TR-No. 14-10, Hiroshima Statistical Research Group, 1–19

High-Dimensional Asymptotic Behaviors of Differences between the Log-Determinants of Two Wishart Matrices

Hirokazu Yanagihara[1,2,3*], Yusuke Hashiyama[1] and Yasunori Fujikoshi[1,2]

[1]Department of Mathematics, Graduate School of Science, Hiroshima University
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8626, Japan
[2]Hiroshima University Statistical Science Research Core
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8626, Japan
[3]Tokyo Kantei Co., Ltd., 2-24-15 Kami-Osaki, Shinagawa-ku, Tokyo 141-0012, Japan

(Last Modified: December 31, 2014)

Abstract

In this paper, we evaluate the asymptotic behaviors of the differences between the log-determinants of two random matrices distributed according to the Wishart distribution by using a high-dimensional asymptotic framework in which the sizes of the matrices and the degrees of freedoms approach ∞ simultaneously. We consider two structures of random matrices: a matrix is completely included in another matrix, and a matrix is partially included in another matrix. As an application of our result, we derive the condition needed to ensure consistency for a given log-likelihood-based information criterion for selecting variables in a canonical correlation analysis.

Key words: Canonical correlation analysis, Consistency of information criterion, High-dimensional asymptotic framework, Information criterion, Model selection.

*Corresponding author
E-mail address: yanagi@math.sci.hiroshima-u.ac.jp (Hirokazu Yanagihara)

1. Introduction

Let $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ be $p \times p$ symmetric random matrices distributed according to the Wishart distribution (Wishart matrices). In this paper, we study the asymptotic behavior of the difference between the log-determinants of two Wishart matrices, i.e.,

$$\log|\boldsymbol{W}_2| - \log|\boldsymbol{W}_1| = \log\frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|}. \tag{1}$$

The difference between the log-determinants of two Wishart matrices plays a key role in multivariate analysis, because many statistics in multivariate analysis (e.g., the log-likelihood ratio statistic under the normality assumption) can be expressed as a difference (see, e.g., Muirhead, 1982; Siotani *et al.*, 1985; Anderson, 2003). Hence, to prove the asymptotic behavior of $\log|\boldsymbol{W}_2|/|\boldsymbol{W}_1|$ is of major interest in multivariate analysis. A common approach to the study of asymptotic behavior is to use a large sample (LS) asymptotic framework such that the sample size n, which is also the number of degrees of freedom of the Wishart matrix, approaches ∞. Since high-dimensional data analysis has been attracting the attention of many researchers in recent years, it is important to study the asymptotic behavior in terms of the following high-dimensional (HD) asymptotic framework.

- HD asymptotic framework: $n - p$ and p/n are approaching ∞ and $c \in [0, 1)$, respectively. It should be emphasized that the ordinary LS asymptotic framework is included in the HD asymptotic framework as a special case.

An aim of this paper is to evaluate the asymptotic behavior of $\log |\boldsymbol{W}_2|/|\boldsymbol{W}_1|$, using the HD asymptotic framework. Let $\boldsymbol{O}_{d,p}$ be a $d \times p$ matrix of zeros. We will consider the following structures for Wishart matrices:

Case 1. The $\boldsymbol{W}_2$ includes $\boldsymbol{W}_1$ completely, i.e.,

$$\boldsymbol{W}_1 \sim W_p(n-k, \boldsymbol{I}_p),\ \boldsymbol{W}_2 = \boldsymbol{W}_1 + \boldsymbol{Z}'\boldsymbol{Z} \sim W_p(n-k+d, \boldsymbol{I}_p), \tag{2}$$

where k and d are positive integers independent of n and p, and $\boldsymbol{W}_1$ and $\boldsymbol{Z}$ are independent random matrices defined by

$$\boldsymbol{Z} \sim N_{d\times p}(\boldsymbol{O}_{d,p}, \boldsymbol{I}_p \otimes \boldsymbol{I}_d).$$

Case 2. The $\boldsymbol{W}_2$ includes $\boldsymbol{W}_2$ partially, i.e.,

$$\boldsymbol{W}_1 = \boldsymbol{U}'\boldsymbol{U} \sim W_p(n-k, \boldsymbol{I}_p),\ \boldsymbol{W}_2 = (\boldsymbol{Z}\boldsymbol{\Gamma}' + \boldsymbol{U})'(\boldsymbol{Z}\boldsymbol{\Gamma}' + \boldsymbol{U}) \sim W_p(n-k, \boldsymbol{I}_p + \boldsymbol{\Gamma}\boldsymbol{\Gamma}'), \tag{3}$$

where $\boldsymbol{\Gamma}$ is a $p \times r$ constant matrix, k and r are positive integers independent of n and p, and $\boldsymbol{U}$ and $\boldsymbol{Z}$ are independent random matrices defined by

$$\boldsymbol{U} \sim N_{(n-k)\times p}(\boldsymbol{O}_{n-k,p}, \boldsymbol{I}_p \otimes \boldsymbol{I}_{n-k}),\ \boldsymbol{Z} \sim N_{(n-k)\times r}(\boldsymbol{O}_{n-k,r}, \boldsymbol{I}_r \otimes \boldsymbol{I}_{n-k}).$$

Cases 1 and 2 correspond to the log-likelihood ratio statistics of models with an inclusion relation and without an inclusion relation, respectively.

As an application of our result, we derive the condition for ensuring the consistency of a log-likelihood-based information criterion (LLBIC) for selecting variables in a canonical correlation analysis (CCA) that analyzes the correlation of two linearly combined variables; note that this is an important method in multivariate analysis. An optimized solution of can be found by solving an eigenvalue problem. CCA has been introduced in many textbooks on applied statistical analysis (see, e.g., Srivastava, 2002, chap. 14.7; Timm, 2002, chap. 8.7), and it is widely used in many applied fields (e.g., Doeswijk *et al.*, 2011; Khalil *et al.*, 2011; and Vahedia, 2011). The family of LLBICs includes many famous information criteria, e.g., Akaike's information criterion (AIC), the bias-corrected AIC ($\mathrm{AIC_c}$), Takeuchi's information criterion (TIC), the Bayesian information criterion (BIC), the consistent AIC (CAIC), and the Hannan and Quinn information criterion (HQC). Under a general model, the AIC, $\mathrm{AIC_c}$, TIC, BIC, CAIC, and HQC were proposed by Akaike (1973;

1974), Hurvich and Tsai (1989), Takeuchi (1976), Schwarz (1978), Bozdogan (1987), and Hannan and Quinn (1979), respectively. The AIC and $\mathrm{AIC_c}$ for selecting variables in CCA were proposed by Fujikoshi (1985), and the TIC for selecting variables in CCA was proposed by Hashiyama *et al.* (2014). By using the AIC for CCA and the definitions of the original information criteria, we formulate the BIC, CAIC, and HQC for selecting variables in CCA. In this paper, if the asymptotic probability that an information criterion selects the true model approaches 1, then we say that the information criterion is consistent. Under the HD asymptotic framework, Yanagihara *et al.* (2012) and Fujikoshi *et al.* (2014) studied the consistency of an information criterion in a multivariate linear regression model. For CCA, there are no results in the literature on the consistency of an information criterion under the HD asymptotic framework, although several authors (e.g., Nishii *et al.*, 1988) have studied consistency under the LS asymptotic framework.

This paper is organized as follows: In Section 2, we present our main results. In Section 3, we show a condition for ensuring the consistency of the LLBIC for CCA. Technical details are provided in the Appendix.

2. Main Results

In this section, we evaluate the asymptotic behavior of $\log |\boldsymbol{W}_2|/|\boldsymbol{W}_1|$ in (1) within the HD asymptotic framework. We begin by considering case 1, and we have the following theorem (the proof is given in Appendix A):

Theorem 1 *Suppose that $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are Wishart matrices given by (2). Then, we have*

$$\log \frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|} = -d \log(1 - p/n) + \frac{\sqrt{p}}{n}\left(1 - \frac{p}{n}\right)\mathrm{tr}(\boldsymbol{V}) + O_p(pn^{-2}), \tag{4}$$

where $\boldsymbol{V}$ is a $d \times d$ random matrix with the order $O_p(1)$, which is given by

$$\boldsymbol{V} = \frac{n}{\sqrt{p}}\left(\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}' - \frac{p}{n-k-p-1}\boldsymbol{I}_d\right), \tag{5}$$

and $\mathrm{tr}(\boldsymbol{V})$ is asymptotically distributed as

$$\mathrm{tr}(\boldsymbol{V}) \xrightarrow{d} \begin{cases} (\chi^2_{dp} - dp)/\sqrt{p} & (p \text{ is bounded}) \\ N(0, 2d/(1-c)^3) & (p \to \infty) \end{cases}. \tag{6}$$

Wakaki (2006) derived a result similar to Theorem 1 by using a property of Wilks' lambda distribution, whereas we used a property of the Wishart distribution to prove it.

Notice that

$$-\frac{n}{p}\log\left(1 - \frac{p}{n}\right) = 1 + O(pn^{-1}),$$

and

$$\frac{\sqrt{p}}{n}\left(1 - \frac{p}{n}\right)\cdot\frac{1}{\sqrt{p}}(X - dp) = \frac{1}{n}(X - dp) + O_p(p^{3/2}n^{-2}),$$

where X is a random variable distributed according to the chi-square distribution with dp degrees of freedom. Hence, from Theorem 1, the following corollary is obtained:

Corollary 1 *Suppose that $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are Wishart matrices given by (2). Then, we have*

$$\log \frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|} = -d \log(1-c) + o_p(1), \tag{7}$$

and

$$\frac{n}{p} \log \frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|} = R + o_p(1),$$

where R is defined by

$$R = \begin{cases} X/p, \ X \sim \chi^2_{dp} & (p \text{ is bounded}) \\ dg(c) & (p \to \infty) \end{cases}.$$

Here, $g(x)$ is a function with domain $x \in [0,1)$, which is given by

$$g(x) = \begin{cases} 1 & (x = 0) \\ -x^{-1}\log(1-x) & (x \in (0,1)) \end{cases}. \tag{8}$$

From elementary calculus calculations, it turns out that $g(0)$ is defined to be equal to $\lim_{x \to 0+}\{-x^{-1}\log(1-x)\}$ and that $g(x)$ is a strictly monotonically increasing function in $x \in [0,1)$. In this paper, we have derived Corollary 1 through Theorem 1. Although it is necessary to clarify the asymptotic distribution of $\mathrm{tr}(\boldsymbol{V})$ in order to prove Theorem 1, (7) can be derived without the asymptotic distribution of $\mathrm{tr}(\boldsymbol{V})$, using only the expectation of $\mathrm{tr}(\boldsymbol{V})$.

Next, we consider case 2, and we have the following theorem (the proof is given in Appendix B):

Theorem 2 *Suppose that $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are Wishart matrices given by (3). Then, we have*

$$\log \frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|} = \log |\boldsymbol{I}_p + \boldsymbol{\Gamma}\boldsymbol{\Gamma}'| + o_p(1). \tag{9}$$

3. Application

3.1. Redundancy Model in CCA

In this section, we show an example of an application of Theorems 1 and 2 to an actual statistical problem. We derive conditions to separately ensure consistency and inconsistency of the LLBIC for selecting variables in CCA.

Let $\boldsymbol{z} = (\boldsymbol{x}', \boldsymbol{y}')' = (x_1, \ldots, x_q, y_1, \ldots, y_p)'$ be a $(q+p)$-dimensional random vector distributed according to the $(q+p)$-variates multivariate normal distribution with the following mean vector and covariance matrix:

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

Suppose that j denotes a subset of $\omega = \{1, \ldots, q\}$ containing q_j elements, and $\boldsymbol{x}_j$ denotes the q_j-dimensional 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 $\boldsymbol{x}$. We will also let $\bar{j}$ denote the complement of the set j, i.e., $\bar{j} = j^c$. Of course, it holds that $\boldsymbol{x}_\omega = \boldsymbol{x}$ and $q_\omega = q$. Without loss of generality, we divide $\boldsymbol{x}$ into two subvectors $\boldsymbol{x} = (\boldsymbol{x}'_j, \boldsymbol{x}'_{\bar{j}})'$, where $\boldsymbol{x}_{\bar{j}}$ is a $(q - q_j)$-dimensional random vector. Expressions of $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{xy}$ corresponding to the division of $\boldsymbol{x}$ are

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

These imply that another expression of $\boldsymbol{\Sigma}$ corresponding to the division is

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{jj} & \boldsymbol{\Sigma}_{j\bar{j}} & \boldsymbol{\Sigma}_{jy} \\ \boldsymbol{\Sigma}'_{j\bar{j}} & \boldsymbol{\Sigma}_{\bar{j}\bar{j}} & \boldsymbol{\Sigma}_{\bar{j}y} \\ \boldsymbol{\Sigma}'_{jy} & \boldsymbol{\Sigma}'_{\bar{j}y} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}. \tag{10}$$

Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{n+1}$ be $(n + 1)$ independent random vectors from $\boldsymbol{z}$, and let $\boldsymbol{S}$ be the usual unbiased estimator of $\boldsymbol{\Sigma}$ given by $\boldsymbol{S} = n^{-1} \sum_{i=1}^{n+1}(\boldsymbol{z}_i - \bar{\boldsymbol{z}})(\boldsymbol{z}_i - \bar{\boldsymbol{z}})'$, where $\bar{\boldsymbol{z}} = (n + 1)^{-1} \sum_{i=1}^{n+1} \boldsymbol{z}_i$. Following the same method that we used for $\boldsymbol{\Sigma}$ in (10), we divide $\boldsymbol{S}$ as

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

One of the interests in CCA is to determine whether $\boldsymbol{x}_{\bar{j}}$ is irrelevant. Fujikoshi (1985) determined that $\boldsymbol{x}_{\bar{j}}$ is irrelevant if the following equation holds:

$$\mathrm{tr}(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}'_{xy}) = \mathrm{tr}(\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}'_{jy}). \tag{11}$$

In particular, we note that (11) is equivalent to

$$\boldsymbol{\Sigma}_{\bar{j}y} - \boldsymbol{\Sigma}'_{j\bar{j}}\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy} = \boldsymbol{O}_{q-q_j,p}. \tag{12}$$

Consequently, the candidate model in which $\boldsymbol{x}_{\bar{j}}$ is irrelevant can be expressed as

$$n\boldsymbol{S} \sim W_{p+q}(n, \boldsymbol{\Sigma}) \ s.t. \ \mathrm{tr}(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}'_{xy}) = \mathrm{tr}(\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}'_{jy}). \tag{13}$$

In CCA, the above model is called the redundancy model j or simply the model j. If the model j is selected as the best model, then we can regard $\boldsymbol{x}_{\bar{j}}$ as irrelevant and $\boldsymbol{x}_j$ as relevant.

An estimator of $\boldsymbol{\Sigma}$ in (13) is given by

$$\hat{\boldsymbol{\Sigma}}_j = \arg\min_{\boldsymbol{\Sigma}} \left\{ F(\boldsymbol{S}, \boldsymbol{\Sigma}) \ s.t. \ \mathrm{tr}(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}'_{xy}) = \mathrm{tr}(\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}'_{jy}) \right\},$$

where $F(\boldsymbol{S}, \boldsymbol{\Sigma})$ is given by

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

It is easy to see that $nF(\boldsymbol{S}, \boldsymbol{\Sigma})$ is the Kullback–Leibler (KL) discrepancy function assessed by the Wishart density. In the analysis of covariance structure, the discrepancy function is frequently called the maximum likelihood discrepancy function (Jöreskog, 1967). Although we do not present it here, $\hat{\boldsymbol{\Sigma}}_j$ can be derived in a closed form (see, e.g., Fujikoshi & Kurata, 2008; Fujikoshi *et al.*, 2010, chap. 11.5).

3.2. A Class of Information Criteria for CCA

Let $\mathcal{J}$ be a set of candidate models denoted by $\mathcal{J} = \{j_1, \dots, j_K\}$, where K is the number of models. We separate $\mathcal{J}$ into two sets such that one is a set of overspecified models, and the other is a set of underspecified models. Let $\mathcal{J}_+$ denote the set of overspecified models, which is defined by

$$\mathcal{J}_+ = \left\{ j \in \mathcal{J} \mid \mathrm{tr}(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{xy}') = \mathrm{tr}(\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{jy}') \right\}.$$

Suppose that the true model is expressed as the overspecified model having the smallest number of elements, i.e.,

$$j_* = \arg\min_{j \in \mathcal{J}_+} q_j.$$

For simplicity, we write q_{j_*} as q_*. On the other hand, let $\mathcal{J}_-$ denote the set of underspecified models, which is defined by

$$\mathcal{J}_- = \overline{\mathcal{J}_+} \cap \mathcal{J}.$$

For the general expression for any $j \in \mathcal{J}$, $\boldsymbol{\Sigma}_{yy\cdot j}$ and $\boldsymbol{S}_{yy\cdot j}$ are defined as

$$\boldsymbol{\Sigma}_{yy\cdot j} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{jy}'\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}, \quad \boldsymbol{S}_{yy\cdot j} = \boldsymbol{S}_{yy} - \boldsymbol{S}_{jy}'\boldsymbol{S}_{jj}^{-1}\boldsymbol{S}_{jy}. \tag{14}$$

In particular, we write $\boldsymbol{\Sigma}_{yy\cdot\omega} = \boldsymbol{\Sigma}_{yy\cdot x}$ and $\boldsymbol{S}_{yy\cdot\omega} = \boldsymbol{S}_{yy\cdot x}$. From Fujikoshi *et al.* (2010, chap. 11.5) and the fact that $\mathrm{tr}(\hat{\boldsymbol{\Sigma}}_j\boldsymbol{S}) = p + q$, the minimum value of $F(\boldsymbol{\Sigma}, \boldsymbol{S})$ under the model j is given by

$$F_{\min}(j) = F(\boldsymbol{S}, \hat{\boldsymbol{\Sigma}}_j) = \log \frac{|\boldsymbol{S}_{yy\cdot j}|}{|\boldsymbol{S}_{yy\cdot x}|}.$$

In the model j, various information criteria can be defined by adding a penalty term for the model complexity $m(j)$ to $nF_{\min}(j)$, i.e., several information criteria are included in the following class of information criteria:

$$\mathrm{IC}_m(j) = nF_{\min}(j) + m(j). \tag{15}$$

By changing $m(j)$, (15) can express the following specific criteria:

$$m(j) = \begin{cases} 2\{pq_j + (q^2 + p^2 + q + p)/2\} & (\mathrm{AIC}) \\ n\{b(q) + b(q_j + p) - b(q_j) - (q + p)\} & (\mathrm{AIC_c}) \\ 2\{pq_j + (q^2 + p^2 + q + p)/2\} + \hat{\kappa}_x + \hat{\kappa}_{(jy)} - \hat{\kappa}_j & (\mathrm{TIC}) \\ \{pq_j + (q^2 + p^2 + q + p)/2\}\log n & (\mathrm{BIC}) \\ \{pq_j + (q^2 + p^2 + q + p)/2\}(\log n + 1) & (\mathrm{CAIC}) \\ 2\{pq_j + (q^2 + p^2 + q + p)/2\}\log\log n & (\mathrm{HQC}) \end{cases}, \tag{16}$$

where $b(q)$ is a function of q defined by $nq/(n - q - 1)$, and $\hat{\kappa}_x$, $\hat{\kappa}_j$, and $\hat{\kappa}_{(jy)}$ are estimators of multivariate kurtoses of $\boldsymbol{x}$, $\boldsymbol{x}_j$, and $(\boldsymbol{x}_j'\boldsymbol{y}')'$, respectively, which are defined by

$$\hat{\kappa}_x = \hat{\kappa}(\boldsymbol{D}_x),\ \hat{\kappa}_j = \hat{\kappa}(\boldsymbol{D}_j),\ \hat{\kappa}_{(j,y)} = \hat{\kappa}(\boldsymbol{D}_{(j,y)}).$$

Here $\hat{\kappa}(\boldsymbol{D})$ is an estimator of multivariate kurtosis of the d-variates extracted from $\boldsymbol{z}$ by $\boldsymbol{D}'\boldsymbol{z}$, which is defined by

$$\hat{\kappa}(\boldsymbol{D}) = \frac{1}{n+1}\sum_{i=1}^{n+1}\left\{(\boldsymbol{z}_i - \bar{\boldsymbol{z}})'\boldsymbol{D}(\boldsymbol{D}'\boldsymbol{S}\boldsymbol{D})^{-1}\boldsymbol{D}'(\boldsymbol{z}_i - \bar{\boldsymbol{z}})\right\}^2 - d(d+2), \tag{17}$$

where $\boldsymbol{D}$ is a $(q+p)\times d$ matrix whose elements are 0 or 1, and satisfies $\boldsymbol{D}'\boldsymbol{D} = \boldsymbol{I}_d$, e.g.,

$$\boldsymbol{D}_x = \begin{pmatrix} \boldsymbol{I}_q \\ \boldsymbol{O}_{p,q} \end{pmatrix},\ \boldsymbol{D}_y = \begin{pmatrix} \boldsymbol{O}_{q,p} \\ \boldsymbol{I}_p \end{pmatrix},\ \boldsymbol{D}_j = \begin{pmatrix} \boldsymbol{I}_{q_j} \\ \boldsymbol{O}_{p+q-q_j,q_j} \end{pmatrix},\ \boldsymbol{D}_{(j,y)} = (\boldsymbol{D}_j, \boldsymbol{D}_y).$$

Additionally, we assume that there exists a nonzero function $a_m(n,p)$ such that

$$\beta_m(j) = \operatorname*{plim}_{n-p\to\infty, p/n\to c} \frac{m(j) - m(j_*)}{a_m(n,p)}, \quad 0 < |\beta_m(j)| < \infty. \tag{18}$$

We regard as the best model the candidate model that makes IC_m the smallest, i.e., the best model chosen by IC_m can be expressed as

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

Let α_m be an asymptotic probability such that $P(\hat{j}_m = j_*) \to \alpha_m$ as $n-p\to\infty$ and $p/n\to c$. The IC_m is consistent if $\alpha_m = 1$, and it is inconsistent if $\alpha_m < 1$.

3.3. Conditions for Consistency in CCA

We begin with the following lemma (the proof is given in Appendix C):

Lemma 1 *The following equations are derived:*

1. *Let*

$$\boldsymbol{W}_1 = \boldsymbol{W}, \quad \boldsymbol{W}_2 = \boldsymbol{W} + \boldsymbol{Z}'\boldsymbol{Z},$$

where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are independent random matrices defined by

$$\boldsymbol{W} \sim W_p(n-q_j, \boldsymbol{I}_p), \quad \boldsymbol{Z} \sim N_{(q_j-q_*)\times p}(\boldsymbol{O}_{q_j-q_*,p}, \boldsymbol{I}_p \otimes \boldsymbol{I}_{q_j-q_*}).$$

Then, for any $j \in \mathcal{J}_+\backslash\{j_*\}$, $F_{\min}(j) - F_{\min}(j_*)$ can be rewritten as

$$F_{\min}(j) - F_{\min}(j_*) = -\log\frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|}. \tag{19}$$

2. *Let*

$$\boldsymbol{W}_1 = \boldsymbol{U}'\boldsymbol{U} = \boldsymbol{W}_3 + \boldsymbol{U}_2'\boldsymbol{U}_2,\ \boldsymbol{W}_2 = (\boldsymbol{Z}\boldsymbol{\Gamma}_j' + \boldsymbol{U})'(\boldsymbol{Z}\boldsymbol{\Gamma}_j' + \boldsymbol{U}),\ \boldsymbol{W}_3 = \boldsymbol{U}_1'\boldsymbol{U}_1,$$

where $\boldsymbol{U} = (\boldsymbol{U}_1', \boldsymbol{U}_2')'$, $\boldsymbol{U}_1$, $\boldsymbol{U}_2$, and $\boldsymbol{Z}$ are mutually independent random matrices defined by

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

and $\mathbf{\Gamma}_j$ *is a* $p \times (q - q_j)$ *nonrandom matrix defined by*

$$\mathbf{\Gamma}_j = \mathbf{\Sigma}_{yy\cdot x}^{-1/2}(\mathbf{\Sigma}_{\bar{j}y} - \mathbf{\Sigma}'_{j\bar{j}}\mathbf{\Sigma}_{jj}^{-1}\mathbf{\Sigma}_{jy})'\mathbf{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1/2}. \tag{20}$$

Here $\mathbf{\Sigma}_{yy\cdot x}$ *is given by* (14)*, and* $\mathbf{\Sigma}_{\bar{j}\bar{j}\cdot j}$ *is defined by*

$$\mathbf{\Sigma}_{\bar{j}\bar{j}\cdot j} = \mathbf{\Sigma}_{\bar{j}\bar{j}} - \mathbf{\Sigma}'_{j\bar{j}}\mathbf{\Sigma}_{jj}^{-1}\mathbf{\Sigma}_{j\bar{j}}. \tag{21}$$

Then, for any $j \in \mathcal{J}_-$*,* $F_{\min}(j) - F_{\min}(\omega)$ *can be rewritten as*

$$F_{\min}(j) - F_{\min}(\omega) = \log\frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|} + \log\frac{|\boldsymbol{W}_1|}{|\boldsymbol{W}_3|}. \tag{22}$$

Let $\delta_j = \log|\boldsymbol{I}_p + \mathbf{\Gamma}_j\mathbf{\Gamma}'_j|$. It is known that $\delta_j \geq 0$ with equality if and only if $j \in \mathcal{J}_+$, because $\mathbf{\Gamma}_j = \boldsymbol{O}_{p,q-q_j}$ holds if and only if $j \in \mathcal{J}_+$. Moreover, δ_j can be rewritten as

$$\delta_j = \log|\boldsymbol{I}_p + \mathbf{\Gamma}_j\mathbf{\Gamma}'_j| = \log\frac{|\mathbf{\Sigma}_{yy\cdot j}|}{|\mathbf{\Sigma}_{yy\cdot x}|}, \tag{23}$$

(the proof is given in Appendix D). Notice that δ_j depends on p not n. It should be kept in mind that sometimes $\lim_{p\to\infty}\delta_j$ becomes infinite and other times it is finite (see examples in Appendix E). The size of a convergent value and the order of δ_j play an important role in deciding whether a criterion is consistent. In addition, we assume that $\lim_{p\to\infty}\delta_j > 0$ for $j \in \mathcal{J}_-$.

When $j \in \mathcal{J}_+\backslash\{j_*\}$, it follows from Corollary 1 and Lemma 1 that $F_{\min}(j) - F_{\min}(j_*) = o_p(1)$. Notice that

$$F_{\min}(j) - F_{\min}(j_*) = F_{\min}(j) - F_{\min}(\omega) + F_{\min}(\omega) - F_{\min}(j_*).$$

Since $\omega \in \mathcal{J}_+$, $F_{\min}(\omega) - F_{\min}(j_*) = o_p(1)$ holds. By using this result, Theorem 2, and Lemma 1, we have

$$F_{\min}(j) - F_{\min}(j_*) = \delta_j + o_p(1) \quad ({}^\forall j \in \mathcal{J}_-). \tag{24}$$

Let us define R_j by

$$R_j = \begin{cases} X_j/p,\ X_j \sim \chi^2_{(q_j-q_*)p} & (p \textit{ is bounded}) \\ (q_j - q_*)g(c) & (p \to \infty) \end{cases}, \tag{25}$$

where $g(x)$ is the function given by (8). By applying Corollary 1 to the case of CCA through Lemma 1, we derive

$$\frac{n}{p}\{F_{\min}(j) - F_{\min}(j_*)\} = -R_j + o_p(1) \quad ({}^\forall j \in \mathcal{J}_+\backslash\{j_*\}). \tag{26}$$

From Lemma A.3 in Yanagihara (2013), we can see that the LLBIC is consistent if the following equations hold:

$$\operatorname*{plim}_{n-p\to\infty, p/n\to c}\frac{1}{p}\{\mathrm{IC}_m(j) - \mathrm{IC}_m(j_*)\} > 0 \quad ({}^\forall j \in \mathcal{J}_+\backslash\{j_*\}),$$
$$\operatorname*{plim}_{n-p\to\infty, p/n\to c}\frac{1}{n}\{\mathrm{IC}_m(j) - \mathrm{IC}_m(j_*)\} > 0 \quad ({}^\forall j \in \mathcal{J}_-).$$

Besides, if there is a model j such that either of the above two inequities is not satisfied, the LLBIC is not consistent. Recall that $\mathrm{IC}_m(j) = nF_{\min}(j) + m(j)$. Hence, from the above equations, (24), and (26), conditions for consistency are obtained as in the following theorem:

Theorem 3 *The IC_m is consistent when $n-p \to \infty$ and $p/n \to c \in [0,1)$ if the following conditions are satisfied simultaneously:*

C1. For any $j \in \mathcal{J}_+ \backslash \{j_*\}$,

$$R_j \left(\lim_{n-p\to\infty, p/n\to c} \frac{p}{a_m(n,p)} \right) < \beta_m(j), \tag{27}$$

where R_j is given by (25), and $a_m(n,p)$ and $\beta_m(j)$ are given in (18).

C2. For any $j \in \mathcal{J}_-$,

$$\lim_{n-p\to\infty, p/n\to c} \frac{n\delta_j}{a_m(n,p)} > -\beta_m(j), \tag{28}$$

where δ_j is given by (23).

If either of the above two conditions is not satisfied, the IC_m is not consistent when $n-p \to \infty$ and $p/n \to c \in [0,1)$.

If p is bounded, $P(|R_j| > \epsilon) \neq 0 \ {}^\forall\epsilon > 0$, because pR_j is a positive random variable distributed according to the chi-square distribution. Hence, when p is bounded, the condition C1 is equivalent to $\lim_{n\to\infty} a_m(n,p) = \infty$.

Although a condition for consistency has been derived, we still do not know which criteria satisfy that condition. Therefore, the conditions for the consistency of specific criteria in (16) are clarified in the following corollary (the proof is given in Appendix F):

Corollary 2 *Let*

$$\mathcal{S}_- = \{j \in \mathcal{J}_- | q_* - q_j > 0\}, \quad c_{\mathrm{a}} = \arg\operatorname*{solve}_{x\in[0,1)}\{-g(x) + 2 = 0\} \approx 0.797. \tag{29}$$

Necessary and sufficient conditions for the consistency of specific criteria are as follows:

- *AIC & TIC*: $p \to \infty$, $c \in [0, c_{\mathrm{a}})$, *and for any* $j \in \mathcal{S}_-$, $\displaystyle\lim_{n-p\to\infty, p/n\to c} \frac{n\delta_j}{2p(q_* - q_j)} > 1.$
- *AIC$_c$*: $p \to \infty$, *and for any* $j \in \mathcal{S}_-$, $\displaystyle\lim_{n-p\to\infty, p/n\to c} \frac{n\delta_j(1-c)^2}{p(q_* - q_j)(2-c)} > 1.$
- *BIC & CAIC*: *for any* $j \in \mathcal{S}_-$, $\displaystyle\lim_{n-p\to\infty, p/n\to c} \frac{n\delta_j}{p(q_* - q_j)\log n} > 1.$
- *HQC*: *for any* $j \in \mathcal{S}_-$, $\displaystyle\lim_{n-p\to\infty, p/n\to c} \frac{n\delta_j}{2p(q_* - q_j)\log\log n} > 1.$

Acknowledgment The first author's research was partially supported by the Ministry of Education, Science, Sports, and Culture, and a Grant-in-Aid for Challenging Exploratory Research, #25540012, 2013–2015. The third author's research was partially supported by the Ministry of Education, Science, Sports, and Culture, a Grant-in-Aid for Scientific Research (C), #25330038, 2013–2015.

References

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

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automatic Control*, **AC-19**, 716–723.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd. ed.). John Wiley & Sons, Hoboken, New Kersey.

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

Doeswijk, T. G., Hageman, J. A., Westerhuis, J. A., Tikunov, Y., Bovy. A. & van Eeuwijk, F. A. (2011). Canonical correlation analysis of multiple sensory directed metabolomics data blocks reveals corresponding parts between data blocks. *Chemometr. Intell. Lab.*, **107**, 371–376.

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.

Fujikoshi, Y. & Kurata, H. (2008). Information criterion for some independence structures. In *New Trends in Psychometrics* (Eds. K. Shigemasu, A. Okada, T. Imaizumi & T. Hoshino), 69–78, Universal Academy Press, Tokyo.

Fujikoshi, Y., Sakurai, T. & Yanagihara, H. (2014). Consistency of high-dimensional AIC-type and C_p-type criteria in multivariate linear regression. *J. Multivariate Anal.*, **123**, 184–200.

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

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *J. Roy. Statist. Soc. Ser.* **B**, **41**, 190–195.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, New York.

Hashiyama, Y., Yanagihara, H. & Fujikoshi, Y. (2014). Jackknife bias correction of the AIC for selecting variables in canonical correlation analysis under model misspecification. *Linear Algebra Appl.*, **455**, 82–106.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and times series model selection in small samples. *Biometrika*, **50**, 226–231.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32**, 443–482.

Khalil, B., Ouarda, T. B. M. J. & St-Hilaire, A. (2011). Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *J. Hydrol.*, **405**, 277–287.

Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā Ser.* **B**, **36**, 115–128.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Inc., New York.

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

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

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

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

Srivastava, M. S. & Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*. North Holland, New York.

Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Math. Sci.*, **153**, 12–18 (in Japanese).

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

Vahedia, S. (2011). Canonical correlation analysis of procrastination, learning strategies and statistics anxiety among Iranian female college students. *Procedia Soc. Behav. Sci.*, **30**, 1620–1624.

Wakaki, H. (2006). Edgeworth expansion of Wilks' lambda statistic. *J. Multivariate Anal.*, **97**, 1958–1964.

Yanagihara, H. (2013). Conditions for consistency of a log-likelihood-based information criterion in normal multivariate linear regression models under the violation of normality assumption. TR No. 13-11, *Statistical Research Group, Hiroshima University*.

Yanagihara, H., Wakaki, H. & Fujikoshi, Y. (2012). A consistency property of the AIC for multivariate linear models when the dimension and the sample size are large. TR No. 12-08, *Statistical Research Group, Hiroshima University*.

Appendix

A. Proof of Theorem 1

It follows from a basic property of a determinant (see, e.g., Harville, 1997, cor. 18.1.2) that

$$\log \frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|} = \log \left|\boldsymbol{I}_p + \boldsymbol{W}_1^{-1/2} \boldsymbol{Z}' \boldsymbol{Z} \boldsymbol{W}^{-1/2}\right| = \log |\boldsymbol{I}_d + \boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}'|. \tag{A.1}$$

Notice that

$$E[\boldsymbol{Z}\boldsymbol{Z}'] = p\boldsymbol{I}_d,\ E\left[\mathrm{tr}\left\{(\boldsymbol{Z}\boldsymbol{Z}')^2\right\}\right] = d(p+d+1)p,\ E\left[\{\mathrm{tr}(\boldsymbol{Z}\boldsymbol{Z}')\}^2\right] = d(dp+2)p.$$

By using the above results, the assumption that $\boldsymbol{Z}$ and $\boldsymbol{W}_1$ are independent, and th. 2.2.8 in Fujikoshi *et al.* (2010), we derive

$$\begin{aligned} E\left[\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}'\right] &= \frac{1}{h_1} E[\boldsymbol{Z}\boldsymbol{Z}'] = \frac{p}{h_1}\boldsymbol{I}_d, \\ E\left[\mathrm{tr}\left\{(\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}')^2\right\}\right] &= \frac{1}{h_0 h_3} E\left[\mathrm{tr}\left\{(\boldsymbol{Z}\boldsymbol{Z}')^2\right\}\right] + \frac{1}{h_0 h_1 h_3} E\left[\{\mathrm{tr}(\boldsymbol{Z}\boldsymbol{Z}')\}^2\right] \\ &= \frac{d(p+d+1)p}{h_0 h_3} + \frac{d(dp+2)p}{h_0 h_1 h_3}, \end{aligned}$$

where $h_i = n - k - p - i$. Hence, the following equation is obtained:

$$E\left[\left\|\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}' - (p/h_1)\boldsymbol{I}_d\right\|^2\right] = \frac{d(d+1)p}{h_0 h_3} + \frac{d\{(d+1)p+2\}p}{h_0 h_1 h_3} + \frac{2dp^2}{h_0 h_1^2 h_3} = O(pn^{-2}).$$

The above equation implies that $\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}' = (p/h_1)\boldsymbol{I}_d + O_p(p^{1/2}n^{-1})$. Hence, $\boldsymbol{V} = O_p(1)$ is derived, where $\boldsymbol{V}$ is given by (5). Moreover, $(1 + p/h_1)^{-1} = O(1)$, because

$$0 < \left(1 + \frac{p}{h_1}\right)^{-1} = \frac{n-k-p-1}{n-k-1} < 1.$$

Substituting (5) into (A.1) yields

$$\begin{aligned} \log \frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|} &= \log |(1+p/h_1)\boldsymbol{I}_d| \left|\boldsymbol{I}_d + \frac{\sqrt{p}}{(1+p/h_1)n}\boldsymbol{V}\right| \\ &= -d\log\left(1 - \frac{p}{n-k-1}\right) + \frac{\sqrt{p}(n-k-p-1)}{n(n-k-1)}\mathrm{tr}(\boldsymbol{V}) + O_p(pn^{-2}). \end{aligned} \tag{A.2}$$

Notice that

$$\begin{aligned} \log\left(1 - \frac{p}{n-k-1}\right) &= \log(1-p/n) + O(pn^{-2}), \\ \frac{\sqrt{p}(n-k-p-1)}{n(n-k-1)} &= \frac{\sqrt{p}}{n}\left(1 - \frac{p}{n}\right) + O(p^{3/2}n^{-3}). \end{aligned}$$

From the above equations and (A.2), (4) is proved.

Next, we prove (6). Notice that for sufficiently large p,

$$\begin{aligned}\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}' &= \left\{(\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}')^{-1}\right\}^{-1}\\ &= (\boldsymbol{Z}\boldsymbol{Z}')^{1/2}\left\{(\boldsymbol{Z}\boldsymbol{Z}')^{1/2}(\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}')^{-1}(\boldsymbol{Z}\boldsymbol{Z}')^{1/2}\right\}^{-1}(\boldsymbol{Z}\boldsymbol{Z}')^{1/2}.\end{aligned}$$

Hence, we have

$$\mathrm{tr}(\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}') = \mathrm{tr}(\boldsymbol{U}_1\boldsymbol{U}_2^{-1}),$$

where $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are $d \times d$ random matrices defined by

$$\boldsymbol{U}_1 = \boldsymbol{Z}\boldsymbol{Z}', \quad \boldsymbol{U}_2 = (\boldsymbol{Z}\boldsymbol{Z}')^{1/2}(\boldsymbol{Z}\boldsymbol{W}_1^{-1}\boldsymbol{Z}')^{-1}(\boldsymbol{Z}\boldsymbol{Z}')^{1/2}.$$

From a basic property of a Wishart distribution and th. 2.3.3 in Fujikoshi *et al.* (2010), we can see that $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are independent, and

$$\boldsymbol{U}_1 \sim W_d(p, \boldsymbol{I}_d), \quad \boldsymbol{U}_2 \sim W_d(n-k-p+d, \boldsymbol{I}_d). \tag{A.3}$$

Let

$$\boldsymbol{V}_1 = \frac{1}{\sqrt{p}}(\boldsymbol{U}_1 - p\boldsymbol{I}_d), \ \boldsymbol{V}_2 = \frac{1}{\sqrt{h_0+d}}\{\boldsymbol{U}_2 - (h_0+d)\boldsymbol{I}_d\}. \tag{A.4}$$

Then, it is easy to see that $\boldsymbol{V}_1 = O_p(1)$ and $\boldsymbol{V}_2 = O_p(1)$, and

$$\mathrm{tr}(\boldsymbol{V}_1) \xrightarrow{d} N(0, 2d) \text{ as } p \to \infty, \quad \mathrm{tr}(\boldsymbol{V}_2) \xrightarrow{d} N(0, 2d) \text{ as } (n-p) \to \infty. \tag{A.5}$$

Notice that

$$\begin{aligned}&\frac{n}{h_0+d} = \frac{n}{n-p} + O(n^{-2}), \quad \sqrt{\frac{p}{h_0+d}} = \sqrt{\frac{p}{n-p}} + O(p^{1/2}n^{-3/2}),\\ &\frac{pd}{h_0+d} - \frac{pd}{h_1} = O(pn^{-2}).\end{aligned}$$

By using the above equations, $\boldsymbol{V}_1$, and $\boldsymbol{V}_2$, $\boldsymbol{V}$ can be expanded as

$$\begin{aligned}\mathrm{tr}(\boldsymbol{V}) &= \frac{n}{\sqrt{p}}\left\{\mathrm{tr}(\boldsymbol{U}_1\boldsymbol{U}_2^{-1}) - \frac{dp}{h_1}\right\}\\ &= \frac{n}{\sqrt{p}}\left[\frac{p}{h_0+d}\mathrm{tr}\left\{\left(\boldsymbol{I}_d + \frac{1}{\sqrt{p}}\boldsymbol{V}_1\right)\left(\boldsymbol{I}_d + \frac{1}{\sqrt{h_0+d}}\boldsymbol{V}_2\right)^{-1}\right\} - \frac{dp}{h_1}\right]\\ &= \left(\frac{n}{n-p}\right)\left\{\mathrm{tr}(\boldsymbol{V}_1) - \sqrt{\frac{p}{n-p}}\mathrm{tr}(\boldsymbol{V}_2)\right\} + O_p(p^{1/2}n^{-1}).\end{aligned} \tag{A.6}$$

Thus, from (A.3), (A.4), (A.5), and (A.6), (6) is proved.

B. Proof of Theorem 2

Let $\boldsymbol{H}(\boldsymbol{\Lambda}, \boldsymbol{O}_{r,p-r})'\boldsymbol{Q}'$ be a singular value decomposition of $\boldsymbol{\Gamma}$, where $\boldsymbol{H}$ is a pth orthogonal matrix satisfying $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_p$, $\boldsymbol{Q}$ is an rth orthogonal matrix satisfying $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{I}_r$, and $\boldsymbol{\Lambda}$ is an rth diagonal matrix whose ath diagonal element is a singular value λ_a, i.e., $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ $(0 \le \lambda_1 \le \cdots \le \lambda_r)$. Then, it is easy to see that $\boldsymbol{V} = \boldsymbol{U}\boldsymbol{H}$ and $\boldsymbol{B} = \boldsymbol{Z}\boldsymbol{Q}$ are independent, and

$$\boldsymbol{V} \sim N_{(n-k)\times p}(\boldsymbol{O}_{n-k,p}, \boldsymbol{I}_p \otimes \boldsymbol{I}_{n-k}), \quad \boldsymbol{B} \sim N_{(n-k)\times r}(\boldsymbol{O}_{n-k,r}, \boldsymbol{I}_r \otimes \boldsymbol{I}_{n-k}).$$

Let us partition $\boldsymbol{V}$ as $\boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2)$, where $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are $(n-k)\times r$ and $(n-k)\times(p-r)$ matrices, respectively. Then, we have

$$\begin{aligned} \boldsymbol{U}'\boldsymbol{U} &= \boldsymbol{H}\begin{pmatrix} \boldsymbol{V}_1'\boldsymbol{V}_1 & \boldsymbol{V}_1'\boldsymbol{V}_2 \\ \boldsymbol{V}_2'\boldsymbol{V}_1 & \boldsymbol{V}_2'\boldsymbol{V}_2 \end{pmatrix}\boldsymbol{H}', \\ (\boldsymbol{Z}\boldsymbol{\Gamma}' + \boldsymbol{U})'(\boldsymbol{Z}\boldsymbol{\Gamma}' + \boldsymbol{U}) &= \boldsymbol{H}\begin{pmatrix} (\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)'(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1) & (\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)'\boldsymbol{V}_2 \\ \boldsymbol{V}_2'(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1) & \boldsymbol{V}_2'\boldsymbol{V}_2 \end{pmatrix}\boldsymbol{H}'. \end{aligned} \tag{B.1}$$

By applying the formula for the determinant of a partitioned matrix (see, e.g., Harville, 1997, th. 13.3.8) to (B.1), we derive

$$\begin{aligned} \boldsymbol{W}_1 &= |\boldsymbol{U}'\boldsymbol{U}| = |\boldsymbol{V}_2'\boldsymbol{V}_2||\boldsymbol{V}_1'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}\boldsymbol{V}_1|, \\ \boldsymbol{W}_2 &= |(\boldsymbol{Z}\boldsymbol{\Gamma}' + \boldsymbol{U})'(\boldsymbol{Z}\boldsymbol{\Gamma}' + \boldsymbol{U})| \\ &= |\boldsymbol{V}_2'\boldsymbol{V}_2||(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)|. \end{aligned} \tag{B.2}$$

It follows from (B.2) that

$$\begin{aligned} T &= \log\frac{|\boldsymbol{W}_2|}{|\boldsymbol{W}_1|} - \log|\boldsymbol{I}_p + \boldsymbol{\Gamma}\boldsymbol{\Gamma}'| \\ &= \log\left\{\frac{|\boldsymbol{V}_1'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}\boldsymbol{V}_1|}{|(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)|}\right\} - \log|\boldsymbol{I}_p + \boldsymbol{\Gamma}\boldsymbol{\Gamma}'|. \end{aligned} \tag{B.3}$$

Notice that $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, and $\boldsymbol{B}$ are mutually independent, and

$$\begin{aligned} &\boldsymbol{V}_1'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}\boldsymbol{V}_1 \sim W_r(n-k-p+r, \boldsymbol{I}_r), \\ &(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1) \sim W_r(n-k-p+r, \boldsymbol{I}_r + \boldsymbol{\Lambda}^2). \end{aligned}$$

Hence, we derive

$$\begin{aligned} &\frac{1}{n-k-p+r}\boldsymbol{V}_1'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}\boldsymbol{V}_1 \xrightarrow{p} \boldsymbol{I}_r, \\ &\frac{1}{n-k-p+r}(\boldsymbol{I}_r + \boldsymbol{\Lambda}^2)^{-1/2}(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)(\boldsymbol{I}_r + \boldsymbol{\Lambda}^2)^{-1/2} \xrightarrow{p} \boldsymbol{I}_r. \end{aligned}$$

These equations imply that

$$\begin{aligned} &\frac{1}{(n-k-p+r)^r}\left|\boldsymbol{V}_1'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}\boldsymbol{V}_1\right| \xrightarrow{p} 1, \\ &\frac{1}{(n-k-p+r)^r|\boldsymbol{I}_r + \boldsymbol{\Lambda}^2|}\left|(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)'\{\boldsymbol{I}_{n-k} - \boldsymbol{V}_2(\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\}(\boldsymbol{B}\boldsymbol{\Lambda} + \boldsymbol{V}_1)\right| \xrightarrow{p} 1. \end{aligned} \tag{B.4}$$

Notice that $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{Q}\boldsymbol{\Lambda}^2\boldsymbol{Q}'$. By using this equation and a basic property of a determinant (see e.g., Harville, 1997, cor. 18.1.2), we derive

$$|\boldsymbol{I}_r + \boldsymbol{\Lambda}^2| = |\boldsymbol{I}_r + \boldsymbol{\Gamma}'\boldsymbol{\Gamma}| = |\boldsymbol{I}_p + \boldsymbol{\Gamma}\boldsymbol{\Gamma}'|. \tag{B.5}$$

By substituting (B.4) and (B.5) into (B.3), $T \xrightarrow{p} 0$ is obtained. This means that equation (9) is proved.

C. Proof of Lemma 1

We first describe a lemma about another expression of $\boldsymbol{S}_{yy\cdot j}$; this is required for proving Lemma 1 (the proof is given in Appendix G).

Lemma A.1 *Let $\boldsymbol{\mathcal{E}}$, $\boldsymbol{A}_j$, and $\boldsymbol{B}$ be mutually independent random matrices, which are defined by*

$$\boldsymbol{\mathcal{E}} \sim N_{n\times q}(\boldsymbol{O}_{n,p}, \boldsymbol{I}_p \otimes \boldsymbol{I}_n),\ \boldsymbol{A}_j \sim N_{n\times(q-q_j)}(\boldsymbol{O}_{n,q-q_j}, \boldsymbol{I}_{q-q_j} \otimes \boldsymbol{I}_n),\ \boldsymbol{B} \sim N_{n\times q}(\boldsymbol{O}_{n,q}, \boldsymbol{\Sigma}_{xx} \otimes \boldsymbol{I}_n),$$

and let $\boldsymbol{B}_j$ denote the $n \times q_j$ matrix consisting of the columns of $\boldsymbol{B}$ indexed by the elements of j. Then, for any $j \in \mathcal{J}$, $n\boldsymbol{S}_{yy\cdot j}$ can be rewritten as

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

where $\boldsymbol{P}_j$ is the projection matrix to the subspace spanned by the columns of $\boldsymbol{B}_j$, i.e., $\boldsymbol{P}_j = \boldsymbol{B}_j(\boldsymbol{B}_j'\boldsymbol{B}_j)^{-1}\boldsymbol{B}_j'$, and $\boldsymbol{\Gamma}_j$ is given by (20). In particular, when $j \in \mathcal{J}_+$,

$$n\boldsymbol{S}_{yy\cdot j} = \boldsymbol{\Sigma}_{yy\cdot x}^{1/2}\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_j)\boldsymbol{\mathcal{E}}\boldsymbol{\Sigma}_{yy\cdot x}^{1/2}. \tag{C.2}$$

When $j \in \mathcal{J}_+\backslash\{j_*\}$, it follows from Lemma A.1 that

$$\begin{aligned} F_{\min}(j) - F_{\min}(j_*) &= -\log\frac{|n\boldsymbol{S}_{yy\cdot j_*}|}{|n\boldsymbol{S}_{yy\cdot j}|} = -\log\frac{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_{j_*})\boldsymbol{\mathcal{E}}|}{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_j)\boldsymbol{\mathcal{E}}|} \\ &= -\log\frac{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_j)\boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{E}}'(\boldsymbol{P}_j - \boldsymbol{P}_{j_*})\boldsymbol{\mathcal{E}}|}{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_j)\boldsymbol{\mathcal{E}}|}. \end{aligned}$$

Notice that $\boldsymbol{I}_n - \boldsymbol{P}_j$ and $\boldsymbol{P}_j - \boldsymbol{P}_{j_*}$ are idempotent matrices, and $(\boldsymbol{I}_n - \boldsymbol{P}_j)(\boldsymbol{P}_j - \boldsymbol{P}_{j_*}) = \boldsymbol{O}_{n,n}$ holds. Hence, (19) is proved.

When $j \in \mathcal{J}_-$, it follows from Lemma A.1 that

$$\begin{aligned} F_{\min}(j) - F_{\min}(\omega) &= \log|n\boldsymbol{S}_{yy\cdot j}| - \log|n\boldsymbol{S}_{yy\cdot x}| \\ &= \log\left|(\boldsymbol{A}_j\boldsymbol{\Gamma}_j' + \boldsymbol{\mathcal{E}})'(\boldsymbol{I}_n - \boldsymbol{P}_j)(\boldsymbol{A}_j\boldsymbol{\Gamma}_j' + \boldsymbol{\mathcal{E}})\right| - \log\left|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_\omega)\boldsymbol{\mathcal{E}}\right| \\ &= \log\frac{|(\boldsymbol{A}_j\boldsymbol{\Gamma}_j' + \boldsymbol{\mathcal{E}})'(\boldsymbol{I}_n - \boldsymbol{P}_j)(\boldsymbol{A}_j\boldsymbol{\Gamma}_j' + \boldsymbol{\mathcal{E}})|}{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_j)\boldsymbol{\mathcal{E}}|} + \log\frac{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_j)\boldsymbol{\mathcal{E}}|}{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_\omega)\boldsymbol{\mathcal{E}}|} \\ &= \log\frac{|(\boldsymbol{A}_j\boldsymbol{\Gamma}_j' + \boldsymbol{\mathcal{E}})'(\boldsymbol{I}_n - \boldsymbol{P}_j)(\boldsymbol{A}_j\boldsymbol{\Gamma}_j' + \boldsymbol{\mathcal{E}})|}{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_j)\boldsymbol{\mathcal{E}}|} \\ &\quad + \log\frac{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_\omega)\boldsymbol{\mathcal{E}} + \boldsymbol{\mathcal{E}}'(\boldsymbol{P}_\omega - \boldsymbol{P}_j)\boldsymbol{\mathcal{E}}|}{|\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_n - \boldsymbol{P}_\omega)\boldsymbol{\mathcal{E}}|}. \end{aligned}$$

Notice that $\boldsymbol{I}_n - \boldsymbol{P}_j$, $\boldsymbol{I}_n - \boldsymbol{P}_\omega$, and $\boldsymbol{P}_\omega - \boldsymbol{P}_j$ are idempotent matrices, and $(\boldsymbol{I}_n - \boldsymbol{P}_\omega)(\boldsymbol{P}_\omega - \boldsymbol{P}_j) = \boldsymbol{O}_{n,n}$ holds. Hence, (22) is proved.

D. Proof of Equation (23)

It follows from the general formula for the inverse of a block matrix, e.g., th. 8.5.11 in Harville

(1997), that

$$\boldsymbol{\Sigma}_{xx}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{jj} & \boldsymbol{\Sigma}_{j\bar{j}} \\ \boldsymbol{\Sigma}_{j\bar{j}}' & \boldsymbol{\Sigma}_{\bar{j}\bar{j}} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{jj}^{-1} + \boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1} & -\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1} \\ -\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1} & \boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}$ are given by (21). By using the above equation, we have

$$\begin{aligned} \boldsymbol{\Sigma}_{xy}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} &= \left(\boldsymbol{\Sigma}_{jy}', \boldsymbol{\Sigma}_{\bar{j}y}'\right)\begin{pmatrix} \boldsymbol{\Sigma}_{jj}^{-1} + \boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1} & -\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1} \\ -\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1} & \boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1} \end{pmatrix}\begin{pmatrix} \boldsymbol{\Sigma}_{jy} \\ \boldsymbol{\Sigma}_{\bar{j}y} \end{pmatrix} \\ &= \boldsymbol{\Sigma}_{jy}'\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy} + (\boldsymbol{\Sigma}_{\bar{j}y} - \boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy})'\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}(\boldsymbol{\Sigma}_{\bar{j}y} - \boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}). \end{aligned} \tag{D.1}$$

It follows from (20) and (D.1) that

$$\boldsymbol{\Sigma}_{yy\cdot x}^{-1/2}\boldsymbol{\Sigma}_{yy\cdot j}\boldsymbol{\Sigma}_{yy\cdot x}^{-1/2} = \boldsymbol{I}_p + \boldsymbol{\Gamma}_j\boldsymbol{\Gamma}_j'.$$

Hence, we have

$$\left|\boldsymbol{I}_p + \boldsymbol{\Gamma}_j\boldsymbol{\Gamma}_j'\right| = \left|\boldsymbol{\Sigma}_{yy\cdot x}^{-1/2}\boldsymbol{\Sigma}_{yy\cdot j}\boldsymbol{\Sigma}_{yy\cdot x}^{-1/2}\right| = \frac{|\boldsymbol{\Sigma}_{yy\cdot j}|}{|\boldsymbol{\Sigma}_{yy\cdot x}|}.$$

By using the above equation, (23) is proved. On the other hand, it follows from (12) that $\boldsymbol{\Gamma}_j = \boldsymbol{O}_{p,q-q_j}$ holds when $j \in \mathcal{J}_+$. Hence, $\delta_j = 0$ holds when $j \in \mathcal{J}_+$.

E. Two Examples of δ_j

Two examples of δ_j are presented in this section. In both examples, we assume that j is the subset of j_* such that $j_* = \{j, \bar{j} \cap j_*\}$.

First, we show the case that a limiting value of δ_j is bounded. Let

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Omega}(q), \quad \boldsymbol{\Sigma}_{yy} = \boldsymbol{\Omega}(p),\ \boldsymbol{\Sigma}_{xy} = \rho\begin{pmatrix} \boldsymbol{1}_{q_*} \\ q_*\rho/\{1+(q_*-1)\rho\}\boldsymbol{1}_{q-q_*} \end{pmatrix}\boldsymbol{1}_p',$$

where $\rho \in (-1, 1)$ and $\boldsymbol{\Omega}(m)$ is an $m \times m$ symmetric matrix defined by

$$\boldsymbol{\Omega}(m) = (1-\rho)\boldsymbol{I}_m + \rho\boldsymbol{1}_m\boldsymbol{1}_m'.$$

From the general formula of the inverse of the sum of two matrices, e.g., cor. 18.2.10 in Harville (1997), we have

$$\boldsymbol{\Omega}(m)^{-1} = \frac{1}{1-\rho}\left\{\boldsymbol{I}_m - \frac{\rho}{1+(m-1)\rho}\boldsymbol{1}_m\boldsymbol{1}_m'\right\}.$$

Then, we can see that

$$\begin{aligned} |\boldsymbol{\Sigma}_{yy\cdot x}| = |\boldsymbol{\Sigma}_{yy\cdot j_*}| &= \left|\boldsymbol{\Omega}(p) - (\rho\boldsymbol{1}_p\boldsymbol{1}_{q_*}')\boldsymbol{\Omega}(q_*)^{-1}(\rho\boldsymbol{1}_{q_*}\boldsymbol{1}_p')\right| \\ &= \left|(1-\rho)\left\{\boldsymbol{I}_p + \frac{\rho}{1+(q_*-1)\rho}\boldsymbol{1}_p\boldsymbol{1}_p'\right\}\right| = \frac{(1-\rho)^p\{1+(p+q_*-1)\}\rho}{1+(q_*-1)\rho}. \end{aligned} \tag{E.1}$$

It follows from the same calculations as in (E.1) that

$$|\boldsymbol{\Sigma}_{yy\cdot j}| = \left|\boldsymbol{\Omega}(p) - (\rho\mathbf{1}_p\mathbf{1}'_{q_j})\boldsymbol{\Omega}(q_j)^{-1}(\rho\mathbf{1}_{q_j}\mathbf{1}'_p)\right| = \frac{(1-\rho)^p\{1+(p+q_j-1)\}\rho}{1+(q_j-1)\rho}. \tag{E.2}$$

Hence, from (E.1) and (E.2), the following equation is derived:

$$\begin{aligned}\delta_j &= \log\frac{|\boldsymbol{\Sigma}_{yy\cdot j}|}{|\boldsymbol{\Sigma}_{yy\cdot x}|} = \log\frac{\{1+(p+q_j-1)\}\{1+(q_*-1)\rho\}}{\{1+(p+q_*-1)\}\{1+(q_j-1)\rho\}}\\ &\to \log\left\{\frac{1+(q_*-1)\rho}{1+(q_j-1)\rho}\right\}.\end{aligned}$$

This equation indicates that the limiting value of δ_j is bounded as $p\to\infty$.

Next, the case that δ_j approaches ∞ is shown. Let

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{I}_q, \quad \boldsymbol{\Sigma}_{yy} = \boldsymbol{I}_p,\ \boldsymbol{\Sigma}_{xy} = \begin{pmatrix}\mathbf{1}_{q_*}\\ \mathbf{0}_{q-q_*}\end{pmatrix}\boldsymbol{\alpha}',$$

where $\boldsymbol{\alpha}$ is a p-dimensional vector defined by

$$\boldsymbol{\alpha} = \sqrt{\frac{1-\rho^2}{q^*\rho^2}}(\rho,\dots,\rho^p)'.$$

Here $\rho\in(-1,1)$. Notice that

$$\boldsymbol{\alpha}'\boldsymbol{\alpha} = \frac{1-\rho^2}{q^*\rho^2}\sum_{k=1}^{p}(\rho^2)^k = \frac{1}{q_*}\left\{1-(\rho^2)^p\right\}.$$

Hence, we can see that

$$\begin{aligned}|\boldsymbol{\Sigma}_{yy\cdot x}| = |\boldsymbol{\Sigma}_{yy\cdot j_*}| &= \left|\boldsymbol{I}_p - (\boldsymbol{\alpha}\mathbf{1}'_{q_*})\boldsymbol{I}_{q_*}(\rho\mathbf{1}_{q_*}\boldsymbol{\alpha}')\right|\\ &= \left|\boldsymbol{I}_p - q_*\boldsymbol{\alpha}\boldsymbol{\alpha}'\right| = 1-\left\{1-(\rho^2)^p\right\} = (\rho^2)^p.\end{aligned} \tag{E.3}$$

It follows from the same calculations as in (E.3) that

$$|\boldsymbol{\Sigma}_{yy\cdot j}| = \left|\boldsymbol{I}_p - (\boldsymbol{\alpha}\mathbf{1}'_{q_j})\boldsymbol{I}_{q_j}(\rho\mathbf{1}_{q_j}\boldsymbol{\alpha}')\right| = \left[1-\frac{q_j}{q_*}\left\{1-(\rho^2)^p\right\}\right]. \tag{E.4}$$

From (E.3) and (E.4), the following equation is derived:

$$\delta_j = \log\frac{|\boldsymbol{\Sigma}_{yy\cdot j}|}{|\boldsymbol{\Sigma}_{yy\cdot x}|} = \log\left[1-\frac{q_j}{q_*}\left\{1-(\rho^2)^p\right\}\right] - p\log(\rho^2).$$

This equation indicates that δ_j approaches ∞ as $p\to\infty$.

F. Proof of Corollary 2

From a simple calculation, we have the following expansion:

$$n\left\{b(q_j+p) - b(q_j) - b(q_*+p) + b(q_*)\right\} = \frac{np(2n-p)}{(n-p)^2}(q_j-q_*) + O(n^{-2}p).$$

It follows from the above equation and (16) that

$$m(j) - m(j_*) = \begin{cases} 2p(q_j - q_*) & \text{(AIC)} \\ p(2 - p/n)(q_j - q_*)/(1 - p/n)^2 + O(pn^{-1}) & (\text{AIC}_\text{c}) \\ 2p(q_j - q_*) + \hat{\kappa}_{(jy)} - \hat{\kappa}_j - \hat{\kappa}_{(j_*y)} + \hat{\kappa}_{j_*} & \text{(TIC)} \\ p(q_j - q_*)\log n & \text{(BIC)} \\ p(q_j - q_*)(1 + \log n) & \text{(CAIC)} \\ 2p(q_j - q_*)\log\log n & \text{(HQC)} \end{cases}. \tag{F.1}$$

On other hand, by using the results in Mardia (1974), the mean and variance of $\hat{\kappa}(\boldsymbol{D})$ in (17) are calculated as

$$E\left[\hat{\kappa}(\boldsymbol{D})\right] = \frac{d(d+2)}{n(n+2)}, \quad Var\left[\hat{\kappa}(\boldsymbol{D})\right] = \frac{8d(d+2)(n-2)(n+1)^4(n-d)(n-d+2)}{n^4(n+2)^2(n+4)(n+6)}.$$

These imply that for any $j \in \mathcal{J}$,

$$\hat{\kappa}_j \xrightarrow{p} 0, \quad \frac{1}{p}\hat{\kappa}_{(jy)} \xrightarrow{p} 0.$$

It follows from the above equations and (F.1) that

$$a_m(n, p) = \begin{cases} p \\ p \\ p\log n \\ p\log\log n \end{cases}, \quad \beta_m(j) = \begin{cases} 2(q_j - q_*) & \text{(AIC, TIC)} \\ \{(2-c)/(1-c)^2\}(q_j - q_*) & (\text{AIC}_\text{c}) \\ q_j - q_* & \text{(BIC, CAIC)} \\ 2(q_j - q_*) & \text{(HQC)} \end{cases}. \tag{F.2}$$

Hence, it is easy to see that

$$\lim_{n-p\to\infty, p/n\to c} \frac{p}{a_m(n,p)} = \begin{cases} 1 & (\text{AIC}, \text{AIC}_\text{c}, \text{TIC}) \\ 0 & \text{(BIC, CAIC, HQC)} \end{cases}.$$

By using the above results and (27), we can see that condition C1 holds for the BIC, CAIC, and HQC. Condition C1 holds for the AIC, AIC_c, and TIC if p goes to ∞ and the following equation is satisfied:

$$g(c) < \begin{cases} 2 & \text{(AIC, TIC)} \\ 1/(1-c) + 1/(1-c)^2 & (\text{AIC}_\text{c}) \end{cases},$$

where $g(x)$ is given by (8). The above equation implies that condition C1 holds for the AIC and TIC if p goes to ∞ and $c \in [0, c_\text{a})$, and it holds for the AIC_c if p goes to ∞, because $-g(x) + 2 > 0$ holds when $x \in [0, c_\text{a})$ and $-g(x) + (1-x)^{-1} + (1-x)^{-2} > 0$ holds when $x \in [0, 1)$, where c_a is given in (29). Moreover, from (28) and (F.2), we can see that condition 2 holds if the following equation is satisfied:

$$q_* - q_j < \lim_{n-p\to\infty, p/n\to c} \begin{cases} n\delta_j/(2p) & \text{(AIC, TIC)} \\ n\delta_j(1-c)^2/\{p(2-c)\} & (\text{AIC}_\text{c}) \\ n\delta_j/(p\log n) & \text{(BIC, CAIC)} \\ n\delta_j/(2p\log\log n) & \text{(HQC)} \end{cases}.$$

It is easy to see that the above equation is satisfied if $q_* - q_j \le 0$, because $\delta_j > 0$. Hence, it is sufficient to consider the case of $j \in \mathcal{S}_-$, where $\mathcal{S}_-$ is given in (29). Consequently, Corollary 2 is proved.

G. Proof of Lemma A.1

Let $\boldsymbol{Y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_{n+1})'$ and $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_{n+1})'$, where $\boldsymbol{y}_i$ and $\boldsymbol{x}_i$ are the ith individuals from $\boldsymbol{y}$ and $\boldsymbol{x}$, respectively, and $\boldsymbol{X}_j$ denotes an $(n+1) \times q_j$ matrix consisting of the columns of $\boldsymbol{X}$ indexed by the elements of j. Then, $n\boldsymbol{S}_{yy\cdot j}$ is expressed as

$$n\boldsymbol{S}_{yy\cdot j} = \boldsymbol{Y}'\boldsymbol{H}_{n+1}\left\{\boldsymbol{I}_{n+1} - \boldsymbol{H}_{n+1}\boldsymbol{X}_j\left(\boldsymbol{X}_j'\boldsymbol{H}_{n+1}\boldsymbol{X}_j\right)^{-1}\boldsymbol{X}_j'\boldsymbol{H}_{n+1}\right\}\boldsymbol{H}_{n+1}\boldsymbol{Y},$$

where $\boldsymbol{H}_n$ is the projection matrix to the orthocomplement of the subspace spanned by $\boldsymbol{1}_n$, i.e., $\boldsymbol{H}_n = \boldsymbol{I}_n - \boldsymbol{1}_n\boldsymbol{1}_n'/n$, and $\boldsymbol{1}_n$ is an n-dimensional vector of ones. Let $\boldsymbol{V}$ be an $n \times p$ random matrix such that

$$(\boldsymbol{B}, \boldsymbol{V}) \sim N_{n\times(q+p)}(\boldsymbol{O}_{n,q+p}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{B}$ is an $n \times q$ random matrix. From a property of a multivariate normal distribution, we can rewrite $n\boldsymbol{S}_{yy\cdot j}$ using $\boldsymbol{V}$ and $\boldsymbol{B}$, as follows:

$$n\boldsymbol{S}_{yy\cdot j} = \boldsymbol{V}'(\boldsymbol{I}_n - \boldsymbol{P}_j)\boldsymbol{V}, \tag{G.1}$$

where $\boldsymbol{P}_j = \boldsymbol{B}_j(\boldsymbol{B}_j'\boldsymbol{B}_j)^{-1}\boldsymbol{B}_j'$, and $\boldsymbol{B}_j$ is an $n \times q_j$ matrix consisting of the columns of $\boldsymbol{B}$ indexed by the elements of j. From a property of a conditional distribution of a multivariate normal distribution, e.g., th. 2.2.7 in Srivastava and Khatri (1979), we have

$$\boldsymbol{V}|\boldsymbol{B} \sim N_{n\times p}(\boldsymbol{B}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}, \boldsymbol{\Sigma}_{yy\cdot x} \otimes \boldsymbol{I}_n), \quad \boldsymbol{B}_{\bar{j}}|\boldsymbol{B}_j \sim N_{n\times q}(\boldsymbol{B}_j\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}, \boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j} \otimes \boldsymbol{I}_n),$$

where $\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}$ is given by (21). Hence, we can express $\boldsymbol{V}$ as

$$\boldsymbol{V} = \boldsymbol{B}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} + \boldsymbol{\mathcal{E}}\boldsymbol{\Sigma}_{yy\cdot x}^{1/2}, \quad \boldsymbol{B}_{\bar{j}} = \boldsymbol{B}_j\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{j\bar{j}} + \boldsymbol{A}_j\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{1/2}, \tag{G.2}$$

where $\boldsymbol{\mathcal{E}}$, $\boldsymbol{A}_j$, and $\boldsymbol{B}_j$ are mutually independent random matrices. It should be kept in mind that any $\boldsymbol{S}_{yy\cdot j}$ ($j \in \mathcal{J}$) can be represented by using the common $\boldsymbol{\mathcal{E}}$, because $\boldsymbol{\mathcal{E}}$ is independent of j. Without loss of generality, we assume that $\boldsymbol{B}$ is arranged as $(\boldsymbol{B}_j, \boldsymbol{B}_{\bar{j}})$. It follows from the same calculation as in (D.1) that

$$\begin{aligned}
\boldsymbol{B}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} &= (\boldsymbol{B}_j, \boldsymbol{B}_{\bar{j}})\begin{pmatrix} \boldsymbol{\Sigma}_{jj}^{-1} + \boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1} & -\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1} \\ -\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}\boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1} & \boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1} \end{pmatrix}\begin{pmatrix} \boldsymbol{\Sigma}_{jy} \\ \boldsymbol{\Sigma}_{\bar{j}y} \end{pmatrix} \\
&= \boldsymbol{B}_j\boldsymbol{\Sigma}_{jj}^{-1}\left\{\boldsymbol{\Sigma}_{jy} - \boldsymbol{\Sigma}_{j\bar{j}}\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}(\boldsymbol{\Sigma}_{\bar{j}y} - \boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy})\right\} + \boldsymbol{B}_{\bar{j}}\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}(\boldsymbol{\Sigma}_{\bar{j}y} - \boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}).
\end{aligned}$$

Substituting $\boldsymbol{B}_{\bar{j}}$ in (G.2) into the above equation yields

$$\begin{aligned}
\boldsymbol{B}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} &= \boldsymbol{B}_j\boldsymbol{\Sigma}_{jj}^{-1}\left\{\boldsymbol{\Sigma}_{jy} - \boldsymbol{\Sigma}_{j\bar{j}}\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}(\boldsymbol{\Sigma}_{\bar{j}y} - \boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy})\right\} \\
&\quad + (\boldsymbol{B}_j\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{j\bar{j}} + \boldsymbol{A}_j\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{1/2})\boldsymbol{\Sigma}_{\bar{j}\bar{j}\cdot j}^{-1}(\boldsymbol{\Sigma}_{\bar{j}y} - \boldsymbol{\Sigma}_{j\bar{j}}'\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy}) \\
&= \boldsymbol{B}_j\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jy} + \boldsymbol{A}_j\boldsymbol{\Gamma}_j'\boldsymbol{\Sigma}_{yy\cdot x}^{1/2},
\end{aligned}$$

where $\boldsymbol{\Gamma}_j$ is the $p \times (q - q_j)$ matrix defined by (20). Substituting the above equation into $\boldsymbol{V}$ in (G.2) yields

$$\boldsymbol{V} = \boldsymbol{B}_j \boldsymbol{\Sigma}_{jj}^{-1} \boldsymbol{\Sigma}_{jy} + \boldsymbol{A}_j \boldsymbol{\Gamma}_j' \boldsymbol{\Sigma}_{yy\cdot x}^{1/2} + \boldsymbol{\mathcal{E}} \boldsymbol{\Sigma}_{yy\cdot x}^{1/2}. \tag{G.3}$$

Notice that $(\boldsymbol{I}_n - \boldsymbol{P}_j)\boldsymbol{B}_j = \boldsymbol{O}_{n,q_j}$. By using (G.1) and (G.3), (C.1) is proved. Moreover, $\boldsymbol{\Gamma}_j = \boldsymbol{O}_{p,q-q_j}$ if $j \in \mathcal{J}_+$, because then (12) holds. This implies that for any $j \in \mathcal{J}_+$

$$\boldsymbol{V} = \boldsymbol{B}_j \boldsymbol{\Sigma}_{jj}^{-1} \boldsymbol{\Sigma}_{jy} + \boldsymbol{\mathcal{E}} \boldsymbol{\Sigma}_{yy\cdot x}^{1/2}. \tag{G.4}$$

Hence, from (G.1) and (G.4), (C.2) is proved.