Toutenburg, Shalabh:

Estimation of Regression Models with Equi-correlated Responses when some Observations on the Response Variable are Missing

Sonderforschungsbereich 386, Paper 174 (1999)

Estimation of Regression Models with Equi-correlated Responses when some Observations on the Response Variable are Missing

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

Shalabh
Department of Statistics
Panjab University
Chandigarh, India

November 11, 1999

Abstract: The present article deals with the problem of estimation of parameters in a linear regression model when some data on response variable is missing and the responses are equicorrelated. The ordinary least squares and optimal homogeneous predictors are employed to find the imputed values of missing observations. Their efficiency properties are analyzed using the small disturbances asymptotic theory. The estimation of regression coefficients using these imputed values is also considered and a comparison of estimators is presented.

1 Introduction

Let us consider the following linear regression model with equi-correlated disturbances:

$$Y_c = X_c\beta + \sigma\epsilon_c \tag{1.1}$$

where Y_c is a $n \times 1$ vector of n observations on the response variable, X is a $n \times K$ full column rank matrix consisting of n observations on K explanatory variables, β is a $K \times 1$ vector of coefficients, σ is an unknown scalar and ϵ_c is a $n \times 1$ vector of disturbances.

It is assumed that disturbances follow a multivariate normal distribution with mean 0, variances 1 and covariances or correlation coefficients ϱ so that we can write

$$\begin{aligned} \mathrm{E}(\epsilon_c) &= 0 \\ \mathrm{E}(\epsilon_c\epsilon_c') = \Sigma_{cc} &= (1-\varrho)I_n + \varrho J_n J_n' \end{aligned} \tag{1.2}$$

where I_n is an identity matrix of order $n \times n$, J_n denotes a column vector with all n elements unity and ϱ is assumed to be known and different from zero.

Such a model provides an interesting framework for the analysis of data in many applications. For example, when data contain measurements on symmetric organs like eyes of persons, the observations are found to be equi-correlated; see, e. g. Mũnoz, Rosner and Carey (1986) and Rosner (1984) for details. Other instances relate to familial data and survey data arising from cluster sampling; see, e. g. , Christensen (1987), King and Evans (1986) and Srivastava (1984).

For the estimation of model parameters, considerable attention has been paid in the literature, see, e. g. Srivastava and Ng (1990) and the references cited therein for a brief review of estimation procedures. A stringent assumption made in all the procedures is that the data has no missing observation. Such a specification may be violated in many practical situations and some observations on the response variable may not be available for one reason or the other. If Y_{mis} denotes a $m \times 1$ vector of m missing values in the response variable and X_* is a $m \times K$ matrix of m observations on the K explanatory variables, we have

$$Y_{mis} = X_*\beta + \sigma\epsilon_* \tag{1.3}$$

where ϵ_* is a $m \times 1$ vector of disturbances having same distributional properties as ϵ_c, i. e. ,

$$\begin{aligned} \mathrm{E}(\epsilon_*\epsilon_*') &= \Sigma_{**} = (1-\varrho)I_m + \varrho J_m J_m' \\ \mathrm{E}(\epsilon_*\epsilon_c') &= \Sigma_{*c} = \varrho J_m J_n' \,. \end{aligned} \tag{1.4}$$

Estimation of β on the basis of $(n+m)$ incomplete observations is the subject matter of this article. In Section 2, we present three sets of inputed values for the missing observations. Utilizing these imputed values, we repair the incomplete data set and use it for the estimation of β. In Section 3, we discuss the properties of imputed values. In the same way, the efficiency properties of estimators of β are analyzed in Section 4.

2 Imputation of Missing Observations And Estimation of Coefficients

Let us assume, following Shalabh (1998), that the regression relationship contains no intercept term and the observations on explanatory variables are taken as deviations from their corresponding means so that $X_c'J_n$ and $X_*'J_m$ are null vectors. An interesting consequence of this specification is that ordinary least squares and generalized least squares estimators of β from (1.1) are identically equal and given by

$$b_c = (X_c'X_c)^{-1}X_c'Y_c \tag{2.1}$$

which clearly does not utilize the m incomplete observations at all; see Toutenburg and Trenkler (1998) for the case when intercept term is present.

A simple alternative for making use of m incomplete observations is to find imputed values for missing observations on response variable and to substitute them in place of missing observations so that the thus repaired data set resembles the complete data set. Treating the problem of finding the imputed values as the problem of predicting the values of response variable outside the sample, Goldberger (1962) has presented the classical predictor of Y_{mis} as

$$P_1 = X_* b_c \tag{2.2}$$

and the optimal homogeneous predictor in the class of unbiased predictors of Y_{mis} as

$$\begin{aligned} P_2 &= X_* b_c + \Sigma_{*c}\Sigma_{cc}^{-1}(Y_c - X_c b_c) \\ &= X_* b_c + \frac{\varrho J_n' Y_c}{1+(n-1)\varrho} J_m \end{aligned} \tag{2.3}$$

which may be serve as the imputed values for missing observations on the response variable; see also Bibby and Toutenburg (1979).

If we relax the constraint of unbiasedness, the optimal homogeneous predictor is given by

$$\begin{aligned} P &= \frac{\beta' X_c' \Sigma_{cc}^{-1} Y_c}{\beta' X_c' \Sigma_{cc}^{-1} X_c \beta + \sigma^2} X_* \beta \\ &\quad + \Sigma_{*c}\Sigma_{cc}^{-1}(Y_c - \frac{\beta' X_c' \Sigma_{cc}^{-1} Y_c}{\beta' X_c' \Sigma_{cc}^{-1} X_c \beta + \sigma^2} X_c \beta) \\ &= \frac{\beta' X_c Y_c}{\beta' X_c' X_c \beta + \sigma^2(1-\varrho)} X_* \beta + \frac{\varrho J_n' Y_c}{1+(n-1)\varrho} J_m \end{aligned} \tag{2.4}$$

see, e. g. ,Rao and Toutenburg (1995, Sec. 6.5).

The predictor (2.4) has no practical utility due to involvement of unknown quantities σ^2 and β. A simple way to obtain a feasible version is to replace them by their unbiased estimators. Thus substituting b_c in place of β and

$$s^2 = \left(\frac{1}{n-K}\right)(Y_c - X_c b_c)' \Sigma_{cc}^{-1} (Y_c - X_c b_c) \tag{2.5}$$

in place of σ^2 in (2.4), we find a feasible predictor for Y_{mis} as follows:

$$\begin{aligned} \hat{P} &= \frac{b_c' X_c' Y_c}{b_c' X_c' X_c b_c + s^2(1-\varrho)} X_* b_c + \frac{\varrho J_n' Y_c}{1+(n-1)\varrho} J_m \\ &= P_2 - \frac{s^2(1-\varrho)}{b_c' X_c' X_c b_c + s^2(1-\varrho)} P_1 \end{aligned} \tag{2.6}$$

which provides another set of imputed values for the missing observations on the response variable.

If we apply the method of ordinary least squares or generalized least squares to (1.1) and (1.3), we get the estimator of β as

$$\beta_* = (X_c' X_c + X_*' X_*)^{-1}(X_c' Y_c + X_*' Y_{mis}) \tag{2.7}$$

which has no utility due to presence of Y_{mis}. Replacing it by the vectors of imputed values, we obtain the following estimators of β:

$$\begin{aligned}
\hat{\beta}_1 &= (X_c'X_c + X_*'X_*)^{-1}(X_c'Y_c + X_*'P_1) \qquad (2.8)\\
&= b_c \\
\hat{\beta}_2 &= (X_c'X_c + X_*'X_*)^{-1}(X_c'Y_c + X_*'P_2) \qquad (2.9)\\
&= b_c \\
\hat{\beta} &= (X_c'X_c + X_*'X_*)^{-1}(X_c'Y_c + X_*'\hat{P}) \qquad (2.10)\\
&= \left[I_k - \frac{s^2(1-\varrho)}{b_c'X_c'X_cb_c + s^2(1-\varrho)}(X_c'X_c + X_*'X_*)^{-1}X_*'X_*\right] b_c\,.
\end{aligned}$$

Thus we observe that the estimators of β employing the unbiased imputed values specified by (2.2) and (2.3) are identically equal to the least squares estimator b_c which ignores the incomplete observations. In other words, the imputation procedure yielding unbiased imputed values does not serve any useful purpose. Such is, however, not the case when biased imputed values given by (2.6) are used. Here we obtain the estimator (2.10) that is clearly a shrunken estimator arising from b_c.

3 Properties Of Imputed Values

From (2.2), (2.3) and (2.6), it is easy to see that P_1 and P_2 are, but $\hat{P}$ is generally not, weakly unbiased for Y_{mis} in the sense that

$$\begin{aligned}
\mathrm{E}(P_1 - Y_{mis}) &= \mathrm{E}(P_2 - Y_{mis}) = 0 \qquad (3.1)\\
\mathrm{E}(\hat{P} - Y_{mis}) &\neq 0\,. \qquad (3.2)
\end{aligned}$$

Further, we have

$$\begin{aligned}
\mathrm{V}(P_1) &= \mathrm{E}(P_1 - Y_{mis})(P_1 - Y_{mis})' \qquad (3.3)\\
&= \sigma^2(1-\varrho)\left[I_m + X_*(X_c'X_c)^{-1}X_*' + \frac{\varrho}{1-\varrho}J_mJ_m'\right]\\
\mathrm{V}(P_2) &= \mathrm{E}(P_2 - Y_{mis})(P_2 - Y_{mis})' \qquad (3.4)\\
&= \sigma^2(1-\varrho)\left[I_m + X_*(X_c'X_c)^{-1}X_*' + \frac{\varrho}{1+(n-1)\varrho}J_mJ_m'\right]
\end{aligned}$$

whence it is clearly seen that the variance covariance matrix of P_1 exceeds the variance covariance matrix of P_2 by a non-negative semi-definite matrix. This implies that P_2 is a better choice in comparison to P_1 for finding the imputed values of missing observations on the response variable.

Next, let us consider $\hat{P}$. The exact expression for the first and second order moments of $(\hat{P} - Y_{mis})$ can be derived but the resulting expressions will be sufficiently complex and it will be hard to draw any clear inference regarding the bias as well as the superiority of $\hat{P}$ over P_1 and P_2 and vice-versa. We therefore consider their approximations using the small disturbance asymptotic theory. Such results are derived in Appendix and presented below.

Theorem 1 *The asymptotic approximation for the bias vector of $\hat{P}$ to order*

$O(\sigma^2)$ is given by

$$
\begin{aligned}
B(\hat{P}) &= \mathrm{E}(\hat{P} - Y_{mis}) \\
&= -\sigma^2 \left(\frac{1-\varrho}{\beta' X_c' X_c \beta} \right) X_* \beta
\end{aligned}
\tag{3.5}
$$

while the difference matrix to order $O(\sigma^4)$ is

$$
\begin{aligned}
D(P_2; \hat{P}) &= \mathrm{E}(P_2 - Y_{mis})(P_2 - Y_{mis})' - \mathrm{E}(\hat{P} - Y_{mis})(\hat{P} - Y_{mis})' \\
&= \sigma^4 \frac{(1-\varrho)^2}{\beta' X_c' X_c \beta} X_* Q X_*
\end{aligned}
\tag{3.6}
$$

where

$$
Q = 2(X_c' X_c)^{-1} - \frac{5(n-K)+2}{(n-K)\beta' X_c' X_c \beta} \beta\beta' \, .
$$

Using Rao and Toutenburg (1995, Theorem A.7, p.303), it is observed that Q cannot be a nonnegative definite matrix so that it follows from (3.6) that $\hat{P}$ does not dominate P_2 with respect to the criterion of mean squared error matrix to the given order of approximation. Similarly, using Rao and Toutenburg (1995, A.59, p. 304), we find that $(-Q)$ cannot be nonnegative definite except in the trivial case $K = 1$. This means that P_2 does not dominate $\hat{P}$. Thus $\hat{P}$ neither dominates P_2 nor is dominated by P_2 according to the mean squared error matrix criterion, at least to the order of our approximation.

Next, let us employ a weak criterion, viz. , the trace of mean squared error matrix, for the comparison of P_2 and $\hat{P}$:

$$
\operatorname{tr} D(P_2; \hat{P}) = \sigma^4 \frac{2(1-\varrho)^2 \beta' X_*' X_* \beta}{(\beta' X_c' X_c \beta)^2} \left[g - \left(\frac{5}{2} + \frac{1}{n-K} \right) \right] \tag{3.7}
$$

where

$$
g = \left(\frac{\beta' X_c' X_c \beta}{\beta' X_*' X_* \beta} \right) \operatorname{tr} (X_c' X_c)^{-1} X_*' X_* \, . \tag{3.8}
$$

Thus we observe that $\hat{P}$ is better than P_2 when

$$
g > \left(\frac{5}{2} + \frac{1}{n-K} \right) \tag{3.9}
$$

while the opposite is true, i. e. , P_2 is better than $\hat{P}$ when

$$
g < \left(\frac{5}{2} + \frac{1}{n-K} \right) \, . \tag{3.10}
$$

The conditions (3.9) and (3.10) are not attractive as they are hard to be verified in practice owing to involvement of β which is known.

If α_{min} and α_{max} denote the minimum and maximum eigenvalues of $X_*' X_*$ in the metric of $X_c' X_c$ and S is the sum of all the eignevalues, we observe that the condition (3.9) is satisfied as long as

$$
\frac{S}{\alpha_{max}} > \left(\frac{5}{2} + \frac{1}{n-K} \right) \tag{3.11}
$$

which is easy to verify in any given application.

Similarly, the condition (3.10) will hold true as long as

$$\frac{S}{\alpha_{min}} < \left(\frac{5}{2} + \frac{1}{n-K}\right) \tag{3.12}$$

that is easy to check.

It is interesting to note that conditions (3.9)-(3.12) are free from values of ϱ.

4 Properties of Estimators of Coefficients

For the estimation of β, we have two distinct estimators b_c and $\hat{\beta}$ specified by (2.1) and (2.10) respectively.

It is easy to see that b_c is unbiased for β while $\hat{\beta}$ is generally biased. The approximate expressions for analyzing the efficiency properties of $\hat{\beta}$ in relation to b_c are obtained in Appendix and presented below:

Theorem 2 *The asymptotic approximation for the bias vector of $\hat{\beta}$ to order $O(\sigma^2)$ is given by*

$$\begin{aligned} B(\hat{\beta}) &= \mathrm{E}(\hat{\beta} - \beta) \\ &= -\sigma^2 \left(\frac{1-\varrho}{\beta' X_c' X_c \beta}\right) W\beta \end{aligned} \tag{4.1}$$

while the difference matrix to order $O(\sigma^4)$ is

$$\begin{aligned} D(b_c; \hat{\beta}) &= \mathrm{E}(b_c - \beta)(b_c - \beta)' - \mathrm{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \\ &= \frac{\sigma^4 (1-\varrho)^2}{\beta' X_c' X_c \beta} \left[W(X_c' X_c)^{-1} + (X_c' X_c)^{-1} W' \right. \\ &\quad \left. - \frac{2}{\beta' X_c' X_c \beta} \{ W\beta\beta' + \beta\beta' W' + \left(\frac{n-K+2}{2(n-K)}\right) W\beta\beta' W' \} \right] \end{aligned} \tag{4.2}$$

where

$$W = (X_c' X_c + X_*' X_*)^{-1} X_*' X_* \, . \tag{4.3}$$

From (4.2), it is difficult to draw any clear inference about the superiority of $\hat{\beta}$ over b_c or vice-versa according to the criterion of mean squared error matrix to the given order of approximation. We therefore consider trace of the mean squared error matrix as the performance criterion. With respect to such a criterion, the estimator $\hat{\beta}$ is better than b_c when

$$\operatorname{tr} W(X_c' X_c)^{-1} > \left[2 \left(\frac{\beta' W \beta}{\beta' X_c' X_c \beta}\right) + \left(\frac{1}{2} + \frac{1}{n-K}\right) \frac{\beta' W' W \beta}{\beta' X_c' X_c \beta} \right] . \tag{4.4}$$

If λ_{max} is the maximum eigenvalue of $W(X_c' X_c)^{-1}$, we have

$$\begin{aligned} \frac{\beta' W \beta}{\beta' X_c' X_c \beta} &\leq \lambda_{max} \\ \frac{\beta' W' W \beta}{\beta' X_c' X_c \beta} &\leq \frac{\beta' W \beta}{\beta' X_c' X_c \beta} \leq \lambda_{max} \, . \end{aligned}$$

Further, if T denotes the total of eigenvalues of $W(X_c'X_c)^{-1}$, the inequality (4.4) is satisfied as long as

$$T > \left(\frac{5}{2} + \frac{1}{n-K}\right)\lambda_{max} \tag{4.5}$$

which is a sufficient condition for the superiority of $\hat{\beta}$ over b_c with respect to the criterion of trace of mean squared error matrix to order $\mathrm{O}(\sigma^4)$. In other words, under the condition (4.5), use of imputation procedure providing biased imputed values for the missing observations of the response variable is worthwhile in comparison to the outright discard of incomplete observations so far as the estimation of coefficients in the model is concerned. An interesting aspect of the condition (4.5) is that it is easy to check in actual practice.

Appendix

From (1.1), (2.1) and (2.5), we observe that

$$\begin{aligned} s^2 &= \left(\frac{\sigma^2}{n-K}\right)\epsilon_c'\left[I_n - X_c(X_c'X_c)^{-1}X_c'\right] \\ &\quad \cdot \left[(1-\varrho)I_n + \varrho J_n J_n'\right]^{-1}\left[I_n - X_c(X_c'X_c)^{-1}X_c'\right]\epsilon_c \\ &= \frac{\sigma^2}{(n-K)(1-\varrho)}\epsilon_c'\left[I_n - X_c(X_c'X_c)^{-1}X_c'\right]\left[I_n - \frac{\varrho}{1+(n-1)\varrho}J_nJ_n'\right] \\ &\quad \cdot \left[I_n - X_c(X_c'X_c)^{-1}X_c'\right]\epsilon_c \\ &= \frac{\sigma^2}{(n-K)(1-\varrho)}\epsilon_c'M\epsilon_c \end{aligned} \tag{A.1}$$

so that

$$\begin{aligned} &\frac{s^2(1-\varrho)}{b_c'X_c'X_cb_c + s^2(1-\varrho)} \\ &\quad = \sigma^2\frac{\epsilon_c'M\epsilon_c}{(n-K)\beta'X_c'X_c\beta}\left[1 + 2\sigma\frac{\beta'X_c'\epsilon_c}{\beta'X_c'X_c\beta} + O_p(\sigma^2)\right]^{-1} \\ &\quad = \sigma^2\frac{\epsilon_c'M\epsilon_c}{(n-K)\beta'X_c'X_c\beta} - 2\sigma^3\frac{\epsilon_c'M\epsilon_c\beta'X_c'\epsilon_c}{(n-K)(\beta'X_c'X_c\beta)^2} + O_p(\sigma^4) \end{aligned} \tag{A.2}$$

where

$$M = I_n - X_c(X_c'X_c)^{-1}X_c' - \frac{\varrho}{1+(n-1)\varrho}J_nJ_n'\,.$$

Using it, we have

$$\begin{aligned} (\hat{P} - Y_{mis}) &= (P_2 - Y_{mis}) - \frac{s^2(1-\varrho)}{b_c'X_c'X_cb_c + s^2(1-\varrho)}X_* \\ &\quad \left[\beta + \sigma(X_c'X_c)^{-1}X_c'\epsilon_c\right] \\ &= \sigma\left[X_*(X_c'X_c)^{-1}X_c'\epsilon_c + \frac{\varrho J_n'\epsilon_c}{1+(n-1)\varrho} - \epsilon_*\right] \\ &\quad -\sigma^2\frac{\epsilon_c'M\epsilon_c}{(n-K)\beta'X_c'X_c\beta}X_*\beta + O_p(\sigma^3) \end{aligned} \tag{A.3}$$

whence it follows that

$$\begin{aligned}\mathrm{E}(\hat{P}-Y_{mis}) &= -\sigma^2\frac{\mathrm{E}(\epsilon_c' M\epsilon_c)}{(n-K)\beta' X_c' X_c\beta}X_*\beta + O(\sigma^3)\\ &= -\sigma^2\frac{\mathrm{tr}\, M\Sigma_{cc}}{(n-K)\beta' X_c' X_c\beta}X_*\beta + O(\sigma^3)\end{aligned}$$

which leads to the result (3.5) of Theorem 1.

Similarly, we have

$$\begin{aligned}D(P_2;\hat{P}) &= \mathrm{E}(P_2-Y_{mis})(P_2-Y_{mis})' - \mathrm{E}(\hat{P}-Y_{mis})(\hat{P}-Y_{mis})'\\ &= \mathrm{E}\left[\frac{s^2(1-\varrho)}{b_c' X_c' X_c b_c + s^2(1-\varrho)}\{X_* b_c(P_2-Y_{mis})' + (P_2-Y_{mis})b_c' X_*'\}\right]\\ &\quad -\mathrm{E}\left[\left\{\frac{s^2(1-\varrho)}{b_c' X_c' X_c b_c + s^2(1-\varrho)}\right\}^2 X_* b_c b_c' X_*'\right].\end{aligned}$$

Observing that

$$\begin{aligned}X_* b_c(P_2-Y_{mis})' &= [\sigma X_*\beta + \sigma^2 X_*(X_c' X_c)^{-1}X_c'\epsilon_c]\\ &\quad [\epsilon_c' X_c'(X_c' X_c)^{-1}X_*' + \frac{\varrho J_n'\epsilon_c}{1+(n-1)\varrho}J_m' - \epsilon_*']\\ X_* b_c b_c' X_*' &= X_*\beta\beta' X_*' + O_p(\sigma)\end{aligned}$$

and using (A.2), we get

$$D(P_2;\hat{P}) = \sigma^3\,\mathrm{E}(F+F') + \sigma^4\,\mathrm{E}(G+G'-H) \tag{A.4}$$

where

$$\begin{aligned}F &= \frac{\epsilon_c' M\epsilon_c}{(n-K)\beta' X_c' X_c\beta}X_*\beta[\epsilon_c' X_c(X_c' X_c)^{-1}X_*' + \frac{\varrho J_m'\epsilon_c}{1+(n-1)\varrho}J_m' - \epsilon_*']\\ G &= \frac{\epsilon_c' M\epsilon_c}{(n-K)\beta' X_c' X_c\beta}X_*[(X_c' X_c)^{-1} - \frac{2}{\beta' X_c' X_c\beta}\beta\beta']X_c'\epsilon_c\\ &\quad [\epsilon_c' X_c(X_c' X_c)^{-1}X_*' + \frac{\varrho J_n'\epsilon_c}{1+(n-1)\varrho}J_m' - \epsilon_*']\\ H &= \left[\frac{\epsilon_c' M\epsilon_c}{(n-K)\beta' X_c' X_c\beta}\right]^2 X_*\beta\beta' X_*'.\end{aligned}$$

Next, by virtue of normality, we have

$$\begin{aligned}\mathrm{E}(\epsilon_c' M\epsilon_c\,\epsilon_c\epsilon_c') &= (\mathrm{tr}\, M\Sigma_{cc})\Sigma_{cc} + 2\Sigma_{cc}M\Sigma_{cc}\\ &= (n-K+2)(1-\varrho)[(1-\varrho)J_n + \varrho J_n J_n')] - 2(1-\varrho)^2 X_c(X_c' X_c)^{-1}X_c'.\end{aligned}$$

Further, ϵ_c and $(\epsilon_* - \Sigma_{*c}\Sigma_{cc}^{-1}\epsilon_c)$ are stochastically independent so that

$$\begin{aligned}\mathrm{E}(\epsilon_c' M\epsilon_c\,\epsilon_c\epsilon_*') &= \mathrm{E}(\epsilon_c' M\epsilon_c\,\epsilon_c\epsilon_c')\Sigma_{cc}^{-1}\Sigma_{*c}'\\ &= (\mathrm{tr}\, M\Sigma_{cc})\Sigma_{*c}' + 2\Sigma_{cc}M\Sigma_{*c}'\\ &= (n-K+2)(1-\varrho)\varrho J_n J_m'.\end{aligned} \tag{A.5}$$

Utilizing these results, it is easy to see that

$$\begin{aligned}
\mathrm{E}(F) &= 0 \\
\mathrm{E}(G) &= \frac{(1-\varrho)^2}{\beta' X_c' X_c \beta} X_* [(X_c' X_c)^{-1} - \frac{2}{\beta' X_c' X_c \beta} \beta\beta'] X_*' \\
\mathrm{E}(H) &= \frac{(1-\varrho)^2 (n-K+2)}{(n-K)(\beta' X_c' X_c \beta)^2} X_* \beta\beta' X_*' .
\end{aligned}$$

Substituting these results in (A.4), we obtain the expression (3.6) of Theorem 1.

For the results of Theorem 2, we observe from (2.10) and (A.2), that

$$(\hat{\beta} - \beta) = \sigma (X_c' X_c)^{-1} X_c' \epsilon_c - \frac{\sigma^2 \epsilon_c M \epsilon_c}{(n-K)\beta' X_c' X_c \beta} W\beta + O_p(\sigma^3)$$

with $W = (X_c' X_c + X_*' X_*)^{-1} X_*' X_*$ so that the bias vector to order $O(\sigma^2)$ is.

$$\begin{aligned}
B(\hat{\beta}) &= -\frac{\sigma^2\, \mathrm{E}(\epsilon_c' M \epsilon_c)}{(n-K)\beta' X_c' X_c \beta} W\beta \\
&= -\sigma^2 \left(\frac{1-\varrho}{\beta' X_c' X_c \beta} \right) W\beta
\end{aligned}$$

leading to the result (4.1).

Similarly, we have

$$\begin{aligned}
D(b_c; \hat{\beta}) &= \mathrm{E}(b_c - \beta)(b_c - \beta)' - \mathrm{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \\
&= \mathrm{E}\left[\frac{s^2(1-\varrho)}{b_c' X_c' X_c b_c + s^2(1-\varrho)} \{ W b_c (b_c - \beta)' + (b_c - \beta) b_c' W' \} \right] \\
&\quad - \mathrm{E}\left[\left\{ \frac{s^2(1-\varrho)}{b_c' X_c' X_c b_c + s^2(1-\varrho)} \right\}^2 W b_c b_c' W' \right] \\
&= \sigma^3\, \mathrm{E}\left[\frac{\epsilon_c' M \epsilon_c}{(n-K)\beta' X_c' X_c \beta} \{ W\beta \epsilon_c' X_c (X_c' X_c)^{-1} + (X_c' X_c)^{-1} X_c' \epsilon_c \beta' W' \} \right] \\
&\quad + \sigma^4\, \mathrm{E}\left[\frac{\epsilon_c' M \epsilon_c}{(n-K)\beta' X_c' X_c \beta} W \left\{ (X_c' X_c)^{-1} - \frac{2}{\beta' X_c' X_c \beta} \beta\beta' \right\} X_c' \epsilon_c \epsilon_c' X_c (X_c' X_c)^{-1} \right. \\
&\quad + \frac{\epsilon_c' M \epsilon_c}{(n-K)\beta' X_c' X_c \beta} (X_c' X_c)^{-1} X_c \epsilon_c \epsilon_c' X_c \left\{ (X_c' X_c)^{-1} - \frac{2}{\beta' X_c' X_c \beta} \beta\beta' \right\} \\
&\quad \left. - \frac{(\epsilon_c' M \epsilon_c)^2}{(n-K)^2 (\beta' X_c' X_c \beta)^2} W\beta\beta' W' \right] + O(\sigma^5) \\
&= \frac{\sigma^4 (1-\varrho)^2}{\beta' X_c' X_c \beta} \left[W (X_c' X_c)^{-1} + (X_c' X_c)^{-1} W' \right. \\
&\quad \left. - \frac{1}{\beta_c' X_c' X_c \beta} \left\{ 2(W\beta\beta' + \beta\beta' W') + \left(\frac{n-K+2}{n-K} \right) W\beta\beta' W' \right\} \right]
\end{aligned}$$

whence we get the result (4.2) of Theorem 2.

References

Bibby, J. M. and Toutenburg, H. (1979). *Prediction and Improved Estimation in Linear Models*, 2 edn, Wiley, New York.

Christensen, R. (1987). The analysis of two stage sampling data by ordinary least squares, *Journal of the American Statistical Association* **82**: 492–498.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized regression model, *Journal of the American Statistical Association* **57**: 369–375.

King, M. L. and Evans, M. A. (1986). Testing for block effects in regression models based on survey data, *Journal of the American Statistical Association* **81**: 677–679.

Mũnoz, A., Rosner, B. and Carey, V. (1986). Regression analysis in the presence of heterogeneous intraclass correlations, *Biometrics* **40**: 653–658.

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

Rosner, B. (1984). Multivariate methods in ophtalmology with application to paired-data situations, *Biometrics* **40**: 1025–1035.

Shalabh (1998). Unbiased prediction in linear regression models with equicorrelated responses, *Statistical Papers* **39**: 237–244.

Srivastava, M. S. (1984). Estimation of intraclass correlation in familial data, *Biometrika* **71**: 177–185.

Srivastava, M. S. and Ng, F. K. L. (1990). Estimation of intraclass correlation in regression models, *Gujrat Statistical Review* **17**: 229–236.

Toutenburg, H. and Trenkler, G. (1998). Using first differences as a device against multicollinearity, *in* R. Galata and H. Küchenhoff (eds), *Econometrics in Theory and Practice, Festschrift für Hans Schneeweiß*, Physica, Heidelberg, pp. 131–136.