Toutenburg, Srivastava:

Impact of Departure from Normality on the Efficiency of Estimating Regression Coefficients when Some Observations are Missing

Sonderforschungsbereich 386, Paper 130 (1998)

Impact of Departure from Normality on the Efficiency of Estimating Regression Coefficients when Some Observations are Missing

H. Toutenburg* V. K. Srivastava**

September 14, 1998

Abstract

This article considers a linear regression model in which some observations on an explanatory variable are missing, and presents three least squares estimators for the regression coefficients vector. One estimator uses complete observations alone while the other two estimators utilize repaired data with nonstochastic and stochastic imputed values for the missing observations. Asymptotic properties of these estimators based on small disturbance asymptotic theory are derived and the impact of departure from normality of disturbances is examined.

1 Introduction

During the process of data collection, we often encounter situations where some observations cannot be recorded for one reason or the other. Such instances occur quite frequently in mail surveys, opinion surveys, crop surveys, socio-economic enquiries and planned experimentation in biological, industrial and medical sciences. Consequently, the traditional statistical analysis cannot be conducted due to some missing observations. Now there are two alternatives. One is to confine attention to complete observations alone and to discard the remaining incomplete observations. The other alternative is to repair the data following some imputation procedure for filling the missing values and then to conduct the analysis. Both the strategies have their own limitations and qualifications.

When few values of some explanatory variables in a linear regression model are missing, there are various ways to find imputed values; see, e.g., Little (1992), Little and Rubin (1987), and Rao and Toutenburg (1995) for an interesting exposition of the subject matter. A popular procedure among them is to employ the method of first order regression which consists of running the regression of an explanatory variable (for which some values are missing) on the remaining explanatory variables in the model utilizing the data set of complete observations and then using the thus obtained estimated equation for finding the predicted values for the missing observations. In this manner, the repaired data

*Institut für Statistik, Universität München, Akademiestr. 1, 80799 München, Germany

**Department of Statistics, Lucknow University, Lucknow 226007, India

set is obtained and is used for conducting the regression analysis; see Toutenburg, Heumann, Fieger and Park (1995) for the efficiency properties of this procedure in relation to the procedure that ignores the incomplete observations all together.

The first order regression method, it may be observed, provides imputed values which are nonstochastic by nature. Introducing a simple modification in this method, Toutenburg, Srivastava and Fieger (1996) have given a procedure which yields stochastic imputed values for missing observations. They have also analyzed the asymptotic properties of the two estimators, arising from the traditional and modified first order regression methods, for the regression coefficients. They have employed the large sample asymptotic theory for which a requirement is that the number of observations is sufficiently large. This specification may not be tenable in many practical situations where it may be hard to have a large data set. We therefore employ the small disturbance asymptotic theory which places no constraint on the number of observations. All that it assumes is that the disturbances are small, or equivalently σ is small and tends to zero meaning thereby that the underlying model tends to be more and more correct. Thus the inferences derived from small disturbance asymptotic theory tend to have larger validity in the neighbourhood of correct model. Obviously, if model is not correctly specified, there is little point in making efforts to draw more efficient inferences.

The plan of this article is as follows. In Section 2, we postulate a linear regression model in which some values of only one explanatory variable are missing. This is done to keep the exposition simple and comprehensible. The traditional first order regression method providing nonstochastic imputed values for missing observations is then described. Stemming from it, we describe an imputation procedure which yields stochastic imputed values for the missing observations. When these imputed values are used to repair the incomplete data set and the least squares method is employed to estimate the vector of regression coefficients, the resulting estimators are presented in Section 3 and their asymptotic properties are analyzed employing the small disturbance asymptotic theory. Finally, some concluding remarks are placed in Section 4.

2 Model Specification and Imputation Procedures

Let us postulate a linear regression relationship connecting a study variable Y and k explanatory variables $X_1, X_2, \ldots X_k$, and assume that some observations on only the last variable X_k are missing in order to keep our exposition simple and comprehensible. Thus we can write the model as follows:

$$y_c = Z_c\delta + \alpha x_c + \sigma\epsilon_c \tag{2.1}$$

$$y_* = Z_*\delta + \alpha x_* + \sigma\epsilon_* \tag{2.2}$$

where y_c and y_* denote column vectors of m_c and m_* observations on the study variable, Z_c and Z_* are matrices of m_c and m_* observations on the first $(k-1)$ explanatory variables on which all the observations are available, x_c and x_* are column vectors of m_c available observations and m_* missing observations on the last explanatory variable, ϵ_c and ϵ_* are column vectors of m_c and m_*

disturbances and δ is a column vector of $(k-1)$ unknown coefficients while α and σ are unknown scalar quantities.

It is assumed that the matrix $(Z_c \ x_c)$ has full column rank. Further, we assume that the elements of disturbance vectors ϵ_c and ϵ_* are independently and identically distributed, following not necessarily a normal distribution, with mean zero, variance unity and third moment γ_1.

In order to find imputed values for the missing observations on the last explanatory variable by the method of first order regression, we run the regression of X_k on $X_1, X_2, \ldots, X_{k-1}$ employing m_c complete observations and then use the estimated equation to obtain the predicted values of X_k for the m_* missing observations. This yields the following vector of imputed values for x_*:

$$x_R = Z_*(Z_c'Z_c)^{-1}Z_c'x_c \tag{2.3}$$

which is obviously a nonstochastic quantity.

Now if we include the study variable Y also as an additional explanatory variable and accordingly run the regression of X_k on $X_1, X_2, \ldots X_{k-1}$ and Y utilizing the m_c complete observations, the predictions for the m_* values in x_* from the estimated equation are given by

$$\hat{x}_* = (Z_* \ y_*)\begin{pmatrix} Z_c'Z_c & Z_c'x_c \\ x_c'Z_c & x_c'x_c \end{pmatrix}\begin{pmatrix} Z_c'y_c \\ x_c'y_c \end{pmatrix}$$

or

$$\hat{x}_* = Z_*(Z_c'Z_c)^{-1}Z_c'x_c + \frac{x_c'My_c}{y_c'My_c}\left[y_* - Z_*(Z_c'Z_c)^{-1}Z_c'y_c\right] \tag{2.4}$$

where $M = I - Z_c(Z_c'Z_c)^{-1}Z_c'$; see Toutenburg et al. (1996).

It is thus seen from (2.4) that the imputed values for x_* are stochastic in nature while the traditional imputed values specified by (2.3) are nonstochastic. Further, the stochastic imputed values appear to be some kind of modified forms of the nonstochastic imputed values.

3 Estimation of Regression Coefficients

If we write

$$\begin{aligned} X_c &= \begin{pmatrix} Z_c & x_c \end{pmatrix} \\ \beta &= \begin{pmatrix} \delta \\ \alpha \end{pmatrix} \\ X_* &= \begin{pmatrix} Z_* & x_* \end{pmatrix} \end{aligned}$$

we can express the model as

$$y_c = X_c\beta + \sigma\epsilon_c \tag{3.1}$$

$$y_* = X_*\beta + \sigma\epsilon_*. \tag{3.2}$$

If we ignore the incomplete observations and restrict our attention to complete cases only, the least squares estimator of β is given by

$$b_c = (X_c'X_c)^{-1}X_c'y_c \tag{3.3}$$

which is unbiased with variance covariance matrix as

$$\mathrm{V}(b_c) = \sigma^2 (X_c' X_c)^{-1}. \tag{3.4}$$

If we substitute x_R and $\hat{x}_*$ for the missing observations x_* in (2.2) and then apply least squares to the thus repaired models using both the equations, we obtain the following two estimators of β:

$$\begin{aligned} b_R &= (X_c' X_c + X_R' X_R)^{-1} (X_c' y_c + X_R' y_*) & (3.5) \\ \hat{\beta} &= (X_c' X_c + \hat{X}_*' \hat{X}_*)^{-1} (X_c' y_c + \hat{X}_*' y_*) & (3.6) \end{aligned}$$

where X_R and $\hat{X}_*$ are same as X_* except that x_* in X_* is replaced by x_R and $\hat{x}_*$ respectively.

The exact properties of b_R in relation to b_c have been studied by Toutenburg et al. (1995) while large sample asymptotic properties of b_c, b_R and $\hat{\beta}$ have been analyzed by Toutenburg et al. (1996).

Now let us consider the asymptotic properties of the estimators b_R and $\hat{\beta}$ when disturbances are small.

From the results reported in Toutenburg et al. (1995), the bias vector and mean squared error matrix of b_R to order $O(\sigma^2)$ are given by

$$\begin{aligned} \mathrm{B}(b_R) &= \mathrm{E}(b_R - \beta) & (3.7) \\ &= \alpha \Omega_R X_R' (x_* - x_R) & \\ \mathrm{M}(b_R) &= \mathrm{E}(b_R - \beta)(b_R - \beta)' & (3.8) \\ &= \alpha^2 \Omega_R X_R' (x_* - x_R)(x_* - x_R)' X_R \Omega_R + \sigma^2 \Omega_R & \end{aligned}$$

where

$$\Omega_R = (X_c' X_c + X_R' X_R)^{-1}. \tag{3.9}$$

These expressions clearly indicate that b_R is not only biased estimator of β but it is inconsistent also according to small disturbance asymptotic theory.

Let us now introduce the following notation:

$$\begin{aligned} \Omega &= (X_c' X_c + X_*' X_*)^{-1} & (3.10) \\ d &= X_*' X_* (X_c' X_c)^{-1} e + \left(\frac{m_c - k - 1}{x_c' M x_c} \right) X_*' (x_* - x_R) & (3.11) \\ A &= Z_* (Z_c' Z_c)^{-1} Z_c' + \frac{1}{x_c' M x_c} (x_* - x_R) x_c' M & (3.12) \\ N &= M - \frac{1}{x_c' M x_c} M x_c x_c' M & (3.13) \end{aligned}$$

$$
\begin{aligned}
f &= (I * A'A)\left[\frac{2}{x_c' M x_c} M x_c - (X_c + 2A'X_*)\Omega e\right] && (3.14)\\
&\quad + [I * A'X_*\Omega(X_c' + X_*'A)]A'X_*\Omega e \\
&\quad + (I * A'X_*\Omega X_*'A)(X_c + A'X_*)\Omega e \\
&\quad - \frac{1}{x_c' M x_c}(I * N)(2I - X_*\Omega X_*')(x_* - x_R) \\
&\quad - \frac{1}{x_c' M x_c}[I * A'X_*\Omega(X_c' + 2X_*'A)]M x_c \\
g &= \frac{1}{x_c' M x_c} M x_c - (X_c + A'X_*)\Omega e && (3.15)\\
F &= (e'\Omega e)X_*'A[I * \{A'X_*\Omega(X_c' + X_*'A) - A'A\}] && (3.16)\\
&\quad + \frac{1}{x_c' M x_c}\Big[X_*'A\{I * (M - 2N)\} + \{e'\Omega X_*'(x_* - x_R)\}X_*'A(I * N) \\
&\qquad - X_*'A\{I * M x_c e'\Omega(X_c' + 2X_*'A)\} \\
&\qquad + X_*'A\{I * A'X_*\Omega e e'\Omega(X_c' + X_*'A)\} \\
&\qquad + \{X_*'(X_* - X_R) + (X_* - X_R)'X_*\}\Omega(X_c' + X_*'A)(I * N)\Big] \\
&\quad - \frac{2}{(x_c' M x_c)^2} X_*'(x_* - x_r)x_c' M(I * N) \\
G &= A'(I - X_*\Omega X_*')A - A'X_*\Omega X_c' && (3.17)
\end{aligned}
$$

where $*$ denotes the Hadamard product operator of matrices and e is a $k \times 1$ vector with first $(k-1)$ elements zero and last element one.

Theorem 1: The asymptotic approximations for the bias vector and mean squared error matrix of the estimator $\hat{\beta}$ to order $O(\sigma^3)$ are given by

$$
\begin{aligned}
\mathrm{B}(\beta) &= \mathrm{E}(\hat{\beta} - \beta) && (3.18)\\
&= \frac{\sigma^2}{\alpha}\Omega d + \sigma^3\left(\frac{\gamma_1}{\alpha^2}\right)[(f'l_c)\Omega e + \Omega F l_c] \\
\mathrm{M}(\hat{\beta}) &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' && (3.19)\\
&= \sigma^2(X_c'X_c)^{-1} - \sigma^3\left(\frac{\gamma_1}{\alpha}\right)\Omega(W + W')\Omega
\end{aligned}
$$

where

$$
W = \left[e l_c'(I * G) + X_*'A(I * l_c g') - \frac{1}{x_c' M x_c} X_*'(x_* - x_R) l_c'(I * N)\right](X_c + A'X_*) \qquad (3.20)
$$

with l_c denoting a $m_c \times 1$ vector having all elements unity.

It is observed from (3.18) and (3.19) that the estimator $\hat{\beta}$ is biased but consistent for β according to small disturbance asymptotic theory. It is, however, difficult to place any definite comment on the bias as well as the efficiency with respect to the estimator b_c due to intricate nature of expressions.

If we consider only the leading term in the expression (3.18), the bias remains same whether the distribution is normal or not. When we include the term of order $O(\sigma^3)$, the impact of departure from normality appears, and an important role is played by the skewness of the distribution of disturbances.

In a similar manner, if we restrict our attention to the leading term in the expression (3.19), we observe that both the estimators b_c and $\hat{\beta}$ are equally efficient and thus it is futile to find imputed values for missing observations and to use them for repairing the model. This, however, does not necessarily remain true when we consider the entire expression upto order $O(\sigma^3)$. And then a dominant role is played by the asymmetry of the distribution of disturbances in ascertaining the efficiency property of $\hat{\beta}$ in relation to b_c. In fact, the sign of γ_1 determines the gain or loss in efficiency while its magnitude scales the gain or loss of efficiency.

It may not be out of place to mention that the effect of fourth (kurtosis) and other higher order moments of the distribution of disturbances will precipitate if we work out higher order asymptotic approximations for the bias vector and mean squared error matrix.

4 Some Remarks

We have considered a linear regression model when some observations on one of the explanatory variables are missing, and have presented three estimators for the regression coefficients vector. The first estimator arises from an application of the least squares procedure using the complete observations alone and discarding the incomplete observations all together. The other two estimators are also least squares estimators but they use the repaired data set in which the missing observations are replaced by imputed values. The second estimator is thus based on nonstochastic imputed values given by the first order regression procedure, while the third estimator is based on stochastic imputed values provided by the modified first order regression procedure.

It is found that first estimator is unbiased while the remaining two estimators are biased. Further, according to small disturbance asymptotic theory, the first and third estimators are consistent and share the same asymptotic properties while the second estimator is not consistent.

Departure from normality of disturbances has no influence on the distributional properties of the first and second estimators. Such is, however, not the case with the third estimator. Its performance properties under normality could be markedly different from those when the distribution of disturbances departs from normality. In this context, the effect of skewness is seen to be more pronounced in comparison to that of kurtosis and other features of the distributions as reflected by fourth and higher order moments.

Incidently, the asymptotic approximations for the bias vector and mean squared error matrix of the third estimator turn out to be sufficiently involved and could not help us in deducing any clear inference in order to appreciate any loss or gain in efficiency. It will be interesting to conduct a numerical exploration using techniques like simulation and bootstrap. Such an exercise may display some meaningful conclusions. Some work in this direction is under a way, and we hope to get back with some interesting findings in the time to come.

Appendix

In order to derive the results stated in Theorem, we first observe from (2.1), (2.2) and (2.4) that

$$\begin{aligned}
\hat{x}_* &= x_R + \frac{\alpha x_c' M x_c + \sigma x_c' M \epsilon_c}{\alpha^2 x_c' M x_c + 2\sigma\alpha x_c' M\epsilon_c + \sigma^2 \epsilon_c' M \epsilon_c} \times \\
&\quad \times [\alpha(x_* - x_R) + \sigma\{\epsilon_* - Z_*(Z_c'Z_c)^{-1}Z_c'\epsilon_c\}] \\
&= x_R + \left(1 + \sigma\frac{x_c' M\epsilon_c}{\alpha x_c' M x_c}\right)\left(1 + 2\sigma\frac{x_c' M \epsilon_c}{\alpha\, x_c' M x_c} + \sigma^2 \frac{\epsilon_c' M \epsilon_c}{\alpha^2\, x_c' M x_c}\right)^{-1} \times \\
&\quad \times \left[(x_* - x_R) + \frac{\sigma}{\alpha}\{\epsilon_* - Z_*(Z_c'Z_c)^{-1}Z_c'\epsilon_c\}\right]
\end{aligned}$$

Expanding in increasing powers of σ, we get

$$\begin{aligned}
\hat{x}_* &= x_* + \frac{\sigma}{\alpha}(\epsilon_* - A\epsilon_c) \\
&\quad - \frac{\sigma^2}{\alpha^2}\left[\left(\frac{x_c' M \epsilon_c}{x_c' M x_c}\right)(\epsilon_* - A\epsilon_c) + \left(\frac{\epsilon_c' N \epsilon_c}{x_c' M x_c}\right)(x_* - x_r)\right] \\
&\quad + \frac{\sigma^3}{\alpha^3}\left[\left(\frac{\epsilon_c'(M - 2N)\epsilon_c}{x_c' M x_c}\right)(\epsilon_* - A\epsilon_c) + 2\left(\frac{\epsilon_c' N \epsilon_c\; x_c' M \epsilon_c}{(x_c' M x_c)^2}\right)(x_* - x_R)\right] \\
&\quad + O_p(\sigma^4)
\end{aligned} \tag{A.1}$$

whence we can express

$$\begin{aligned}
\hat{X}_* &= X_* + \sigma U - \frac{\sigma^2}{\alpha^2}\left[\alpha\left(\frac{x_c' M\epsilon_c}{x_c' M x_c}\right)U + \left(\frac{\epsilon_c' N \epsilon_c}{x_c' M x_c}\right)(X_* - X_R)\right] \\
&\quad + \frac{\sigma^3}{\alpha^3}\left[\alpha\left(\frac{\epsilon_c'(M-2N)\epsilon_c}{x_c' M x_c}\right)U + 2\left(\frac{\epsilon_c' N\epsilon_c\; x_c' M \epsilon_c}{(x_c' M x_c)^2}\right)(X_* - X_R)\right] \\
&\quad + O_p(\sigma^4).
\end{aligned} \tag{A.2}$$

Using it, we observe that

$$\begin{aligned}
&(X_c'X_c + \hat{X}_*'\hat{X}_*)^{-1} \\
&= \left[\Omega^{-1} + \sigma(X_*'U + U'X_*) + \sigma^2\left\{U'U - \frac{x_c' M\epsilon_c}{\alpha\, x_c' M x_c}(X_*'U + U'X_*)\right.\right. \\
&\quad \left.\left. - \frac{\epsilon_c' N \epsilon_c}{\alpha^2\, x_c' M x_c}(X_*'(X_* - X_R) + (X_* - X_R)'X_*)\right\} + O_p(\sigma^3)\right]^{-1} \\
&= \Omega - \sigma\Omega(X_*'U + U'X_*)\Omega \\
&\quad + \sigma^2\big[\Omega(X_*'U + U'X_*)\Omega(X_*'U + U'X_*)\Omega - \Omega U'U\Omega \\
&\quad + \frac{x_c' M \epsilon_c}{\alpha\, x_c' M x_c}\Omega(X_*'U + U'X_*)\Omega \\
&\quad + \frac{\epsilon_c' N\epsilon_c}{\alpha^2\, x_c' M x_c}\Omega(X_*'(X_* - X_R) + (X_* - X_R)'X_*)\Omega\big] \\
&\quad + O_p(\sigma^3)
\end{aligned} \tag{A.3}$$

$$
\begin{aligned}
\hat{X}_*'\epsilon_* &= X_*'\epsilon_* + \sigma U'\epsilon_* && \text{(A.4)}\\
&\quad - \frac{\sigma^2}{\alpha^2\ x_c'Mx_c}\left[\alpha(x_c'M\epsilon_c)U'\epsilon_* + (\epsilon_c'N\epsilon_c)(X_* - X_R)'\epsilon_*\right]\\
&\quad + O_p(\sigma^3)\\
&= X_*'\epsilon_* + \frac{\sigma}{\alpha}(\epsilon_*'\epsilon_* - \epsilon_*'A\epsilon_c)e && \text{(A.5)}\\
&\quad - \frac{\sigma^2}{\alpha^2\ x_c'Mx_c}\left[(x_c'M\epsilon_c)(\epsilon_*'\epsilon_* - \epsilon_*'A\epsilon_c) + (\epsilon_c'N\epsilon_c)(x_* - x_R)'\epsilon_*\right]e\\
&\quad + O_p(\sigma^3)
\end{aligned}
$$

$$
\begin{aligned}
&\hat{X}_*'(X_* - \hat{X}_*)\beta && \text{(A.6)}\\
&= -\sigma X_*'U\beta + \sigma^2\left[\frac{x_c'M\epsilon_c}{\alpha\ x_c'Mx_c}X_*'U\beta + \frac{\epsilon_c'N\epsilon_c}{\alpha^2\ x_c'Mx_c}X_*'(X_* - X_R)\beta - U'U\beta\right]\\
&\quad + \sigma^3\left[\frac{2\ x_c'M\epsilon_c}{\alpha\ x_c'Mx_c}U'U\beta - \frac{\epsilon_c'(M-2N)\epsilon_c}{\alpha^2\ x_c'Mx_c}X_*'U\beta\right.\\
&\qquad + \frac{\epsilon_c'N\epsilon_c}{\alpha^2\ x_c'Mx_c}\{U'(X_* - X_R)\beta + (X_* - X_R)'U\beta\}\\
&\qquad \left. - \frac{2\ x_c'M\epsilon_c\ \epsilon_c'N\epsilon_c}{\alpha^3\ (x_c'Mx_c)^2}X_*'(X_* - X_R)\beta\right]\\
&\quad + O_p(\sigma^4)\\
&= -\sigma X_*'(\epsilon_* - A\epsilon_c) + \frac{\sigma^2}{\alpha}\left[\frac{x_c'M\epsilon_c}{x_c'Mx_c}X_*'(\epsilon_* - A\epsilon_c)\right.\\
&\qquad \left. + \frac{\epsilon_c'N\epsilon_c}{x_c'Mx_c}X_*'(x_* - x_R) - (\epsilon_* - A\epsilon_c)'(\epsilon_* - A\epsilon_c)e\right]\\
&\quad + \frac{\sigma^3}{\alpha^2\ x_c'Mx_c}\left[2\{(x_c'M\epsilon_c)(\epsilon_* - A\epsilon_c)'(\epsilon_* - A\epsilon_c)\right.\\
&\qquad + (\epsilon_c'N\epsilon_c)(\epsilon_* - A\epsilon_c)'(x_* - x_R)\}e\\
&\qquad \left. - \epsilon_c'(M-2N)\epsilon_c\ X_*'(\epsilon_* - A\epsilon_c) - \frac{2\ x_c'M\epsilon_c\ \epsilon_c'N\epsilon_c}{x_c'Mx_c}X_*'(x_* - x_R)\right]\\
&\quad + O_p(\sigma^4)
\end{aligned}
$$

with e denoting a column vector with first $(k-1)$ elements as zero and last element as one.

Utilizing these results, we can express

$$(\hat{\beta} - \beta) = \sigma\xi_1 + \sigma^2\xi_2 + \sigma^3\xi_3 + O_p(\sigma^4) \qquad \text{(A.7)}$$

where

$$\begin{aligned}
\xi_1 &= \Omega(X_c' + X_*'A)\epsilon_c && \text{(A.8)}\\
\xi_2 &= \frac{1}{\alpha}(\epsilon_* - A\epsilon_c)'A\epsilon_c\ \Omega e + \frac{x_c'M\epsilon_c}{\alpha\ x_c'Mx_c}\Omega X_*'(\epsilon_* - A\epsilon_c) && \text{(A.9)}\\
&\quad + \frac{\epsilon_c'N\epsilon_c}{\alpha\ x_c'Mx_c}\Omega X_*'(x_* - x_R) - \Omega(X_*'U + U'X_*)\xi_1\\
\xi_3 &= \frac{1}{\alpha^2\ x_c'Mx_c}\{(x_c'M\epsilon_c)(\epsilon_* - A\epsilon_c)'(\epsilon_* - 2A\epsilon_c) && \text{(A.10)}\\
&\quad + (\epsilon_c'N\epsilon_c)(x_* - x_R)(\epsilon_* - 2A\epsilon_c)\}\Omega e\\
&\quad - \frac{\epsilon_c'(M-2N)\epsilon_c}{\alpha^2\ x_c'Mx_c}\Omega X_*'(\epsilon_* - A\epsilon_c)\\
&\quad - \frac{2\ x_c'M\epsilon_c\ \epsilon_c'N\epsilon_c}{\alpha^2\ (x_c'Mx_c)^2}\Omega X_*'(x_* - x_R)\\
&\quad + (\epsilon_c'N\epsilon_c)(x_* - x_R)'(\epsilon_* - 2A\epsilon_c)\}\Omega e\\
&\quad - \frac{(\epsilon_* - A\epsilon_c)'A\epsilon_c}{\alpha}\Omega(X_*'U + U'X_*)\Omega e\\
&\quad + \frac{x_c'M\epsilon_c}{\alpha\ x_c'Mx_c}\Omega(X_*'U + U'X_*)\Omega(X_c'\epsilon_c - X_*'(\epsilon_* - 2A\epsilon_c))\\
&\quad - \frac{\epsilon_c'N\epsilon_c}{\alpha\ x_c'Mx_c}\Omega(X_*'U + U'X_*)\Omega X_*'(x_* - x_R)\\
&\quad + \Omega(X_*'U + U'X_*)\Omega(X_*'U + U'X_*)\Omega(X_c' + X_*'A)\epsilon_c\\
&\quad + \frac{\epsilon_c'N\epsilon_c}{\alpha^2\ x_c'Mx_c}\Omega[X_*'(X_* - X_R) + (X_* - X_R)'X_*]\Omega(X_c' + X_*'A)\epsilon_c\\
&\quad - \Omega U'U\Omega(X_c' + X_*'A)\epsilon_c.
\end{aligned}$$

Let us first note the following results which are repeatedly employed in finding the expectations:

$$\begin{aligned}
AX_c &= (AZ_c \quad Ax_c)\\
&= (Z_* \qquad x_*) \quad = \quad X_*
\end{aligned} \tag{A.11}$$

$$\begin{aligned}
&X_*(X_c'X_c)^{-1}X_* && \text{(A.12)}\\
&= (Z_* \quad x_*)\begin{pmatrix} Z_c'Z_c & Z_c'x_c \\ x_c'Z_c & x_c'x_c \end{pmatrix}^{-1}\begin{pmatrix} Z_*' \\ x_*\prime \end{pmatrix}\\
&= \frac{1}{x_c'Mx_c}[Z_*(Z_c'Z_c)^{-1}Z_c'x_cx_c'Z_c(Z_c'Z_c)^{-1}Z_*' - Z_*(Z_c'Z_c)^{-1}Z_c'x_cx_*'\\
&\qquad - x_*x_c'Z_c(Z_c'Z_c)^{-1}Z_*' + x_*x_*'] + Z_*(Z_c'Z_c)^{-1}Z_*'\\
&= \frac{1}{x_c'Mx_c}(x_* - x_R)(x_* - x_R)' + Z_*(Z_c'Z_c)^{-1}Z_*'\\
&= AA'
\end{aligned}$$

$$\begin{aligned}
&X_*'A(X_c + A'X_*) && \text{(A.13)}\\
&= X_*'AX_c + X_*'AA'X_*\\
&= X_*'X_* + X_*'X_*(X_c'X_c)^{-1}X_*'X_*\\
&= (\Omega X_c'X_c\Omega)^{-1} - \Omega^{-1}
\end{aligned}$$

$$
\begin{aligned}
&(X_c' + X_*'A)(X_c + A'X_*) && \text{(A.14)}\\
&\quad = X_c'X_c + X_*'AX_c + X_c'A'X_* + X_*'AA'X_* \\
&\quad = X_c'X_c + 2X_*'X_* + X_*'S_*(X_c'X_c)^{-1}X_*'X_* && \text{(A.15)}\\
&\quad = (\Omega X_c'X_c\Omega)^{-1}.
\end{aligned}
$$

If D denotes a square matrix and d is any row vector, then we have

$$
\begin{aligned}
\mathrm{E}(\epsilon_c' D\epsilon_c\ \epsilon_c') &= \gamma_1 l_c'(I * D) && \text{(A.16)}\\
\mathrm{E}(d\epsilon_c\ \epsilon_c\epsilon_c') &= \gamma_1(I * l_c d) && \text{(A.17)}
\end{aligned}
$$

where '$*$' is the Hadamard product operator, I is an identity matrix of order $m_c \times m_c$ and l_c is a $m_c \times 1$ vector with all elements unity.

Now it is easy to see that

$$
\begin{aligned}
\mathrm{E}(\epsilon_*'A\epsilon_c - \epsilon_c'A'A\epsilon_c) &= -\operatorname{tr} A'A && \text{(A.18)}\\
&= -\operatorname{tr}(X_c'X_c)^{-1}X_*'X_* \\
\mathrm{E}[(x_c'M\epsilon_c)\Omega X_*'(\epsilon_* - A\epsilon_c)] &= -\Omega X_*'AMx_c && \text{(A.19)}\\
&= -\Omega X_*'(x_* - x_R) \\
\mathrm{E}(\epsilon_c'N\epsilon_c) &= \operatorname{tr} N && \text{(A.20)}\\
&= (m_c - k) \\
\mathrm{E}(\Omega X_*'U\xi_1) &= \frac{1}{\alpha}\,\mathrm{E}[\xi_1'e\ \Omega X_*'(\epsilon_* - A\epsilon_c)] && \text{(A.21)}\\
&= \frac{1}{\alpha}\,\mathrm{E}[\epsilon_c'(X_c + A'X_*)\Omega e\ \Omega X_*'(\epsilon_* - A\epsilon_c)] \\
&= -\frac{1}{\alpha}\Omega X_*'A(X_c + A'X_*)\Omega e \\
&= -\frac{1}{\alpha}[(X_c'X_c)^{-1} - \Omega]e \\
&= -\frac{1}{\alpha}\Omega X_*'X_*(X_c'X_c)^{-1}e \\
\mathrm{E}(\Omega U'X_*\xi_1) &= \frac{1}{\alpha}\,\mathrm{E}[(\epsilon_* - A\epsilon_c)'X_*\xi_1]\Omega e && \text{(A.22)}\\
&= \frac{1}{\alpha}\,\mathrm{E}[(\epsilon_* - A\epsilon_c)'X_*\Omega(X_c' + X_*'A)\epsilon_c]\Omega e \\
&= -\frac{1}{\alpha}[\operatorname{tr}\Omega X_*'A(X_c + A'X_*)]\Omega e \\
&= -\frac{1}{\alpha}[\operatorname{tr}(X_c'X_c)^{-1}X_*'X_*]\Omega e
\end{aligned}
$$

whence it follows that

$$
\begin{aligned}
\mathrm{E}(\xi_2) &= \frac{1}{\alpha}\Omega[X_*'X_*(X_c'X_c)^{-1}e + \frac{m_c - k - 1}{x_c'Mx_c}X_*'(x_* - X_R)] && \text{(A.23)}\\
&= \frac{1}{\alpha}\Omega d
\end{aligned}
$$

In a similar manner, the following results can be obtained

$$\begin{aligned} \mathrm{E}[(x_c' M\epsilon_c)(\epsilon_* - A\epsilon_c)'(\epsilon_* - 2A\epsilon_c)] &= 2\,\mathrm{E}[x_c' M\epsilon_c\ \epsilon_c' A' A\epsilon_c] \\ &= 2\gamma_1\ x_c' M(I * A'A)l_c \end{aligned} \tag{A.24}$$

$$\mathrm{E}[(\epsilon_c' N\epsilon_c)(x_* - x_R)'(\epsilon_* - 2A\epsilon_c)] = -2\gamma_1\ (x_* - x_R)'(I * N)l_c \tag{A.25}$$

$$\mathrm{E}[\epsilon_c'(M - 2N)\epsilon_c \Omega X_*'(\epsilon_* - A\epsilon_c)] = -\gamma_1\ \Omega X_*' A\{I * (M - 2N)\}l_c \tag{A.26}$$

$$\mathrm{E}[x_c' M\epsilon_c\ \epsilon_c' N\epsilon_c] = \gamma_1\ x_c' M(I * N)l_c \tag{A.27}$$

$$\begin{aligned} &\mathrm{E}[(\epsilon_* - A\epsilon_c)' A\epsilon_c\ \Omega(X_*'U + U'X_*)\Omega e] \\ &= \frac{1}{\alpha}\,\mathrm{E}[(\epsilon_* - A\epsilon_c)' A\epsilon_c \Omega\{e'\Omega e\ X_*'(\epsilon_* - A\epsilon_c) + e'\Omega X_*'(\epsilon_* - A\epsilon_c)e\}] \\ &= \frac{\gamma_1}{\alpha}[(e'\Omega e)\Omega X_*' A(I * A'A)l_c + e'\Omega X_*' A(I * A'A)l_c\ \Omega e] \end{aligned} \tag{A.28}$$

$$\begin{aligned} &\mathrm{E}[(x_c' M\epsilon_c)\Omega(X_*'U + U'X_*)\Omega\{X_c'\epsilon_c - X_*'(\epsilon_* - 2A\epsilon_c)\}] \\ &= \frac{1}{\alpha}\,\mathrm{E}[(x_c' M\epsilon_c)e'\Omega\{X_c'\epsilon_c - X_*'(\epsilon_* - 2A\epsilon_c)\}\Omega X_*'(\epsilon_* - A\epsilon_c)] \\ &\quad + \frac{1}{\alpha}\,\mathrm{E}[(x_c' M\epsilon_c)(\epsilon_* - A\epsilon_c)' X_*\Omega\{X_c'\epsilon_c - X_*(\epsilon_* - 2A\epsilon_c)\}]\Omega e \\ &= -\frac{1}{\alpha}\,\mathrm{E}[\epsilon_c' M x_c e'\Omega(X_c' + 2X_*')\epsilon_c\ \Omega X_*' A\epsilon_c] \\ &\quad - \frac{1}{\alpha}\,\mathrm{E}[(x_c' M\epsilon_c)\epsilon_c' A' X_*\Omega(X_c' + 2X_*')\epsilon_c]\Omega e \\ &= -\frac{\gamma_1}{\alpha}\Omega X_*' A[I * M x_c e'\Omega(X_c' + 2X_*')]l_c \\ &\quad - \frac{\gamma_1}{\alpha} x_c' M[I * A' X_*\Omega(X_c' + 2X_*')]l_c\ \Omega e \end{aligned} \tag{A.29}$$

$$\begin{aligned} &\mathrm{E}[(\epsilon_c' N\epsilon_c)\Omega(X_*'U + U'X_*)\Omega X_*'(x_* - x_R)] \\ &= \frac{1}{\alpha}\,\mathrm{E}[(\epsilon_c' N\epsilon_c)e'\Omega X_*'(x_* - x_R)\ \Omega X_*'(\epsilon_* - A\epsilon_c)] \\ &\quad + \frac{1}{\alpha}\,\mathrm{E}[(\epsilon_c' N\epsilon_c)(\epsilon_* - A\epsilon_c)' X_*\Omega X_*'(x_* - x_R)]\Omega e \\ &= -\frac{\gamma_1}{\alpha}\{e'\Omega X_*'(x_* - x_R)\}\Omega X_*' A(I * N)l_c \\ &\quad - \frac{\gamma_1}{\alpha}(x_* - x_R)' X_*\Omega X_*' A(I * N)l_c\ \Omega e \end{aligned} \tag{A.30}$$

$$
\begin{aligned}
&\mathrm{E}[\Omega(X_*'U + U'X_*)\Omega(X_*'U + U'X_*)\Omega(X_c' + X_*'A)\epsilon_c] && \text{(A.31)}\\
&= \frac{1}{\alpha^2}\,\mathrm{E}[e'\Omega(X_c' + X_*'A)\epsilon_c\; e'\Omega X_*'(\epsilon_* - A\epsilon_c)\;\Omega X_*'(\epsilon_* - A\epsilon_c)]\\
&\quad + \frac{1}{\alpha^2}(e'\Omega e)\,\mathrm{E}[(\epsilon_* - A\epsilon_c)'X_*\Omega(X_c' + X_*'A)\epsilon_c\;\Omega X_*'(\epsilon_* - A\epsilon_c)]\\
&\quad + \frac{1}{\alpha^2}\,\mathrm{E}[e'\Omega(X_c' + X_*'A)\epsilon_c\;(\epsilon_* - A\epsilon_c)'X_*\Omega X_*'(\epsilon_* - A\epsilon_c)]\Omega e\\
&\quad + \frac{1}{\alpha^2}\,\mathrm{E}[e'\Omega X_*'(\epsilon_* - A\epsilon_c)\;(\epsilon_* - A\epsilon_c)'X_*\Omega(X_c' + X_*'A)\epsilon_c]\Omega e\\
&= \frac{\gamma_1}{\alpha^2}[\Omega X_*'A\{I * A'X_*\Omega ee'\Omega(X_c' + X_*'A)\}l_c\\
&\qquad + (e'\Omega e)\Omega X_*'A\{I * A'X_*\Omega(X_c' + X_*'A)\}l_c\\
&\qquad + e'\Omega(X_c' + X_*'A)(I * A'X_*\Omega X_*'A)l_c\;\Omega e\\
&\qquad + e'\Omega X_*'A\{I * A'X_*\Omega(X_c' + X_*'A)\}l_c\;\Omega e]
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{E}[(\epsilon_c'N\epsilon_c)\Omega\{X_*'(X_* - X_R) + (X_* - X_R)'X_*\}\Omega(X_c' + X_*'A)\epsilon_c]\\
&= \gamma_1\Omega\{X_*'(X_* - X_R) + (X_* - X_R)'X_*\}\Omega(X_c' + X_*'A)(I * N)l_c && \text{(A.32)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{E}[\Omega U'U\Omega(X_c' + X_*'A)\epsilon_c] && \text{(A.33)}\\
&= \frac{1}{\alpha^2}\,\mathrm{E}[(\epsilon_* - A\epsilon_c)'(\epsilon_* - A\epsilon_c)\Omega ee'\Omega(X_c' + X_*'A)\epsilon_c]\\
&= \frac{\gamma_1}{\alpha^2}[e'\Omega(X_c' + X_*'A)(I * A'A)l_c]\Omega e.
\end{aligned}
$$

Combining these results, we find

$$
\mathrm{E}(\xi_3) = (f'l_c)\Omega e + \Omega F l_c \tag{A.34}
$$

where f and F are specified by (3.14) and (3.16) respectively.

Now the bias vector of $\hat{\beta}$ to order $O(\sigma^3)$ is given by

$$
\begin{aligned}
\mathrm{B}(\hat{\beta}) &= \mathrm{E}(\hat{\beta} - \beta) && \text{(A.35)}\\
&= \sigma\,\mathrm{E}(\xi_1) + \sigma^2\,\mathrm{E}(\xi_2) + \sigma^3\,\mathrm{E}(\xi_3)\\
&= \sigma^2\,\mathrm{E}(\xi_2) + \sigma^3\,\mathrm{E}(\xi_3).
\end{aligned}
$$

Substituting (A.23) and (A.34), we obtain the result (3.18) stated in Theorem.

For the mean squared error matrix to order $O(\sigma^3)$, we observe that

$$
\mathrm{M}(\hat{\beta}) = \sigma^2\,\mathrm{E}(\xi_1\xi_1') + \sigma^3\,\mathrm{E}(\xi_2\xi_1' + \xi_1\xi_2'). \tag{A.36}
$$

It is easy to see that

$$
\begin{aligned}
\mathrm{E}(\xi_1\xi_1\prime) &= \Omega(X_c' + X_*A)(X_c + A'X_*)\Omega && \text{(A.37)}\\
&= (X_c'X_c)^{-1}.
\end{aligned}
$$

Similarly, using (A.16) and (A.17), we have

$$\begin{aligned}
\mathrm{E}[(\epsilon_* - A\epsilon_c)' A\epsilon_c \Omega e \epsilon_c'] &= -\Omega e\, \mathrm{E}(\epsilon_c' A' A \epsilon_c\, \epsilon_c) && \text{(A.38)}\\
&= -\gamma_1 \Omega e l_c'(I * A'A) \\
\mathrm{E}[(x_c' M\epsilon_c)\Omega X_*'(\epsilon_* - A\epsilon_c)\epsilon_c'] &= -\Omega X_*' A\, \mathrm{E}(x_c' M \epsilon_c\, \epsilon_c \epsilon_c') && \text{(A.39)}\\
&= -\gamma_1 \Omega X_*' A(I * l_c x_c' M) \\
\mathrm{E}[(\epsilon_c' N \epsilon_c)\epsilon_c'] &= \gamma_1 l_k'(I * N) && \text{(A.40)}\\
\mathrm{E}[X_*' U \xi_1 \epsilon_c'] &= X_*'\, \mathrm{E}[e'\Omega(X_c' + X_*' A)\epsilon_c\, (\epsilon_* - A\epsilon_c)\epsilon_c'] \\
&= -\gamma_1 X_*' A(I * l_c e' \Omega (X_c' + X_*' A)) && \text{(A.41)}\\
\mathrm{E}[U' X_* \xi_1 \epsilon_c'] &= e\, \mathrm{E}[(\epsilon_* - A\epsilon_c)' X_* \Omega (X_c' + X_*' A)\epsilon_c\, \epsilon_c'] \\
&= -\gamma_1 e l_c'(I * A' X_* \Omega (X_c' + X_*' A)). && \text{(A.42)}
\end{aligned}$$

Utilizing these results, we obtain from (A.8) and (A.9) that

$$\begin{aligned}
&\mathrm{E}(\xi_2 \xi_1') \\
&= -\frac{\gamma_1}{\alpha}\Omega \left[e l_c'(I * G) + X_*' A(I * l_c g') - \frac{1}{x_c' M x_c} X_*'(x_* - x_R) l_c'(I * N) \right] \times \\
&\quad \times (X_c + A' X_*)\Omega \qquad \text{(A.43)}
\end{aligned}$$

where the matrix G and the vector g are specified by (3.17) and (3.15) respectively.

Substituting (A.37) and (A.43) in (A.36), we obtain the expression (3.19) for the mean squared error matrix of $\hat{\beta}$ to the desired order of approximation.

References

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.

Toutenburg, H., Srivastava, V. K. and Fieger, A. (1996). Estimation of parameters in multiple regression with missing X-observations using first order regression procedure, *SFB386 – Discussion Paper 38*, Ludwig-Maximilians-Universität München.