similar way, the mean squarred error matrix to order $O(\sigma^4)$ is given by

$$
M(b) = \sigma^2\,\mathrm{E}(\xi_1\xi_1') - \sigma^3\,\mathrm{E}(\xi_2\xi_1' + \xi_1\xi_2') - \sigma^4\,\mathrm{E}(\xi_3\xi_1' + \xi_1\xi_3' - \xi_2\xi_2').
$$

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

$$
\begin{aligned}
\mathrm{E}(\xi_1\xi_1') &= (X_c'X_c)^{-1} \\
\mathrm{E}(\xi_2\xi_1') &= 0 \\
\mathrm{E}(\xi_1\xi_2') &= 0 \\
\mathrm{E}(\xi_3\xi_1') &= \frac{1}{\beta'Q\beta}(X_c'X_c)^{-1}S\left(I - \frac{2}{\beta'Q\beta}\beta\beta'Q\right)(X_c'X_c)^{-1} \\
\mathrm{E}(\xi_1\xi_3') &= \frac{1}{\beta'Q\beta}(X_c'X_c)^{-1}\left(I - \frac{2}{\beta'Q\beta}Q\beta\beta'\right)S(X_c'X_c)^{-1} \\
\mathrm{E}(\xi_2\xi_2') &= \frac{m_c-p+2}{(\beta'Q\beta)^2}(x_c'X_c)^{-1}S\beta\beta'S(X_c'X_c)^{-1}.
\end{aligned}
$$

Substituting these expressions, we obtain the result (A.2).

References

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

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*, Springer, New York.

Yancey, T. A., Judge, G. G. and Bock, M. E. (1974). A mean square error test when stochastic restrictions are used in regression, *Communications in Statistics, Part A–Theory and Methods* **3**: 755–768.