Ann. Inst. Statist. Math.
Vol. 55, No. 3, 537–553 (2003)

CORRECTED VERSIONS OF CROSS-VALIDATION CRITERIA FOR SELECTING MULTIVARIATE REGRESSION AND GROWTH CURVE MODELS

YASUNORI FUJIKOSHI[1], TAKAFUMI NOGUCHI[1]*, MEGU OHTAKI[2] AND HIROKAZU YANAGIHARA[3]**

[1] *Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1, Kagamiyama, Higashi-Hiroshima 739-8526, Japan*
[2] *Department of Environmetrics and Biometrics, Research Institute for Radiation Biology and Medicine, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan*
[3] *Department of Statistical Methodology, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan*

(Received April 4, 2001; revised October 18, 2002)

Abstract. This paper is concerned with cross-validation (CV) criteria for choice of models, which can be regarded as approximately unbiased estimators for two types of risk functions. One is AIC type of risk or equivalently the expected Kullback-Leibler distance between the distributions of observations under a candidate model and the true model. The other is based on the expected mean squared error of prediction. In this paper we study asymptotic properties of CV criteria for selecting multivariate regression models and growth curve models under the assumption that a candidate model includes the true model. Based on the results, we propose their corrected versions which are more nearly unbiased for their risks. Through numerical experiments, some tendency of the CV criteria will be also pointed.

Key words and phrases: CV criterion, corrected versions, growth curve models, model selection, multivariate regression models, risk.

1. Introduction

This paper is concerned with cross-validation (CV) criterion for choice of model (see, e.g., Stone (1974)) which could be considered as an approximation for a risk. We consider two types of risks for selecting multivariate regression and growth curve models. One is (i) AIC type of risk or equivalently the expected Kullback-Leibler distance between the distributions of observations under a candidate model and the true model. The other is (ii) the expected mean squared error of prediction.

CV criteria for the risks (i) and (ii) have used as alternatives to AIC (Akaike (1973)) and C_p (Mallows (1973)), respectively. It is known (Stone (1977)) that the AIC is asymptotically equivalent to the CV criterion for the risk (i) in the i.i.d. case. Some corrected versions of AIC and C_p have been proposed in multivariate regression models by Sugiura (1978), Berdrick and Tsai (1994), Fujikoshi and Satoh (1997), and in the

*Now at Iki High School, 88 Katabarufure, Gounouracho, Ikigun, Nagasaki 811-5136, Japan.

**Now at Institute of Policy and Planning Sciences, University of Tsukuba, 1-1-1 Tandy, Tsukuba, Ibaraki 305-8573, Japan.

growth curve model by Satoh *et al.* (1997), etc. These corrections are intended to reduce bias in the estimation of risks. The purpose of this paper is to study some refinement on asymptotic behaviors of the CV criteria in the two important multivariate models, i.e., multivariate regression models and growth curve model. More precisely, we derive asymptotic expansions for the bias terms in the CV criteria for overspecified model including the true model. The results reveal some tendency of the CV criteria. Further, using the results we propose corrected versions of the CV criteria, which are more nearly unbiased and which provided better model selections in small samples. Through numerical experiments it is shown that our corrected versions are similar to asymptotic behaviors of the corrected AIC and C_p. In general, CV criteria might be used for more complicated models. The tendency of CV criteria pointed in this paper is expected to be useful for such models.

The present paper is organized in the following way. In Section 2 we state two types of risks and the corresponding CV criteria. In Section 3 we obtain corrections of CV criteria for selecting multivariate regression models. In Section 4 we obtain corrections of CV criteria for selecting growth curve models. Some numerical studies are also given to see how well our corrections work.

2. Risk functions and CV criteria

Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ be independent p-dimensional random variables, and let $Y = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$. Suppose that under a candidate model M, $(-2)\log$-likelihood can be expressed as

$$\begin{aligned} \ell(\Theta) &= \sum_{i=1}^{n}(-2)\log f(\boldsymbol{y}_i; \boldsymbol{\eta}_i, \Sigma) \\ &= \sum_{i=1}^{n} \ell_i(\Theta), \end{aligned}$$

where Θ is the set of unknown parameters under a candidate model M, and $\mathrm{E}[\boldsymbol{y}_i \mid \mathrm{M}] = \boldsymbol{\eta}_i$, $\mathrm{Var}[\boldsymbol{y}_i \mid \mathrm{M}] = \Sigma$. Consider

$$\Delta_A(\Theta) = \mathrm{E}^*[\ell(\Theta)], \tag{2.1}$$

where E^* denotes the expectation with respect to the true distribution of Y. Let $g(Y)$ be the density function of Y under the true model M^*. Then, note that 2 times the Kullback-Leibler distance between the distributions of Y under the true model M^* and a candidate model M can be expressed as

$$\Delta_A(\Theta) + 2\mathrm{E}^*[\log\{g(Y)\}].$$

The second term in the above expression is common for different candidate models, and so we may ignore the second term when we are interesting in comparison with different models. Let $\mathrm{E}^*[\boldsymbol{y}_i] = \boldsymbol{\eta}_i^*$, $\mathrm{Var}^*[\boldsymbol{y}_i] = \Sigma^*$. When M^* is normal, we have

$$\begin{aligned} \Delta_A(\Theta) &= n\log|\Sigma| + n\,\mathrm{tr}(\Sigma^{-1}\Sigma^*) \\ &\quad + \sum_{i=1}^{n}\mathrm{tr}\{\Sigma^{-1}(\boldsymbol{\eta}_i^* - \boldsymbol{\eta}_i)(\boldsymbol{\eta}_i^* - \boldsymbol{\eta}_i)'\} + np\log 2\pi. \end{aligned}$$

A well known AIC type of risk is defined by

$$R_A = \mathrm{E}^*[\Delta_A(\hat{\Theta})], \tag{2.2}$$

where $\hat{\Theta}$ is the maximum likelihood of Θ under a candidate M. Akaike (1973) proposed AIC $= \ell(\hat{\Theta}) + 2m$, where m is the dimension of Θ under a candidate M, as an estimation for (2.2). Here, we are interesting in examining the corresponding CV criterion

$$\mathrm{CV}_A = \sum_{i=1}^{n} \ell_i(\hat{\Theta}_{[-i]}),$$

where $\hat{\Theta}_{[-i]}$ is the maximum likelihood estimation of Θ under M, based on the observation matrix $Y_{(-i)}$ obtained from Y by deleting $\boldsymbol{y}_i$. Similarly we use the notations, $\hat{\boldsymbol{\eta}}_{i[-i]}$, $\hat{\Sigma}_{[-i]}$, etc.

On the other hand, we may consider the risk R_P based on the standardized mean squared error of prediction. Let

$$\begin{aligned}\Delta_P(\Theta) &= \mathrm{E}^*\left[\sum_{i=1}^{n} \mathrm{tr}\{\Sigma^{*-1}(\boldsymbol{y}_i - \boldsymbol{\eta}_i)(\boldsymbol{y}_i - \boldsymbol{\eta}_i)'\}\right] \\ &= \sum_{i=1}^{n} \mathrm{tr}\{\Sigma^{*-1}(\boldsymbol{\eta}_i^* - \boldsymbol{\eta}_i)(\boldsymbol{\eta}_i^* - \boldsymbol{\eta}_i)'\} + np.\end{aligned}$$

Then the expected mean squared error, R_P is defined by

$$R_P = \mathrm{E}^*[\Delta_P(\hat{\Theta})]. \tag{2.3}$$

Mallows' C_p in the usual univariate regression model can be regarded as an estimation for (2.3). Let M_F be the full model of Y, which uses all the explanatory variables in the observed data set. Then we define the corresponding CV-criterion as follows:

$$\mathrm{CV}_P = \sum_{i=1}^{n} \mathrm{tr}\{\tilde{\Sigma}_{F[-i]}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{\eta}}_{i[-i]})(\boldsymbol{y}_i - \hat{\boldsymbol{\eta}}_{i[-i]})'\},$$

where $\tilde{\Sigma}_{F[-i]}$ is the unbiased estimator for Σ^{*-1} under the full model M_F based on the observation matrix $Y_{(-i)}$. Note that we have used $\tilde{\Sigma}_{F[-i]}$ not $\hat{\Sigma}_{F[-i]}$ which is the maximum likelihood estimation of Σ^* under M_F based on the observation matrix $Y_{(-i)}$, because if we use $\hat{\Sigma}_{F[-i]}^{-1}$, CV_P is not asymptotically equivalent to C_p.

One of our interests is to study unbiased properties of CV_A and CV_P as estimators of R_A and R_P, respectively, in two important multivariate models.

3. Multivariate regression models

Suppose that a candidate model M is defined by multivariate normal regression model with k explanatory variables, i.e.,

$$\mathrm{M} : Y \sim \mathrm{N}_{n\times p}(X\mathcal{B}, \Sigma \otimes I_n),$$

where $\mathcal{B}$ is a $k \times p$ unknown parameter matrix and $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ is an $n \times k$ matrix with rank k. Then, the maximum likelihood estimators of $\mathcal{B}$ and Σ are defined by

$$\hat{\mathcal{B}} = (X'X)^{-1}X'Y, \qquad \hat{\Sigma} = \frac{1}{n}Y'(I_n - P_X)Y,$$

where $P_X = X(X'X)^{-1}X'$ denotes the projection matrix to the space spanned by the columns of X. Let $Y_{(-i)}$ and $X_{(-i)}$ be the matrices obtained from Y and X by deleting their i-th rows, and $\hat{\mathcal{B}}_{[-i]}$ and $\hat{\Sigma}_{[-i]}$ be the maximum likelihood estimators of $\mathcal{B}$ and Σ, based on the observation matrices $Y_{(-i)}$ and $X_{(-i)}$, i.e.,

$$\hat{\mathcal{B}}_{[-i]} = (X'_{(-i)}X_{(-i)})^{-1}X'_{(-i)}Y_{(-i)}, \qquad \hat{\Sigma}_{[-i]} = \frac{1}{n-1}Y'_{(-i)}(I_{n-1} - P_{X_{(-i)}})Y_{(-i)}.$$

Letting $\hat{\boldsymbol{y}}_{i[-i]} = \hat{\mathcal{B}}'_{[-i]}\boldsymbol{x}_i$, we can write the CV criteria as

$$\text{(3.1)} \qquad \mathrm{CV}_A = \sum_{i=1}^{n}[\log|\hat{\Sigma}_{[-i]}| + \mathrm{tr}\{\hat{\Sigma}_{[-i]}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})'\}] + np\log 2\pi,$$

$$\text{(3.2)} \qquad \mathrm{CV}_P = \sum_{i=1}^{n}\mathrm{tr}\{\tilde{\Sigma}_{F[-i]}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})'\},$$

respectively. Here, $\tilde{\Sigma}_{F[-i]}$ is defined by

$$\tilde{\Sigma}_{F[-i]} = \frac{1}{n - k_F - p - 2}Y'_{(-i)}(I_{n-1} - P_{X_{(-i)}})Y_{(-i)}.$$

In this case, we assume the full model which uses all the explanatory variables in the observed data set is expressed as

$$\mathrm{M}_F : Y \sim \mathrm{N}_{n\times p}(X_F\mathcal{B}_F, \Sigma_F \otimes I_n),$$

where, without loss of generality, X_F may be an $n \times k_F$ matrix decomposed as $X_F = (X, X_R)$. Note that $\mathrm{E}^*[\tilde{\Sigma}_{F[-i]}^{-1}] = \Sigma^{*-1}$ (see Lemma 3.2) when the full model includes the true model. Using relations $X'_{(-i)}X_{(-i)} = X'X - \boldsymbol{x}_i\boldsymbol{x}_i'$, $X'_{(-i)}Y_{(-i)} = X'Y - \boldsymbol{x}_i\boldsymbol{y}_i'$ and a general formula of inverse matrix (see, e.g. Siotani *et al.* (1985)), we can rewrite $\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]}$ as follows.

$$\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]} = \frac{1}{\{1 - (P_X)_{ii}\}}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i),$$

where $\hat{\boldsymbol{y}}_i = \hat{\mathcal{B}}\boldsymbol{x}_i$ and $(P_X)_{ii}$ is the i-th diagonal element of P_X. Moreover, using the same reductions, we can rewrite $\hat{\Sigma}_{[-i]}$ as

$$\text{(3.3)} \qquad \hat{\Sigma}_{[-i]} = \frac{1}{n-1}\left[n\hat{\Sigma} - \frac{1}{\{1 - (P_X)_{ii}\}}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)'\right].$$

Therefore,

$$\text{(3.4)} \qquad \sum_{i=1}^{n}\mathrm{tr}\{\hat{\Sigma}_{[-i]}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})'\}$$

$$= \sum_{i=1}^{n} \frac{(n-1)(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)'\hat{\Sigma}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)}{\{1-(P_X)_{ii}\}[n\{1-(P_X)_{ii}\} - (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)'\hat{\Sigma}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)]}.$$

$$\sum_{i=1}^{n} \mathrm{tr}\{\tilde{\Sigma}_{F[-i]}^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})'\} \tag{3.5}$$
$$= \sum_{i=1}^{n} \frac{n-k_F-p-2}{(n-k_F-p-1)\{1-(P_X)_{ii}\}^2}$$
$$\times \left[(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)'\tilde{\Sigma}_F^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i) + \frac{\{(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)'\tilde{\Sigma}_F^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{Fi})\}^2}{(n-k_F-p-1)\{1-(P_{X_F})_{ii}\} - (\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{Fi})'\tilde{\Sigma}_F^{-1}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{Fi})}\right].$$

Here, $(P_{X_F})_{ii}$ is the i-th diagonal element of P_{X_F}, and $\hat{\boldsymbol{y}}_{Fi} = \hat{\mathcal{B}}_F'\boldsymbol{x}_{Fi}$ and $\tilde{\Sigma}_F = (Y - \hat{\mathcal{B}}_F X_F)'(Y - \hat{\mathcal{B}}_F X_F)/(n-k_F-p-1)$, where $\hat{\mathcal{B}}_F = (X_F'X_F)^{-1}X_F'Y$ and $X_F = (\boldsymbol{x}_{F1}, \ldots, \boldsymbol{x}_{Fn})'$. Note that we can express (3.1) and (3.2) as the ones with less computation by substituting (3.3), (3.4) and (3.5) to (3.1) and (3.2), since it is not necessary to calculate $\hat{\mathcal{B}}_{[-i]}$, $\hat{\Sigma}_{[-i]}$, $\hat{\mathcal{B}}_F$ and $\tilde{\Sigma}_{F[-i]}$, repeatedly.

Now we examine biasedness properties of these criteria under the assumption that the candidate model M includes the true model M*. From our assumption we can write as

$$\mathrm{M}^* : Y \sim \mathrm{N}_{n\times p}(X\mathcal{B}^*, \Sigma^* \otimes I_n).$$

Under this assumption, the risks (2.2) and (2.3) can be calculated as

$$R_A = \mathrm{E}^*[n \log|\hat{\Sigma}|] + \frac{np(n+k)}{n-p-k-1} + np\log 2\pi,$$
$$R_P = p(n+k).$$

We use the following lemma:

LEMMA 3.1. *Suppose that* Y *is distributed as* $\mathrm{N}_{n\times p}(X\mathcal{B}^*, \Sigma^* \otimes I_n)$. *Then*

(i) $\hat{\mathcal{B}}_{[-i]} \sim \mathrm{N}_{k\times p}(\mathcal{B}^*, \Sigma^* \otimes (X_{(-i)}'X_{(-i)})^{-1})$,

(ii) $(n-1)\hat{\Sigma}_{[-i]} \sim \mathrm{W}_p(n-k-1, \Sigma^*)$,

(iii) $\hat{\mathcal{B}}_{[-i]}$, $\hat{\Sigma}_{[-i]}$ *and* $\boldsymbol{y}_i$ *are mutually independent, and similarly* $\hat{\mathcal{B}}_{[-i]}$, $\tilde{\Sigma}_{F[-i]}$ *and* $\boldsymbol{y}_i$ *are mutually independent.*

(iv) $\mathrm{E}^*[\log|\hat{\Sigma}| - \log|\hat{\Sigma}_{[-i]}|] = \frac{1}{n^2}\{pk + \frac{1}{2}p(p+1)\} + \mathrm{O}(n^{-3})$.

PROOF. The first three results (i)~(iii) follows for a general result in multivariate normal regression model, see, e.g., Anderson (1984). As for the result (iv), it can be obtained in the following way. Let S be distributed as $\mathrm{W}_p(m, I_p)$, and let

$$S = mI_p + \sqrt{m}V.$$

Then we have

$$\mathrm{E}[\log|S|] = \mathrm{E}\left[\frac{1}{\sqrt{m}}\mathrm{tr}(V) - \frac{1}{2}\left(\frac{1}{\sqrt{m}}\right)^2 \mathrm{tr}(V^2)\right.$$

$$+\frac{1}{3}\left(\frac{1}{\sqrt{m}}\right)^3 \mathrm{tr}(V^3) - \frac{1}{4}\left(\frac{1}{\sqrt{m}}\right)^4 \mathrm{tr}(V^4)\Bigg] + \mathrm{O}(m^{-3})$$
$$= -\frac{1}{2m}p(p+1) - \frac{1}{12m^2}p(2p^2+3p-1) + \mathrm{O}(m^{-3}).$$

Using the above results, we obtain (iv).

Moreover, we use the following lemma (see, e.g., Siotani *et al.* (1985)) in this and next sections.

LEMMA 3.2. *Suppose that U is distributed as $W_p(m, \Sigma)$. Then the expectation of U^{-1} can be expressed as follow.*

$$\mathrm{E}(U^{-1}) = \frac{1}{m-p-1}\Sigma^{-1} \qquad (m > p+2).$$

By using Lemmas 3.1 and 3.2, we can obtain the following expressions for two biases.

$$\begin{aligned} B_A &= R_A - \mathrm{E}^*[\mathrm{CV}_A] \\ &= \frac{1}{2n}\{2kp + p(p+1)\} \\ &\quad + \frac{np(n+k)}{n-p-k-1} - \frac{(n-1)p}{n-p-k-2}\sum_{i=1}^{n}\frac{1}{\{1-(P_X)_{ii}\}} + \mathrm{O}(n^{-2}), \\ B_P &= R_P - \mathrm{E}^*[\mathrm{CV}_P] \\ &= p(n+k) - p\sum_{i=1}^{n}\frac{1}{\{1-(P_X)_{ii}\}}, \end{aligned}$$

where $(A)_{ii}$ denote the i-th diagonal element of a matrix A. In general, $0 \le (P_X)_{ii} \le 1$. Here it is assumed that $0 \le (P_X)_{ii} < 1$. Therefore, we propose new corrections of CV_A and CV_P in multivariate linear model defined by

$$\begin{aligned} \mathrm{CCV}_A = \mathrm{CV}_A &+ \frac{1}{2n}\{2kp + p(p+1)\} \\ &+ \frac{np(n+k)}{n-d_1} - \frac{(n-1)p}{n-d_1-1}\sum_{i=1}^{n}\frac{1}{\{1-(P_X)_{ii}\}}, \end{aligned} \tag{3.6}$$

$$\mathrm{CCV}_P = \mathrm{CV}_P + p(n+k) - p\sum_{i=1}^{n}\frac{1}{\{1-(P_X)_{ii}\}}, \tag{3.7}$$

where $d_1 = k+p+1$. From our construction it holds that if a candidate model M is an overspecified model, then

$$\mathrm{E}^*[\mathrm{CCV}_A] = R_A + \mathrm{O}(n^{-2}), \qquad \mathrm{E}^*[\mathrm{CCV}_A] = R_P.$$

Next, we consider the properties of biases of CV_A and CV_P. Noting that $\sum_{i=1}^{n} 1/\{1-(P_X)_{ii}\} \ge \sum_{i=1}^{n}\{1+(P_X)_{ii}\} = n+k$, we have

$$\frac{1}{2n}\{2kp + p(p+1)\} + \frac{np(n+k)}{n-d_1} - \frac{(n-1)p}{n-d_1-1}\sum_{i=1}^{n}\frac{1}{\{1-(P_X)_{ii}\}} \le 0,$$

Table 1. Risks and average biases by eight criteria in 1,000 repetitions; $n = 30$, $p = 6$, the true model $= \{1, 2, 3\}$.

AIC type Criteria					
	Risks	Biases & Frequencies (%)			
Model	R_A	CV_A	CCV_A	AIC	CAIC
{1}	594.41	−8.17	−5.23	14.78	−4.86
		(18.6)	(9.5)	(0.0)	(4.8)
{1, 2}	605.04	−12.49	−7.46	21.03	−7.26
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2,3}*	582.23	−9.97	−1.14	37.38	−1.62
		(81.1)	(89.3)	(82.3)	(95.0)
{1,2,3,4}*	599.72	−13.89	−0.64	50.78	−1.33
		(0.3)	(1.2)	(17.7)	(0.2)
C_p type Criteria					
	Risks	Biases & Frequencies (%)			
Model	R_P	CV_P	CCV_P	C_p	CC_p
{1}	279.90	−11.03	−10.59	−42.71	−36.08
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2}	281.35	−18.94	−17.55	−38.78	−34.35
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2,3}*	198.03	−3.21	0.01	−2.24	−0.03
		(90.5)	(87.6)	(79.8)	(86.9)
{1,2,3,4}*	204.01	−5.03	0.05	0.01	0.01
		(9.5)	(12.4)	(20.2)	(13.1)

*denotes models including the true model.

$$p(n+k) - p\sum_{i=1}^{n} \frac{1}{\{1-(P_X)_{ii}\}} \leq 0.$$

From these results, we can see that CV_A overestimates asymptotically for R_A and CV_P overestimates exactly for R_P under overspecified models.

Tables 1 and 2 give simulation results of risks, average biases and frequencies of model selected in two cases. Let $X = (\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(k)})$. A candidate model $\{j\}$ means the model when we use $\boldsymbol{x}_{(j)}$ as the regressors. Similarly, the candidate model $\{i, j\}$ means when we use $(\boldsymbol{x}_{(i)}, \boldsymbol{x}_{(j)})$. The first case in Table 1 considers only the hierarchical models. In this case, the true model is $\{1, 2, 3\}$. The next case in Table 2 is a normal multivariate model whose true model is $\{1, 2\}$. In both cases $n = 30$ and $p = 6$. In order to compare with the four well known criteria: AIC, CAIC, C_p and CC_p. These criteria are defined by

$$\mathrm{AIC} = n\log|\hat{\Sigma}| + np(\log 2\pi + 1) + 2\left\{pk + \frac{1}{2}p(p+1)\right\},$$

$$\mathrm{CAIC} = n\log|\hat{\Sigma}| + np\log 2\pi + \frac{n(n+k)p}{n-k-p-1},$$

$$C_p = (n - k_F)\,\mathrm{tr}(\hat{\Sigma}_F^{-1}\hat{\Sigma}) + 2kp,$$

Table 2. Risks and average biases by eight criteria in 1,000 repetitions; $n = 30$, $p = 6$, the true model $= \{1, 2\}$.

AIC type Criteria					
	Risks	Biases & Frequencies (%)			
Model	R_A	CV_A	CCV_A	AIC	CAIC
{1}	620.17	−5.74	−2.98	15.20	−4.43
		(0.0)	(0.0)	(0.0)	(0.0)
{2}	588.07	−5.99	−3.14	15.35	−4.29
		(1.7)	(1.1)	(0.0)	(0.1)
{3}	623.84	−5.22	−2.39	15.61	−4.02
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2}*	567.59	−5.81	−1.02	26.97	−1.31
		(97.6)	(97.4)	(86.1)	(99.1)
{1,3}	631.65	−10.18	−5.43	21.82	−6.47
		(0.0)	(0.0)	(0.0)	(0.0)
{2,3}	599.36	−9.69	−4.72	22.36	−5.93
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2,3}*	583.11	−8.26	−0.09	38.20	−0.80
		(0.7)	(1.5)	(13.9)	(0.8)
C_p type Criteria					
	Risks	Biases & Frequencies (%)			
Model	R_P	CV_P	CCV_P	C_p	CC_p
{1}	445.68	−12.53	−12.22	−85.23	−81.03
		(0.0)	(0.0)	(0.0)	(0.0)
{2}	254.51	−5.20	−4.81	−27.97	−23.77
		(0.0)	(0.0)	(0.0)	(0.0)
{3}	479.85	−10.45	−10.09	−97.05	−92.85
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2}*	191.88	−1.50	−0.28	−2.42	−0.32
		(91.4)	(87.1)	(81.7)	(88.1)
{1,3}	445.42	−37.70	−36.51	−80.85	−78.75
		(0.0)	(0.0)	(0.0)	(0.0)
{2,3}	257.17	−11.05	−9.70	−24.37	−22.27
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2,3}*	197.97	−2.66	0.14	−0.03	−0.03
		(8.6)	(12.9)	(18.3)	(11.9)

*denotes models including the true model.

$$CC_p = (n - k_F)\,\mathrm{tr}(\hat{\Sigma}_F^{-1}\hat{\Sigma}) + \frac{\{2(n-k_F)k - (p+1)(k_F+k)\}}{n - k_F - p - 1},$$

where $\hat{\Sigma}_F = Y'(I_n - P_{X_F})Y/n$. From these tables, we can see that CCV_A and CCV_P are better estimators to their risks than CV_A and CV_P when a candidate model includes the true model, respectively. Furthermore, CCV_A improves the biases even if a candidate model does not include the true model. On the other hand, it notes that CV_A and

CV_P become overestimations for their risks. From our simulation results, it can be also understood that our assertion is right. Making an additional remark, we can see that CCV_A and CAIC, CCV_P and CC_p, have similar performances, respectively.

4. Growth curve models

Suppose that a candidate model M is defined by the growth curve model, which was proposed Potthoff and Roy (1964), with q within-individual explanatory variables, i.e.,

$$\mathrm{M} : Y \sim \mathrm{N}_{n\times p}(A\Xi B, \Sigma \otimes I_n),$$

where Ξ is a $k \times q$ unknown parameter matrix and B is a $q \times p$ within-individual design matrix with rank q. Here, $A = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)'$ is an $a \times k$ between-individual matrix with rank k indicating whether each of individuals belongs to the j-th ($j = 1, 2, \dots, k$) population. Namely, if $\boldsymbol{y}_i$ belongs to the j-th population, then the j-th element of $\boldsymbol{a}_i$ is 1 and the others are 0. Therefore, without loss of generality, let n_j denote the sample size of the j-th population, A is given by

$$A = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_k} \end{pmatrix}.$$

Then the maximum likelihood estimators of Ξ and Σ are defined by

$$\hat{\Xi} = (A'A)^{-1}A'YS^{-1}B'(BS^{-1}B')^{-1},$$
$$\hat{\Sigma} = \frac{1}{n}(Y - A\hat{\Xi}B)'(Y - A\hat{\Xi}B),$$

respectively, where $S = Y'(I_n - P_A)Y/(n-k)$. Let $\hat{\Xi}_{[-i]}$ and $\hat{\Sigma}_{[-i]}$ be the maximum likelihood estimators of Ξ and Σ, based on the observation matrices $Y_{(-i)}$ and $A_{(-i)}$, i.e.,

$$\hat{\Xi}_{[-i]} = (A'_{(-i)}A_{(-i)})^{-1}A'_{(-i)}Y_{(-i)}S^{-1}_{[-i]}B'(BS^{-1}_{[-i]}B')^{-1},$$
$$\hat{\Sigma}_{(-i)} = \frac{1}{n-1}(Y_{(-i)} - A_{(-i)}\hat{\Xi}_{[-i]}B)'(Y_{(-i)} - A_{(-i)}\hat{\Xi}_{[-i]}B),$$

respectively, where $S_{[-i]} = Y'_{(-i)}(I_{n-1} - P_{A_{(-i)}})Y_{(-i)}/(n-k-1)$. Let $\hat{\boldsymbol{y}}_{i[-i]} = B'\hat{\Xi}'_{i[-i]}\boldsymbol{a}_i$. In this model, CV-criteria can be written as

$$\mathrm{CV}_A = \sum_{i=1}^{n}[\log|\hat{\Sigma}_{[-i]}| + \mathrm{tr}\{\hat{\Sigma}^{-1}_{[-i]}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})'\}] + np\log 2\pi, \tag{4.1}$$

$$\mathrm{CV}_P = \sum_{i=1}^{n}\mathrm{tr}\{\tilde{S}^{-1}_{[-i]}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_{i[-i]})'\}, \tag{4.2}$$

respectively, where

$$\tilde{S}_{[-i]} = \frac{1}{n-p-k-2}Y'_{(-i)}(I_{n-1} - P_{A_{(-i)}})Y_{(-i)} = \frac{n-k-1}{n-p-k-2}S_{[-i]}.$$

By using the relations $Y'_{(-i)}Y_{(-i)} = Y'Y - \boldsymbol{y}_i\boldsymbol{y}'_i$, $A'_{(-i)}A_{(-i)} = A'A - \boldsymbol{a}_i\boldsymbol{a}'_i$ and $A'_{(-i)}Y_{(-i)} = A'Y - \boldsymbol{a}_i\boldsymbol{y}'_i$ in the former sections, we can rewrite $S_{[-i]}$, $\widehat{\Xi}_{[-i]}$ and $\hat{\Sigma}_{[-i]}$ as in the following forms.

$$\begin{aligned}
S_{[-i]} &= \frac{n-k}{n-k-1}\left[S - \frac{1}{(n-k)\{1-(P_A)_{ii}\}}(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)(\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i)'\right],\\
\widehat{\Xi}_{[-i]} &= (A'A - \boldsymbol{a}_i\boldsymbol{a}'_i)^{-1}(A'Y - \boldsymbol{a}_i\boldsymbol{y}'_i)S^{-1}_{[-i]}B'(BS^{-1}_{[-i]}B')^{-1},\\
\hat{\Sigma}_{[-i]} &= \frac{1}{n-1}\{(Y - A\widehat{\Xi}_{[-i]}B)'(Y - A\widehat{\Xi}_{[-i]}B)\\
&\qquad - (\boldsymbol{y}_i - B'\widehat{\Xi}'_{[-i]}\boldsymbol{a}_i)(\boldsymbol{y}_i - B'\widehat{\Xi}'_{[-i]}\boldsymbol{a}_i)'\}.
\end{aligned}$$

Now we examine biasedness properties of these criteria under the assumption that candidate model M includes the true model M*. Since M includes M*, we can write as

$$\mathrm{M}^* : Y \sim \mathrm{N}_{n\times p}(A\Xi^*B, \Sigma^* \otimes I_n).$$

Then the risk functions (2.2) and (2.3) can be calculated as

$$\begin{aligned}
R_A &= \mathrm{E}^*[n\log|\hat{\Sigma}|] + np\log 2\pi\\
&\quad + \frac{n^2(p-q)}{n-p+q-1} + \frac{nq(n-k-1)(n+k)}{(n-k-p-1)(n-k-p+q-1)},\\
R_P &= np + \frac{kq(n-k-1)}{n-k-p+q-1}.
\end{aligned}$$

For the derivation, see, e.g., Satoh *et al.* (1997).

We use the following lemma (see, e.g., Siotani *et al.* (1985)):

LEMMA 4.1. *Suppose that W is distributed as $W_p(n, \Sigma)$ and W and Σ are decomposed as*

$$W = \begin{array}{c} \\ q \\ p-q \end{array}\!\!\begin{array}{c} \begin{array}{cc} q & p-q \end{array} \\ \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \end{array}, \qquad \Sigma = \begin{array}{c} \\ q \\ p-q \end{array}\!\!\begin{array}{c} \begin{array}{cc} q & p-q \end{array} \\ \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \end{array}.$$

Then it holds that:

(i) $W_{11\cdot 2} = W_{11} - W_{12}W_{22}^{-1}W_{21} \sim \mathrm{W}_q(n-p+q, \Sigma_{11\cdot 2})$, $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$,

(ii) $W_{22} \sim \mathrm{W}_{p-q}(n, \Sigma_{22})$,

(iii) *If* $\Sigma_{12} = O_{q\times(p-q)}$ *then* $W_{12}W_{22}^{-1}W_{21}$, $W_{11\cdot 2}$ *and* W_{22} *are mutually independent and* $W_{12}W_{22}^{-1}W_{21} \sim \mathrm{W}_q(p-q, \Sigma_{11})$,

(iv) *When* W_{22} *is given,* $W_{12}W_{22}^{-1/2} \sim \mathrm{N}_{(p-q)\times q}(\Sigma_{12}\Sigma_{22}^{-1}W_{22}^{-1/2}, \Sigma_{11\cdot 2} \otimes I_{p-q})$.

Using the canonical form as in Gleser and Olkin (1970), results of Satoh *et al.* (1997) and Lemmas 3.1, 3.2 and 4.1, we obtain the following results.

$$\sum_{i=1}^{n}\mathrm{E}^*[\log|\hat{\Sigma}_{[-i]}|] = \mathrm{E}^*[n\log|\hat{\Sigma}|] - \frac{1}{2}\{p(p+1) + 2kq\} + \mathrm{O}(n^{-2}), \tag{4.3}$$

$$\sum_{i=1}^{n}\mathrm{E}^*[\mathrm{tr}(\hat{\Sigma}^{-1}_{[-i]}\Sigma^*)] = \frac{n(n-1)(p-q)}{n-p+q-2} \tag{4.4}$$

$$+\frac{n(n-1)(n-k-2)q}{(n-k-p-2)(n-k-p+q-2)},$$

$$\sum_{i=1}^{n}\mathrm{E}^*[\mathrm{tr}\{\hat{\Sigma}_{[-i]}^{-1}B'(\hat{\Xi}_{[-i]}-\Xi^*)'\boldsymbol{a}_i\boldsymbol{a}_i'(\hat{\Xi}_{[-i]}-\Xi^*)B\}]$$
$$=\frac{(n-1)(n-k-2)q}{(n-k-p-2)(n-k-p+q-2)}\left(k+\sum_{j=1}^{k}\frac{1}{n_j-1}\right), \tag{4.5}$$

$$\sum_{i=1}^{n}\mathrm{E}^*[\mathrm{tr}\{\tilde{S}_{[-i]}^{-1}B'(\hat{\Xi}_{[-i]}-\Xi^*)'\boldsymbol{a}_i\boldsymbol{a}_i'(\hat{\Xi}_{[-i]}-\Xi^*)B\}]$$
$$=\frac{(n-k-2)q}{(n-k-p+q-2)}\left(k+\sum_{j=1}^{k}\frac{1}{n_j-1}\right). \tag{4.6}$$

Outlines of evaluations are shown in the Appendix. Suppose that $n/n_j=\mathrm{O}(1)$ $(1\le j\le k)$. Then the biases B_A and B_P can be evaluated as

$$\begin{aligned}
B_A &= R_A-\mathrm{E}^*[\mathrm{CV}_A]\\
&=\frac{1}{2n}\{p(p+1)+2kq\}-\frac{n(p-q)(p-q+1)}{(n-p+q-1)(n-p+q-2)}\\
&\quad-(n+k)q\left\{\frac{(n-1)(n-k-2)}{(n-k-p-2)(n-k-p+q-2)}\right.\\
&\qquad\left.-\frac{n(n-k-1)}{(n-k-p-1)(n-k-p+q-1)}\right\}\\
&\quad-\frac{(n-1)(n-k-2)q}{(n-k-p-2)(n-k-p+q-2)}\sum_{j=1}^{k}\frac{1}{n_j-1}+\mathrm{O}(n^{-2}),\\
B_P &= R_P-\mathrm{E}^*[\mathrm{CV}_P]\\
&=-\frac{kq(p-q)}{(n-k-p+q-1)(n-k-p+q-2)}-\frac{(n-k-2)q}{(n-k-p+q-2)}\sum_{j=1}^{k}\frac{1}{n_j-1}.
\end{aligned}$$

Therefore, we propose new corrections of CV_A and CV_P in growth curve model defined by

$$\begin{aligned}
\mathrm{CCV}_A &= \mathrm{CV}_A+\frac{1}{2n}\{p(p+1)+2kq\}-\frac{n(p-q)d_2}{(n-d_2)(n-d_2-1)}\\
&\quad-\frac{(n-1)(n-k-2)q}{(n-d_1-1)(n-d_3-1)}\left(n+k+\sum_{j=1}^{k}\frac{1}{n_j-1}\right)\\
&\quad+\frac{n(n+k)(n-k-1)q}{(n-d_1)(n-d_3)},
\end{aligned}\tag{4.7}$$

$$\mathrm{CCV}_P=\mathrm{CV}_P-\frac{q}{(n-d_3-1)}\left\{\frac{k(p-q)}{(n-d_3)}+(n-k-2)\sum_{j=1}^{k}\frac{1}{n_j-1}\right\}, \tag{4.8}$$

where $d_1=k+p+1$, $d_2=p-q+1$ and $d_3=k+p-q+1$. In order to obtain more

Table 3. Risks and average biases by eight criteria in 1,000 repetitions; $n = 30$, $p = 6$, $k = 1$, the true model $= \{1, 2, 3\}$.

AIC type Criteria					
	Risks	Biases & Frequencies (%)			
Model	R_A	CV_A	CCV_A	AIC	CAIC
{1}	594.78	−2.58	−0.55	12.82	−2.14
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2}	597.07	−2.57	−0.36	14.70	−1.94
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2,3}*	549.73	−2.82	−0.46	17.64	−0.24
		(88.4)	(87.7)	(81.5)	(89.8)
{1,2,3,4}*	551.44	−2.99	−0.53	18.46	−0.31
		(11.6)	(12.3)	(18.5)	(10.2)
C_p type Criteria					
	Risks	Biases & Frequencies (%)			
Model	R_P	CV_P	CCV_P	C_p	CC_p
{1}	319.99	11.56	11.61	−29.91	−28.83
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2}	324.29	21.84	21.93	−17.00	−16.50
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2,3}*	183.35	−0.42	−0.29	−0.06	0.06
		(85.1)	(84.8)	(79.9)	(81.9)
{1,2,3,4}*	184.27	−0.50	−0.34	0.07	−0.01
		(14.9)	(15.2)	(20.1)	(18.1)

*denotes models including the true model.

simple expressions for CCV_A and CCV_P, we can omit the n^{-2} terms in the expressions, since they may be considered to be small in comparison with the corrections of the order $\mathrm{O}(n^{-1})$. So,

$$\mathrm{CCV}'_A = \mathrm{CV}_A - \frac{1}{2n}p(p+1) - q\sum_{j=1}^{n}\frac{1}{n_j},$$

$$\mathrm{CCV}'_P = \mathrm{CV}_P - q\sum_{j=1}\frac{1}{n_j}.$$

Next, we consider the properties of biases of CV_A and CV_P. From (4.8), it can see that the bias of CV_P is a negative valued. Noting that $\sum_{i=1}^{n} 1/(n_j-1) \geq k/n$, we can see that the bias of CV_A is a negative valued. Therefore, we can see that CV_A overestimates asymptotically for R_A and CV_P overestimates exactly for R_P under overspecified models.

Tables 3 and 4 give simulation results of risks, average biases and frequencies of model selected in the cases of nested models. The first case is $k = 1$ and the second case is $k = 3$. Let $B = (\boldsymbol{b}_{(1)}, \ldots, \boldsymbol{b}_{(q)})'$. The candidate model $\{1\}$ means the model when we use $\boldsymbol{b}'_{(1)}$ as the regressors. Similarly, the candidate model $\{1, 2\}$ means when we use $(\boldsymbol{b}_{(1)}, \boldsymbol{b}_{(2)})'$. In both cases $n = 30$, $p = 6$, $q = 4$ and the true model $= \{1, 2, 3\}$. In order to compare with other standard criteria, we prepared the four ones : AIC and C_p, and

Table 4. Risks and average biases by eight criteria in 1,000 repetitions; $n = 30$, $p = 6$, $k = 3$, the true model $= \{1, 2, 3\}$.

AIC type Criteria					
	Risks	Biases & Frequencies (%)			
Model	R_A	CV_A	CCV_A	AIC	CAIC
{1}	589.23	−3.25	−0.11	19.24	−1.54
		(3.4)	(2.1)	(0.0)	(0.3)
{1,2}	595.92	−4.20	0.03	25.71	−1.29
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2,3}*	567.69	−5.51	−0.38	31.20	−0.52
		(91.9)	(91.4)	(81.2)	(95.1)
{1,2,3,4}*	574.02	−5.53	0.33	35.29	0.12
		(4.7)	(6.5)	(18.8)	(4.6)
C_p type Criteria					
	Risks	Biases & Frequencies (%)			
Model	R_P	CV_P	CCV_P	C_p	CC_p
{1}	262.65	7.66	8.12	−20.14	−16.57
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2}	262.31	12.64	13.48	−9.84	−8.20
		(0.0)	(0.0)	(0.0)	(0.0)
{1,2,3}*	190.23	−1.32	−0.13	−0.38	0.01
		(88.7)	(87.3)	(78.8)	(82.4)
{1,2,3,4}*	193.25	−1.22	0.28	0.71	0.46
		(11.3)	(12.7)	(21.2)	(17.6)

*denotes models including the true model.

corrected criteria CAIC and CC_p which were proposed by Satoh *et al.* (1997). These criteria are defined by

$$\mathrm{AIC} = n\log|\hat{\Sigma}| + np(\log 2\pi + 1) + 2\left\{kq + \frac{1}{2}p(p+1)\right\},$$

$$\mathrm{CAIC} = n\log|\hat{\Sigma}| + np\log 2\pi + \frac{n^2(p-q)}{n-d_1} + \frac{nq(n-k-1)(n+k)}{(n-d_1)(n-d_3)},$$

$$C_p = n\,\mathrm{tr}(S^{-1}\hat{\Sigma}) + 2kq,$$

$$CC_p = n\,\mathrm{tr}(S^{-1}\hat{\Sigma}) + k(p+q) - \frac{k(p-q)(n-k-q)}{(n-d_3)}.$$

From these tables, we can see that CCV_A and CCV_P are better estimators to their risk functions than CV_A and CV_P when a candidate model includes the true model, respectively. Furthermore, CCV_A improves the biases even if a candidate model does not include the true model. On the other hand, it may be noted that CV_A and CV_P become overestimations for their risks. From our simulation results, it can be also understood that our assertion is right. Making an additional remark, we can see that CCV_A and CAIC, CCV_P and CC_p, have similar performances, respectively.

Through the simulating experiments in Sections 3 and 4, we can see that the corrections are necessary for CV criteria when the sample size n is small. The bias of C_p under overspecified model is small as $\mathrm{tr}(\hat{\Sigma}_F^{-1}\hat{\Sigma}) = p$, but, the bias of CV_P is not so. Under overspecified model, the correction CV_P is more important than one of C_p. Further, CV criteria have a tendency to overestimate their risks, even if it is in case of risk based on the Kullback-Leibler distance. This property is different from the one of AIC which is based on the Kullback-Leibler distance, because AIC has a tendency to underestimate for a risk. Moreover, the corrected versions of CV criteria have the same performances as the ones of other standard adjusted criteria, i.e., CAIC and CC_p. Our conclusions are limited in the multivariate regression and growth curve models. However, we can expect that there are such tendencies for other models. Making an additional remark, we can see that CV_P is asymptotically equivalent to C_p by using $\tilde{\Sigma}_F$ not $\hat{\Sigma}_F$. Therefore, we must be careful when CV_P is used, because it has the possibility that the constant bias term is left.

Acknowledgements

The authors would like to thank the referees for several valuable comments.

Appendix

The aim of this section is to state outlines of calculations of (4.3), (4.4), (4.5) and (4.6).

A.1 *Derivation of* (4.3)

Using the canonical reductions as in Gleser and Olkin (1970), we can write

$$Y = \Gamma_1 \begin{pmatrix} T_1 \Xi T_2 & O_{k\times(p-q)} \\ O_{(n-k)\times q} & O_{(n-k)\times(p-q)} \end{pmatrix} \Gamma_2 + \mathcal{E},$$

under model M, where $T_1 : k \times k$ and $T_2 : q \times q$ are certain non-singular matrices and $\Gamma_1 : n \times n$ and $\Gamma_2 : p \times p$ are certain orthogonal matrices. Let $Z = \Gamma_1' Y \Gamma_2'$, then

$$\mathrm{E}[Z] = \begin{pmatrix} \Theta & O_{k\times(p-q)} \\ O_{(n-k)\times q} & O_{(n-k)\times(p-q)} \end{pmatrix}, \qquad \mathrm{Var}[\mathrm{vec}(Z)] = \Lambda \otimes I_n,$$

where $\Theta = T_1 \Xi T_2$ and $\Lambda = \Gamma_2 \Sigma \Gamma_2'$. Let us decompose Z and Λ as

$$Z = \begin{array}{c} \\ k \\ n-k \end{array} \!\!\begin{array}{c} \begin{array}{cc} q & p-q \end{array} \\ \begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} \end{array}, \qquad \Lambda = \begin{array}{c} \\ q \\ p-q \end{array} \!\!\begin{array}{c} \begin{array}{cc} q & p-q \end{array} \\ \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \end{array},$$

and let

$$V = \begin{pmatrix} Z_{21}' \\ Z_{22}' \end{pmatrix} (Z_{21}\ Z_{22}) = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}.$$

From Gleser and Olkin (1970), the maximum likelihood estimators of Θ and Λ are

$$\hat{\Theta} = Z_{11} - Z_{12} V_{22}^{-1} V_{21},$$
$$n\hat{\Lambda} = V + \begin{pmatrix} V_{12} V_{22}^{-1} \\ I_{p-q} \end{pmatrix} Z_{12}' Z_{12} \left(V_{22}^{-1} V_{21} \quad I_{p-q} \right).$$

Further,

$$|\hat{\Sigma}| = |\Gamma_2 \hat{\Lambda} \Gamma_2'| = \frac{1}{n^p} |V_{11\cdot 2}| |V_{22} + Z_{12}' Z_{12}|,$$

where $V_{11\cdot 2} = V_{11} - V_{12} V_{22}^{-1} V_{21}$. Note that V is distributed as $\mathrm{W}_p(n-k, \Lambda)$. Then, from Lemma 4.1 (i) and (ii),

$$\tilde{V}_{11\cdot 2} \sim \mathrm{W}_q(n-k-p+q, I_q), \qquad \tilde{V}_{22} \sim \mathrm{W}_{p-q}(n, I_{p-q}),$$

where $\tilde{V}_{11\cdot 2} = \Lambda_{11\cdot 2}^{-1/2} V_{11\cdot 2} \Lambda_{11\cdot 2}^{-1/2}$, $\Lambda_{11\cdot 2} = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}$ and $\tilde{V}_{22} = \Lambda^{-1/2}(V_{22} + Z_{12}'Z_{12})\Lambda^{-1/2}$. Further, $|\hat{\Sigma}|$ can be rewritten as

$$|\hat{\Sigma}| = \frac{(n-k-p+q)^q}{n^q} \left| \frac{1}{n-k-p+q} \tilde{V}_{11\cdot 2} \right| \left| \frac{1}{n} \tilde{V}_{22} \right| |\Sigma|.$$

Therefore

$$\begin{aligned} \log|\hat{\Sigma}| &= \log|\Sigma| + \log\left| \frac{1}{n-k-p+q} \tilde{V}_{11\cdot 2} \right| \\ &\quad + \log\left| \frac{1}{n} \tilde{V}_{22} \right| + q \log\left(\frac{n-k-p+q}{n} \right). \end{aligned}$$

From Lemma 3.1 (iv),

$$\begin{aligned} &\mathrm{E}^*\left[\log\left| \frac{1}{n-k-p+q} \tilde{V}_{11\cdot 2} \right| \right] \\ &\quad = -\frac{1}{2n} q(q+1) \\ &\qquad - \frac{1}{12n^2}\{q(2q^2+3q-1) + 6q(q+1)(k+p-q)\} + \mathrm{O}(n^{-3}), \\ &\mathrm{E}^*\left[\log\left| \frac{1}{n} \tilde{V}_{22} \right| \right] \\ &\quad = -\frac{1}{2n}(p-q)(p-q+1) \\ &\qquad - \frac{1}{12n^2}(p-q)\{2(p-q)^2 + 3(p-q) - 1\} + \mathrm{O}(n^{-3}). \end{aligned}$$

These imply that

$$\mathrm{E}^*[\log|\hat{\Sigma}|] = \log|\Sigma^*| - \frac{1}{2n}\{p(p+1) + 2kq\} - \frac{c}{12n^2} + \mathrm{O}(n^{-3}),$$

where

$$\begin{aligned} c &= q(2q^2+3q-1) + 6q(q+1)(k+p-q) \\ &\quad + (p-q)\{2(p-q)^2 + 3(p-q) - 1\} + 12q(k+p-q)^2. \end{aligned}$$

Replacing n by $n-1$ in the above result yields

$$\begin{aligned} \mathrm{E}^*[\log|\hat{\Sigma}_{[-i]}|] &= \log|\Sigma^*| - \frac{1}{2n}\{p(p+1) + 2kq\} \\ &\quad - \frac{1}{12n^2}\{c + 6p(p+1) + 12kq\} + \mathrm{O}(n^{-3}), \end{aligned}$$

and hence

$$\mathrm{E}^*[\log|\hat{\Sigma}_{[-i]}|] = \mathrm{E}^*[\log|\hat{\Sigma}|] - \frac{1}{2n^2}\{p(p+1)+2kq\} + \mathrm{O}(n^{-3}).$$

This implies (4.3).

A.2 *Derivation of* (4.4)

From Satoh *et al.* ((1997), pp. 283), we have

$$\mathrm{E}^*[\mathrm{tr}(\hat{\Sigma}^{-1}\Sigma^*)] = \frac{n(p-q)}{n-p+q-1} + \frac{(n+k)(n-1)q}{(n-k-p-1)(n-k-p+q-1)}.$$

Therefore, replacing n by $n-1$ and summing from 1 to n yield the result (4.4).

A.3 *Derivation of* (4.5)

In the computation of (4.5), we use notations in Satoh *et al.* (1997) as

$$W = (n-k)H'\Sigma^{*-1/2}S\Sigma^{*-1/2}H = \begin{array}{cc} & \begin{array}{cc} q & p-q \end{array} \\ \begin{array}{c} q \\ p-q \end{array} & \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \end{array},$$

$$Z = (A'A)^{-1/2}A'(Y - A\Xi^*B)\Sigma^{*-1/2} = (Z_1\ Z_2),$$

where $H = (H_1\ H_2)$ is a $p \times p$ orthogonal matrix and H_1 is defined by

$$H_1 = \Sigma^{*-1/2}B'(B\Sigma^{*-1}B')^{-1/2}.$$

Then W and Z are independent each other and distributed as

$$W \sim \mathrm{W}_p(n-k, I_p), \qquad Z \sim \mathrm{N}_{k\times p}(O_{k\times p}, I_{kp}).$$

We can express $\hat{\Sigma}^{-1}$ and $(A'A)^{1/2}(\hat{\Xi} - \Xi^*)B\Sigma^{*-1/2}$ as

$$\hat{\Sigma}^{-1} = n\Sigma^{*-1/2}HU^{-1}H'\Sigma^{*-1/2},$$

$$(A'A)^{1/2}(\hat{\Xi} - \Xi^*)B\Sigma^{*-1/2} = Z\begin{pmatrix} I_q \\ -W_{22}^{-1}W_{21} \end{pmatrix}H_1',$$

respectively, where

$$U^{-1} = \begin{pmatrix} W_{11\cdot 2}^{-1} & -W_{11\cdot 2}^{-1}W_{12}W_{22}^{-1} \\ -W_{22}^{-1}W_{21}W_{11\cdot 2}^{-1} & (W_{22}+Z_2'Z_2)^{-1} + W_{22}^{-1}W_{21}W_{11\cdot 2}^{-1}W_{12}W_{22}^{-1} \end{pmatrix}.$$

On the other hand, note that

$$\begin{aligned}
&\mathrm{E}^*[\mathrm{tr}\{\hat{\Sigma}_{[-i]}^{-1}B'(\hat{\Xi} - \Xi^*)'\boldsymbol{a}_i\boldsymbol{a}_i'(\hat{\Xi} - \Xi^*)B\}] \\
&\quad = \mathrm{E}^*[\mathrm{tr}\{\hat{\Sigma}_{[-i]}^{-1}B'(\hat{\Xi} - \Xi^*)'(A_{(-i)}'A_{(-i)})^{1/2}Q_{i(-i)} \\
&\qquad\qquad \times (A_{(-i)}'A_{(-i)})^{1/2}(\hat{\Xi} - \Xi^*)B\}],
\end{aligned}$$

where

$$Q_{i(-i)} = (A_{(-i)}'A_{(-i)})^{-1/2}\boldsymbol{a}_i\boldsymbol{a}_i'(A_{(-i)}'A_{(-i)})^{-1/2}.$$

Let $W_{(-i)}$ and $Z_{(-i)}$ be the ones defined from W and Z by replacing n, A and Y to $n-1$, $A_{(-i)}$ and $Y_{(-i)}$. Then the previous expression can be computed as

$$(n-1)\mathrm{E}^*\left[\mathrm{tr}\left\{\begin{pmatrix} I_q \\ -W_{(-i)22}^{-1}W_{(-i)21}\end{pmatrix} W_{(-i)11\cdot 2}^{-1}(I_q \;\; -W_{(-i)12}W_{(-i)22}^{-1}) \times Z'_{(-i)}Q_{i(-i)}Z_{(-i)}\right\}\right].$$

If $\boldsymbol{y}_i$ belongs to the j-th population, then

$$Q_{i(-i)} = \mathrm{diag}\{0,\dots,0,(n_j-1)^{-1},0,\dots,0\}.$$

Using this result and Lemma 4.1 and yields the result (4.5).

A.4 *Derivation of* (4.6)

Using the same notations as in Subsection A.3, we can get

$$\begin{aligned}&\mathrm{E}^*[\mathrm{tr}\{\tilde{S}_{[-i]}^{-1}B'(\widehat{\Xi}-\Xi^*)'\boldsymbol{a}_i\boldsymbol{a}_i'(\widehat{\Xi}-\Xi^*)B\}]\\ &= \frac{n-k-p-2}{n-1}\mathrm{E}^*[\mathrm{tr}\{\widehat{\Sigma}_{[-i]}^{-1}B'(\widehat{\Xi}-\Xi^*)'(A'_{(-i)}A_{(-i)})^{1/2}Q_{i(-i)}\\ &\qquad \times (A'_{(-i)}A_{(-i)})^{1/2}(\widehat{\Xi}-\Xi^*)B\}].\end{aligned}$$

Therefore, we can obtain the result (4.6).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csáki), 267–281, Akadémia Kiado, Budapest.

Anderson, T. W. (1984). *An Introduction to Multivariate Analysis*, 2nd ed., John Wiley & Sons, New York.

Bedrick, E. J. and Tsai, C. L. (1994). Model selection for multivariate regression in small samples, *Biometrics*, **76**, 226–231.

Fujikoshi, Y. and Satoh, K. (1997). Modified AIC and C_p in multivariate linear regression, *Biometrika*, **84**, 707–716.

Gleser, L. J. and Olkin, I. (1970). Linear models in multivariate analysis, *Essays in Probability and Statistics* (eds. R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao and K. J. C. Smith), University of North Carolina Press, Chapel Hill, North Carolina.

Mallows, C. L. (1973). Some comments on C_p, *Technometrics*, **15**, 661–675.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, **51**, 313–325.

Satoh, K., Kobayashi, M. and Fujikoshi, Y. (1997). Variable selection for the growth curve model, *J. Multivariate Anal.*, **60**, 277–292.

Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985). *Modern Multivariate Statistical Analysis: A Graduate Course and Handbook*, American Sciences Press, Columbus, Ohio.

Stone, M. (1974). Cross-validation and multinomial prediction, *Biometrika*, **61**, 509–515.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, *J. Roy. Statist. Soc. Ser. B*, **39**, 44–47.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections, *Comm. Statist. Theory Methods*, **7**, 13–26.