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 $\infty$ 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 $W_1$ and $W_2$ 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.,

$$\log |W_2| - \log |W_1| = \log \frac{|W_2|}{|W_1|}. \quad (1)$$

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 $\log |W_2|/|W_1|$ 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 \( \infty \). 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:** \( n \to p \) and \( p/n \) are approaching \( \infty \) 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 \( \log |W_2|/|W_1| \), using the HD asymptotic framework. Let \( O_{d,p} \) be a \( d \times p \) matrix of zeros. We will consider the following structures for Wishart matrices:

**Case 1.** The \( W_2 \) includes \( W_1 \) completely, i.e.,

\[
W_1 \sim W_p(n - k, I_p), \quad W_2 = W_1 + Z'Z \sim W_p(n - k + d, I_p),
\]

where \( k \) and \( d \) are positive integers independent of \( n \) and \( p \), and \( W_1 \) and \( Z \) are independent random matrices defined by

\[
Z \sim N_{d\times p}(O_{d,p}, I_p \otimes I_d).
\]

**Case 2.** The \( W_2 \) includes \( W_2 \) partially, i.e.,

\[
W_1 = U'U \sim W_p(n - k, I_p), \quad W_2 = (Z\Gamma' + U)'(Z\Gamma' + U) \sim W_p(n - k, I_p + \Gamma\Gamma'),
\]

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

\[
U \sim N_{(n-k)\times p}(O_{n-k,p}, I_p \otimes I_{n-k}), \quad Z \sim N_{(n-k)\times r}(O_{n-k,r}, I_r \otimes I_{n-k}).
\]

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 (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, AIC\(_c\), TIC, BIC, CAIC, and HQC were proposed by Akaike (1973;
Yanagihara, H., Hashiyama, Y. & Fujikoshi, Y.

1974), Hurvich and Tsai (1989), Takeuchi (1976), Schwarz (1978), Bozdogan (1987), and Hannan and Quinn (1979), respectively. The AIC and 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 $\log |W_2|/|W_1|$ 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 $W_1$ and $W_2$ are Wishart matrices given by (2). Then, we have

$$
\log \frac{|W_2|}{|W_1|} = -d \log(1 - p/n) + \frac{\sqrt{p}}{n} \left(1 - \frac{p}{n}\right) \text{tr}(V) + O_p(pn^{-2}),
$$

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

$$
V = \frac{n}{\sqrt{p}} \left( Z W_1^{-1} Z' - \frac{p}{n-k-p-1} I_d \right),
$$

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

$$
\text{tr}(V) \xrightarrow{d} \begin{cases} 
\frac{(\chi_{dp}^2 - dp)/\sqrt{p}}{O_p(1)} & (p \text{ is bounded}) \\
N(0, 2d/(1 - c^2)) & (p \to \infty)
\end{cases}.
$$

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 $W_1$ and $W_2$ are Wishart matrices given by (2). Then, we have
\[ \log \frac{|W_2|}{|W_1|} = -d \log(1 - c) + o_p(1), \]  
and
\[ \frac{n}{p} \log \frac{|W_2|}{|W_1|} = R + o_p(1), \]
where $R$ is defined by
\[ R = \begin{cases} \frac{X}{p}, & X \sim \chi^2_{dp} \quad (p \text{ bounded}) \\ dg(c) & (p \to \infty) \end{cases} \]
Here, $g(x)$ is a function with domain $x \in [0, 1)$, which is given by
\[ g(x) = \begin{cases} 1 & (x = 0) \\ -x^{-1} \log(1 - x) & (x \in (0, 1)) \end{cases}. \]
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 $\text{tr}(V)$ in order to prove Theorem 1, (7) can be derived without the asymptotic distribution of $\text{tr}(V)$, using only the expectation of $\text{tr}(V)$.

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

Theorem 2  Suppose that $W_1$ and $W_2$ are Wishart matrices given by (3). Then, we have
\[ \log \frac{|W_2|}{|W_1|} = \log |I_p + \Gamma \Gamma'| + o_p(1). \]  

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 $z = (x', y')' = (x_1, \ldots, x_q, y_1, \ldots, y_p)'$ be a $(q + p)$-dimensional random vector distributed according to the $(q + p)$-variate multivariate normal distribution with the following mean vector and covariance matrix:
\[ E[z] = \mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \text{Cov}(z) = \Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}. \]
Suppose that $j$ denotes a subset of $\omega = \{1, \ldots, q\}$ containing $q_j$ elements, and $x_j$ denotes the $q_j$-dimensional vector consisting of $x$ indexed by the elements of $j$. For example, if $j = \{1, 2, 4\}$, then $x_j$ consists of the first, second, and fourth elements of $x$. We will also let $\bar{j}$ denote the complement of the set $j$, i.e., $\bar{j} = \bar{j}$. Of course, it holds that $x_\omega = x$ and $q_\omega = q$. Without loss of generality,
we divide $x$ into two subvectors $x = (x', x')'$, where $x_j$ is a $(q - q_j)$-dimensional random vector. Expressions of $\Sigma_{xx}$ and $\Sigma_{xy}$ corresponding to the division of $x$ are

$$
\Sigma_{xx} = \begin{pmatrix}
\Sigma_{jj} & \Sigma_{ij} \\
\Sigma_{ij}^\prime & \Sigma_{yy}
\end{pmatrix},
\Sigma_{xy} = \begin{pmatrix}
\Sigma_{ij} \\
\Sigma_{yy}
\end{pmatrix}.
$$

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

$$
\Sigma = \begin{pmatrix}
\Sigma_{jj} & \Sigma_{ij} & \Sigma_{ij}^\prime \\
\Sigma_{ij} & \Sigma_{yy} & \Sigma_{yy} \\
\Sigma_{ij}^\prime & \Sigma_{yy} & \Sigma_{yy}
\end{pmatrix}.
$$ (10)

Let $z_1, \ldots, z_{n+1}$ be $(n + 1)$ independent random vectors from $z$, and let $S$ be the usual unbiased estimator of $\Sigma$ given by $S = n^{-1}\sum_{i=1}^{n+1}(z_i - \bar{z})(z_i - \bar{z})'$, where $\bar{z} = (n + 1)^{-1}\sum_{i=1}^{n+1}z_i$. Following the same method that we used for $\Sigma$ in (10), we divide $S$ as

$$
S = \begin{pmatrix}
S_{xx} & S_{xy} \\
S_{xy}^\prime & S_{yy}
\end{pmatrix} = \begin{pmatrix}
S_{ij} & S_{ij} & S_{ij}^\prime \\
S_{ij} & S_{yy} & S_{yy} \\
S_{ij}^\prime & S_{yy} & S_{yy}
\end{pmatrix}.
$$

One of the interests in CCA is to determine whether $x_j$ is irrelevant. Fujikoshi (1985) determined that $x_j$ is irrelevant if the following equation holds:

$$
\text{tr}(\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yy}') = \text{tr}(\Sigma_{jj}^{-1}\Sigma_{jj}^\prime\Sigma_{yy}^{-1}\Sigma_{yy}').
$$ (11)

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

$$
\Sigma_{jj}^\prime - \Sigma_{jj}'^\prime \Sigma_{jj}^{-1} \Sigma_{jj} = O_{q - q_j, p}.
$$ (12)

Consequently, the candidate model in which $x_j$ is irrelevant can be expressed as

$$
nS \sim W_{pq}(n, \Sigma) \text{ s.t. } \text{tr}(\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yy}') = \text{tr}(\Sigma_{jj}^{-1}\Sigma_{jj}'^\prime\Sigma_{yy}^{-1}\Sigma_{yy}').
$$ (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 $x_j$ as irrelevant and $x_j$ as relevant.

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

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

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

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

It is easy to see that $nF(S, \Sigma)$ 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, $\hat{\Sigma}_j$ 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 $\mathcal{F}$ be a set of candidate models denoted by $\mathcal{F} = \{j_1, \ldots, j_K\}$, where $K$ is the number of models. We separate $\mathcal{F}$ into two sets such that one is a set of overspecified models, and the other is a set of underspecified models. Let $\mathcal{F}_+$ denote the set of overspecified models, which is defined by

$$J_+ = \left\{ j \in J \mid \text{tr}(\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}) = \text{tr}(\Sigma_{jj}^{-1}\Sigma_{jj}\Sigma_{yy}^{-1}\Sigma_{jj}^\prime) \right\}.$$

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

$$j_* = \arg\min_{j \in J_+} q_j.$$

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

$$\mathcal{F}_- = \mathcal{F}_+ \cap \mathcal{F}.$$

For the general expression for any $j \in J$, $\Sigma_{yy,j}$ and $S_{yy,j}$ are defined as

$$\Sigma_{yy,j} = \Sigma_{yy} - \Sigma_{yy,j}\Sigma_{yy,j}^\prime \Sigma_{yy,j}, \quad S_{yy,j} = S_{yy} - S_{yy,j}\Sigma_{yy,j}^\prime \Sigma_{yy,j}.$$

In particular, we write $\Sigma_{yy,o} = \Sigma_{yy,x}$ and $S_{yy,o} = S_{yy,x}$. From Fujikoshi et al. (2010, chap. 11.5) and the fact that $\text{tr}(\Sigma_jS) = p + q$, the minimum value of $F(\Sigma, S)$ under the model $j$ is given by

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

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:

$$IC_m(j) = nF_{\min}(j) + m(j).$$

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

$$m(j) = \begin{cases} 
2[pq_j + (q^2 + p^2 + q + p)/2] & \text{(AIC)} \\
\frac{n(b(q) + b(q_j + p) - b(q_j) - (q + p))}{2} & \text{(AIC$_2$)} \\
2[pq_j + (q^2 + p^2 + q + p)/2] + \hat{k}_x + \hat{k}_{(j)} - \hat{k}_{i} & \text{(TIC)} \\
[pq_j + (q^2 + p^2 + q + p)/2] \log n & \text{(BIC)} \\
\{pq_j + (q^2 + p^2 + q + p)/2\} \log n + 1 & \text{(CAIC)} \\
2[pq_j + (q^2 + p^2 + q + p)/2] \log \log n & \text{(HQC)}
\end{cases},$$

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

$$\hat{k}_x = \hat{k}(D_x), \quad \hat{k}_j = \hat{k}(D_j), \quad \hat{k}_{(j)} = \hat{k}(D_{(j)}).$$

Here $\hat{k}(D)$ is an estimator of multivariate kurtosis of the $d$-variates extracted from $z$ by $D^\prime z$, which
Yanagihara, H., Hashiyama, Y. & Fujikoshi, Y.

is defined by

\[
\hat{k}(D) = \frac{1}{n+1} \sum_{i=1}^{n+1} \left( (z_i - \bar{z})^\top D(D^\top SD)^{-1}D'(z_i - \bar{z}) \right)^2 - d(d + 2), \tag{17}
\]

where \( D \) is a \((q+p) \times d \) matrix whose elements are 0 or 1, and satisfies \( D' D = I_d \), e.g.,

\[
D_4 = \begin{pmatrix} I_q \\ O_{p,q} \end{pmatrix}, \quad D_5 = \begin{pmatrix} O_{q,p} \\ I_p \end{pmatrix}, \quad D_6 = \begin{pmatrix} I_q \\ O_{p+q-q,q_j} \end{pmatrix}, \quad D_{(j\gamma)} = (D_j, D_{\gamma}).
\]

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

\[
\beta_m(j) = \lim_{n-p \to \infty, p/n \to c} \frac{m(j) - m(j_\star)}{a_m(n, p)}, \quad 0 < |\beta_m(j)| < \infty. \tag{18}
\]

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

\[
\hat{j}_m = \arg \min_{j \in J} IC_m(j).
\]

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

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

\[
W_1 = W, \quad W_2 = W + Z'Z,
\]

where \( W_1 \) and \( W_2 \) are independent random matrices defined by

\[
W \sim W_p(n - q_j, I_p), \quad Z \sim N_{(q_j, q_j, p)}(O_{q_j, q_j, p}, I_p \otimes I_{q_j, q_j}).
\]

Then, for any \( j \in J \setminus \{j_\star\} \), \( F_{\min}(j) - F_{\min}(j_\star) \) can be rewritten as

\[
F_{\min}(j) - F_{\min}(j_\star) = -\log \frac{|W_2|}{|W_1|}. \tag{19}
\]

2. Let

\[
W_1 = U'U = W_3 + U_1'U_2, \quad W_2 = (Z\Gamma_1' + U)(Z\Gamma_1' + U)', \quad W_3 = U_1'U_1,
\]

where \( U = (U_1', U_2'), \quad U_1, U_2, \text{ and } Z \) are mutually independent random matrices defined by

\[
U_1 \sim N_{(n-q) \times p}(O_{n-q,p}, I_{n-q} \otimes I_p), \quad U_2 \sim N_{(q-q) \times p}(O_{q-q,p}, I_{q-q} \otimes I_p),
\]

\[
Z \sim N_{(n-q) \times (q-q)}(O_{n-q,q-q}, I_{q-q} \otimes I_{n-q}).
\]
Let $\Delta_j = \log |I_p + \Gamma_j^j|$. It is known that $\Delta_j \geq 0$ with equality if and only if $j \in \mathcal{F}_+$, because $\Gamma_j = O_{p,q-q_i}$ holds if and only if $j \in \mathcal{F}_+$. Moreover, $\Delta_j$ can be rewritten as

$$\Delta_j = \log |I_p + \Gamma_j^j| = \log \frac{\Sigma_{yy}^{j}}{\Sigma_{yy}^{j+}}. \quad (23)$$

(23) (proof is given in Appendix D). Notice that $\Delta_j$ 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 $\Delta_j$ play an important role in deciding whether a criterion is consistent. In addition, we assume that $\lim_{p \to \infty} \Delta_j > 0$ for $j \in \mathcal{F}_-$. When $j \in \mathcal{F}_+ \setminus \{j_1\}$, it follows from Corollary 1 and Lemma 1 that $F_{\min}(j) - F_{\min}(j_1) = o_p(1)$.

Notice that

$$F_{\min}(j) - F_{\min}(j_1) = F_{\min}(j) - F_{\min}(\omega) + F_{\min}(\omega) - F_{\min}(j_1).$$

Since $\omega \in \mathcal{F}_+$, $F_{\min}(\omega) - F_{\min}(j_1) = o_p(1)$. By using this result, Theorem 2, and Lemma 1, we have

$$F_{\min}(j) - F_{\min}(j_1) = \Delta_j + o_p(1) \quad (\forall j \in \mathcal{F}_-). \quad (24)$$

Let us define $R_j$ by

$$R_j = \begin{cases} \frac{X_j}{p}, & X_j \sim \chi^2_{p(q_i-q_j)p} \text{ (} p \text{ is bounded)} \\ g(c)(p \to \infty) & \end{cases}, \quad (25)$$

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

$$\frac{n}{p} \{F_{\min}(j) - F_{\min}(j_1)\} = -R_j + o_p(1) \quad (\forall j \in \mathcal{F}_+ \setminus \{j_1\}). \quad (26)$$

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

$$\lim_{n \to \infty, p \to \infty, n/p \to 0} \frac{1}{p} |\text{IC}_m(j) - \text{IC}_m(j_1)| > 0 \quad (\forall j \in \mathcal{F}_+ \setminus \{j_1\}),$$

$$\lim_{n \to \infty, p \to \infty, n/p \to 0} \frac{1}{p} |\text{IC}_m(j) - \text{IC}_m(j_1)| > 0 \quad (\forall j \in \mathcal{F}_-).$$
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 $j \in J_+ \setminus \{j_c\}$,

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

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

C2. For any $j \in J_-$,

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

where $\delta_j$ 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, $P(|R_j| > \epsilon) \neq 0$, $\epsilon > 0$, 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

$$S_+ = \{ j \in J_+ | q_* - q_j > 0 \}, \quad c_* = \arg \max_{x \in [0, 1]} [-q(x) + 2 = 0] = 0.797.$$ (29)

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

- **AIC & TIC:** $p \to \infty$, $c \in (0, c_*)$, and for any $j \in S_+$,

$$\lim_{n-p \to \infty, p/n \to c} \frac{n\delta_j}{2p(q_* - q_j)} > 1.$$

- **AIC:** $p \to \infty$, and for any $j \in S_+$,

$$\lim_{n-p \to \infty, p/n \to c} \frac{n\delta_j(1-c)^2}{p(q_* - q_j)(2-c)} > 1.$$

- **BIC & CAIC:** for any $j \in S_+$,

$$\lim_{n-p \to \infty, p/n \to c} \frac{n\delta_j}{p(q_* - q_j)\log n} > 1.$$

- **HQC:** for any $j \in S_+$,

$$\lim_{n-p \to \infty, p/n \to c} \frac{n\delta_j}{2p(q_* - q_j)\log \log n} > 1.$$

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


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
\[
\log \left| \frac{W_2}{W_1} \right| = \log \left| I_d + W_1^{-1/2} Z' Z W_1^{-1/2} \right| = \log \left| I_d + Z W_1^{-1} Z' \right|. \tag{A.1}
\]

Notice that
\[
E[ZZ'] = p I_d, \quad E \left[ \log \left( (ZZ')^2 \right) \right] = d(p + d + 1)p, \quad E \left[ \log \left( (ZZ')^2 \right) \right] = d(dp + 2)p.
\]

By using the above results, the assumption that \( Z \) and \( W_1 \) are independent, and th. 2.2.8 in Fujikoshi et al. (2010), we derive
\[
E \left[ Z W_1^{-1} Z' \right] = \frac{1}{h_1} E[ZZ'] = \frac{p}{h_1} I_d,
\]
\[
E \left[ \log \left( (Z W_1^{-1} Z')^2 \right) \right] = \frac{1}{h_0 h_3} E \left[ \log \left( (Z Z')^2 \right) \right] + \frac{1}{h_0 h_1 h_3} E \left[ \log \left( (Z Z')^2 \right) \right]
= \frac{d(p + d + 1)p}{h_0 h_3} + \frac{d(dp + 2)p}{h_0 h_1 h_3},
\]
where \( h_i = n - k - p - i \). Hence, the following equation is obtained:
\[
E \left[ \|Z W_1^{-1} Z' - (p/h_1) I_d\|^2 \right] = \frac{d(d + 1)p}{h_0 h_3} + \frac{d((d + 1)p + 2)p}{h_0 h_1 h_3} + \frac{2dp^2}{h_0 h_1 h_3} = O(p n^{-2}).
\]
The above equation implies that \( Z W_1^{-1} Z' = (p/h_1) I_d + O_p(p^{1/2} n^{-1}) \). Hence, \( V = O_p(1) \) is derived, where \( V \) 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
\[
\log \left| \frac{W_2}{W_1} \right| = \log \left| (1 + p/h_1) I_d + \frac{\sqrt{p}}{(1 + p/h_1)n} V \right|
\]
\[
= -d \log \left( 1 - \frac{p}{n - k - 1} \right) + \frac{\sqrt{p}(n - k - p - 1)}{n(n - k - 1)} \log(V) + O(p n^{-2}). \tag{A.2}
\]

Notice that
\[
\log \left( 1 - \frac{p}{n - k - 1} \right) = \log(1 - p/n) + O(p n^{-2}),
\]
\[
\frac{\sqrt{p}(n - k - p - 1)}{n(n - k - 1)} = \frac{\sqrt{p}}{n} \left( 1 - \frac{p}{n} \right) + O(p^3 n^{-3}).
\]

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

Next, we prove (6). Notice that for sufficiently large \( p \),
Asymptotic Behaviors of Log-Determinants of Two Wishart Matrices

\[
ZW_1^{-1}Z' = \left((ZW_1^{-1}Z')^{-1}\right)^{-1} = (ZZ')^{1/2}\left((ZW_1^{-1}Z')(ZZ')^{-1}\right)^{-1/2}.
\]

Hence, we have

\[
\text{tr}(ZW_1^{-1}Z') = \text{tr}(U_1U_2^{-1}),
\]

where \(U_1\) and \(U_2\) are \(d \times d\) random matrices defined by

\[
U_1 = ZZ', \quad U_2 = (ZZ')^{1/2}(ZW_1^{-1}Z')(ZZ')^{-1/2}.
\]

From a basic property of a Wishart distribution and th. 2.3.3 in Fujikoshi et al. (2010), we can see that \(U_1\) and \(U_2\) are independent, and

\[
U_1 \sim W_d(p, I_d), \quad U_2 \sim W_d(n-k-p+d, I_d).
\]

Let

\[
V_1 = \frac{1}{\sqrt{p}}(U_1 - pI_d), \quad V_2 = \frac{1}{\sqrt{h_0 + d}}[U_2 - (h_0 + d)I_d].
\]

Then, it is easy to see that \(V_1 = O_p(1)\) and \(V_2 = O_p(1)\), and

\[
\text{tr}(V_1) \xrightarrow{d} N(0, 2d) \text{ as } p \to \infty, \quad \text{tr}(V_2) \xrightarrow{d} N(0, 2d) \text{ as } (n-p) \to \infty.
\]

Notice that

\[
\frac{n}{h_0 + d} = \frac{n}{n-p} + O(n^{-2}), \quad \sqrt{\frac{p}{h_0 + d}} = \frac{p}{n-p} + O(p^{1/2}n^{-3/2}),
\]

\[
\frac{pd}{h_0 + d} = \frac{pd}{h_1} = O(pn^{-2}).
\]

By using the above equations, \(V_1\), and \(V_2\), \(V\) can be expanded as

\[
\text{tr}(V) = \frac{n}{\sqrt{p}} \left\{ \text{tr}(U_1U_2^{-1}) - \frac{dp}{h_1} \right\}
\]

\[
= \frac{n}{\sqrt{p}} \left\{ \frac{p}{h_0 + d} \text{tr} \left\{ \left( I_d + \frac{1}{\sqrt{p}} V_1 \right) \left( I_d + \frac{1}{\sqrt{h_0 + d}} V_2 \right)^{-1} \right\} - \frac{dp}{h_1} \right\}
\]

\[
= \left( \frac{n}{n-p} \right) \left( \text{tr}(V_1) - \frac{p}{n-p} \text{tr}(V_2) \right) + O(p^{1/2}n^{-1}).
\]

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

B. Proof of Theorem 2

Let \(H(\Lambda, O_{r,p}, \cdot)'Q'\) be a singular value decomposition of \(\Gamma\), where \(H\) is a \(p\)th orthogonal matrix satisfying \(H'H = H' = I_p\), \(Q\) is an \(r\)th orthogonal matrix satisfying \(Q'Q = QQ' = I_r\), and \(\Lambda\) is an \(r\)th diagonal matrix whose \(r\)th diagonal element is a singular value \(\lambda_r\), i.e., \(\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r)\) \((0 \leq \lambda_1 \leq \cdots \leq \lambda_r)\). Then, it is easy to see that \(V = UH\) and \(B = ZQ\) are independent, and
Let us partition \( V \) as \( V = (V_1, V_2) \), where \( V_1 \) and \( V_2 \) are \((n - k) \times r\) and \((n - k) \times (p - r)\) matrices, respectively. Then, we have

\[
U'U = H \begin{pmatrix} V'_1 V_1 & V'_1 V_2 \\ V'_2 V_1 & V'_2 V_2 \end{pmatrix} H'.
\]

By substituting (B.4) and (B.5) into (B.3), we derive

\[
W_1 = |U'U| = |V'_2 V_2||V'_1|^r(I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2)V_1|,
\]

\[
W_2 = |ZI' + U|^r(ZI' + U)|
\]

\[
= |V'_2 V_2||B\Lambda + V_1'|I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2|(B\Lambda + V_1)|.
\]

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

\[
T = \log \frac{|W_2|}{|W_1|} - \log |I_p + \Gamma'\Gamma|
\]

\[
= \log \left( \frac{|V'_1|^r(I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2)V_1|}{|B\Lambda + V_1|^r(I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2)(B\Lambda + V_1)|} \right) - \log |I_p + \Gamma'\Gamma|.
\]

Notice that \( V_1', V_2, \) and \( B \) are mutually independent, and

\[
\begin{align*}
V'_1(I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2)V_1 \sim W_r(n - k - p + r, I_r), \\
(B\Lambda + V_1')(I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2)(B\Lambda + V_1) \sim W_r(n - k - p + r, I_r + \Lambda^2).
\end{align*}
\]

Hence, we derive

\[
\frac{1}{n - k - p + r} V'_1(I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2)V_1 \xrightarrow{p} I_r,
\]

\[
\frac{1}{n - k - p + r} (I_r + \Lambda^2)^{-1/2} (B\Lambda + V_1')(I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2)(B\Lambda + V_1)(I_r + \Lambda^2)^{-1/2} \xrightarrow{p} I_r.
\]

These equations imply that

\[
\frac{1}{(n - k - p + r)} |V'_1(I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2)V_1| \xrightarrow{p} 1,
\]

\[
\frac{1}{(n - k - p + r) |I_r + \Lambda^2|} |(B\Lambda + V_1')(I_{n-k} - V_2(V'_2 V_2)^{-1} V'_2)(B\Lambda + V_1)| \xrightarrow{p} 1.
\]

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

\[
|I_r + \Lambda^2| = |I_r + \Gamma'\Gamma| = |I_p + \Gamma'\Gamma|.
\]

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 $S_{yy}$; this is required for proving Lemma 1 (the proof is given in Appendix G).

**Lemma A.1** Let $E$, $A_j$, and $B$ be mutually independent random matrices, which are defined by

$$E \sim N_{nq}(O_{nq}, I_p \otimes I_n), \quad A_j \sim N_{n(q-q_j)}(O_{n(q-q_j), I_q} \otimes I_n), \quad B \sim N_{nq}(O_{nq}, \Sigma_{xx} \otimes I_n),$$

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

$$nS_{yy,j} = \Sigma_{yy,x}^{1/2}(A_j \Gamma'_j + E)(I_n - P_j)(A_j \Gamma'_j + E)^\dagger \Sigma_{yy,x}^{1/2}, \quad (C.1)$$

where $P_j$ is the projection matrix to the subspace spanned by the columns of $B_j$, i.e., $P_j = B_j(B_j B_j)^{-1} B_j$, and $\Gamma_j$ is given by (20). In particular, when $j \in J^*$,

$$nS_{yy,j} = \Sigma_{yy,x}^{1/2}E'(I_n - P_j)E \Sigma_{yy,x}^{1/2}. \quad (C.2)$$

When $j \in J \setminus \{i\}$, it follows from Lemma A.1 that

$$F_{\text{min}}(j) - F_{\text{min}}(i) = -\log \frac{|nS_{yy,j}|}{|nS_{yy,i}|} = -\log \frac{|E'(I_n - P_j)E|}{|E'(I_n - P_i)E|}$$

$$= -\log \frac{|E'(I_n - P_j)E + E'(P_j - P_i)E|}{|E'(I_n - P_i)E|}. \quad (21)$$

Notice that $I_n - P_i$ and $P_j - P_i$ are idempotent matrices, and $(I_n - P_j)(P_j - P_i) = O_{n,n}$ holds. Hence, (19) is proved.

When $j \in J^*$, it follows from Lemma A.1 that

$$F_{\text{min}}(j) - F_{\text{min}(\omega)} = \log |nS_{yy,j}| - \log |nS_{yy,i}|$$

$$= \log \frac{|(A_j \Gamma'_j + E')(I_n - P_j)(A_j \Gamma'_j + E)|}{|E'(I_n - P_j)E|} - \log \frac{|E'(I_n - P_\omega)E|}{|E'(I_n - P_i)E|}$$

$$= \log \frac{|(A_j \Gamma'_j + E')(I_n - P_j)(A_j \Gamma'_j + E)|}{|E'(I_n - P_j)E|} + \log \frac{|E'(I_n - P_\omega)E|}{|E'(I_n - P_\omega)E|}$$

$$= \log \frac{|(A_j \Gamma'_j + E')(I_n - P_j)(A_j \Gamma'_j + E)|}{|E'(I_n - P_j)E|} + \log \frac{|E'(I_n - P_\omega)E + E'(P_\omega - P_j)E|}{|E'(I_n - P_\omega)E|}.$$

Notice that $I_n - P_j$, $I_n - P_\omega$, and $P_\omega - P_j$ are idempotent matrices, and $(I_n - P_\omega)(P_\omega - P_j) = O_{n,n}$ 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
By using the above equation, (23) is proved. On the other hand, it follows from (12) that

\[
\text{(1997), that}
\]

\[
\begin{align*}
\Sigma^{-1}_{xx} &= \begin{pmatrix}
\Sigma_{jj} & \Sigma_{jj} \\
\Sigma_{jj} & \Sigma_{jj}
\end{pmatrix}^{-1} = \begin{pmatrix}
\Sigma_{jj}^{-1} + \Sigma_{jj}^{-1} \Sigma_{jj} \Sigma_{jj}^{-1} - \Sigma_{jj}^{-1} \Sigma_{jj}^{-1} \\
-\Sigma_{jj}^{-1} \Sigma_{jj}^{-1} \Sigma_{jj}^{-1} & \Sigma_{jj}^{-1}
\end{pmatrix},
\end{align*}
\]

where \(\Sigma_{jj}\) are given by (21). By using the above equation, we have

\[
\Sigma'_{xy} \Sigma^{-1}_{xx} \Sigma_{xy} = \begin{pmatrix}
\Sigma'_{jj} \\
\Sigma'_{jj}
\end{pmatrix} \begin{pmatrix}
\Sigma_{jj}^{-1} + \Sigma_{jj}^{-1} \Sigma_{jj} \Sigma_{jj}^{-1} - \Sigma_{jj}^{-1} \Sigma_{jj}^{-1} \\
-\Sigma_{jj}^{-1} \Sigma_{jj}^{-1} \Sigma_{jj}^{-1} & \Sigma_{jj}^{-1}
\end{pmatrix} \begin{pmatrix}
\Sigma_{jj} \\
\Sigma_{jj}
\end{pmatrix} = \Sigma'_{jj} \Sigma_{jj}^{-1} + (\Sigma_{jj} - \Sigma_{jj}^{-1} \Sigma_{jj} \Sigma_{jj}^{-1}) \Sigma_{jj}^{-1} (\Sigma_{jj} - \Sigma_{jj}^{-1} \Sigma_{jj} \Sigma_{jj}^{-1}).
\]

(D.1)

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

\[
\Sigma'_{yy-x} \Sigma_{yy}^{-1/2} \Sigma'_{yy-x} = I_p + \Gamma_j' \Gamma_j.
\]

Hence, we have

\[
|I_p + \Gamma_j \Gamma_j'| = |\Sigma_{yy-x}^{-1/2} \Sigma_{yy}^{-1/2} \Sigma_{yy-x}^{-1/2}| = \frac{|\Sigma_{yy}|}{|\Sigma_{yy-x}|}.
\]

By using the above equation, (23) is proved. On the other hand, it follows from (12) that \(\Gamma_j = O_{p-q-q'}\) holds when \(j \in \mathcal{J}_r\). Hence, \(\delta_j = 0\) holds when \(j \in \mathcal{J}_r\).

E. Two Examples of \(\delta_j\)

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

First, we show the case that a limiting value of \(\delta_j\) is bounded. Let

\[
\Sigma_{xx} = \Omega(q), \quad \Sigma_{yy} = \Omega(p), \quad \Sigma_{xy} = \rho \begin{pmatrix}
I_q, \\
q, q/(1 + (q - 1)\rho)I_{q-q'}
\end{pmatrix} T_p,
\]

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

\[
\Omega(m) = (1 - \rho)I_m + \rho I_m T_m.
\]

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

\[
\Omega(m)^{-1} = \frac{1}{1 - \rho} \left\{I_m - \frac{\rho}{1 + (m - 1)\rho} I_m T_m \right\}.
\]

Then, we can see that

\[
|\Sigma_{yy-x}| = |\Sigma_{yy}| - |\Omega(p) - (\rho I_{1-q'} \Omega(q_{-1})^{-1}(\rho I_{1-q'} I_{1-q'})|
\]

\[
= \left|1 - \rho \left(I_p + \frac{\rho}{1 + (q - 1)\rho} I_{1-q'} \right) \right| = \frac{(1 - \rho)^{p}(1 + (p - 1)\rho)}{1 + (q - 1)\rho}.
\]

(E.1)

It follows from the same calculations as in (E.1) that
Asymptotic Behaviors of Log-Determinants of Two Wishart Matrices

\[
|\Sigma_{yy}| = \left| \Omega(p) - (\rho_1 \mathbf{1}_{q_j})' \Omega(q_j)^{-1} (\rho_1 \mathbf{1}_{q_j}) \right| = \frac{(1 - \rho)^p (1 + (p + q_j - 1) \rho)}{1 + (q_j - 1) \rho}.
\]  
\[(E.2)\]

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

\[
\delta_j = \log \frac{|\Sigma_{yy}|}{|\Sigma_{yy,s}|} = \log \frac{(1 + (p + q_j - 1)) [1 + (q_j - 1) \rho]}{[1 + (p + q_j - 1)] [1 + (q_j - 1) \rho]} 
\rightarrow \log \left\{ \frac{1 + (q_j - 1) \rho}{1 + (q_j - 1) \rho} \right\}.
\]

This equation indicates that the limiting value of \( \delta_j \) is bounded as \( p \to \infty \).

Next, the case that \( \delta_j \) approaches \( \infty \) is shown. Let

\[
\Sigma_{xx} = I_q, \quad \Sigma_{yy} = I_p, \quad \Sigma_{xy} = \left( \begin{array}{c} \mathbf{1}_q \\ 0_{q \times q_j} \end{array} \right) \alpha',
\]

where \( \alpha \) is a \( p \)-dimensional vector defined by

\[
\alpha = \sqrt{1 - \rho^2} (\sqrt{\rho, \ldots, \rho^p})'.
\]

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

\[
\alpha' \alpha = \frac{1 - \rho^2}{q^2} \sum_{k=1}^{p} (\rho^k)^2 = \frac{1}{q} \left[ 1 - (\rho^2)^p \right].
\]

Hence, we can see that

\[
|\Sigma_{yy,s}| = |\Sigma_{yy,s}| = \left| I_p - (\alpha \mathbf{1}_{q_j}')(\rho_1 \mathbf{1}_{q_j}) \right| = \left| I_p - q_j \alpha \alpha' \right| = 1 - \left[ 1 - (\rho^2)^p \right] = (\rho^2)^p.
\]

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

\[
|\Sigma_{yy}| = \left| I_p - (\alpha \mathbf{1}_{q_j}')(\rho_1 \mathbf{1}_{q_j}) \right| = \left[ 1 - \frac{q_j}{q} \left[ 1 - (\rho^2)^p \right] \right].
\]

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

\[
\delta_j = \log \frac{|\Sigma_{yy}|}{|\Sigma_{yy,s}|} = \log \left[ 1 - \frac{q_j}{q} \left[ 1 - (\rho^2)^p \right] \right] - p \log(\rho^2).
\]

This equation indicates that \( \delta_j \) approaches \( \infty \) as \( p \to \infty \).

**F. Proof of Corollary 2**

From a simple calculation, we have the following expansion:

\[
n \left( b(q_j + p) - b(q_j) - b(q_*) + b(q_1) \right) = \frac{np(2n - p)}{(n - p)^2} (q_j - q_*) + O(n^{-2} p).
\]

16
It follows from the above equation and (16) that

\[
m(j) - m(j_j) = \begin{cases} 
2p(q_j - q_s) & \text{(AIC)} \\
p(2 - p/n)(q_j - q_s)/(1 - p/n)^2 + O(p/n) & \text{(AIC)} \\
2p(q_j - q_s) + \hat{k}_{(j)} - \hat{k}_j - \hat{k}_{(j,j)} + \hat{k}_j & \text{(TIC)} \\
pq_j - q_s) \log n & \text{(BIC)} \\
q_j - q_s)(1 + \log n) & \text{(CAIC)} \\
2p(q_j - q_s) \log \log n & \text{(HQC)} 
\end{cases} \tag{F.1}
\]

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

\[
E[\hat{k}(D)] = \frac{d(d + 2)}{n(n + 2)}, \quad \text{Var}[\hat{k}(D)] = \frac{8d(d + 2)(n - 2)(n + 1)^2(n - d)(n - d + 2)}{n^4(n + 2)^2(n + 4)(n + 6)}.
\]

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

\[
\hat{k}_j \xrightarrow{p} 0, \quad \frac{1}{p} \hat{k}_{(j)} \xrightarrow{p} 0.
\]

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

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

Hence, it is easy to see that

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

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, and TIC if \(p\) goes to \(\infty\) and the following equation is satisfied:

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

where \(g(c)\) is given by (8). The above equation implies that condition C1 holds for the AIC and TIC if \(p\) goes to \(\infty\) and \(c \in [0, c_a]\), and it holds for the AIC\(_c\) if \(p\) goes to \(\infty\), because \(-g(x) + 2 > 0\) holds when \(x \in [0, c_a]\) 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:

\[
q_s - q_j \leq \lim_{n \to \infty, p/n \to c} \begin{cases} 
\frac{n\delta_j}{(2p)} & \text{(AIC, TIC)} \\
\frac{n\delta_j}{(1 - c)^2/(p(2 - c))} & \text{(AIC)} \\
\frac{n\delta_j}{p \log n} & \text{(BIC, CAIC)} \\
\frac{n\delta_j}{(2p \log n)} & \text{(HQC)} 
\end{cases}.
\]

It is easy to see that the above equation is satisfied if \(q_s - q_j \leq 0\), because \(\delta_j > 0\). Hence, it is sufficient to consider the case of \(j \in S_+\), where \(S_+\) is given in (29). Consequently, Corollary 2 is proved.
G.  Proof of Lemma A.1

Let \( Y = (y_1, \ldots, y_n)' \) and \( X = (x_1, \ldots, x_{n+1})' \), where \( y_i \) and \( x_i \) are the \( i \)th individuals from \( y \) and \( x \), respectively, and \( X_j \) denotes an \((n + 1) \times q_j \) matrix consisting of the columns of \( X \) indexed by the elements of \( j \). Then, \( nS_{yy,j} \) is expressed as

\[
nS_{yy,j} = Y' H_{n+1} \left( I_{n+1} - H_{n+1} X_j (X_j' H_{n+1} X_j)^{-1} X_j' H_{n+1} \right) H_{n+1} Y,
\]

where \( H_n \) is the projection matrix to the orthocomplement of the subspace spanned by \( 1_n \), i.e., \( H_n = I_n - 1_1 1_1' / n \), and \( 1_n \) is an \( n \)-dimensional vector of ones. Let \( V \) be an \( n \times p \) random matrix such that

\[
(B, V) \sim N_{n \times (q + p)}(O_{n \times (q + p)}, \Sigma),
\]

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

\[
nS_{yy,j} = V'(I_n - P_j)V, \tag{G.1}
\]

where \( P_j = B_j' B_j^{-1} B_j' \), and \( B_j \) is an \( n \times q_j \) matrix consisting of the columns of \( B \) 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

\[
V|B \sim N_{n \times p}(B \Sigma_{xx}^{-1} \Sigma_{xy}, \Sigma_{yy} \otimes I_n), \quad B_j|B_j \sim N_{n \times q}(B_j \Sigma_{jj}^{-1} \Sigma_{jj}, \Sigma_{jj} \otimes I_n),
\]

where \( \Sigma_{jj} \) is given by (21). Hence, we can express \( V \) as

\[
V = B \Sigma_{xx}^{-1} \Sigma_{xy} + E \Sigma_{yy}^{1/2}, \quad B_j = B_j \Sigma_{jj}^{-1} \Sigma_{jj} + A_j \Sigma_{jj}^{1/2}, \tag{G.2}
\]

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

\[
B \Sigma_{xx}^{-1} \Sigma_{xy} = (B_j, B_j) \begin{pmatrix} \Sigma_{jj}^{-1} + \Sigma_{jj}^{-1} \Sigma_{jj} \Sigma_{jj}^{-1} - \Sigma_{jj}^{-1} \Sigma_{jj}^{-1} & -\Sigma_{jj}^{-1} \\ -\Sigma_{jj}^{-1} & \Sigma_{jj}^{-1} \end{pmatrix}
\]

\[
= B_j \Sigma_{jj}^{-1} \{ \Sigma_{jj} - \Sigma_{jj}^{-1} \Sigma_{jj} \Sigma_{jj}^{-1} \Sigma_{jj} \} + B_j \Sigma_{jj}^{-1} \{ \Sigma_{jj} - \Sigma_{jj}^{-1} \Sigma_{jj} \Sigma_{jj}^{-1} \Sigma_{jj} \}.
\]

Substituting \( B_j \) in (G.2) into the above equation yields

\[
B \Sigma_{xx}^{-1} \Sigma_{xy} = B_j \Sigma_{jj}^{-1} \{ \Sigma_{jj} - \Sigma_{jj}^{-1} \Sigma_{jj} \} + (B_j \Sigma_{jj}^{-1} + A_j \Sigma_{jj}^{1/2}) \{ \Sigma_{jj} - \Sigma_{jj}^{-1} \Sigma_{jj} \} = B_j \Sigma_{jj}^{-1} \Sigma_{jj} + A_j \Gamma_j^{1/2} \Sigma_{yy}^{1/2},
\]

where \( \Gamma_j \) is the \( p \times (q - q_j) \) matrix defined by (20). Substituting the above equation into \( V \) in (G.2)
Yanagihara, H., Hashiyama, Y. & Fujikoshi, Y.

yields

\[ V = B_j \Sigma^{-1}_{jj} \Sigma_{yy} + A_j \Gamma_j \Sigma^{1/2}_{yy} + E \Sigma^{1/2}_{yy, xx}. \]  

(G.3)

Notice that \((P_j - I)B_j = O_{p-q}\). By using (G.1) and (G.3), (C.1) is proved. Moreover, \(\Gamma_j = O_{p-q-q_j}\) if \(j \in J_+\), because then (12) holds. This implies that for any \(j \in J_+\)

\[ V = B_j \Sigma^{-1}_{jj} \Sigma_{yy} + E \Sigma^{1/2}_{yy, xx}. \]  

(G.4)

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