Strong consistency of log-likelihood-based information criterion in high-dimensional canonical correlation analysis

Ryoya Oda[*] , **Hirokazu Yanagihara and Yasunori Fujikoshi**

Department of Mathematics, Graduate School of Science, Hiroshima University

1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8626, Japan

(Last Modified: January 7, 2019)

Abstract

We consider the strong consistency of a log-likelihood-based information criterion in a normality-assumed canonical correlation analysis between q- and p-dimensional random vectors for a high-dimensional case such that the sample size n and number of dimensions p are large but p/n is less than 1. In general, strong consistency is a stricter property than weak consistency; thus, sufficient conditions for the former do not always coincide with those for the latter. We derive the sufficient conditions for the strong consistency of this log-likelihood-based information criterion for the high-dimensional case. It is shown that the sufficient conditions for strong consistency of several criteria are the same as those for weak consistency obtained by Yanagihara *et al.* (2017).

1 Introduction

Let $\boldsymbol{x} = (x_1, \ldots, x_q)'$ and $\boldsymbol{y} = (y_1, \ldots, y_p)'$ be q- and p-dimensional random vectors. Investigating relationships between $\boldsymbol{x}$ and $\boldsymbol{y}$ is central to multivariate analysis. Canonical correlation analysis (CCA) is one such multivariate method which has been considered widely in the theoretical and applied peer-reviewed literature, as well as textbooks aimed at under- and post-graduate students (see, e.g., Srivastava, 2002, chap. 14.7; Timm, 2002, chap. 8.7).

In actual empirical contexts, there are cases where some of the variables in a given model may be redundant. It is important to be able to effectively identify and remove such variables. This paper focusses on removing redundant variables in $\boldsymbol{x}$. The problem is regarded as selecting the best subset of $\boldsymbol{x}$, and this has hitherto been investigated in many studies (e.g., McKay, 1977; Fujikoshi, 1982, 1985; Ogura, 2010). As one approach to model selection, the method of minimizing an information criterion is well known. The most widely applied of these criteria is Akaike's (1973, 1974) information criterion (AIC). Fujikoshi (1985) applied Akaike's idea to the issue of selection in CCA. Nishii *et al.* (1988) proposed a generalized information criterion (GIC). Fukui (2015) and Yanagihara *et al.* (2017) considered a log-likelihood-based information criterion (LLIC). GIC and LLIC are essentially the same and are defined by adding a penalty term to

[*] Corresponding author. Email: oda.stat@gmail.com

a negative twofold maximum log-likelihood. GIC and LLIC include several information criteria as special cases: AIC, a bias-corrected AIC (AIC$_{\mathrm{c}}$) proposed by Fujikoshi (1985), the Bayesian information criterion (BIC) proposed by Schwarz (1978), a consistent AIC (CAIC) proposed by Bozdogan (1987), the Hannan-Quinn information criterion (HQC) proposed by Hannan and Quinn (1979), and so on.

An important property to consider regarding information criteria is their consistency. In fact, there are two properties in this respect, that is weak consistency and strong consistency. Let $\hat{j}$ and j_* be the best subset identified by minimizing respectively an information criterion and the true subset, i.e., the minimum subset including the true model. Weak consistency means that the asymptotic probability of selecting the true subset approaches one, i.e., $P(\hat{j} = j_*) \to 1$ $(n \to \infty)$, where n is the sample size. In the context of CCA, Yanagihara *et al.* (2017) obtained sufficient conditions for weak consistency of LLIC when the sample size n tends to ∞ and the number of dimensions p may tend to ∞, assuming that the true distribution of the observation vectors is the multivariate normal distribution. Moreover, they derived sufficient conditions for several specific criteria. Relaxing the normality assumption, Fukui (2015) derived sufficient conditions for weak consistency of LLIC when both n and p tend to ∞. On the other hand, strong consistency means that the probability that the best subset approaches the true subset is one, i.e., $P(\hat{j} \to j_*) = 1$. Since each subset is discrete, strong consistency ensures that there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $\hat{j} = j_*$ with probability 1. Hence, strong consistency is stricter than weak consistency for selecting the true subset. Again in the context of CCA, Nishii *et al.* (1988) obtained the sufficient conditions for strong consistency of GIC when only the sample size n tends to ∞. However, the conditions for strong consistency have not hitherto been derived for the case where both n and p tend to ∞. Moreover, since strong consistency is stricter than weak consistency, it is not currently known whether several criteria satisfying sufficient conditions for weak consistency according to Yanagihara *et al.* (2017) are strongly consistent in high-dimensional cases.

The aim of this paper is to obtain sufficient conditions for strong consistency of LLIC and several other criteria when both the sample size n and the number of dimensions p tend to ∞ but p does not exceed n. We assume that the number of dimensions p is a function of n, that is we write $p = p(n)$, and use the following high-dimensional (HD) asymptotic framework:

$$p = p(n),\ n \to \infty,\ \frac{p}{n} \to c \in [0, 1).$$

Based on sufficient conditions for strong consistency of LLIC, we show that the conditions for strong consistency of AIC, AIC$_{\mathrm{c}}$, BIC, CAIC, and HQC are equivalent to those for weak consistency put forward by Yanagihara *et al.* (2017).

The remainder of the paper is organized as follows. In section 2, we introduce redundancy models and LLIC. In section 3, we present our key lemmas and main results to derive the sufficient conditions for strong consistency. Technical details are relegated to the Appendix.

2 Preliminaries

In this section, we introduce redundancy models and LLIC in the context of CCA. Let $\boldsymbol{z} = (\boldsymbol{x}', \boldsymbol{y}')'$ be a $(q+p)$-dimensional random vector distributed according to a $(q+p)$-variate normal distribution with

$$E[\boldsymbol{z}] = \boldsymbol{\mu} = (\boldsymbol{\mu}_x', \boldsymbol{\mu}_y')' \ , \ Cov[\boldsymbol{z}] = \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{xy}' & \boldsymbol{\Sigma}_{yy} \end{pmatrix},$$

where $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ are q- and p-dimensional mean vectors of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{yy}$ are $q \times q$ and $p \times p$ covariance matrices of $\boldsymbol{x}$ and $\boldsymbol{y}$, and $\boldsymbol{\Sigma}_{xy}$ is the $q \times p$ covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{y}$. Suppose that j denotes a subset of $\omega = \{1, \dots, q\}$ containing q_j elements, and $\boldsymbol{x}_j$ denotes the q_j-dimensional random vector consisting of $\boldsymbol{x}$ indexed by the elements of j. For example, if $j = \{1, 2, 4\}$, then $\boldsymbol{x}_j$ consists of the first, second, and fourth elements of $\boldsymbol{x}$. Without loss of generality, we can express $\boldsymbol{x}$ as $\boldsymbol{x} = (\boldsymbol{x}_j', \boldsymbol{x}_{\bar{j}}')'$, where $\boldsymbol{x}_{\bar{j}}$ is a $(q - q_j)$-dimensional random vector. Then, for a subset j, the covariance matrices $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{xy}$ are expressed as follows:

$$\boldsymbol{\Sigma}_{xx} = \begin{pmatrix} \boldsymbol{\Sigma}_{jj} & \boldsymbol{\Sigma}_{j\bar{j}} \\ \boldsymbol{\Sigma}_{j\bar{j}}' & \boldsymbol{\Sigma}_{\bar{j}\bar{j}} \end{pmatrix}, \ \ \boldsymbol{\Sigma}_{xy} = \begin{pmatrix} \boldsymbol{\Sigma}_{jy} \\ \boldsymbol{\Sigma}_{\bar{j}y} \end{pmatrix},$$

where the sizes of $\boldsymbol{\Sigma}_{jj}$, $\boldsymbol{\Sigma}_{j\bar{j}}$, $\boldsymbol{\Sigma}_{jy}$, and $\boldsymbol{\Sigma}_{\bar{j}y}$ are $q_j \times q_j$, $q_j \times (q - q_j)$, $q_j \times p$, and $(q - q_j) \times p$. Let $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n$ be n independent random vectors from $\boldsymbol{z}$, and let $\bar{\boldsymbol{z}}$ be the sample mean of $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n$ given by $\bar{\boldsymbol{z}} = n^{-1} \sum_{i=1}^n \boldsymbol{z}_i$ and $\boldsymbol{S}$ be the usual unbiased estimator of $\boldsymbol{\Sigma}$ given by $\boldsymbol{S} = (n-1)^{-1} \sum_{i=1}^n (\boldsymbol{z}_i - \bar{\boldsymbol{z}})(\boldsymbol{z}_i - \bar{\boldsymbol{z}})'$. In the same way as $\boldsymbol{\Sigma}$, we also partition $\boldsymbol{S}$ as follows:

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{xx} & \boldsymbol{S}_{xy} \\ \boldsymbol{S}_{xy}' & \boldsymbol{S}_{yy} \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{jj} & \boldsymbol{S}_{j\bar{j}} & \boldsymbol{S}_{jy} \\ \boldsymbol{S}_{j\bar{j}}' & \boldsymbol{S}_{\bar{j}\bar{j}} & \boldsymbol{S}_{\bar{j}y} \\ \boldsymbol{S}_{jy}' & \boldsymbol{S}_{\bar{j}y}' & \boldsymbol{S}_{yy} \end{pmatrix}.$$

From Fujikoshi (1982), $\boldsymbol{x}_{\bar{j}}$ is redundant if the following equation holds:

$$\mathrm{tr}(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}) = \mathrm{tr}(\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{jy}'). \tag{1}$$

The left-hand side in (1) expresses the sum of squares of the canonical correlation coefficients between $\boldsymbol{x}$ and $\boldsymbol{y}$, and the right-hand side expresses the sum of squares of the canonical correlation coefficients between $\boldsymbol{x}_j$ and $\boldsymbol{y}$. In particular, we note that (1) is equivalent (see, Fujikoshi, 1982) to

$$\boldsymbol{\Sigma}_{\bar{j}y \cdot j} = \boldsymbol{O}_{q-q_j,p}, \tag{2}$$

where $\boldsymbol{\Sigma}_{ab\cdot c} = \boldsymbol{\Sigma}_{ab} - \boldsymbol{\Sigma}_{ac}\boldsymbol{\Sigma}_{cc}^{-1}\boldsymbol{\Sigma}_{cb}$, and $\boldsymbol{O}_{q-q_j,p}$ is the $(q - q_j) \times p$ matrix of zeros.

We regard a subset j as the candidate model such that $\boldsymbol{x}_{\bar{j}}$ is redundant. Following Fujikoshi (1985), the candidate model j such that $\boldsymbol{x}_{\bar{j}}$ is redundant is expressed as

$$j : (n-1)\boldsymbol{S} \sim W_{p+q}(n-1, \boldsymbol{\Sigma}) \ s.t. \ \mathrm{tr}(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}) = \mathrm{tr}(\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{jy}'). \tag{3}$$

Let $\mathcal{J}$ be a set of candidate models. We then separate $\mathcal{J}$ into two sets, one is a set of overspecified models $\mathcal{J}_+$ and the other is a set of underspecified models $\mathcal{J}_-$, which are defined by

$$\mathcal{J}_+ = \{j \in \mathcal{J} | \ \mathrm{tr}(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}) = \mathrm{tr}(\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{jy}')\}, \ \mathcal{J}_- = \mathcal{J} \backslash \mathcal{J}_+.$$

Then, the true model (or subset) j_* can be regarded as the smallest overspecified model, i.e., $j_* = \arg\min_{j \in \mathcal{J}_+} q_j$. For simplicity, we write q_{j_*} as q_*. An estimator of $\boldsymbol{\Sigma}$ in (3) is given by

$$\hat{\boldsymbol{\Sigma}}_j = \arg\min_{\boldsymbol{\Sigma}} F(\boldsymbol{S}, \boldsymbol{\Sigma}) \text{ subject to } \mathrm{tr}(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx}) = \mathrm{tr}(\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{jy}'),$$

where $F(\boldsymbol{S}, \boldsymbol{\Sigma})$ is the discrepancy function based on Stein's loss function given by

$$F(\boldsymbol{S}, \boldsymbol{\Sigma}) = (n-1)\{\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}) - \log|\boldsymbol{\Sigma}^{-1}\boldsymbol{S}| - (p+q)\}.$$

By using $F(\boldsymbol{S}, \hat{\boldsymbol{\Sigma}}_j)$, LLIC in (3) is defined as

$$\mathrm{LLIC}(j) = F(\boldsymbol{S}, \hat{\boldsymbol{\Sigma}}_j) + m(j) = (n-1)\log\frac{|\boldsymbol{S}_{yy\cdot j}|}{|\boldsymbol{S}_{yy\cdot\omega}|} + m(j),$$

where $\boldsymbol{S}_{yy\cdot j} = \boldsymbol{S}_{yy} - \boldsymbol{S}'_{jy}\boldsymbol{S}_{jj}^{-1}\boldsymbol{S}_{jy}$, and $m(j)$ is the penalty term in (3). For simplicity, we write $\boldsymbol{S}_{yy\cdot\omega}$ as $\boldsymbol{S}_{yy\cdot x}$. By choosing $m(j)$ in various quantities, we can express the following criteria as special cases of LLIC:

$$m(j) = \begin{cases} p^2 + q^2 + p + q + 2pq_j & \text{(AIC)} \\ (n-1)^2\left(\dfrac{p+q_j}{n-p-q_j-2} + \dfrac{q}{n-q-2} - \dfrac{q_j}{n-q_j-2} - \dfrac{p+q}{n-1}\right) & (\mathrm{AIC_c}) \\ \left\{\dfrac{(p+q)(p+q+1)}{2} - p(q-q_j)\right\}\log n & \text{(BIC)} \\ \left\{\dfrac{(p+q)(p+q+1)}{2} - p(q-q_j)\right\}(1+\log n) & \text{(CAIC)} \\ 2\left\{\dfrac{(p+q)(p+q+1)}{2} - p(q-q_j)\right\}\log\log n & \text{(HQC)} \end{cases}.$$

Note that LLIC is the same as GIC when $m(j)$ is expressed as the number of parameters in (3) multiplied by the strength of the penalty. The best subset $\hat{j}$ selected by LLIC is given by

$$\hat{j} = \arg\min_{j\in\mathcal{J}} \mathrm{LLIC}(j).$$

3 Main results

In this section, we give the sufficient conditions for strong consistency of LLIC under the HD asymptotic framework. First, we present some lemmas which are required for deriving the strong consistency conditions, with proofs provided in the Appendix.

Lemma 3.1 *Suppose that p is fixed or $p = p(n)$. Let $\hat{j} = \arg\min_{j\in\mathcal{J}} \mathrm{LLIC}(j)$, and let $h_{j,\ell}$ be some positive constant not converging to 0 for $j, \ell \in \mathcal{J}$. Then, we have*

$$\forall \ell \in \mathcal{J}\backslash\{j\},\ \frac{1}{h_{j,\ell}}\{\mathrm{LLIC}(\ell) - \mathrm{LLIC}(j)\} \to \tau_{j,\ell} > 0,\ a.s.\ \Rightarrow P(\hat{j} \to j) = 1.$$

From Lemma 3.1, to obtain sufficient conditions such that LLIC is strongly consistent, we may derive the almost sure convergence of $h_{j,\ell}^{-1}\{\mathrm{LLIC}(j) - \mathrm{LLIC}(j_*)\}$ for all $j \in \mathcal{J}\backslash\{j_*\}$. Hence, we derive the convergence by using the following lemma.

Lemma 3.2 *Let $p = p(n)$ and let r be a natural number not relying on n. Suppose that t_1 and $\boldsymbol{T}_2$ are a random variable and an $r \times r$ matrix satisfying*

$$E\left[(t_1 - E[t_1])^{2k}\right] = O(p^k n^{-2k})\ (\forall k \in \mathbb{N}),\ E\left[||\boldsymbol{T}_2 - E[\boldsymbol{T}_2]||^4\right] = O(n^2),$$

where $||\boldsymbol{A}||$ is the Frobenius norm for a matrix $\boldsymbol{A}$. Then for all $\varepsilon > 0$, we have

$$np^{-1/2-\varepsilon}(t_1 - E[t_1]) = o(1),\ a.s., \tag{4}$$

$$n^{-3/4-\varepsilon}(\boldsymbol{T}_2 - E[\boldsymbol{T}_2]) = o(1),\ a.s. \tag{5}$$

Before giving the strong consistency conditions, let us prepare some notation. For a subset $j \in \mathcal{J}$, let a non-centrality parameter and a $p \times (q - q_j)$ matrix be denoted by

$$\delta_j = \log |\boldsymbol{I}_p + \boldsymbol{\Gamma}_j \boldsymbol{\Gamma}_j'|, \ \boldsymbol{\Gamma}_j = \boldsymbol{\Sigma}_{yy\cdot\omega}^{-1/2} \boldsymbol{\Sigma}_{\bar{j}y\cdot j}' \boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1/2}. \tag{6}$$

As well as $\boldsymbol{S}_{yy\cdot x}$, we write $\boldsymbol{\Sigma}_{yy\cdot\omega}$ as $\boldsymbol{\Sigma}_{yy\cdot x}$. From (2), we observe that $\delta_j = 0$ and $\boldsymbol{\Gamma}_j = \boldsymbol{O}_{q-q_j,p}$ hold if and only if $j \in \mathcal{J}_+$. Hence, we derive the sufficient conditions in each case of $j \in \mathcal{J}_+ \backslash \{j_*\}$ and $j \in \mathcal{J}_-$. By using this notation and these lemmas, we derive the sufficient conditions for strong consistency of LLIC (the proof is given in Appendix C).

Theorem 3.1 GIC *is strongly consistent as* $p = p(n),\ n \to \infty,\ p/n \to c \in [0, 1)$*, if the following two conditions are satisfied simultaneously*:

$$\mathrm{C1} : \lim_{n\to\infty,\ p/n\to c} p^{1/2-\varepsilon} \left\{ (q_j - q_*) \frac{n}{p} \log \left(1 - \frac{p}{n}\right) + m(j) - m(j_*) \right\} > 0 \ for\ some\ \varepsilon\ (0 < \varepsilon \leq 1/2),$$

$$\mathrm{C2} : \forall j \in \mathcal{J}_-, \lim_{n\to\infty,\ p/n\to c} \left[\delta_j + (q_j - q_*) \log \left(1 - \frac{p}{n}\right) + \frac{1}{n}\{m(j) - m(j_*)\} \right] > 0,$$

where δ_j *is defined in* (6).

We note that the sufficient conditions for strong consistency by Theorem 3.1 are similar to those for weak consistency according to Yanagihara *et al.* (2017) under the HD asymptotic framework. By using Theorem 3.1, we derive the sufficient conditions for strong consistency of several criteria. The proof is omitted because it can be found in Yanagihara *et al.* (2017).

Corollary 3.1 *As* $p = p(n),\ n \to \infty,\ p/n \to c \in [0, 1)$*, the sufficient conditions for strong consistency of several criteria are giving as follows:*

(1) *When* $c = 0$*, AIC, AIC*$_c$*, BIC, CAIC, and HQC are strongly consistent.*

(2) *When* $c > 0$*,*

(i) *AIC is strongly consistent, if* $c < c_a$ *and*

$$q_* - q_j < \frac{1}{2} \left\{ \frac{1}{c} \left(\lim_{n\to\infty,\ p/n\to c} \delta_j \right) + (q_j - q_*) \frac{1}{c} \log(1 - c) \right\} \ (\forall j \in \mathcal{J}_-), \tag{7}$$

where c_a (≈ 0.797) *is the solution of* $x^{-1} \log(1 - x) + 2 = 0$*. Especially, if* $\delta_j \to \infty$*, (7) holds.*

(ii) *AIC*$_c$ *is strongly consistent if*

$$q_* - q_j < \frac{(1-c)^2}{(2-c)} \left\{ \frac{1}{c} \left(\lim_{n\to\infty,\ p/n\to c} \delta_j \right) + (q_j - q_*) \frac{1}{c} \log(1 - c) \right\} \ (\forall j \in \mathcal{J}_-). \tag{8}$$

Especially, if $\delta_j \to \infty$*, (8) holds.*

(iii) *BIC and CAIC are strongly consistent if*

$$q_* - q_j < \frac{1}{c} \left(\lim_{n\to\infty,\ p/n\to c} \frac{\delta_j}{\log n} \right) \ (\forall j \in \mathcal{J}_-). \tag{9}$$

Especially, if $\delta_j / \log n \to \infty$*, (9) holds.*

(iv) *HQC is strongly consistent if*

$$q_* - q_j < \frac{1}{2c}\left(\lim_{(n\to\infty,\ p/n\to c}\frac{\delta_j}{\log\log n}\right)\ (\forall j\in\mathcal{J}_-). \tag{10}$$

Especially, if $\delta_j/\log\log n \to \infty$, (10) *holds.*

From Corollary 3.1, under the HD asymptotic framework we observe that the conditions for strong consistency of AIC, AIC_c, BIC, CAIC, and HQC are equivalent to those for weak consistency derived by Yanagihara *et al.* (2017).

Acknowledgments

Ryoya Oda was supported by a Research Fellowship for Young Scientists from the Japan Society for the Promotion of Science. Hirokazu Yanagihara and Yasunori Fujikoshi were partially supported by Grants-in-Aid for Scientific Research (C) from the Ministry of Education, Science, Sports, and Culture, #18K03415 and #16K00047, respectively.

Appendix

A Proof of Lemma 3.1

From the assumption of Lemma 3.1, the following reductions can be derived:

$$\begin{aligned}
1 &= P\left(\bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\left\{\left|\frac{1}{h_{j,\ell}}(\mathrm{LLIC}(\ell)-\mathrm{LLIC}(j))-\tau_{j,\ell}\right|<\frac{1}{k}\right\}\right)\\
&= P\left(\bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\left\{-\frac{1}{k}+\tau_{j,\ell}<\frac{1}{h_{j,\ell}}(\mathrm{LLIC}(\ell)-\mathrm{LLIC}(j))<\frac{1}{k}+\tau_{j,\ell}\right\}\right)\\
&\le P\left(\bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\left\{-\frac{1}{k}+\tau_{j,\ell}<\frac{1}{h_{j,\ell}}(\mathrm{LLIC}(\ell)-\mathrm{LLIC}(j))\right\}\right)\\
&\le P\left(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\left\{\frac{1}{h_{j,\ell}}(\mathrm{LLIC}(\ell)-\mathrm{LLIC}(j))>\tau_{j,\ell}\right\}\right)\\
&\le P\left(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\left\{\frac{1}{h_{j,\ell}}(\mathrm{LLIC}(\ell)-\mathrm{LLIC}(j))>0\right\}\right).
\end{aligned}$$

Hence, we have

$$\begin{aligned}
P(\hat{j} \to j) &= P\left(\bigcap_{k=1}^{\infty}\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}\left\{|\hat{j}-j| < \frac{1}{k}\right\}\right) \\
&= 1 - P\left(\bigcup_{k=1}^{\infty}\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\left\{|\hat{j}-j| \geq \frac{1}{k}\right\}\right) \\
&= 1 - P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{\hat{j} \neq j\}\right) \\
&= 1 - P\left(\bigcup_{\ell \in \mathcal{J}\backslash\{j\}}\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{\mathrm{LLIC}(\ell) < \mathrm{LLIC}(j)\}\right) \\
&\geq 1 - \sum_{\ell \in \mathcal{J}\backslash\{j\}} P\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty}\{\mathrm{LLIC}(\ell) - \mathrm{LLIC}(j) < 0\}\right) \\
&= 1.
\end{aligned}$$

This completes the proof of Lemma 3.1.

B Proof of Lemma 3.2

Let us take an arbitrary $\varepsilon > 0$, and let k be a natural number such that $k > (2\varepsilon)^{-1}$. By using Markov's inequality, for all $\delta > 0$, we have

$$\begin{aligned}
P(np^{-1/2-\varepsilon}|t_1 - E[t_1]| > \delta) &\leq \frac{1}{(n^{-1}p^{1/2+\varepsilon}\delta)^{2k}} E[(t_1 - E[t_1])^{2k}] \\
&= O(p^{-2k\varepsilon}), \\
P(n^{-3/4-\varepsilon}||\boldsymbol{T} - E[\boldsymbol{T}]|| > \delta) &\leq \frac{1}{(n^{3/4+\varepsilon}\delta)^4} E[(||\boldsymbol{T} - E[\boldsymbol{T}]||^4] \\
&= O(n^{-1-\varepsilon}).
\end{aligned}$$

Then, since $p = p(n)$ and $k > (2\varepsilon)^{-1}$, it holds that $\sum_{n=1}^{\infty} p^{-2k\varepsilon} < \infty$ and $\sum_{n=1}^{\infty} n^{-1-\varepsilon} < \infty$. These equations and the Borel-Cantelli lemma complete the proof of Lemma 3.2.

C Proof of Theorem 3.1

To prove Theorem 3.1, we use three lemmas from Yanagihara *et al.* (2017) and Oda & Yanagihara (2019). Before Lemma C.1 is introduced, let $\boldsymbol{Q}$ be an $n \times (n-1)$ matrix satisfying $\boldsymbol{I}_n - n^{-1}\boldsymbol{1}_n\boldsymbol{1}_n' = \boldsymbol{Q}\boldsymbol{Q}'$ and $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}_{n-1}$, where $\boldsymbol{1}_n$ is the n-dimensional vector of ones. Further, let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$, where $\boldsymbol{x}_i$ is the i-th individual from $\boldsymbol{x}$. The following lemma is Lemma C.1 by Yanagihara *et al.* (2017).

Lemma C.1 *For a subset $j \in \mathcal{J}$, let $\boldsymbol{\mathcal{E}}$, $\boldsymbol{A}_j$, and $\boldsymbol{B}_j$ be mutually independent random matrices, which are distributed according to*

$$\begin{aligned}
&\boldsymbol{\mathcal{E}} \sim N_{(n-1)\times p}(\boldsymbol{O}_{n-1,p}, \boldsymbol{I}_p \otimes \boldsymbol{I}_{n-1}),\ \ \boldsymbol{A}_j \sim N_{(n-1)\times(q-q_j)}(\boldsymbol{O}_{n-1,q-q_j}, \boldsymbol{I}_{q-q_j} \otimes \boldsymbol{I}_{n-1}), \\
&\boldsymbol{B} = \boldsymbol{Q}'\boldsymbol{X} = (\boldsymbol{B}_j, \boldsymbol{B}_{\bar{j}}) \sim N_{(n-1)\times q}(\boldsymbol{O}_{n-1,q}, \boldsymbol{\Sigma}_{xx} \otimes \boldsymbol{I}_{n-1}),
\end{aligned}$$

where $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{B}$ are independent and do not rely on j, and $\boldsymbol{B}_j : (n-1) \times q_j$. Then, we have

$$(n-1)\boldsymbol{S}_{yy\cdot x} = \boldsymbol{\Sigma}_{yy\cdot x}^{1/2}\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_{n-1} - \boldsymbol{P})\boldsymbol{\mathcal{E}}\boldsymbol{\Sigma}_{yy\cdot x}^{1/2},$$
$$(n-1)\boldsymbol{S}_{yy\cdot j} = \boldsymbol{\Sigma}_{yy\cdot x}^{1/2}(\boldsymbol{A}_j\boldsymbol{\Gamma}_j' + \boldsymbol{\mathcal{E}})'(\boldsymbol{I}_{n-1} - \boldsymbol{P}_j)(\boldsymbol{A}_j\boldsymbol{\Gamma}_j' + \boldsymbol{\mathcal{E}})\boldsymbol{\Sigma}_{yy\cdot x}^{1/2},$$

where $\boldsymbol{P} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'$, $\boldsymbol{P}_j = \boldsymbol{B}_j(\boldsymbol{B}_j'\boldsymbol{B}_j)^{-1}\boldsymbol{B}_j'$, and $\boldsymbol{\Gamma}_j$ is defined in (6).

The following lemma is given by using (23) and (B.6) in Yanagihara *et al.* (2017).

Lemma C.2 *For a subset $j \in \mathcal{J}$, let $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ be independent random matrices distributed according to*

$$\boldsymbol{U}_1 \sim N_{(n-q-1)\times p}(\boldsymbol{O}_{(n-q-1)\times p}, \boldsymbol{I}_p \otimes \boldsymbol{I}_{n-q-1}),\ \boldsymbol{U}_2 \sim N_{(q-q_j)\times p}(\boldsymbol{O}_{(q-q_j)\times p}, \boldsymbol{I}_p \otimes \boldsymbol{I}_{q-q_j}).$$

Further, let $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ be random matrices distributed according to

$$\boldsymbol{W}_1 \sim W_{q-q_j}(n-p+q-2q_j-1, \boldsymbol{I}_{q-q_j}),\ \boldsymbol{W}_2 \sim W_{q-q_j}(n-p+q-2q_j-1, \boldsymbol{I}_{q-q_j}).$$

Then, we have

$$\log\frac{|\boldsymbol{S}_{yy\cdot j}|}{|\boldsymbol{S}_{yy\cdot x}|} = \delta_j + \log\frac{|\boldsymbol{U}_1'\boldsymbol{U}_1 + \boldsymbol{U}_2'\boldsymbol{U}_2|}{|\boldsymbol{U}_1'\boldsymbol{U}_1|} + \log\frac{|\boldsymbol{W}_1|}{|\boldsymbol{W}_2|},$$

where δ_j is defined in (6).

The following lemma is Lemma C.2 in Oda & Yanagihara (2019).

Lemma C.3 *Suppose that $N - 4k > 0$ for $k \in \mathbb{N}$. Let u and v be independent random variables distributed according to $u \sim \chi^2(N)$ and $v \sim \chi^2(p)$. Then, we have*

$$E\left[\left(\frac{v}{u} - \frac{p}{N-2}\right)^{2k}\right] = O(p^k N^{-2k}).$$

First, we consider the case of $j \in \mathcal{J}_+\backslash\{j_*\}$. The distinct elements of $j\backslash j_*$ denote $a_1, \dots, a_{q_j - q_*}$. Let $j_0 = j$, $j_i = j_{i-1}\backslash\{a_i\}$ $(1 \le i \le q_j - q_*)$. Then, $j_{q_j-q_*} = j$ holds, and we can express LLIC(j) – LLIC(j_*) as follows:

$$\begin{aligned}\mathrm{LLIC}(j) - \mathrm{LLIC}(j_*) &= (n-1)\log\frac{|\boldsymbol{S}_{yy\cdot j}|}{|\boldsymbol{S}_{yy\cdot j_*}|} + m(j) - m(j_*) \\ &= (n-1)\sum_{i=1}^{q_j-q_*}\log\frac{|\boldsymbol{S}_{yy\cdot j_{i-1}}|}{|\boldsymbol{S}_{yy\cdot j_i}|} + m(j) - m(j_*). \qquad \text{(C.1)}\end{aligned}$$

Then, from Lemma C.1, $\boldsymbol{S}_{yy\cdot j_{i-1}}$ can be expressed as follows:

$$(n-1)\boldsymbol{S}_{yy\cdot j_{i-1}} = \boldsymbol{\Sigma}_{yy\cdot x}^{1/2}\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_{n-1} - \boldsymbol{P}_{j_{i-1}})\boldsymbol{\mathcal{E}}\boldsymbol{\Sigma}_{yy\cdot x}^{1/2}, \qquad \text{(C.2)}$$

where $\boldsymbol{\mathcal{E}} \sim N_{(n-1)\times p}(\boldsymbol{O}_{n-1,p}, \boldsymbol{I}_p \otimes \boldsymbol{I}_{n-1})$, $\boldsymbol{P}_{j_{i-1}} = \boldsymbol{B}_{j_{i-1}}(\boldsymbol{B}_{j_{i-1}}'\boldsymbol{B}_{j_{i-1}})^{-1}\boldsymbol{B}_{j_{i-1}}'$,
$\boldsymbol{B}_{j_{i-1}} \sim N_{(n-1)\times(q_j-i+1)}(\boldsymbol{O}_{n-1,q_j-i+1}, \boldsymbol{\Sigma}_{j_{i-1}j_{i-1}} \otimes \boldsymbol{I}_{n-1})$, and $\boldsymbol{\mathcal{E}}$ is independent of $\boldsymbol{B}_{j_{i-1}}$. Moreover, by applying Lemma C.1 to $\boldsymbol{S}_{yy\cdot j_i}$, we have

$$(n-1)\boldsymbol{S}_{yy\cdot j_i} = \boldsymbol{\Sigma}_{yy\cdot x}^{1/2}\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_{n-1} - \boldsymbol{P}_{j_i})\boldsymbol{\mathcal{E}}\boldsymbol{\Sigma}_{yy\cdot x}^{1/2}, \qquad \text{(C.3)}$$

where $\boldsymbol{P}_{j_i} = \boldsymbol{B}_{j_i}(\boldsymbol{B}_{j_i}'\boldsymbol{B}_{j_i})^{-1}\boldsymbol{B}_{j_i}'$, and $\boldsymbol{B}_{j_i}$ is the $(n-1)\times(q_j-i)$ sub matrix of $\boldsymbol{B}_{j_{i-1}} = (\boldsymbol{B}_{j_i}, \boldsymbol{b}_{j_i})$. Let

$$\boldsymbol{V}_{i,1} = \boldsymbol{\mathcal{E}}'(\boldsymbol{I}_{n-1} - \boldsymbol{P}_{j_{i-1}})\boldsymbol{\mathcal{E}},\ \boldsymbol{V}_{i,2} = \boldsymbol{\mathcal{E}}'(\boldsymbol{P}_{j_{i-1}} - \boldsymbol{P}_{j_i})\boldsymbol{\mathcal{E}}. \tag{C.4}$$

Since $(\boldsymbol{I}_{n-1} - \boldsymbol{P}_{j_{i-1}})(\boldsymbol{P}_{j_{i-1}} - \boldsymbol{P}_{j_i}) = \boldsymbol{O}_{n-1,n-1}$ holds, we observe that $\boldsymbol{V}_{i,1}$ and $\boldsymbol{V}_{i,2}$ are independent, and $\boldsymbol{V}_{i,1} \sim W_p(n-q_j+i-2, \boldsymbol{I}_p)$, $\boldsymbol{V}_{i,2} \sim W_p(1, \boldsymbol{I}_p)$ from a property of the Wishart distribution and Cochran's Theorem (see, e.g., Fujikoshi *et al.*, 2010, Theorem 2.4.2). By using (C.2), (C.3), and (C.4), we have

$$\begin{aligned}\frac{|\boldsymbol{S}_{yy\cdot j_i}|}{|\boldsymbol{S}_{yy\cdot j_{i-1}}|} &= \frac{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_{n-1} - \boldsymbol{P}_{j_i})\boldsymbol{\mathcal{E}}|}{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_{n-1} - \boldsymbol{P}_{j_{i-1}})\boldsymbol{\mathcal{E}}|} \\ &= \frac{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_{n-1} - \boldsymbol{P}_{j_{i-1}})\boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{E}}'(\boldsymbol{P}_{j_{i-1}} - \boldsymbol{P}_{j_i})\boldsymbol{\mathcal{E}}|}{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_{n-1} - \boldsymbol{P}_{j_{i-1}})\boldsymbol{\mathcal{E}}|} \\ &= \frac{|\boldsymbol{V}_{i,1} + \boldsymbol{V}_{i,2}|}{|\boldsymbol{V}_{i,1}|}. \end{aligned} \tag{C.5}$$

Since $\boldsymbol{V}_{i,2} \sim W_p(1, \boldsymbol{I}_p)$, we can express $\boldsymbol{V}_{i,2} = \boldsymbol{v}_i\boldsymbol{v}_i'$, where $\boldsymbol{v}_i \sim N_p(\boldsymbol{0}_p, \boldsymbol{I}_p)$ and $\boldsymbol{v}_i$ is independent of $\boldsymbol{V}_{i,1}$. Then, (C.5) is calculated as

$$\begin{aligned}\frac{|\boldsymbol{S}_{yy\cdot j_i}|}{|\boldsymbol{S}_{yy\cdot j_{i-1}}|} &= |\boldsymbol{I}_p + \boldsymbol{V}_{i,1}^{-1}\boldsymbol{V}_{i,2}| \\ &= 1 + \boldsymbol{v}_i'\boldsymbol{V}_{i,1}^{-1}\boldsymbol{v}_i \\ &= 1 + \frac{||\boldsymbol{v}_i||^2}{\left(||\boldsymbol{v}_i||^{-1}\boldsymbol{v}_i'\boldsymbol{V}_{i,1}^{-1}\boldsymbol{v}_i||\boldsymbol{v}_i||^{-1}\right)^{-1}}. \end{aligned} \tag{C.6}$$

Let $\tilde{v}_i = ||\boldsymbol{v}_i||^2$ and $\tilde{u}_i = \left(||\boldsymbol{v}_i||^{-1}\boldsymbol{v}_i'\boldsymbol{V}_{i,1}^{-1}\boldsymbol{v}_i||\boldsymbol{v}_i||^{-1}\right)^{-1}$. Then, from a property of the Wishart distribution (see, e.g., Fujikoshi *et al.*, 2010, Theorem 2.3.3), we see that $\tilde{v}_i$ and $\tilde{u}_i$ are independent, and $\tilde{v}_i \sim \chi^2(p)$ and $\tilde{u}_i \sim \chi^2(n-p-q_j+i-1)$. Then, (C.6) is expressed as

$$\frac{|\boldsymbol{S}_{yy\cdot j_i}|}{|\boldsymbol{S}_{yy\cdot j_{i-1}}|} = \frac{\tilde{v}_i}{\tilde{u}_i}.$$

From Lemma C.3, by applying (4) in Lemma 3.2 to the above equation, for all $\varepsilon > 0$ $(0 < \varepsilon \le 1/2)$, the following equation can be derived:

$$\begin{aligned}\log\frac{|\boldsymbol{S}_{yy\cdot j_i}|}{|\boldsymbol{S}_{yy\cdot j_{i-1}}|} &= \log\left(1 + \frac{p}{n-p-q_j+i-3} + o(p^{1/2+\epsilon}n^{-1})\right) \\ &= \log\frac{n}{n-p} + o(p^{1/2+\varepsilon}n^{-1}),\ a.s.\end{aligned}$$

From the above equation, we have

$$\begin{aligned}\log\frac{|\boldsymbol{S}_{yy\cdot j}|}{|\boldsymbol{S}_{yy\cdot j_*}|} &= -\sum_{i=1}^{q_j-q_*}\log\frac{|\boldsymbol{S}_{yy\cdot j_i}|}{|\boldsymbol{S}_{yy\cdot j_{i-1}}|} \\ &= (q_j - q_*)\log\left(1 - \frac{p}{n}\right) + o(p^{1/2+\varepsilon}n^{-1}),\ a.s. \end{aligned} \tag{C.7}$$

Therefore, from (C.1) and (C.7), we can expand $p^{-1}\{\mathrm{LLIC}(j) - \mathrm{LLIC}(j_*)\}$ as follows:

$$\begin{aligned}&\frac{1}{p}\{\mathrm{LLIC}(j) - \mathrm{LLIC}(j_*)\} \\ &= (q_j - q_*)\frac{n}{p}\log\left(1 - \frac{p}{n}\right) + m(j) - m(j_*) + o(p^{-1/2+\varepsilon}),\ a.s.\end{aligned} \tag{C.8}$$

Next, we consider the case of $j \in \mathcal{J}_-$. By using Lemma C.2, we have

$$\log \frac{|\boldsymbol{S}_{yy\cdot j}|}{|\boldsymbol{S}_{yy\cdot x}|} = \delta_j + \log \frac{|\boldsymbol{U}_1'\boldsymbol{U}_1 + \boldsymbol{U}_2'\boldsymbol{U}_2|}{|\boldsymbol{U}_1'\boldsymbol{U}_1|} + \log \frac{|\boldsymbol{W}_1|}{|\boldsymbol{W}_2|}, \tag{C.9}$$

where $\boldsymbol{U}_1$, $\boldsymbol{U}_2$, $\boldsymbol{W}_1$, and $\boldsymbol{W}_2$ are defined in Lemma C.2. Let

$$\tilde{\boldsymbol{U}} = (\boldsymbol{U}_2\boldsymbol{U}_2')^{1/2}\{\boldsymbol{U}_2(\boldsymbol{U}_1'\boldsymbol{U}_1)^{-1}\boldsymbol{U}_2'\}^{-1}(\boldsymbol{U}_2\boldsymbol{U}_2')^{1/2}.$$

From a property of the Wishart distribution, we observe that $\tilde{\boldsymbol{U}}$ and $\boldsymbol{U}_2$ are independent and $\tilde{\boldsymbol{U}} \sim W_{q-q_j}(n-p-q_j-1, \boldsymbol{I}_{q-q_j})$. Then, (C.9) is expressed as

$$\log \frac{|\boldsymbol{S}_{yy\cdot j}|}{|\boldsymbol{S}_{yy\cdot x}|} = \delta_j + \log|\boldsymbol{I}_{q-q_j} + \tilde{\boldsymbol{U}}^{-1}\boldsymbol{U}_2\boldsymbol{U}_2'| + \log \frac{|\boldsymbol{W}_1|}{|\boldsymbol{W}_2|}. \tag{C.10}$$

By a simple calculation, we can note that $E[||\tilde{\boldsymbol{U}} - E[\tilde{\boldsymbol{U}}]||^4] = O(n^2)$, $E[||\boldsymbol{U}_2\boldsymbol{U}_2' - E[\boldsymbol{U}_2\boldsymbol{U}_2']||^4] = O(p^2)$, and $E[||\boldsymbol{W}_1 - E[\boldsymbol{W}_1]||^4] = O(n^2)$. Hence, we can apply (5) in Lemma 3.2 to $\tilde{\boldsymbol{U}}$, $\boldsymbol{U}_2\boldsymbol{U}_2'$, $\boldsymbol{W}_1$, and $\boldsymbol{W}_2$. From Taylor expansion, for all $\delta > 0$ $(0 < \delta < 1/4)$ the following equations can be derived:

$$\log|\boldsymbol{I}_{q-q_j} + \tilde{\boldsymbol{U}}^{-1}\boldsymbol{U}_2\boldsymbol{U}_2'| = (q-q_j)\log\frac{n}{n-p} + o\left(p^{3/4+\delta}n^{-1}\right) + o\left(pn^{-5/4+\delta}\right), \ a.s., \tag{C.11}$$

$$\log\frac{|\boldsymbol{W}_1|}{|\boldsymbol{W}_2|} = o(n^{-1/4+\delta}), \ a.s. \tag{C.12}$$

From (C.10)-(C.12), we have

$$\log \frac{|\boldsymbol{S}_{yy\cdot j}|}{|\boldsymbol{S}_{yy\cdot x}|} = \delta_j + (q-q_j)\log\frac{n}{n-p} + o(1), \ a.s. \tag{C.13}$$

Therefore, from (C.7) and (C.13), we can expand $n^{-1}\{\mathrm{LLIC}(j) - \mathrm{LLIC}(j_*)\}$ as follows:

$$\begin{aligned}
&\frac{1}{n}\{\mathrm{LLIC}(j) - \mathrm{LLIC}(j_*)\} \\
&= \frac{1}{n}\left\{(n-1)\log\frac{|\boldsymbol{S}_{yy\cdot j}|}{|\boldsymbol{S}_{yy\cdot x}|} + (n-1)\log\frac{|\boldsymbol{S}_{yy\cdot x}|}{|\boldsymbol{S}_{yy\cdot j_*}|} + m(j) - m(j_*)\right\} \\
&= \left(1-\frac{1}{n}\right)\delta_j + (q_j - q_*)\log\left(1-\frac{p}{n}\right) + \frac{1}{n}\{m(j)-m(j_*)\} + o(1), \ a.s.
\end{aligned} \tag{C.14}$$

Lemma 3.1, (C.8), and (C.14) complete the proof of Theorem 3.1.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory* (eds. B. N. Petrov & F. Csáki), pp. 995–1010. Akadémiai Kiadó, Budapest.

[2] Akaike, H. (1974). A new look at the statistical model identification. *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control* **AC − 19**, 716–723.

[3] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, **52**, 345–370.

[4] Fukui, K. (2015). Consistency of log-likelihood-based information criteria for selecting variables in high-dimensional canonical correlation analysis under nonnormality. *Hiroshima Math. J.*, **45**, 175–205.

[5] Fujikoshi, Y. (1982). A test for additional information in canonical correlation analysis. *Ann. Inst. Statist. Math.*, **34**, 523–530.

[6] Fujikoshi, Y. (1985). Selection of variables in discriminant analysis and canonical correlation analysis. In *Multivariate Analysis VI* (ed. P. R. Krishnaiah), 219-236, North-Holland, Amsterdam.

[7] Fujikoshi, Y., Shimizu, R. & Ulyanov, V. V. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations.* John Wiley & Sons, Inc., Hoboken, New Jersey.

[8] Hannan, E. J. & Quinn, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc. Ser. B*, **26**, 270–273.

[9] McKay, R. J. (1977). Variable selection in multivariate regression: an application of simultaneous test procedures. *J. Roy. Statist. Soc., Ser. B*, **39**, 371–380.

[10] Nishii, R., Bai, Z. D. & Krishnaiah, P. R. (1988). Strong consistency information criterion for model selection in multivariate analysis. *Hiroshima Math. J.*, **18**, 451–462.

[11] Oda, R. & Yanagihara, H. (2019). A fast and consistent variable selection method for high-dimensional multivariate linear regression with a large number of explanatory variables. TR No. 19–1, *Statistical Research Group*, Hiroshima University.

[12] Ogura, T. (2010). A variable selection method in principal canonical correlation analysis. *Comput. Statist. Data Anal.*, **54**, 1117–1123.

[13] Srivastava, M. S. (2002). *Methods of Multivariate Statistics.* John Wiley & Sons, New York.

[14] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.

[15] Timm, N. H. (2002). *Applied Multivariate Analysis.* Springer-Verlag, New York.

[16] Yanagihara, H., Oda, R., Hashiyama, Y. & Fujikoshi, Y. (2017). High-Dimensional asymptotic behaviors of differences between the log-determinants of two Wishart matrices. *J. Multivariate Anal.*, **157**, 70–86.