HIROSHIMA MATH. J.
16 (1986), 269–277

Selection of variables in a multivariate inverse regression problem

Yasunori FUJIKOSHI and Ryuei NISHII

(Received May 20, 1985)

1. Introduction

We consider a multivariate inverse regression problem with an aim of estimating an unknown $\boldsymbol{x}$ vector from an observed $\boldsymbol{y}$ vector, where $\boldsymbol{y}=(y_1,\dots, y_p)'$ and $\boldsymbol{x}=(x_1,\dots, x_q)'$ are the vectors of p response variables and q explanatory variables, respectively. It is assumed that

$$\text{(1.1)} \qquad \boldsymbol{y} = \boldsymbol{\alpha} + \beta'\boldsymbol{x} + \boldsymbol{e},$$

where $\boldsymbol{\alpha}$: $p\times 1$ is the vector of unknown parameters, β: $q\times p$ is the matrix of unknown parameters and $\boldsymbol{e}$ is an error vector having a p-variate normal distribution with mean zero and unknown covariance matrix Σ. Suppose that the $\boldsymbol{y}$ value in (1.1) has been observed, but the corresponding $\boldsymbol{x}$ value is unknown. Further, suppose that the N independent observations $\boldsymbol{y}_i$, $\boldsymbol{x}_i$ $(i=1,\dots, N)$ with the relationship (1.1) have been given. The data thus consist of the array:

$$\text{(1.2)} \qquad \begin{bmatrix} \boldsymbol{y} \\ \cdot \end{bmatrix}, \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{x}_1 \end{bmatrix}, \begin{bmatrix} \boldsymbol{y}_2 \\ \boldsymbol{x}_2 \end{bmatrix},\dots,\begin{bmatrix} \boldsymbol{y}_N \\ \boldsymbol{x}_N \end{bmatrix},$$

where the dot represents the unknown $\boldsymbol{x}$ value to be estimated. The observations $Y=[\boldsymbol{y}_1,\dots, \boldsymbol{y}_N]'$ and $X=[\boldsymbol{x}_1,\dots, \boldsymbol{x}_N]'$ satisfy

$$Y = \boldsymbol{j}_N\boldsymbol{\alpha}' + X\beta + E,$$

where $\boldsymbol{j}_N=(1,\dots, 1)'$: $N\times 1$, $E=[\boldsymbol{e}_1,\dots, \boldsymbol{e}_N]'$, and $\boldsymbol{e}_i$ $(i=1,\dots, N)$ are independently and identically distributed as $N_p[\boldsymbol{0}, \Sigma]$. Brown [2] has discussed various problems in a general formulation as well as in this formulation.

In this paper we consider the problem of selecting the "best" subset of response variables in the situation where we want to estimate $\boldsymbol{x}$ corresponding to the $\boldsymbol{y}$ value in (1.1). Various methods for selection of variables have been proposed, especially in the area of regression analysis and discriminant analysis (for a summary of the methods, see, e.g., Thompson [12], [13], McKay and Campbell [7], [8]). In the case of multivariate inverse regression Brown [2] has proposed a procedure, based on a test of the redundancy of a subset of response variables. This paper presents two methods for selection of the *best* subset.

One is based on an estimate of the asymptotic mean squared error of the classical estimate. For a realization of this method, we need to obtain asymptotic bias and mean squared error of the classical estimate, which are given in Section 2. The other method is obtained by applying Akaike's information criterion. Numerical performances of the two methods are examined by applying them to the wheat quality data analyzed by Brown [2].

2. Asymptotic bias and mean squared error of the classical estimate

If $\boldsymbol{\alpha}$, β and Σ are known, a natural estimate $\boldsymbol{x}$ would be defined by minimizing

$$(\boldsymbol{y} - \boldsymbol{\alpha} - \beta'\boldsymbol{x})'\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\alpha} - \beta'\boldsymbol{x})$$

with respect to $\boldsymbol{x}$. The estimate is uniquely defined as

$$\hat{\boldsymbol{x}}_0 = (\beta\Sigma^{-1}\beta')^{-1}\beta\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\alpha}) \tag{2.1}$$

under the assumption of rank $(\beta)=q\leq p$. This assumption is put on throughout this paper. So, we treat the case of $p\geq q$. For the case when $\boldsymbol{\alpha}$, β and Σ are unknown, the estimates of these parameters based on (1.2) are used to construct estimates of $\boldsymbol{x}$. The usual estimates of $\boldsymbol{\alpha}$, β and Σ are defined as follows:

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{a} = \bar{\boldsymbol{y}} - B'\bar{\boldsymbol{x}}, \quad \hat{\beta} = B = S_{xx}^{-1}S_{xy},$$

$$\hat{\Sigma} = S_e = \frac{1}{n}\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{a} - B'\boldsymbol{x}_i)(\boldsymbol{y}_i - \boldsymbol{a} - B'\boldsymbol{x}_i)'$$

where $n=N-q-1$, $\bar{\boldsymbol{x}}=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i$, $\bar{\boldsymbol{y}}=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{y}_i$ and

$$S = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{N}\begin{pmatrix} \boldsymbol{x}_i - \bar{\boldsymbol{x}} \\ \boldsymbol{y}_i - \bar{\boldsymbol{y}} \end{pmatrix}\begin{pmatrix} \boldsymbol{x}_i - \bar{\boldsymbol{x}} \\ \boldsymbol{y}_i - \bar{\boldsymbol{y}} \end{pmatrix}'.$$

Here we assume the usual restrictions of rank $(X)=q$ and $n\geq p$. Then an estimate of $\boldsymbol{x}$ based on the natural estimate (2.1) is

$$\hat{\boldsymbol{x}} = (BS_e^{-1}B')^{-1}BS_e^{-1}(\boldsymbol{y}-\boldsymbol{a}) \tag{2.2}$$

which is called the classical estimate. The estimate in the case of $p=q=1$ has been studied in literature (Ott and Myers [9], Williams [14], Shukla [11], Lwin and Maritz [6], etc.). It may be noted that the mean and mean squared error of the estimate are infinite, but the asymptotic expressions of the quantities based on an asymptotic expansion of the distribution of the estimate are finite and could be used as measures of performance for the estimate. The asymptotic results can be obtained by expanding the estimate by Taylor's series and taking expectations term by term, which is known as delta method. We shall denote the expectations of a function $g(\hat{x})$ of $\hat{x}$ in this sense and in ordinary sense by $\mathscr{E}[g(\hat{x})]$ and $\mathrm{E}[g(\hat{x})]$, respectively.

The conditional distribution of $\hat{x}$ given B is closely related to the distribution of an estimate in a growth curve model. Following Gleser and Olkin [5] we use a nonsingular matrix T of order q and an orthogonal matrix Γ of order p such that

$$\bar{B} = B\Sigma^{-1/2} = T[I_q, 0]\Gamma' = T\Gamma_1',$$

where $\Gamma = [\Gamma_1, \Gamma_2]$ and $\Gamma_1 : p \times q$. Let

$$W = n\Gamma' \Sigma^{-1/2} S_e \Sigma^{-1/2} \Gamma = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}, \quad W_{11} : q \times q,$$

$$z = \Gamma' \{\Sigma^{-1/2}(y - \bar{y}) - \bar{B}'(x - \bar{x})\} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \quad z_1 : q \times 1.$$

Then z and W are independently distributed as a normal distribution $N_p[\boldsymbol{\zeta}, (1+N^{-1})I_p]$ and a Wishart distribution $W_p(I_p, n)$, respectively, where $\bar{\beta} = \beta \Sigma^{-1/2}$,

$$\boldsymbol{\zeta} = \Gamma'(\bar{\beta} - \bar{B})'(x - \bar{x}) = (\boldsymbol{\zeta}_1', \boldsymbol{\zeta}_2')', \quad \boldsymbol{\zeta}_1 : q \times 1.$$

Further, we have

$$(2.3) \qquad (BS_e^{-1}B')^{-1} = n^{-1} T'^{-1} (W_{11} - W_{12} W_{22}^{-1} W_{21}) T^{-1}.$$

LEMMA 1. *Let* $\boldsymbol{b} = z_1 - W_{12} W_{122}^{-1} z_2$ *and* $\ell = (1+N^{-1})^{-1} z_2' W_{22}^{-1} z_2$. *Then*

$$(2.4) \qquad \hat{x} = x + T'^{-1}\boldsymbol{b},$$

and the conditional distribution of $\boldsymbol{b}$ *given* ℓ *and* B *is* $N_q[\boldsymbol{\zeta}_1, (1+N^{-1})(1+\ell)I_q]$.

PROOF. This results is obtained by the same lines as in Gleser and Olkin [5] or Fujikoshi and Nishii [4].

LEMMA 2. *It holds that*

$$(2.5) \qquad \mathrm{E}[(\hat{x} - x) \mid \bar{B}] = g(\bar{B}),$$

$$(2.6) \qquad \mathrm{E}[(\hat{x} - x)' S_{xx}^{-1} (\hat{x} - x) \mid \bar{B}] = h(\bar{B}),$$

where $g(\bar{B}) = (\bar{B}\bar{B}')^{-1}\bar{B}d$, $d = (\bar{\beta} - \bar{B})'(x - \bar{x})$,

$$h(\bar{B}) = (1+N^{-1})[1 + \{n - (p-q) - 1\}^{-1}\{p - q + (1+N^{-1})^{-1} d'(I_p - \bar{B}'(\bar{B}\bar{B}')^{-1}\bar{B})d\}] \cdot \mathrm{tr}\, S_{xx}^{-1}(\bar{B}\bar{B}')^{-1} + \mathrm{tr}\, S_{xx}^{-1}(\bar{B}\bar{B}')^{-1}\bar{B}dd'\bar{B}'(\bar{B}\bar{B}')^{-1}.$$

PROOF. From Lemma 1 it is easily seen that

$$\begin{aligned}
&\mathrm{E}\,[(\hat{\boldsymbol{x}}-\boldsymbol{x})|\bar{B}] = T'^{-1}\boldsymbol{\zeta}_1, \\
&\mathrm{E}\,[(\hat{\boldsymbol{x}}-\boldsymbol{x})'S_{xx}^{-1}(\hat{\boldsymbol{x}}-\boldsymbol{x})|\bar{B}] = \mathrm{tr}\, S_{xx}^{-1}\, T'^{-1}\{(1+N^{-1}) \\
&\qquad \cdot (1+\mathrm{E}\,[\ell|\bar{B}])I_q + \boldsymbol{\zeta}_1\boldsymbol{\zeta}_1'\}\, T^{-1}, \\
&\mathrm{E}\,[\ell|\bar{B}] = \{n-(p-q)-1\}^{-1}\{p-q + (1+N^{-1})^{-1})\boldsymbol{\zeta}_2'\boldsymbol{\zeta}_2\}.
\end{aligned}$$

These imply Lemma 2.

The asymptotic bias and standardized mean squared error of $\hat{\boldsymbol{x}}$ are given in the following.

THEOREM 1. *Assume that* $\bar{\boldsymbol{x}}=O(1)$ *and* $S_{xx}=O(1)$. *Then it holds that*

$$\begin{aligned}
\mathscr{E}[(\hat{\boldsymbol{x}}-\boldsymbol{x})] &= -\frac{1}{n}(p-q-1)\Theta S_{xx}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}}) + o(n^{-1}), \\
r &= \mathscr{E}[(\hat{\boldsymbol{x}}-\boldsymbol{x})'S_{xx}^{-1}(\hat{\boldsymbol{x}}-\boldsymbol{x})] \\
&= \left(1+\frac{1}{N}\right)\cdot\frac{n-1}{n-p+q-1}\Big[\mathrm{tr}\, S_{xx}^{-1}\Theta + \frac{1}{n}\{(\mathrm{tr}\, S_{xx}^{-1}\Theta)^2 \\
&\quad - (p-q-2)\,\mathrm{tr}\,(S_{xx}^{-1}\Theta)^2\}\Big] \\
&\quad + \frac{1}{n}(\boldsymbol{x}-\bar{\boldsymbol{x}})'S_{xx}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}})\,\mathrm{tr}\, S_{xx}^{-1}\Theta + o(n^{-1}),
\end{aligned}$$

where $\Theta=(\bar{\beta}\bar{\beta}')^{-1}$ *and* $\bar{\beta}=\beta\,\Sigma^{-1/2}$.

PROOF. By considering the conditional distribution of $\hat{\boldsymbol{x}}$ given B we can write

$$\mathscr{E}[(\hat{\boldsymbol{x}}-\boldsymbol{x})] = \mathscr{E}[\boldsymbol{g}(\bar{B})],\; r = \mathscr{E}[h(\bar{B})],$$

where $\bar{B}=B\,\Sigma^{-1/2}$. The explicit formulas for $\boldsymbol{g}(\bar{B})$ and $h(\bar{B})$ are given in Lemma 2. The expectations of $\boldsymbol{g}(\bar{B})$ and $h(\bar{B})$ with respect to $\bar{B}$ can be carried out by delta method as follows. Let

$$\bar{B} = \bar{\beta} + \frac{1}{\sqrt{n}}S_{xx}^{-1/2}U. \tag{2.7}$$

Then the elements of U: $q\times p$ are indepdnently distributed as $N[0, 1]$. Substituting (2.7) to $\boldsymbol{g}(\bar{B})$ and expanding the resultant expressions, we obtain the first formula. Similarly we have

$$\begin{aligned}
r &= (1+N^{-1})\frac{n-1}{n-p+q-1}\mathscr{E}[\mathrm{tr}\, S_{xx}^{-1}(\bar{B}\bar{B}')^{-1}] \\
&\quad + \frac{1}{n}(\boldsymbol{x}-\bar{\boldsymbol{x}})'S_{xx}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}})\,\mathrm{tr}\, S_{xx}^{-1}(\bar{\beta}\bar{\beta}')^{-1} + o(n^{-1})
\end{aligned} \tag{2.8}$$

which is equal to the right-hand side of the second formula.

We note that the asymptotic bias and mean squared error in the case of $p=q=1$ agree with the results obtained by Ott and Myers [9] and Shukla [11]. An asymptotically unbiased estimate of r is obtained as follows. Using

$$\mathscr{E}[\operatorname{tr} S_{xx}^{-1}(\bar{B}\bar{B}')^{-1}] = \frac{n}{n-p+q}\mathscr{E}[\operatorname{tr} S_{xx}^{-1}(BS_e^{-1}B')^{-1}]$$

which follows from (2.3), we can write

$$r = (1+N^{-1})\left[\frac{n(n-1)}{(n-p+q-1)(n-p+q)} + \frac{1}{n}(\boldsymbol{x}-\bar{\boldsymbol{x}})'S_{xx}^{-1}(\boldsymbol{x}-\bar{\boldsymbol{x}})\right]\operatorname{tr} S_{xx}^{-1}(BS_e^{-1}B')^{-1} + o(n^{-1}) . \tag{2.9}$$

Replacing the $\boldsymbol{x}$ in (2.9) by the estimate $\hat{\boldsymbol{x}}$ in (2.2) we obtain an asymptotically unbiased estimate

$$\hat{r} = (1+N^{-1})\left[\frac{n(n-1)}{(n-p+q-1)(n-p+q)} + \frac{1}{n}(\hat{\boldsymbol{x}}-\bar{\boldsymbol{x}})'S_{xx}^{-1}(\hat{\boldsymbol{x}}-\bar{\boldsymbol{x}})\right]\operatorname{tr} S_{xx}^{-1}(BS_e^{-1}B')^{-1} \tag{2.10}$$

which satisfies $\mathscr{E}[\hat{r}]=r+o(n^{-1})$.

3. Two methods for selection of variables

We shall derive three criteria for determining the *best* subset of response variables $y_1,\dots, y_p$. Let $j=\{j_1, j_2,\dots, j_{k(j)}\}$ be a subset of $\{1,\dots, p\}$ and consider the subset of response variables

$$\boldsymbol{y}(j) = (y_{j_1}, y_{j_2},\dots, y_{j_{k(j)}})'$$

specified by j. We consider only all the subsets such that $k(j)\geq q$.

First we derive two criteria, based on estimates of the asymptotic mean squared error of the classical estimate. The classical estimate of $\boldsymbol{x}$ based on $\boldsymbol{y}(j)$ is given by

$$\hat{\boldsymbol{x}}_j = \{B(j)S_e(j)^{-1}B(j)'\}^{-1}B(j)S_e(j)^{-1}\{\boldsymbol{y}(j)-\bar{\boldsymbol{y}}(j)\}, \tag{3.1}$$

where $B(j)$, $S_e(j)$ and $\bar{\boldsymbol{y}}(j)$ denote the submatrices and subvector of B, S_e and $\bar{\boldsymbol{y}}$, respectively. For estimating $\boldsymbol{x}$, it is natural to select the subset $\boldsymbol{y}(j)$ minimizing the standardized mean squared error

$$r_j = \mathscr{E}[(\hat{\boldsymbol{x}}_j-\boldsymbol{x})'S_{xx}^{-1}(\hat{\boldsymbol{x}}_j-\boldsymbol{x})] . \tag{3.2}$$

The quantity r_j can be interpreted as a risk when the response variable $y(j)$ is selected. By the same way as the derivation of (2.9) we can write

$$(3.3) \qquad r_j = (1+N^{-1})\mathscr{E}\left[\left\{c_j + \frac{1}{n}(x-\bar{x})'S_{xx}^{-1}(x-\bar{x})\right\}D_j\right] + o(n^{-1}),$$

where

$$(3.4) \qquad c_j = n(n-1)/[\{n-k(j)+q\}\{n-k(j)+q-1\}],$$

$$(3.5) \qquad D_j = \operatorname{tr} S_{xx}^{-1}\{B(j)S_e(j)^{-1}B(j)'\}^{-1}.$$

The quantity in the square brackets of (3.3) can be regarded as an estimate for r_j. Of course the x appearing in the terms of order n^{-1} is unknown and must be estimated. Replacing the x by the estimate $\hat{x}$ in (2.2), i.e., the classical estimate based on the full set of response variables, we obtain a criterion

$$(3.6) \qquad \hat{r}_j = \left\{c_j + \frac{1}{n}(\hat{x}-x)'S_{xx}^{-1}(\hat{x}-\bar{x})\right\}D_j.$$

Using this criterion, we select the subset $y(j)$ minimizing $\hat{r}_j$ with respect to j such that $k(j) \geq q$. If x lies very near to $\bar{x}$, then we may use

$$(3.7) \qquad \tilde{r}_j = c_j D_j$$

instead of $\hat{r}_j$. We can expand c_j as

$$c_j = 1 + 2(k(j)-q)/n + O(n^{-2}).$$

So, the quantity c_j can be regarded as a correction term when we estimate r_j by D_j. It is easy to see that if $i \supset j$, then $D_i \leq D_j$ and if $k(i) \geq k(j)$, then $c_i \geq c_j$.

Next we derive an alternative criterion based on a natural family of models $H(j)$ relating to selection of variables, and a model selection criterion in a predictive approach. We shall introduce a natural family of $H(j)$ in terms of the natural estimate x_0 in (2.1). For simplicity, we consider the case of $j = \bar{k} = \{1, \ldots, k\}$. The natural estimate $\hat{x}_0$ in (2.1) can be expressed as

$$(3.8) \qquad \hat{x}_0 = \Xi(y-\alpha) = \Xi_1(y_1-\alpha_1) + \Xi_2(y_2-\alpha_2),$$

where $y = (y_1', y_2')'$, $\alpha = (\alpha_1', \alpha_2')'$, $\alpha_1: k \times 1$, $y_1: k \times 1$, $\Xi = (\beta\Sigma^{-1}\beta')^{-1}\beta\Sigma^{-1} = [\Xi_1, \Xi_2]$, $\Xi_1: q \times k$. If $\Xi_2 = 0$, we can say that y_2 has no additional information in estimating x, in the presense of y_1. We define $H(\bar{k})$ by the restriction "$\Xi_2 = 0$". The model $H(j)$ for general subset j is similarly defined. Let β and Σ be partitioned as

$$\beta = [\beta_1, \beta_2], \quad \beta_1: q \times k, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Sigma_{11}: k \times k.$$

Then it is easy to see that

(3.9)
$$\Xi_2 = 0 \Longleftrightarrow \beta_2 = \beta_1 \Sigma_{11}^{-1} \Sigma_{12}.$$

The latter statement is called Rao's additional information hypothesis (Rao [10]) in the multivariate linear model (1.1) and is interpreted as the hypothesis that $\boldsymbol{y}_2$ supplies no additional information about departures from nullity of the hypothesis "$\beta=0$", independently of $\boldsymbol{y}_1$. The equivalence (3.9) shows that the family of the models $H(j)$ based on Rao's additional information hypothesis in the multivariate linear model is also useful for our problem. For the selection of models $H(j)$, we apply Akaike's information criterion (Akaike [1])

(3.10)
$$\mathrm{AIC}(j) = -2\log L(\hat{\Theta}(j)) + 2p(j)$$

where $L(\Theta)$ is the likelihood function of observations $\boldsymbol{y}_1,\dots, \boldsymbol{y}_N$, $\hat{\Theta}(j)$ is the maximum likelihood estimate of $\Theta=\{\boldsymbol{\alpha}, \beta, \Sigma\}$ under $H(j)$, and $p(j)$ is the dimensionality of Θ under $H(j)$. The Akaike's information criterion for Rao's additional information hypothesis in the multivariate linear model has been obtained by Fujikoshi [3]. Therefore, we obtain

(3.11)
$$\begin{aligned} \mathrm{A}_j &= \mathrm{AIC}(j) - \mathrm{AIC}(\{1,\dots, p\}) \\ &= -N\log\left\{\frac{|S_e|\,|S_e(j)+B(j)'S_{xx}B(j)|}{|S_e+B'S_{xx}B|\,|S_e(j)|}\right\} - 2q\,\{p-k(j)\}. \end{aligned}$$

We select the subset of response variables $\boldsymbol{y}(j)$ minimizing A_j. Here the subsets j selected are restricted to the subsets j such that $k(j)\geq q$.

We note that the criteria $\hat{r}_j$ and $\tilde{r}_j$ depend on the value of $\boldsymbol{y}$ in (1.1), but the criterion A_j does not depend on its value. Brown [2] has given a procedure, based on a test of additional information for the hypothesis $\beta\boldsymbol{x}_0=\boldsymbol{0}$, where $\boldsymbol{x}_0$ is a fixed vector.

4. A numerical example

In this section we shall examine the numerical performances of the three criteria $\hat{r}_j$, $\tilde{r}_j$ and A_j, by applying them to the wheat quality data analyzed by Brown [2]. The data consist of 21 samples of the four response variables $\boldsymbol{y}=(y_1,\dots, y_4)'$ and the two explanatory variables $\boldsymbol{x}=(x_1, x_2)'$. Here $y_1,\dots, y_4$ denote the four infrared reflectance measurements, and x_1 and x_2 denote the percentages of water and protein, respectively. To examine the performances of criteria, we assume that the h-th sample is used for prediction purposes, and the other 20 samples are used to estimate the relationship between $\boldsymbol{y}$ and $\boldsymbol{x}$. So, we assume that the $\boldsymbol{y}$ value of the h-th sample is known, but the corresponding $\boldsymbol{x}$ value is unknown. Our problem is to find the *best* subset of $\{y_1,\dots, y_4\}$ in the situation where we want to estimate the $\boldsymbol{x}$ value of the h-th sample. The subsets selected by the three criteria $\hat{r}_j$, $\tilde{r}_j$ and A_j are given in Table 1.

TABLE 1
The subsets selected by the criteria $\hat{r}_j$, $\tilde{r}_j$ and A_j for $h=1,\ldots,21$

	$h=1, 3,\ldots, 8, 10,\ldots, 21$	$h=2$	$h=9$
$\hat{r}_j$	{1, 2, 3, 4}	{1, 2, 4}	{1, 2, 3, 4}
$\tilde{r}_j$	{1, 2, 3, 4}	{1, 2, 4}	{1, 2, 3, 4}
A_j	{1, 2, 3, 4}	{1, 2, 4}	{1, 2, 4}

The criteria $\hat{r}_j$ and $\tilde{r}_j$ choose the same subsets for all h, and the three criteria select the full model {1, 2, 3, 4} except for $h=2$ and 9. In the case of $h=2$, all criteria select the model {1, 2, 4}. However, the difference between $\hat{r}_{\{1,2,4\}}$ and $\hat{r}_{\{1,2,3,4\}}$ is very small, and this is also true for $\tilde{r}_j$ and A_j. In the case of $h=9$, $\hat{r}_j(\tilde{r}_j)$ and A_j select the full model and {1, 2, 4} respectively. But the difference between $\mathrm{A}_{\{1,2,4\}}$ and $\mathrm{A}_{\{1,2,3,4\}}$ is also negligible. To see the behaviours of the three criteria for different subsets, the values of the three criteria for $h=1$ and $h=2$ are given in Table 2.

TABLE 2
The values of $\hat{r}_j$, $\tilde{r}_j$ and A_j for $h=1$ and 2

j	$h=1$			$h=2$		
	$\hat{r}_j$	$\tilde{r}_j$	A_j	$\hat{r}_j$	$\tilde{r}_j$	A_j
{1, 2, 3, 4}	0.0237*	0.0236*	0.00*	0.0248	0.0247	0.00
{1, 2, 3}	0.0278	0.0276	9.55	0.0285	0.0284	8.20
{1, 2, 4}	0.0246	0.0245	0.80	0.0247*	0.0246*	−0.56*
{1, 3, 4}	0.0562	0.0560	28.40	0.0477	0.0474	23.70
{2, 3, 4}	0.0345	0.0344	17.46	0.0324	0.0322	14.27
{1, 2}	51.82	51.56	87.90	38.58	38.36	85.88
{1, 3}	0.3503	0.3485	68.20	0.3622	0.3602	66.30
{1, 4}	0.3338	0.3321	64.30	0.3404	0.3385	63.35
{2, 3}	0.0583	0.0580	29.14	0.0578	0.0575	27.14
{2, 4}	0.0547	0.0544	31.06	0.0538	0.0535	29.78
{3, 4}	0.6825	0.6791	57.49	0.7333	0.7291	55.23

(* denotes the minimum value of each of the criteria)

The behaviours of each of the criteria for other h are similar to the one for $h=1$. It may be noted that the criteria $\hat{r}_j$, $\tilde{r}_j$ and A_j have similar performances.

References

[1] H. Akaike, Information theory and an extension of the maximum likelihood principle, 2nd International Symposium on Information Theory (B. N. Petrov and F. Czáki, eds.), Akadémiai Kiadó, Budapest, 1973, 267–281.

[2] P. J. Brown, Multivariate calibration, J. R. Statist. Soc., B, **44** (1982), 287–321.

[3] Y. Fujikoshi, A criterion for variable selection in multiple discriminant analysis, Hiroshima Math. J., **13** (1983), 203–214.

[4] Y. Fujikoshi and R. Nishii, On the distribution of a statistic in multivariate inverse regressiona analysis, Hiroshima Math. J., **14** (1984), 215–225.

[5] L. J. Gleser and I. Olkin, Estimation for a regression model with unknown covariance matrix, Proc. Sixth Berkeley Symp. Math. Statist. Prob., **1** (1972), 541–568.

[6] T. Lwin and J. S. Maritz, An analysis of the linear-calibration controversy from the perspective of compound estimation, Technometrics, **24** (1982), 235–242.

[7] R. J. McKay and N. A. Campbell, Variable selection techniques in discriminant analysis I. Description, Brit. J. Math. Psych., **35** (1982), 1–29.

[8] R. J. McKay and N. A. Campbell, Variable selection techniques in discriminant analysis II. Allocation, Brit. J. Math. Psych., **35** (1982), 30–41.

[9] R. L. Ott and R. H. Myers, Optimal experimental designs for estimating the independent variable in regression, Technometrics, **10** (1968), 811–823.

[10] C. R. Rao, Linear Statistical Inference and its Applications, New York, Wiley, 1965.

[11] G. K. Shukla, On the problem of calibration, Technometrics, **14** (1969), 547–553.

[12] M. L. Thompson, Selection of variables in multiple regression: Part I. A review and evaluation, Internat Statist. Rev., **46** (1978), 1–19.

[13] M. L. Thompson, Selection of variables in multiple regression: Part II. Chosen procedures, computations and examples, Internat Statist. Rev., **46** (1978), 129–146.

[14] E. J. Williams, A note on regression methods in calibration, Technometrics, **11** (1969), 189–192.

Department of Mathematics,
Faculty of Science,
Hiroshima University
and
Department of Information and Behavioral Sciences,
Faculty of Integrated Arts and Sciences,
Hiroshima University