Ann. Inst. Statist. Math.
Vol. 48, No. 1, 89–95 (1996)

r-k CLASS ESTIMATION IN REGRESSION MODEL WITH CONCOMITANT VARIABLES

MASAYUKI JIMICHI[1] AND NOBUO INAGAKI[2]

[1] *School of Business Administration, Kwansei Gakuin University, 1-155, Uegaharaichiban-cho, Nishinomiya, Hyogo 662, Japan*
[2] *Department of Mathematical Science, Faculty of Engineering Science, Osaka University, Machikaneyama-cho, Toyonaka, Osaka 560, Japan*

(Received August 8, 1994; revised April 25, 1995)

Abstract. We treat with the r-k class estimation in a regression model, which includes the ordinary least squares estimator, the ordinary ridge regression estimator and the principal component regression estimator as special cases of the r-k class estimator. Many papers compared total mean square error of these estimators. Sarkar (1989, *Ann. Inst. Statist. Math.*, **41**, 717–724) asserts that the results of this comparison are still valid in a misspecified linear model. We point out some confusions of Sarkar and show additional conditions under which his assertion holds.

Key words and phrases: Concomitant variables, multicollinearity, ordinary least squares estimator, ordinary ridge regression estimator, principal component regression estimator, r-k class estimator.

1. Introduction

It is well-known in the regression model :

$$Y = X\beta + \varepsilon \tag{1.1}$$

that the ordinary least square (OLS) estimator $\hat{\beta}$ of the regression coefficient vector β has large variance under the multicollinearity of explanatory vectors

$$X = (x_1, \ldots, x_p).$$

In order to take care of the trouble of this variance divergence, we discuss the following two devices to solve the normal equation of the least square method although these cause their estimators to be biased:

(L1) Reduction of the $p \times p$ information matrix $X'X$ to $r \times r$ one with $r < p$ eliminates the multicollinearity.

(L2) Addition of the dummy diagonal matrix kI_p to $X'X$ ensures the stability of the solution and its variance.

From the point of view of (L1), Kendall (1957) proposed and Marquardt (1970) investigated the principal component regression (PCR) estimator $\hat{\boldsymbol{\beta}}(r,\cdot)$ (say), where r is the number of principal components selected. From the point of view of (L2), Hoerl and Kennard (1970) proposed the ordinary ridge regression (ORR) estimator $\hat{\boldsymbol{\beta}}(\cdot,k)$ (say). See also Lawless and Wang (1976).

Now, we face to the r-k class estimator derived from a combination with the above two devices, which is suggested by Marquardt (1970) and examined in detail by Baye and Parker (1984). Let us denote the r-k class estimator with r selected principal components and the dummy constant k by $\hat{\boldsymbol{\beta}}(r,k)$. Then, the OLS, PCR and the ORR estimators are rewritten with this notation as follows:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(p,0), \qquad \hat{\boldsymbol{\beta}}(r,\cdot) = \hat{\boldsymbol{\beta}}(r,0), \qquad \hat{\boldsymbol{\beta}}(\cdot,k) = \hat{\boldsymbol{\beta}}(p,k), \tag{1.2}$$

respectively.

Baye and Parker (1984) showed that the total mean square error (TMSE) of the r-k class estimator is less than one of the PCR estimator under the common r and adaptive k:

$$\mathrm{TMSE}(\hat{\boldsymbol{\beta}}(r,k)) \le \mathrm{TMSE}(\hat{\boldsymbol{\beta}}(r,\cdot)),$$

where, denoting the Euclidean norm by $\|\cdot\|$,

$$\mathrm{TMSE}(\hat{\boldsymbol{\beta}}(r,k)) = E\{(\hat{\boldsymbol{\beta}}(r,k)-\boldsymbol{\beta})'(\hat{\boldsymbol{\beta}}(r,k)-\boldsymbol{\beta})\} = E\{\|\hat{\boldsymbol{\beta}}(r,k)-\boldsymbol{\beta}\|^2\}.$$

In the same way, Nomura and Ohkubo (1985) also showed the superiority of the r-k class estimators with adaptive r, k over the OLS, ORR estimators. See also Martinez (1990).

Sarkar (1989) treated with the following misspecified linear regression model:

$$\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \qquad \boldsymbol{u} \equiv \boldsymbol{C}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{1.3}$$

which is caused by omission of some explanatory vectors

$$\boldsymbol{C} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_q),$$

whereas Rao (1973) discussed the same model from the point of view that concomitant vectors $\boldsymbol{C}$ are added to the explanatory vectors and are available to estimation of regression coefficients. Sarkar asserted, unfortunately with some confusions, that the superiority of the r-k class estimator over the OLS, PCR, ORR estimators is still kept in the misspecified model under the same conditions as in Baye and Parker and Nomura and Ohkubo. Our main aim is to point out his confusions and to correct the conditions so that the Sarkar's assertion could be affirmed.

2. Preliminary

Let us consider the model (1.3) where $\boldsymbol{Y}^*$ is an n dimensional column vector of response variables, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ are p, q dimensional column vectors of unknown regression coefficient parameters, respectively, and $\boldsymbol{X}$, $\boldsymbol{C}$, $\boldsymbol{\varepsilon}$ satisfy the following assumptions:

ASSUMPTIONS. (A1) $\boldsymbol{X}$ is an $n \times p$ matrix of explanatory vectors standardized in such way $\boldsymbol{X}'\boldsymbol{X}$ is a correlation matrix and $\boldsymbol{C}$ an $n \times q$ matrix of concomitant vectors, which are observable and satisfy

$$\mathrm{rank}(\boldsymbol{X}, \boldsymbol{C}) = p + q (\leq n), \qquad \boldsymbol{X}'\boldsymbol{C} \neq \boldsymbol{O}.$$

(A2) $\boldsymbol{\varepsilon}$ is an n dimensional column vector of errors with the following mean vector and variance matrix:

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}, \qquad \mathrm{V}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$$

where $\boldsymbol{I}_n$ is the n dimensional identity matrix and σ^2 is an unknown variance of error.

In this model, we consider the r-k class estimator for regression coefficients without using concomitant variables explicitly: letting $N_p = \{1, \ldots, p\}$,

$$\hat{\boldsymbol{\beta}}^*(r,k) = \boldsymbol{T}_r \boldsymbol{\Lambda}_r(k)^{-1} \boldsymbol{T}_r' \boldsymbol{X}' \boldsymbol{Y}^*, \quad \text{for} \quad r \in N_p, \ k \geq 0, \tag{2.1}$$

which includes the OLS, PCR, ORR estimators as special cases, respectively:

$$\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{Y}^*, \tag{2.2}$$

$$\hat{\boldsymbol{\beta}}^*(r, \cdot) = \boldsymbol{T}_r \boldsymbol{\Lambda}_r^{-1} \boldsymbol{T}_r' \boldsymbol{X}' \boldsymbol{Y}^*, \quad r \in N_p, \tag{2.3}$$

$$\hat{\boldsymbol{\beta}}^*(\cdot, k) = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p)^{-1} \boldsymbol{X}' \boldsymbol{Y}^*, \quad k \geq 0. \tag{2.4}$$

In the above equations, we use the notations defined as follows : $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ is the diagonal matrix of eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ with

$$\lambda_1 \geq \cdots \geq \lambda_r \geq \lambda_{r+1} \geq \cdots \geq \lambda_p > 0$$

and $\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_p)$ is a $p \times p$ orthogonal matrix satisfing $\boldsymbol{T}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T} = \boldsymbol{\Lambda}$. Furthermore, $\boldsymbol{T}_r$ and $\boldsymbol{T}_{p\backslash r}$ are $p \times r$ and $p \times (p-r)$ orthogonal submatrices of $\boldsymbol{T}$, respectively:

$$\boldsymbol{T} = (\boldsymbol{T}_r; \boldsymbol{T}_{p\backslash r}) = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_r; \boldsymbol{t}_{r+1}, \ldots, \boldsymbol{t}_p)$$

satisfing

$$\boldsymbol{T}_r'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}_r = \boldsymbol{\Lambda}_r \equiv \mathrm{diag}(\lambda_1, \ldots, \lambda_r),$$
$$\boldsymbol{T}_{p\backslash r}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{T}_{p\backslash r} = \boldsymbol{\Lambda}_{p\backslash r} \equiv \mathrm{diag}(\lambda_{r+1}, \ldots, \lambda_p),$$

and $\boldsymbol{\Lambda}_r(k) = \boldsymbol{\Lambda}_r + k\boldsymbol{I}_r$.

Sarkar (1989) discussed the r-k class estimator in the misspecified regression model which is the same as the above concomitant regression model without the fact that $\boldsymbol{C}$ is unobservable in the former model but observable in the latter. Then, he asserted that the superiority of r-k class estimator holds still in the misspecified regression model under the same conditions as in the usual regression model (1.1). However, we have the following equation of the total mean square error of r-k estimators $\hat{\boldsymbol{\beta}}(r,k)$ and $\hat{\boldsymbol{\beta}}^*(r,k)$ in the above two models (1.1) and (1.3), respectively:

$$\begin{aligned}(2.5)\quad \mathrm{TMSE}(\hat{\boldsymbol{\beta}}^*(r,k)) &= \mathrm{TMSE}(\hat{\boldsymbol{\beta}}(r,k)) \\ &\quad + \boldsymbol{\delta}'\boldsymbol{T}_r\boldsymbol{\Lambda}_r(k)^{-2}\boldsymbol{T}_r'\boldsymbol{\delta} + 2\boldsymbol{\delta}'\boldsymbol{T}_r\boldsymbol{\Lambda}_r(k)^{-1}\boldsymbol{T}_r'(\boldsymbol{T}_r\boldsymbol{\Lambda}_r(k)^{-1}\boldsymbol{\Lambda}_r\boldsymbol{T}_r' - \boldsymbol{I}_p)\boldsymbol{\beta},\end{aligned}$$

where $\boldsymbol{\delta} = \boldsymbol{X}'\boldsymbol{C}\boldsymbol{\gamma}$ and

$$(2.6)\quad \mathrm{TMSE}(\hat{\boldsymbol{\beta}}(r,k)) = \sigma^2 \operatorname{tr} \boldsymbol{\Lambda}_r(k)^{-1}\boldsymbol{\Lambda}_r\boldsymbol{\Lambda}_r(k)^{-1} + \|(\boldsymbol{T}_r\boldsymbol{\Lambda}_r(k)^{-1}\boldsymbol{\Lambda}_r\boldsymbol{T}_r' - \boldsymbol{I}_p)\boldsymbol{\beta}\|^2.$$

Unfortunately, Sarkar lost the last term in the above equation (2.5), although this cross term can not be ignored.

In the next section, we investigate the effects of this cross term for TMSE comparisons in the concomitant regression model and show additional conditions for concomitant variables under which the superiority of r-k class estimator is confirmed to be kept.

3. Main results

First, we rearrange the results and their attended conditions of comparison among the TMSE's of the r-k class estimators in the usual regression model. We have the following four conditions:

$$\begin{aligned}&(C1)\quad 0 < k \le \frac{2\sigma^2}{\|\boldsymbol{\beta}\|^2},\\ &(C2)\quad \frac{\sigma^2}{\lambda_i} - \alpha_i^2 > 0, \qquad \text{for} \quad i \in N_{p\backslash r},\\ &(C3)\quad 0 < k \le \frac{2\sigma^2}{\|\boldsymbol{\alpha}_r\|^2},\\ &(C4)\quad 0 \le k < \min_{i\in N_{p\backslash r}} \left\{ \frac{\lambda_i}{2\alpha_i^2}\left(\frac{\sigma^2}{\lambda_i} - \alpha_i^2\right)\right\},\end{aligned}$$

where for $r = 1, \dots, p$, we put sets of integers:

$$N_r = \{1, \dots, r\}, \qquad N_{p\backslash r} = \{r+1, \dots, p\}, \qquad N_{p\backslash p} \equiv \emptyset,$$

and vectors:

$$\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_p]' = \begin{bmatrix} \boldsymbol{\alpha}_r \\ \boldsymbol{\alpha}_{p\backslash r} \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{T}_r'\boldsymbol{\beta} \\ \boldsymbol{T}_{p\backslash r}'\boldsymbol{\beta} \end{bmatrix} = \boldsymbol{T}'\boldsymbol{\beta},$$

$$\boldsymbol{\eta} = [\eta_1, \dots, \eta_r, \eta_{r+1}, \dots, \eta_p]' = \begin{bmatrix} \boldsymbol{\eta}_r \\ \boldsymbol{\eta}_{p\backslash r} \end{bmatrix} \equiv \begin{bmatrix} \boldsymbol{T}_r'\boldsymbol{\delta} \\ \boldsymbol{T}_{p\backslash r}'\boldsymbol{\delta} \end{bmatrix} = \boldsymbol{T}'\boldsymbol{\delta}.$$

In the sense of TMSE, Hoerl and Kennard (1970) and Marquardt (1970) showed the superiority of the ORR estimator under Condition ($C1$) and that of the PCR estimator under ($C2$) as compared to the OLS estimator, respectively. Baye and Parker (1984) and Nomura and Ohkubo (1985) also obtained the superiority of the r-k class estimator as compared to the PCR estimator under ($C3$) and the ORR estimator under ($C4$), respectively. Note that ($C1$) is included in ($C3$) and ($C2$) is done in ($C4$), as a special case, respectively.

Now, we propose the following conditions which guarantee Sarkar's assertion in the misspecified regression model :

$$\begin{aligned}
&(C1^*) : (C1) \quad \text{and} \quad \alpha_i\eta_i \geq 0, \quad \text{for any} \quad i \in N_p,\\
&(C2^*) \equiv (C2),\\
&(C3^*) : (C3) \quad \text{and} \quad \alpha_i\eta_i \geq 0, \quad \text{for any} \quad i \in N_r,\\
&(C4^*) : 0 \leq k < \min_{i\in N_p\setminus r}\left\{\frac{\lambda_i}{|\alpha_i|}\left(\sqrt{\frac{\sigma^2}{\lambda_i}} - |\alpha_i|\right)\right\}.
\end{aligned}$$

Since $(C2^*) \equiv (C2)$ is included in $(C4^*)$ and $(C1^*)$ is done in $(C3^*)$ as a special case, respectively, it is sufficient to discuss cases for Conditions $(C3^*)$ and $(C4^*)$.

First, we consider the comparison between the TMSE of the r-k class estimator $\hat{\boldsymbol{\beta}}^*(r,k)$ and the PCR estimator $\hat{\boldsymbol{\beta}}^*(r,\cdot)$.

THEOREM 3.1. *Under Condition* $(C3^*)$ *for a fixed* $r \in N_p$, *the total mean square error of the r-k class estimator is less than that of the PCR estimator*:

$$TMSE(\hat{\boldsymbol{\beta}}^*(r,k)) < TMSE(\hat{\boldsymbol{\beta}}^*(r,\cdot)).$$

PROOF. It follows from the relation (1.2) and the equation (2.5) that

$$\begin{aligned}
&\mathrm{TMSE}(\hat{\boldsymbol{\beta}}^*(r,k)) - \mathrm{TMSE}(\hat{\boldsymbol{\beta}}^*(r,\cdot))\\
&\quad = \{\mathrm{TMSE}(\hat{\boldsymbol{\beta}}(r,k)) - \mathrm{TMSE}(\hat{\boldsymbol{\beta}}(r,\cdot))\} + \boldsymbol{\delta}'\boldsymbol{T}_r(\boldsymbol{\Lambda}_r(k)^{-2} - \boldsymbol{\Lambda}_r^{-2})\boldsymbol{T}_r'\boldsymbol{\delta}\\
&\qquad + 2\boldsymbol{\delta}'\boldsymbol{T}_r\boldsymbol{\Lambda}_r(k)^{-1}\boldsymbol{T}_r'(\boldsymbol{T}_r\boldsymbol{\Lambda}_r(k)^{-1}\boldsymbol{\Lambda}_r\boldsymbol{T}_r' - \boldsymbol{I}_p)\boldsymbol{\beta}\\
&\quad \equiv R_1 + R_2 + R_3 \qquad \text{(say)},
\end{aligned}$$

where R_3 was omitted in Sarkar (1989). Let us check the sign of R_1, R_2 and R_3. We have, already under ($C3$),

$$R_1 = \mathrm{TMSE}(\hat{\boldsymbol{\beta}}(r,k)) - \mathrm{TMSE}(\hat{\boldsymbol{\beta}}(r,\cdot)) < 0,$$

and, for $k > 0$,

$$R_2 = -\sum_{i\in N_r}\frac{(2\lambda_i + k)k\eta_i^2}{\lambda_i^2(\lambda_i + k)^2} < 0.$$

Furthermore, we see

$$R_3 = -2 \sum_{i \in N_r} \frac{k}{(\lambda_i + k)^2} \alpha_i \eta_i \leq 0,$$

under additional condition :

$$\alpha_i \eta_i \geq 0, \qquad \text{for} \quad i \in N_r. \qquad \square$$

Since $(C3^*)$ for $r = p$ means $(C4^*)$, we have the following corollary by Theorem 3.1 together with the relation (1.2).

COROLLARY 3.1. *Under Condition* $(C1^*)$, *the total mean square error of the ORR estimator is less than that of the OLS estimator in concomitant regression model:*

$$TMSE(\hat{\boldsymbol{\beta}}^*(\cdot, k)) < TMSE(\hat{\boldsymbol{\beta}}^*).$$

Next, we consider the comparison between the r-k class estimator $\hat{\boldsymbol{\beta}}^*(r, k)$ and the ORR estimator $\hat{\boldsymbol{\beta}}^*(\cdot, k)$ in the concomitant regression model (1.3).

THEOREM 3.2. *Under Condition* $(C4^*)$ *for a fixed* $r \in N_{p-1}$, *the total mean square error of the* r-k *class estimator is less than that of the ORR estimator:*

$$TMSE(\hat{\boldsymbol{\beta}}^*(r, k)) < TMSE(\hat{\boldsymbol{\beta}}^*(\cdot, k)).$$

PROOF. Note that we can rewrite the TMSE of $\hat{\boldsymbol{\beta}}^*(r, k)$ as follows:

$$\text{(3.1)} \qquad \begin{aligned} \mathrm{TMSE}(\hat{\boldsymbol{\beta}}^*(r, k)) = {} & \sigma^2 \operatorname{tr} \boldsymbol{\Lambda}_r(k)^{-1} \boldsymbol{\Lambda}_r \boldsymbol{\Lambda}_r(k)^{-1} + \boldsymbol{\beta}'(\boldsymbol{I}_p - \boldsymbol{T}_r \boldsymbol{T}_r')\boldsymbol{\beta} \\ & + (\boldsymbol{\delta} - k\boldsymbol{\beta})' \boldsymbol{T}_r \boldsymbol{\Lambda}_r(k)^{-2} \boldsymbol{T}_r'(\boldsymbol{\delta} - k\boldsymbol{\beta}). \end{aligned}$$

This equation and the relation (1.2) leads to

$$\begin{aligned} & \mathrm{TMSE}(\hat{\boldsymbol{\beta}}^*(r, k)) - \mathrm{TMSE}(\hat{\boldsymbol{\beta}}^*(\cdot, k)) \\ & \quad = \sigma^2 \operatorname{tr} \boldsymbol{\Lambda}_r(k)^{-1} \boldsymbol{\Lambda}_r \boldsymbol{\Lambda}_r(k)^{-1} + \boldsymbol{\beta}'(\boldsymbol{I}_p - \boldsymbol{T}_r \boldsymbol{T}_r')\boldsymbol{\beta} \\ & \qquad + (\boldsymbol{\delta} - k\boldsymbol{\beta})' \boldsymbol{T}_r \boldsymbol{\Lambda}_r(k)^{-2} \boldsymbol{T}_r'(\boldsymbol{\delta} - k\boldsymbol{\beta}) \\ & \qquad - \{\sigma^2 \operatorname{tr} \boldsymbol{\Lambda}(k)^{-1} \boldsymbol{\Lambda} \boldsymbol{\Lambda}(k)^{-1} + (\boldsymbol{\delta} - k\boldsymbol{\beta})' \boldsymbol{T} \boldsymbol{\Lambda}(k)^{-2} \boldsymbol{T}'(\boldsymbol{\delta} - k\boldsymbol{\beta})\} \\ & \quad = (\boldsymbol{\alpha}_{p\setminus r}' \boldsymbol{\alpha}_{p\setminus r} - \sigma^2 \operatorname{tr} \boldsymbol{\Lambda}_{p\setminus r}(k)^{-1} \boldsymbol{\Lambda}_{p\setminus r} \boldsymbol{\Lambda}_{p\setminus r}(k)^{-1}) \\ & \qquad - (\boldsymbol{\delta} - k\boldsymbol{\beta})' \boldsymbol{T}_{p\setminus r} \boldsymbol{\Lambda}_{p\setminus r}(k)^{-2} \boldsymbol{T}_{p\setminus r}'(\boldsymbol{\delta} - k\boldsymbol{\beta}) \\ & \quad \equiv \tilde{R}_1 + \tilde{R}_2 \qquad \text{(say)}. \end{aligned}$$

Let us check the signs of $\tilde{R}_1$ and $\tilde{R}_2$. Since the matrix $\boldsymbol{\Lambda}_r(k)^{-2}$ is positive definite for $k \geq 0$, we see $\tilde{R}_2 \leq 0$, with equality only for $\boldsymbol{T}_{p\setminus r}'(\boldsymbol{\delta} - k\boldsymbol{\beta}) = 0$.

Furthermore, we have

$$\tilde{R}_1 = \sum_{i \in N_{p \setminus r}} \left\{ \alpha_i^2 - \frac{\sigma^2 \lambda_i}{(\lambda_i + k)^2} \right\} < 0,$$

under Condition $(C4^*)$. □

We have the following corollary by Theorem 3.2 with $k = 0$:

COROLLARY 3.2. *Under Condition* $(C2^*)$ *for a fixed* $r \in N_{p-1}$, *the total mean square error of the PCR estimator is less than that of the OLS estimator*:

$$TMSE(\hat{\boldsymbol{\beta}}^*(r, \cdot)) < TMSE(\hat{\boldsymbol{\beta}}^*).$$

4. Discussion

Our results is not really a comment for the Sarkar's assertion. Because the conditions $(C1^*)$, $(C3^*)$ can not be used under the misspecified model proposed by Sarkar (1989). If the conditions could be recomposed without depending on $\boldsymbol{C}$ in (1.3) like the conditions $(C2^*)$, $(C4^*)$, we would use his model from (in) which $\boldsymbol{C}$ is eliminated (unknown). Unfortunately, such conditions couldn't be made up. Rao (1973) discussed mathematically same model from the point of view that (known) concomitant vectors $\boldsymbol{C}$ are available to estimation of regression coefficients $\boldsymbol{\beta}$. Since the conditions was well-reformed under the model, we introduced the framework. Hence, our results do not hold without the idea which $\boldsymbol{C}$ is the known matrix of concomitant variables.

Acknowledgements

The authors wish to thank the referee for his helpful comments.

REFERENCES

Baye, M. R. and Parker, D. F. (1984). Combining ridge and principal component regression: A money demand illustration, *Comm. Statist. Theory Methods*, **13**, 197–205.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55–67.

Kendall, M. G. (1957). *A Course in Multivariate Analysis*, Griffin, London.

Lawless, J. F. and Wang, P. (1976). A simulation study of ridge and other regression estimators, *Comm. Statist. Theory Methods*, **5**, 307–323.

Marquardt, D. W. (1970). Generalized inverses, ridge regression biased linear estimation, and nonlinear estimation, *Technometrics*, **12**, 591–612.

Martinez, J. R. (1990). An optimal property of the "(r,k) class estimator", *Comm. Statist. Theory Methods*, **19**, 1281–1289.

Nomura, M. and Ohkubo, T. (1985). A note on combining ridge and principal component regression, *Comm. Statist. Theory Methods*, **14**, 2489–2493.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York.

Sarkar, N. (1989). Comparisons among some estimators in misspecified models with multicollinearity, *Ann. Inst. Statist. Math.*, **41**, 717–724.