

Article

High-Dimensionality-Adjusted Asymptotically Loss- and Mean-Efficient GC_p Criterion for Normal Multivariate Linear Regression Models

Hirokazu Yanagihara ^{1,2,3}¹ Osaka Central Advanced Mathematical Institute, Osaka Metropolitan University, 3-3-138 Sugimoto, Sumiyoshi-ku, Osaka 558-8585, Japan; yanagi-hiro@omu.ac.jp² Department of Medical Statistics, Research & Development Center, Osaka Medical and Pharmaceutical University, 2-7 Daigaku-machi, Takatsuki, Osaka 569-8686, Japan³ Mathematical Risk Analysis, Risk Analysis Research Center, The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

Abstract

A variable selection method is put forward for multivariate linear regression models which obey normality. This method hinges on minimizing a generalized C_p (GC_p) criterion which is defined by adding a positive constant value (the product of α and the number of parameters in the mean structure) to the minimum value of the multivariate residual sum of squares. The paper seeks to clarify the sufficient condition for α to simultaneously satisfy asymptotically loss- and mean-efficient properties in an asymptotic framework such that the sample size always goes to ∞ , but the dimension of the vector of response variables can be either fixed or infinite. Based on this, we propose an asymptotically loss- and mean-efficient GC_p criterion by using α , which satisfies the obtained sufficient condition even with high dimensionality of the vector of response variables.

Keywords: loss efficiency; generalized C_p criterion; high-dimensionality; mean efficiency; selection probability; variable selection

MSC: 62H12; 62F12; 62J05

1. Introduction

The multivariate linear regression model is central to theoretical and applied inferential analysis. This model is introduced in many statistical textbooks (see, e.g., chap. 9 [1]; chap. 4 [2]) and is widely used in chemometrics, engineering, econometrics, psychometrics, and many other fields for the prediction of response variables to a set of explanatory variables (see, e.g., [3–6]). Numerous practical guides for applying this model are also available, such as [7,8]. Let $Y = (y_1, \dots, y_n)^\top$ be an $n \times p$ matrix of p response variables, and let $X = (x_1, \dots, x_n)^\top$ be an $n \times k$ matrix of non-stochastic k explanatory variables, where n is the sample size. To ensure both the possibility of estimating the model and the existence of variable selection criteria, we assume that $\text{rank}(X) = k$ ($< n$) and $n - p - k - 1 > 0$. Let j denote a subset of $\omega = \{1, \dots, k\}$ containing k_j elements and X_j denote the $n \times k_j$ matrix consisting of the columns of X indexed by the elements of j , where k_A denotes the number of elements of a set A , i.e., $k_A = \#(A)$. For example, if $j = \{1, 2, 4\}$, then X_j consists of the first, second, and fourth columns of X . For convenience, elements of j are arranged in ascending order. We then consider the following multivariate linear regression model with k_j explanatory variables as the candidate model:



Academic Editors: Sungsu Kim, Chi Seng Pun and Brani Vidakovic

Received: 30 April 2026

Revised: 5 July 2026

Accepted: 15 July 2026

Published: 16 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

$$Y \sim \mathcal{N}_{n \times p}(X_j \Theta_j, \Sigma_j \otimes I_n), \quad (1)$$

where Θ_j is a $k_j \times p$ unknown matrix of regression coefficients, and Σ_j is a $p \times p$ unknown covariance matrix with $\text{rank}(\Sigma_j) = p$. We identify the candidate model by set j and denote the candidate model in (1) as model j . Importantly, the model with X_ω (namely, X) is called the full model. We will assume that the data are generated from the following true model:

$$Y \sim \mathcal{N}_{n \times p}(X_{j_*} \Theta_*, \Sigma_* \otimes I_n), \quad (2)$$

where j_* is a set of integers indicating the subset of explanatory variables in the true model, Θ_* is a $k_{j_*} \times p$ matrix of true regression coefficients, and Σ_* is a $p \times p$ true covariance matrix with $\text{rank}(\Sigma_*) = p$. We denote the model in (2) as the true model j_* . Henceforth, for expository purposes, we simply represent X_{j_*} and k_{j_*} as X_* and k_* , respectively.

In what follows, we focus on a variable selection method by minimizing a generalized C_p (GC_p) criterion, when p may be large but still smaller than n . The GC_p criterion is defined by adding a positive constant value to the minimum value of the multivariate residual sum of squares. This positive value is the product of α and the number of parameters in the mean structure, where α expresses a penalty for the complexity of the candidate model. The GC_p criterion in univariate and multivariate linear regression model contexts was proposed by Atkinson [9] and Nagai et al. [10], respectively. The family of GC_p criteria contains many widely known variable selection criteria, e.g., the Mallows C_p criterion proposed by Sparks et al. [11] (the original C_p was proposed by Mallows [12], within the univariate linear regression model, and Mallows [13] further discussed its interpretation in terms of prediction error and model selection) and the modified C_p (MC_p) criterion proposed by Fujikoshi and Satoh [14], which is a completely bias-corrected version of the C_p criterion. Given that our focus is on multivariate linear regression models where the dimension p may be large, the following asymptotic framework is used for assessing the asymptotic property of the variable-selection method:

$$n \rightarrow \infty \text{ and } p/n \rightarrow c_0 \in [0, 1). \quad (3)$$

For simplicity, we will use “ $n \rightarrow \infty, p/n \rightarrow c_0$ ” to refer to this asymptotic framework. It should be emphasized that we are not concerned whether p goes to ∞ or not in this asymptotic framework. If p increases, the columns of Θ_* and the rows and columns of Σ_* increase, so it is necessary to clarify how true parameters increase. We assume that as p increases by 1, a new column is added to the right of the current Θ_* , and a new column and row are added to the right of and the bottom of the current Σ_* , respectively.

There are two important properties of variable selection criteria. One is a consistency property whereby the selection probability of the true model by a variable selection criterion converges to 1 asymptotically. The second is an efficiency property which can be divided into two sub-properties. The first is asymptotic loss efficiency whereby the ratio of the minimum loss function to the loss function of the model selected by a variable selection criterion goes to 1 asymptotically. The second is asymptotic mean efficiency whereby the ratio of the minimum expected loss function to the expected loss function of the model selected by a variable selection criterion goes to 1 asymptotically (see, e.g., [15–17]).

Yanagihara [18] clarified the sufficient condition for α in the GC_p criterion to satisfy a consistency property under the asymptotic framework in (3) and proposed the high-dimensionality-adjusted consistent GC_p ($HCGC_p$) by using the sufficient condition. An aim of this paper is to derive the sufficient condition for α in the GC_p criterion to simultaneously satisfy loss and mean efficiency properties under the asymptotic framework in (3). We

also propose the asymptotically loss- and mean-efficient GC_p criterion by using α , which satisfies the obtained sufficient condition even with high dimensionality of the vector of response variables. In our setting, the true model is included in the set of all possible candidate models. For asymptotic loss and mean efficiency for settings different from ours, the true model is not included in the set of candidate models; see [19].

It has been widely recognized that the Akaike information criterion (AIC) [20] is not consistent, whereas the Bayesian information criterion (BIC) [21] is consistent for model selection (see, e.g., [22]). However, Yanagihara et al. [23] showed that, in variable selection for multivariate regression models, the consistency properties may be reversed when $p \rightarrow \infty$, that is, AIC can become consistent while BIC becomes inconsistent. A similar phenomenon was also reported by Fujikoshi et al. [24], which appeared earlier due to differences in the publication process. Such a reversal has also been reported for C_p -type criteria. Fujikoshi et al. [24] showed that the C_p criterion, which is inconsistent under the ordinary asymptotic framework in which only $n \rightarrow \infty$, becomes consistent when $p \rightarrow \infty$ provided that $c_0 < 0.5$. On the other hand, the BIC-type C_p (BC_p) criterion, which is the GC_p criterion with $\alpha = \log n$, is consistent under the ordinary asymptotic framework. In the univariate case, this follows directly from the results of Rao and Wu [25] and Bai et al. [26]. In contrast to the C_p criterion, Yanagihara [18] showed that the BC_p criterion may become inconsistent when $p \rightarrow \infty$. These examples demonstrate that reexamining existing model selection criteria with $p \rightarrow \infty$ may lead to new insights that differ substantially from those obtained under the ordinary asymptotic framework.

The remainder of the paper is organized as follows. In Section 2, we present the necessary notation and assumptions for assessing the loss- and mean-efficient properties of the GC_p criterion in model j (1). The main results are shown in Section 3. In Section 4, we present the results of numerical experiments and compare the variable selection performance of the proposed criterion with those of existing four variable selection criteria, i.e., C_p , MC_p , BC_p and $HCGC_p$ criteria. Technical details are provided in the Appendix A.

2. Formulation of Loss- and Mean-Efficient Properties

First, we describe several classes of j which express subsets of X in the candidate model. Let $\mathcal{J}$ be a set of all possible candidate models denoted by $\mathcal{J} = \wp(\omega)$, where $\wp(A)$ is the powerset of set A . We assume the following regarding $\mathcal{J}$:

A1. The true model is included in the set of candidate models, i.e., $j_* \in \mathcal{J}$.

Moreover, we separate $\mathcal{J}$ into two sets, one a set of overspecified models wherein the explanatory variables contain all the explanatory variables of the true model j_* in (2) and the other a set of underspecified models, i.e.,

$$\mathcal{J}_+ = \{j \in \mathcal{J} | j_* \subseteq j\}, \quad \mathcal{J}_- = \mathcal{J}_+^c \cap \mathcal{J}, \quad (4)$$

where A^c denotes the complement of a set A . We use the terminology “overspecified model” and “underspecified model” in the same sense as [14].

Let S_j be the unbiased estimator of Σ_j in model j (1), i.e.,

$$S_j = \frac{1}{n-k} Y^\top (I_n - P_j) Y, \quad (5)$$

where P_j is the projection matrix to the subspace spanned by the columns of X_j , i.e., $P_j = X_j(X_j^\top X_j)^{-1}X_j^\top$. Let $d(A, B)$ and $\hat{d}(A, B)$ be the squared Mahalanobis' distances between two $n \times p$ matrices A and B as

$$d(A, B) = \text{tr}\left\{(A - B)\Sigma_*^{-1}(A - B)^\top\right\} = \|\Sigma_*^{-1/2}(A - B)\|^2,$$

$$\hat{d}(A, B) = \text{tr}\left\{(A - B)S_\omega^{-1}(A - B)^\top\right\} = \|S_\omega^{-1/2}(A - B)\|^2.$$

The matrix of regression coefficients Θ_j in model j (1) is estimated using the least squares approach; i.e., Θ_j is estimated by minimizing the multivariate RSS and then the estimator of Θ_j is given by

$$\hat{\Theta}_j = (X_j^\top X_j)^{-1}X_j^\top Y = \arg \min_{\Theta_j \in \mathbb{M}(k_j, p)} \hat{d}(Y, X_j \Theta_j),$$

where $\mathbb{M}(k, p)$ denotes a set of $k \times p$ real matrices.

Let a $p \times p$ noncentrality matrix be denoted by

$$\Delta_j = \frac{1}{n}\Sigma_*^{-1/2}\Theta_*^\top X_*^\top (I_n - P_j)X_*\Theta_*\Sigma_*^{-1/2},$$

where Θ_* , Σ_* , and X_* are the matrix of the true regression coefficients, the true covariance matrix, and the matrix of the true explanatory variables, respectively, given by (2). It is straightforward to discern from the definition of $\mathcal{J}_+$ that $\Delta_j = O_{p,p}$ if and only if $j \in \mathcal{J}_+$, where $O_{n,p}$ is an $n \times p$ matrix of zeros. For the noncentrality matrix, we assume that

A2. $R_0 = \lim_{n \rightarrow \infty} \frac{1}{n}X^\top X$ exists and is positive definite.

Assumption A2 is a commonly used condition in studies of the asymptotic behaviors of regression coefficient estimators. A weaker condition that is frequently adopted in the literature is that the limit inferior of the smallest eigenvalue of $X^\top X/n$ is strictly positive. Although this weaker condition is more general, the present study requires the exact convergence of the differences between GC_p criteria. Since such exact convergence is guaranteed by Assumption A2 but not necessarily by the weaker condition, we adopt Assumption A2 throughout this paper. Let

$$\delta_j = \text{tr}(\Delta_j). \quad (6)$$

If Assumption A2 holds, then we have

$$\forall j \in \mathcal{J}_-, \quad \inf_{n > p+k, p \geq 1} \delta_j > 0.$$

Even though δ_j diverges, we do not assume a specific order of the divergence. Instead, we use the following assumption:

A3. $\exists \epsilon \in \mathbb{N}$ s.t. $\lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n^\epsilon} \Theta_*^\top \Sigma_*^{-1} \Theta_*^\top = O_{k_*, k_*}$.

Assumption A3 specifies the asymptotic behavior of the newly added components of the true coefficient matrix and covariance matrix as p increases. The assumption allows these components to increase or decrease with p , provided that their effect on $\Theta_*^\top \Sigma_*^{-1} \Theta_*^\top$ is of at most polynomial order. In contrast, Assumption A3 excludes extreme situations in which $\Theta_*^\top \Sigma_*^{-1} \Theta_*^\top$ grows exponentially fast. Such situations correspond to cases where the

signal-to-noise ratio increases exponentially and the noise component becomes negligible as p increases. If Assumptions A2 and A3 hold, then we have

$$\exists \epsilon \in \mathbb{N} \quad \text{s.t.} \quad \forall j \in \mathcal{J}_-, \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{\delta_j}{n^\epsilon} = 0.$$

We determine the best subset of ω by minimizing a GC_p criterion. The GC_p criterion for model j is defined by

$$GC_p(j|\alpha) = (n - k_j) \text{tr}(\mathbf{S}_j \mathbf{S}_\omega^{-1}) + \alpha p k_j, \quad (7)$$

where α is a positive value expressing a penalty for the complexity of the candidate model. A variable selection criterion included in the family of GC_p criteria is specified by an individual α . This family contains the C_p , MC_p , BC_p and $HCGC_p$ criteria as special cases, i.e.,

$$\alpha = \begin{cases} 2 & (C_p) \\ 2c_M^{-1} & (MC_p) \\ \log n & (BC_p) \\ a_C(\beta) & (HCGC_p) \end{cases},$$

where c_M is a positive value given by

$$c_M = \frac{n - p - k - 1}{n - k} = 1 - \frac{p + 1}{n - k}, \quad (8)$$

$a_C(\beta)$ is a positive value expressed as a function of β and defined by

$$a_C(\beta) = \frac{n}{n - p} + \beta, \quad (9)$$

and β is a positive value satisfying

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} \sqrt{p} \beta = \infty, \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{p}{n} \beta = 0. \quad (10)$$

The best subset of ω , which is chosen by minimizing the GC_p criterion, is written as

$$\hat{j}_\alpha = \arg \min_{j \in \mathcal{J}} GC_p(j|\alpha). \quad (11)$$

Herein, the selected model is denoted as $\hat{j}_\alpha$.

The squared loss function between the true mean and the fitted value, i.e., $\mathbf{X}_* \boldsymbol{\Theta}_*$ and $\mathbf{X}_j \hat{\boldsymbol{\Theta}}_j$, is defined by

$$\mathcal{L}(j) = d(\mathbf{X}_* \boldsymbol{\Theta}_*, \mathbf{X}_j \hat{\boldsymbol{\Theta}}_j) = n \delta_j + \text{tr}(\boldsymbol{\mathcal{E}}^\top \mathbf{P}_j \boldsymbol{\mathcal{E}}), \quad (12)$$

where $\boldsymbol{\mathcal{E}} = (\mathbf{Y} - \mathbf{X}_* \boldsymbol{\Theta}_*) \boldsymbol{\Sigma}_*^{-1/2} \sim \mathcal{N}_{n \times p}(\mathbf{O}_{n,p}, \mathbf{I}_p \otimes \mathbf{I}_n)$. Notice that $E[\text{tr}(\boldsymbol{\mathcal{E}}^\top \mathbf{P}_j \boldsymbol{\mathcal{E}})] = p k_j$. Hence, the expectation of $\mathcal{L}(j)$, denoted by $\mathcal{R}(j)$, is given by

$$\mathcal{R}(j) = E[\mathcal{L}(j)] = n \delta_j + k_j p. \quad (13)$$

Subsets of ω minimizing $\mathcal{L}(j)$ and $\mathcal{R}(j)$ are written as

$$\hat{j}_{LE} = \arg \min_{j \in \mathcal{J}} \mathcal{L}(j), \quad \hat{j}_{ME} = \arg \min_{j \in \mathcal{J}} \mathcal{R}(j), \quad (14)$$

Next, $\hat{j}_{LE}$ and j_{ME} are the loss optimal model and the mean optimal model in the sense of minimizing the loss function and the mean of the loss function, respectively. By using the loss function, we formulate the asymptotically loss-efficient property.

Definition 1. If the following equation holds, we say that the GC_p criterion is asymptotically loss-efficient:

$$\text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{\mathcal{L}(\hat{j}_\alpha)}{\mathcal{L}(\hat{j}_{LE})} = 1,$$

where plim is a probability limit operator, i.e., denotes a convergence in probability.

It should be kept in mind that

$$\mathcal{L}(\hat{j}_\alpha) = \sum_{j \in \mathcal{J}} \mathcal{L}(j) I(\hat{j}_\alpha = j), \quad \mathcal{L}(\hat{j}_{LE}) = \sum_{j \in \mathcal{J}} \mathcal{L}(j) I(\hat{j}_{LE} = j), \quad (15)$$

where $I(A)$ is the indicator function, i.e., $I(A) = 1$ if A is true and $I(A) = 0$ if A is not true. By using the mean of the loss function, we formulate the asymptotically mean-efficient property.

Definition 2. If the following equation holds, we say that the GC_p criterion is asymptotically mean-efficient:

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{E[\mathcal{L}(\hat{j}_\alpha)]}{\mathcal{R}(j_{ME})} = 1.$$

3. Main Results

3.1. Asymptotically Loss-Efficient Property

To clarify the sufficient condition for α in (7) for loss efficiency, we consider asymptotic behaviors of the loss function $\mathcal{L}(j)$, the loss optimal model $\hat{j}_{LE}$, and the selected model $\hat{j}_\alpha$, given by (12), (14), and (11), respectively.

First, we prepare three conditions for α which play an important role in determining whether the GC_p criterion is efficient.

$$\text{C1.} \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \sqrt{p} \left(-\frac{n}{n-p} + \alpha \right) = \infty.$$

$$\text{C2.} \quad \forall j \in \mathcal{J}_-,$$

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{p}{n\delta_j} \left(-\frac{n}{n-p} + \alpha \right) = 0. \quad (16)$$

$$\text{C3.} \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{p}{n} \alpha = \frac{2c_0}{1-c_0}.$$

Conditions C1 and C2 are sufficient conditions for α to satisfy consistency property (see [18]). Roughly speaking, C1 controls the growth rate of the penalty term and is used to eliminate overspecified models, whereas C2 is required to eliminate underspecified models. As shown later lemma, for certain settings, consistency property implies efficiency property, and therefore these conditions are also required for satisfying efficiency property. When the relationship between consistency and efficiency properties breaks down, the additional Condition C3 becomes necessary. Condition C3 ensures that the asymptotic limit of the difference between the GC_p criteria coincides, up to a constant multiple, with that of the corresponding risk functions, which is essential for satisfying efficiency property.

Let θ_a^* ($a \in j_*$) be the p -dimensional vector of regression coefficients for the a th explanatory variable. Then we prepare the following subsets of j_* and $\mathcal{J}_-$:

$$\xi_* = \left\{ a \in j_* \mid \theta_a^{*\top} \Sigma_*^{-1} \theta_a^* = O(1) \text{ as } p \rightarrow \infty \right\}, \quad \mathcal{S}_- = \{j \in \mathcal{J}_- \mid j \supseteq (\xi_*^c \cap j_*)\}, \quad (17)$$

where $\mathcal{J}_-$ is the set of underspecified models given by (4). See the simulation settings in Section 4 for concrete examples of ξ_* . The set ξ_* is introduced because the asymptotic behavior of δ_j given by (6) plays a key role in the condition for α to satisfy efficiency properties. Specifically, the conditions differ according to whether some δ_j converge or all δ_j diverge. This distinction is determined by the existence of explanatory variables $a \in j_*$ not satisfying $\theta_a^{*\top} \Sigma_*^{-1} \theta_a^* = O(1)$ as $p \rightarrow \infty$. Based on the distinction, the following Lemma 1 shows the convergence or divergence of δ_j (the proof is given in Appendix A.1).

Lemma 1. Suppose that Assumption A2 holds.

(1) When $j \in \mathcal{S}_-$, δ_j converges to a positive value as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$, i.e.,

$$\forall j \in \mathcal{S}_-, \quad \exists \delta_{j,0} > 0 \quad \text{s.t.} \quad \delta_{j,0} = \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \delta_j.$$

(2) When $j \in \mathcal{S}_-^c \cap \mathcal{J}_-$, δ_j diverges to ∞ as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$, i.e.,

$$\forall j \in \mathcal{S}_-^c \cap \mathcal{J}_-, \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \delta_j = \infty.$$

Lemma 1 indicates that δ_j in $j \in \mathcal{J}_-$ is not convergent if $\xi_* = \emptyset$ because $\mathcal{S}_- = \emptyset$ when $\xi_* = \emptyset$. Here, let $\delta_{j,0} = 0$ for $j \in \mathcal{J}_+$, where $\mathcal{J}_+$ is the set of overspecified models given by (4). By using Lemma 1 and the definition of $\mathcal{L}(j)$, we show that $\mathcal{L}(j)$ can converge to a positive value as follows (the proof is given in Appendix A.2).

Lemma 2. Suppose that Assumption A2 holds. If $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$, $\mathcal{L}(j)$ for any $j \in \mathcal{S}_- \cup \mathcal{J}_+$ converges to a positive value as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$, i.e.,

$$\forall j \in \mathcal{S}_- \cup \mathcal{J}_+, \quad \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{\mathcal{L}(j)}{n} = \delta_{j,0} + c_0 k_j = \lambda(j) > 0. \quad (18)$$

Let j_0 be the model defined by

$$j_0 = \begin{cases} \arg \min_{j \in \{j_*\} \cup \mathcal{S}_-} \lambda(j) & (\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}) \\ j_* & (\{\xi_* = \emptyset\} \cup \{c_0 = 0\}) \end{cases}. \quad (19)$$

By using j_0 and Lemma 2, the following lemma concerning behaviors of the loss optimal model is obtained (the proof is given in Appendix A.3).

Lemma 3. Suppose that Assumptions A1 and A2 hold.

(1) The loss optimal model is the true model or an underspecified model with probability 1, i.e.,

$$P(\hat{j}_{\text{LE}} \in \{j_*\} \cup \mathcal{J}_-) = 1.$$

(2) The probability of $\hat{j}_{\text{LE}} = j_0$ converges to 1 as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$, i.e.,

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} P(\hat{j}_{\text{LE}} = j_0) = 1. \quad (20)$$

Since Equation (20) holds, we call j_0 in (19) the asymptotically loss optimal model.

The following key lemma to prove the efficient property of the GC_p criterion is obtained (the proof is given in Appendix A.4).

Lemma 4. Suppose that

$$\exists \ell \in \mathcal{J} \quad \text{s.t.} \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} P(\hat{j}_{LE} = \ell) = 1, \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} P(\hat{j}_\alpha = \ell) = 1.$$

Then, the GC_p criterion is asymptotically loss-efficient when $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$.

From Lemmas 3 and 4, we can discern that the GC_p criterion is asymptotically loss-efficient if $P(\hat{j}_\alpha = j_0) \rightarrow 1$ as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Probability convergence is ensured by the following lemma (the proof is given in Appendix A.5).

Lemma 5. Suppose that Assumptions A1 and A2 hold. The selection probability of j_0 by the GC_p criterion converges to 1 as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$, i.e.,

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} P(\hat{j}_\alpha = j_0) = 1,$$

if α satisfies Conditions C1 and C2 when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$, or Condition C3 when $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$, where j_0 is the asymptotically loss optimal model given by (19).

From Lemmas 3–5, we derive the following main theorem concerning the asymptotically loss-efficient property.

Theorem 1. Suppose that Assumptions A1 and A2 hold. The GC_p criterion is asymptotically loss-efficient when $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$ if α satisfies Conditions C1 and C2 when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$, or Condition C3 when $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$.

3.2. Asymptotically Mean-Efficient Property

To clarify the sufficient condition for α in (7) for mean efficiency, asymptotic behaviors of the mean of the loss function $\mathcal{R}(j)$, the mean optimal model j_{ME} , and the selected model $\hat{j}_\alpha$, given by (13), (14), and (11), respectively, are considered.

First, we present a lemma concerning convergence of $\mathcal{R}(j_{ME})/p$ (the proof is given in Appendix A.6).

Lemma 6. Suppose that Assumptions A1 and A2 hold. The minimum of $\mathcal{R}(j)/p$ is expanded as

$$\frac{1}{p} \mathcal{R}(j_{ME}) = \frac{1}{p} \mathcal{R}(j_0) + o(1), \quad \text{as } n \rightarrow \infty, \quad p/n \rightarrow c_0,$$

where j_0 is the asymptotically loss optimal model given by (19).

We derive the order of the selection probability of some underspecified model by the GC_p criterion (the proof is given in Appendix A.7).

Lemma 7. Suppose that Assumptions A1, A2, and A3 hold. If α satisfies Condition C2, $np^{-1}\delta_j P(\hat{j}_\alpha = j)$ for any $j \in \mathcal{J}_-$ converges to 0 as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$. If α satisfies Condition C3, $np^{-1}\delta_j P(\hat{j}_\alpha = j)$ for any $j \in \mathcal{S}_-^c \cap \mathcal{J}_-$ converges to 0 as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$.

By using Lemma 7, the following theorem concerning convergence of $E[\mathcal{L}(\hat{j}_\alpha)]/p$ is derived (the proof is given in Appendix A.8).

Lemma 8. Suppose that Assumptions A1, A2, and A3 hold. Then $E[\mathcal{L}(\hat{j}_\alpha)]/p$ is expanded as

$$\frac{1}{p}E[\mathcal{L}(\hat{j}_\alpha)] = \frac{1}{p}\mathcal{R}(j_0) + o(1), \quad \text{as } n \rightarrow \infty, \quad p/n \rightarrow c_0,$$

if α satisfies Conditions C1 and C2 when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$, or Condition C3 when $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$, where j_0 is the asymptotically loss optimal model given by (19).

Finally, from Lemmas 6 and 8, we obtain the following main theorem concerning the asymptotically mean-efficient property.

Theorem 2. Suppose that Assumptions A1, A2, and A3 hold. The GC_p criterion is asymptotically mean-efficient when $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$ if α satisfies Conditions C1 and C2 when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$ or Condition C3 when $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$.

3.3. A High-Dimensionality-Adjusted Efficient GC_p Criterion

From Theorems 1 and 2, we can discern that sufficient conditions for α to satisfy asymptotic loss efficiency are equal to those to satisfy asymptotic mean efficiency. By using the α conditions, an asymptotically loss- and mean-efficient GC_p criterion can be proposed. Unfortunately, the conditions for α in Theorems 1 and 2 are not verifiable in practical terms because they depend on the true model. Hence, we attempt to derive sufficient conditions for α that do not depend on the true model. This is achieved by invoking the following corollary (the proof is given in Appendix A.9).

Corollary 1. Suppose that Assumptions A1, A2, and A3 hold. The GC_p criterion is asymptotically loss- and mean-efficient when $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$ if α satisfies Conditions C1 and C3.

Let $a_E(\beta)$ be a positive value expressed as a function of positive value β in (10), which is defined by

$$a_E(\beta) = \frac{2n}{n-p} + \beta. \quad (21)$$

The conditions of β imply that

$$\begin{aligned} \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \sqrt{p} \left(-\frac{n}{n-p} + a_E(\beta) \right) &= \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \left(\frac{\sqrt{pn}}{n-p} + \sqrt{p}\beta \right) = \infty, \\ \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{p}{n} a_E(\beta) &= \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \left(\frac{2p}{n-p} + \frac{p}{n}\beta \right) = \frac{2c_0}{1-c_0}. \end{aligned}$$

The above equations indicate that $a_E(\beta)$ satisfies Conditions C1 and C3. This means that the GC_p criterion with $\alpha = a_E(\beta)$ is asymptotically loss- and mean-efficient under the asymptotic framework in (3). We call the GC_p criterion with $\alpha = a_E(\beta)$ a high-dimensionality-adjusted efficient GC_p ($HEGC_p$) criterion. The existing $HCGC_p$ criterion was developed with the same high-dimensional setting as the proposed $HEGC_p$ criterion, $\alpha = a_C(\beta)$, where $a_C(\beta)$ is given by (9). The relationship between the $HEGC_p$ and $HCGC_p$ criteria is given by

$$a_E(\beta) = \frac{n}{n-p} + a_C(\beta).$$

Due to the additional term $n/(n-p)$, $a_E(\beta)$ satisfies Condition C3. Consequently, the $HEGC_p$ criterion achieves both loss efficiency and mean efficiency. The proposed $HEGC_p$ criterion is also related to the existing MC_p criterion. We consider the criterion obtained by replacing α in the MC_p criterion with

$$2c_M^{-1} + \beta,$$

where c_M is the constant given by (8). Let

$$r_M = c_M^{-1} - \frac{n}{n-p} = \frac{n+kp}{(n-p)(n-k-p-1)}.$$

Since $r_M > 0$ holds, we have

$$2c_M^{-1} + \beta = a_E(2r_M + \beta).$$

Furthermore,

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} \sqrt{p}(2r_M + \beta) = \infty, \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{p}{n}(2r_M + \beta) = 0.$$

Hence, regarding $2r_M + \beta$ as a new β , the criterion with $\alpha = 2c_M^{-1} + \beta$ is also an $HEGC_p$ criterion. Therefore, a subclass of $HEGC_p$ criteria can be interpreted as MC_p criteria with an additional adjustment term β . Among the existing GC_p criteria, no criterion employs an α that guarantees efficiency in all settings considered in this paper. Nevertheless, the MC_p criterion is unique in that, although it is not always efficient, it does achieve efficiency when p diverges. This is because the choice $\alpha = 2c_M^{-1}$ used in the MC_p criterion satisfies Conditions C2 and C3; however, it satisfies Condition C1 only when p approaches ∞ . This property makes MC_p the closest existing GC_p criterion to being universally efficient.

Since β is required to satisfy two convergence conditions in (10), infinitely many choices are possible. In this paper, however, β is determined according to the following considerations.

(1) Choice of the base term

We first adopt a term that is commonly used in consistent information criteria under the ordinary asymptotic framework in which only $n \rightarrow \infty$. Such a term should diverge more slowly than any positive power of n . Typical examples include the $\log n$ term used in the BIC and the Consistent AIC (CAIC) [27]. Indeed, the $HCGC_p$ criterion employs $\log n$. A smaller penalty can be obtained by replacing $\log n$ with $\log \log n$, which was originally used in the Hannan–Quinn Information Criterion (HQC) [28].

(2) Adjustment for the convergence conditions on β

The base term is then adjusted so that the convergence conditions imposed on β are satisfied. The simplest approach is to divide it by $\sqrt{p}$. Although other adjustments are possible, this modification automatically satisfies both convergence conditions.

(3) Adjustment based on the size of c_0

Since this paper treats c_0 as a constant, under the asymptotic framework (3) used here, the asymptotic result remains unchanged regardless of the value of c_0 , provided it is between 0 and 1. However, in the finite sample size, the size of c_0 contributes significantly to the actual results. This is evident from the fact that the asymptotic standard deviation of the difference in GC_p criteria between two overspecified models is proportional to $\{n/(n-p)\}^{3/2}$ (see [18]). Therefore, we consider it preferable to adjust α using $n/(n-p)$ so that a larger penalty is imposed as c_0 approaches 1. In fact, [18] proposes an $HCGC_p$ criterion that satisfies consistency using β , employing the asymptotic standard deviation itself.

As mentioned earlier, MC_p is the closest existing GC_p criterion to being universally efficient. This fact motivates our consideration of a specific β . First, let us consider a criterion where an adjusting term is added to the α of the MC_p criterion, i.e., defined as $\alpha = 2c_M^{-1} + \beta$. Recall that $2c_M^{-1} + \beta = a_E(2r_M + \beta)$. We note that the proposed $HEGC_p$ criterion satisfies $a_E(\beta) > a_C(\beta)$. Hence, to avoid imposing an excessively large penalty, we employ the more slowly diverging term $\log \log n$ but $\log n$. Furthermore, the term r_M already depends on c_0 and increases as c_0 approaches 1. Therefore, the effect of c_0

is partially incorporated through r_M . For this reason, it is preferable not to impose an additional adjustment that is excessively sensitive to c_0 through β . Hence, instead of using $\{n/(n-p)\}^{3/2}$, we employ the milder adjustment $\{n/(n-p)\}^{1/2}$. Based on these considerations, we propose

$$b_E = \sqrt{\frac{n}{n-p} \frac{\log \log n}{\sqrt{p}}}, \quad (22)$$

as a practical choice of β . In particular, we call a GC_p criterion with $\alpha = 2c_M^{-1} + b_E$ as the high-dimensionality-adjusted MC_p ($HEMC_p$) criterion.

4. Numerical Study

In this section, we present the results of a simulation study that compared the performance of variable selection using the proposed $HEGC_p$ criterion as well as four existing criteria. We compared the ratios of loss function expectations in selected models according to the five GC_p criteria:

$$\begin{aligned} \text{Criterion 1 (HEMC}_p\text{): } & \alpha = 2c_M^{-1} + b_E, & \text{Criterion 2 (C}_p\text{): } & \alpha = 2, \\ \text{Criterion 3 (MC}_p\text{): } & \alpha = 2c_M^{-1}, & \text{Criterion 4 (BC}_p\text{): } & \alpha = \log n, \\ \text{Criterion 5 (HGC}_p\text{): } & \alpha = a_C(b_C), \end{aligned}$$

where c_M , b_E and $a_C(\beta)$ are given by (8), (22) and (9), respectively. The b_C is positive value proposed by [18], which is given by

$$b_C = \left(\frac{n}{n-p} \right)^{3/2} \frac{\log n}{\sqrt{p}}.$$

We conducted Monte Carlo simulations based on 50,000 replications, using several values of n and p . The set of candidate models was $\mathcal{J} = \{j_1, \dots, j_{15}\}$, where $j_a = \{1, \dots, a\}$ ($a = 1, \dots, 15$). We generated $z_1, \dots, z_n$ independently from $U(-1, 1)$. Using $z_1, \dots, z_n$, we constructed an $n \times k$ matrix of explanatory variables X , in which the (a, b) th element was given by z_a^{b-1} ($a = 1, \dots, n; b = 1, \dots, 15$). The true model was determined by $\Theta_* = \eta_* \mathbf{1}_p^\top$, $j_* = \{1, 2, 3, 4, 5\}$, and $\Sigma_* = \tau \Psi_*(0.8)$, where $\eta_* = (1, -2, 3, -4, 5)^\top$ and $\mathbf{1}_p$ is the p -dimensional vector of ones. Thus, $j_1, \dots, j_4$ were underspecified models, $j_5, \dots, j_{15}$ were overspecified models, and $j_* = j_5$. To specify the form of $\Psi_*(\rho)$, two types of correlation matrices were prepared as follows.

Type 1. Autocorrelation matrix, i.e., the (a, b) th element of $\Psi_*(\rho)$ is $\rho^{|a-b|}$.

Type 2. Uniform correlation matrix, i.e., $\Psi_*(\rho) = (1 - \rho)I_p + \rho \mathbf{1}_p \mathbf{1}_p^\top$.

Specifically, $\tau = 0.4$ and 1.2 were used, corresponding to strong- and weak-signal settings, respectively.

It follows from the equation $\Theta_* = \eta_* \mathbf{1}_p^\top$ that for $a = 1, \dots, 5$,

$$\theta_a^{*\top} \Sigma_*^{-1} \theta_a^* = \frac{a^2}{\tau} \mathbf{1}_p^\top \Psi_*(\rho)^{-1} \mathbf{1}_p = \begin{cases} \frac{a^2 \{p(1-\rho) + 2\rho\}}{\tau(1+\rho)} = O(p) & (\Psi_*(\rho) \text{ is of Type 1}) \\ \frac{a^2 p}{\tau \{1 + \rho(p-1)\}} = O(1) & (\Psi_*(\rho) \text{ is of Type 2}) \end{cases}.$$

Hence, in the simulation models, ξ_* in (17) was

$$\xi_* = \begin{cases} \emptyset & (\Psi_*(\rho) \text{ is of Type 1}) \\ \{1, 2, 3, 4, 5\} & (\Psi_*(\rho) \text{ is of Type 2}) \end{cases}.$$

From the conditions stated in Theorems 1 and 2 and the set ξ_* , we can determine whether a GC_p criterion is efficient. In the setting considered in this paper, a consistent GC_p criterion

is also efficient whenever $j_0 = j_*$ holds. For the simulation settings considered here, $j_0 = j_*$ holds if $\Psi_*(\rho)$ is of Type 1, or if $c_0 = 0$ when $\Psi_*(\rho)$ is of Type 2. Therefore, the C_p , BC_p , and $HCGC_p$ criteria are efficient if $\Psi_*(\rho)$ is of Type 1, or if $c_0 = 0$ when $\Psi_*(\rho)$ is of Type 2. Furthermore, as discussed at the end of Section 3, the MC_p criterion is efficient when $p \rightarrow \infty$. Based on these results, the asymptotic loss and mean efficiency properties of the criteria considered in the simulation studies are summarized in Table 1. The symbol $\bigcirc$ denotes an efficient criterion, whereas the symbol $\triangle$ indicates that the criterion is efficient if and only if the condition given in parentheses is satisfied.

Table 1. Efficiency results.

Type of Σ_*	Type of p	$HEMC_p$	C_p	MC_p	BC_p	$HCGC_p$
Auto-correlation	$p < \infty$	$\bigcirc$			$\bigcirc$	$\bigcirc$
	$p \rightarrow \infty$	$\bigcirc$	$\triangle$ ($c_0 < 0.5$)	$\bigcirc$	$\bigcirc$	$\bigcirc$
Uniform correlation	$p < \infty$	$\bigcirc$			$\bigcirc$	$\bigcirc$
	$p \rightarrow \infty$	$\bigcirc$	$\triangle$ ($c_0 = 0$)	$\bigcirc$	$\triangle$ ($c_0 = 0$)	$\triangle$ ($c_0 = 0$)

Note: The symbol $\bigcirc$ denotes an efficient criterion, and the symbol $\triangle$ denotes that the criterion is efficient if and only if the equation within the brackets is satisfied.

Tables 2 and 3 give the probabilities that each criterion will select $\hat{j}_{LE}$ when the true covariance matrices are Types 1 and 2, respectively. Tables 4 and 5 give the relative expected loss functions selected by each criterion, i.e., $E[\mathcal{L}(\hat{j}_\alpha)]/R(j_{ME})$, when the true covariance matrices are Types 1 and 2, respectively. In each table, the left-hand side shows the results when p is fixed, the right-hand side shows the results when p increases with n while maintaining a fixed ratio p/n , the upper part shows the results for $\tau = 0.4$, and the lower part shows the results for $\tau = 1.2$. The bottom row of each 2×2 block denotes the average value of that column. In all tables, we omitted results for $(n, p) = (100, 90)$. This is because the MC_p criterion is defined only for $n - p - k - 1 > 0$. Since this condition is not satisfied when $(n, p) = (100, 90)$, the MC_p and $HEMC_p$ criteria cannot be computed.

From these tables, it is clear that variable selection methods based on efficient criteria performed very well. In particular, the selection probability of the $HEMC_p$ criterion was always high, and the relative expected loss function selected by the $HEMC_p$ criterion was always small, even for a moderately sized n . By contrast, the inferior performance of variable selection methods based on criteria that are not efficient is to be expected from the theoretical properties of these criteria. The variable selection method based on the C_p criterion performs the worst as it is not efficient in most cases. In fact, it is not efficient for about 7/8 of all cases. The variable selection methods based on BC_p and $HCGC_p$ criteria are not efficient in about 3/4 of all cases. However, while the method based on the BC_p criterion is theoretically efficient when $\Psi_*(\rho)$ is of Type 1 and $p \rightarrow \infty$, when considering a finite sample size, it lacks the consistency property unless the expression before taking the limit of Condition C2 is positive; therefore, the condition $n > e^{1/(1-c_0)}$ must be satisfied. When $c_0 = 0.9$, this results in $n > 22,026$, meaning that in numerical studies with $n \leq 1000$, the method rarely selects the true model. Consequently, its performance was inferior to that of the $HCGC_p$ criterion. The performance of the variable selection method based on the MC_p criterion was comparable to those based on the $HEMC_p$ criterion. In particular, in the strong-signal setting, when $p \rightarrow \infty$, the performance exceeded that of the $HEMC_p$ criterion. However, since the MC_p criterion is not asymptotically loss- and mean-efficient when p is fixed, the performance of the variable selection method based on the MC_p criterion was slightly inferior to that based on the $HEMC_p$ criterion when p was small.

Table 2. Probabilities of selecting $\hat{J}_{LE}$ by each of the five criteria when the true variance–covariance matrix is an autocorrelation matrix.

Selection Probability (%)							Selection Probability (%)					
n	p	$HEMC_p$	C_p	MC_p	BC_p	$HCGC_p$	p	$HEMC_p$	C_p	MC_p	BC_p	$HCGC_p$
Case of $\tau = 0.4$												
100	5	97.11	86.36	89.87	98.17	98.55	5	97.19	86.59	90.00	98.22	98.47
300	5	98.12	89.48	90.38	99.99	99.61	15	99.78	97.58	98.54	100.00	99.84
500	5	98.36	90.20	90.75	100.00	99.78	25	99.98	99.46	99.73	100.00	99.95
1000	5	98.61	90.55	90.82	100.00	99.91	50	100.00	99.98	99.99	100.00	99.99
100	10	98.67	90.63	95.61	96.05	98.70	10	98.66	90.42	95.54	96.14	98.69
300	10	99.41	95.64	96.58	100.00	99.77	30	99.98	99.12	99.80	100.00	99.88
500	10	99.50	96.32	96.81	100.00	99.89	50	100.00	99.88	99.99	100.00	99.97
1000	10	99.62	96.55	96.79	100.00	99.95	100	100.00	100.00	100.00	100.00	100.00
100	30	99.63	73.28	99.05	98.85	97.78	30	99.60	73.01	99.09	98.91	97.81
300	30	99.98	99.12	99.82	100.00	99.89	90	100.00	96.22	100.00	100.00	99.90
500	30	99.99	99.58	99.84	100.00	99.95	150	100.00	99.19	100.00	100.00	99.97
1000	30	99.99	99.82	99.88	100.00	99.99	300	100.00	99.98	100.00	100.00	100.00
100	50	99.19	0.78	98.80	97.82	92.11	50	99.27	0.73	98.90	97.88	92.19
300	50	100.00	99.31	99.97	100.00	99.90	150	100.00	4.08	100.00	100.00	99.65
500	50	100.00	99.89	99.99	100.00	99.97	250	100.00	5.86	100.00	100.00	99.94
1000	50	100.00	99.98	99.99	100.00	99.99	500	100.00	8.69	100.00	100.00	99.99
100	70	94.88	0.00	94.72	1.60	43.24	70	95.06	0.00	94.82	1.62	43.37
300	70	100.00	98.73	99.98	100.00	99.88	210	99.99	0.00	99.97	98.11	98.21
500	70	100.00	99.91	100.00	100.00	99.97	350	100.00	0.00	100.00	99.99	99.69
1000	70	100.00	100.00	100.00	100.00	100.00	700	100.00	0.00	100.00	100.00	99.98
300	90	100.00	96.38	99.99	100.00	99.87	270	96.64	0.00	96.51	0.00	32.60
500	90	100.00	99.91	100.00	100.00	99.98	450	99.67	0.00	99.63	0.00	87.97
1000	90	100.00	100.00	100.00	100.00	100.00	900	99.99	0.00	99.99	0.00	99.15
Ave.		99.26	87.06	97.38	95.33	96.90		99.38	50.47	98.80	82.21	93.36
Case of $\tau = 1.2$												
100	5	70.70	76.64	77.64	32.84	62.10	5	70.41	76.60	77.60	32.13	61.71
300	5	98.08	89.50	90.36	97.81	99.55	15	99.79	97.46	98.45	85.51	99.83
500	5	98.39	90.26	90.78	100.00	99.78	25	99.97	99.37	99.70	99.89	99.95
1000	5	98.52	90.45	90.66	100.00	99.88	50	100.00	99.99	100.00	100.00	100.00
100	10	62.46	77.87	75.42	13.58	62.09	10	62.48	78.00	75.49	13.66	62.19
300	10	99.42	95.59	96.61	91.60	99.74	30	99.98	99.09	99.75	77.39	99.86
500	10	99.49	96.10	96.66	99.95	99.89	50	100.00	99.85	99.98	99.98	99.96
1000	10	99.55	96.42	96.68	100.00	99.94	100	100.00	100.00	100.00	100.00	99.99
100	30	51.20	66.81	63.53	17.85	72.46	30	51.04	66.97	63.34	17.96	72.22
300	30	99.97	99.05	99.79	77.54	99.87	90	100.00	96.32	100.00	97.82	99.87
500	30	99.98	99.62	99.86	99.88	99.95	150	100.00	99.20	100.00	100.00	99.97
1000	30	99.99	99.83	99.90	100.00	99.99	300	100.00	99.97	100.00	100.00	100.00
100	50	48.55	0.62	55.58	64.27	75.28	50	48.35	0.59	55.63	64.58	75.26
300	50	100.00	99.35	99.97	82.69	99.91	150	100.00	4.09	100.00	100.00	99.68
500	50	100.00	99.87	99.98	99.98	99.96	250	100.00	6.12	100.00	100.00	99.93
1000	50	100.00	99.98	99.99	100.00	99.99	500	100.00	8.47	100.00	100.00	100.00
100	70	42.77	0.00	44.87	1.49	37.66	70	42.66	0.00	45.09	1.46	37.74
300	70	100.00	98.68	100.00	92.26	99.89	210	99.99	0.00	99.98	98.08	98.17
500	70	100.00	99.94	100.00	99.99	99.98	350	100.00	0.00	100.00	99.99	99.72
1000	70	100.00	100.00	100.00	100.00	100.00	700	100.00	0.00	100.00	100.00	99.99
300	90	100.00	96.13	100.00	98.06	99.86	270	95.94	0.00	95.86	0.00	32.37
500	90	100.00	99.93	100.00	100.00	99.97	450	99.66	0.00	99.64	0.00	88.04
1000	90	100.00	100.00	100.00	100.00	100.00	900	100.00	0.00	100.00	0.00	99.09
Ave.		89.96	85.77	90.36	81.30	91.64		90.01	49.22	91.76	69.06	88.07

Table 3. Probabilities of selecting $\hat{J}_{LE}$ by each of the five criteria when the true variance–covariance matrix is a uniform correlation matrix.

Selection Probability (%)							Selection Probability (%)					
n	p	$HEMC_p$	C_p	MC_p	BC_p	$HCGC_p$	p	$HEMC_p$	C_p	MC_p	BC_p	$HCGC_p$
Case of $\tau = 0.4$												
100	5	96.78	86.43	89.97	94.20	97.73	5	96.88	86.62	90.02	94.29	97.83
300	5	98.09	89.22	90.19	99.99	99.63	15	99.80	97.62	98.55	99.69	99.85
500	5	98.29	90.10	90.58	100.00	99.74	25	99.96	99.31	99.67	99.91	99.93
1000	5	98.47	90.54	90.81	100.00	99.87	50	100.00	99.98	99.99	99.99	99.99
100	10	94.60	89.86	94.27	63.00	94.57	10	94.75	90.05	94.51	63.30	94.72
300	10	99.47	95.58	96.62	100.00	99.81	30	99.97	99.12	99.78	55.32	99.85
500	10	99.56	96.10	96.65	100.00	99.88	50	100.00	99.87	99.99	35.49	99.98
1000	10	99.64	96.57	96.82	100.00	99.97	100	100.00	100.00	100.00	10.85	100.00
100	30	37.76	58.37	48.41	13.78	57.35	30	37.50	58.50	48.18	13.66	57.49
300	30	99.98	99.11	99.80	55.70	99.88	90	76.05	95.71	87.38	0.34	98.30
500	30	99.98	99.58	99.85	98.89	99.94	150	86.89	99.17	94.59	0.01	99.93
1000	30	99.99	99.84	99.90	100.00	99.98	300	98.38	99.98	99.61	0.00	100.00
100	50	34.40	0.06	36.11	36.40	28.41	50	34.40	0.07	36.24	36.31	28.61
300	50	99.66	99.20	99.91	4.11	99.85	150	68.17	0.43	61.02	54.44	18.13
500	50	100.00	99.90	99.99	35.80	99.97	250	75.81	0.43	66.90	47.50	9.06
1000	50	100.00	99.99	99.99	100.00	99.99	500	75.02	0.64	65.08	18.20	7.87
100	70	43.29	0.00	41.72	0.07	6.07	70	42.84	0.00	41.33	0.07	6.13
300	70	94.00	98.68	98.00	0.33	99.65	210	63.43	0.00	66.33	35.37	36.20
500	70	100.00	99.93	100.00	1.94	99.97	350	85.21	0.00	86.47	67.90	28.29
1000	70	100.00	100.00	100.00	93.91	100.00	700	97.76	0.00	97.28	97.13	6.51
300	90	75.66	95.84	87.13	0.39	98.26	270	38.23	0.00	38.11	0.00	6.38
500	90	99.96	99.92	100.00	0.07	99.98	450	45.61	0.00	46.61	0.00	36.27
1000	90	100.00	100.00	100.00	34.16	100.00	900	66.27	0.00	67.86	0.00	48.38
Ave.		89.98	86.30	89.42	57.95	90.46		77.52	49.02	77.63	40.42	59.99
Case of $\tau = 1.2$												
100	5	57.58	69.30	69.19	20.02	47.35	5	57.48	69.16	68.98	19.71	47.43
300	5	97.98	89.56	90.48	92.11	99.02	15	93.27	96.46	96.74	6.81	92.24
500	5	98.31	90.03	90.56	99.86	99.76	25	98.61	99.33	99.52	1.09	99.03
1000	5	98.44	90.69	90.92	100.00	99.91	50	99.99	99.97	99.98	0.01	99.98
100	10	23.11	41.68	35.14	5.19	22.89	10	23.06	42.29	35.41	5.09	22.84
300	10	97.93	95.39	96.34	38.98	96.88	30	51.46	81.80	71.29	0.65	66.40
500	10	99.52	96.20	96.78	86.94	99.88	50	64.51	92.98	83.54	0.06	88.04
1000	10	99.62	96.42	96.70	100.00	99.94	100	86.26	99.42	96.06	0.00	99.75
100	30	69.37	9.90	59.27	83.61	47.87	30	69.74	9.55	59.64	84.03	48.07
300	30	51.69	81.99	71.47	0.61	66.76	90	33.62	19.34	34.27	43.15	37.46
500	30	96.09	99.30	99.08	0.10	98.12	150	41.16	26.62	55.24	13.83	59.92
1000	30	100.00	99.83	99.89	22.44	99.99	300	72.33	20.43	85.94	0.63	39.52
100	50	91.59	0.01	88.10	81.59	58.94	50	91.66	0.02	88.18	81.57	59.21
300	50	53.81	26.79	45.26	0.13	37.18	150	95.41	0.00	92.04	99.93	43.55
500	50	64.36	92.64	83.14	0.09	87.86	250	96.50	0.00	93.29	100.00	19.46
1000	50	99.99	99.97	99.99	0.01	99.99	500	98.15	0.00	95.85	100.00	1.02
100	70	88.69	0.00	87.29	0.27	22.12	70	88.80	0.00	87.51	0.29	21.81
300	70	54.72	35.32	65.63	4.22	61.22	210	97.44	0.00	96.44	58.15	59.18
500	70	30.65	48.41	33.90	0.02	44.15	350	99.02	0.00	98.49	86.72	51.67
1000	70	99.33	99.98	99.92	0.00	99.98	700	99.85	0.00	99.75	99.71	25.07
300	90	33.85	19.28	34.33	43.09	37.31	270	88.15	0.00	87.75	0.00	12.34
500	90	79.53	23.62	67.87	0.00	32.90	450	94.47	0.00	94.11	0.00	47.01
1000	90	93.50	99.79	98.56	0.00	99.88	900	98.75	0.00	98.61	0.00	63.84
Ave.		77.38	65.48	78.25	33.88	72.17		79.99	32.93	83.42	34.84	52.38

Table 4. Relative expected loss function selected by each of the five criteria when the true variance-covariance matrix is an autocorrelation matrix.

Relative Expected Loss							Relative Expected Loss					
n	p	$HEMC_p$	C_p	MC_p	BC_p	$HCGC_p$	p	$HEMC_p$	C_p	MC_p	BC_p	$HCGC_p$
Case of $\tau = 0.4$												
100	5	1.019	1.095	1.067	1.024	1.010	5	1.020	1.095	1.068	1.025	1.012
300	5	1.014	1.070	1.064	1.001	1.004	15	1.000	1.010	1.006	0.999	1.000
500	5	1.013	1.066	1.062	1.002	1.003	25	1.000	1.002	1.001	1.000	1.000
1000	5	1.009	1.062	1.060	1.000	1.001	50	1.000	1.000	1.000	1.000	1.000
100	10	1.007	1.049	1.022	1.040	1.007	10	1.007	1.050	1.022	1.038	1.007
300	10	1.004	1.022	1.018	1.001	1.002	30	1.001	1.005	1.002	1.001	1.002
500	10	1.002	1.018	1.015	0.999	1.000	50	1.000	1.000	1.000	1.000	1.000
1000	10	1.001	1.016	1.015	0.999	0.999	100	1.000	1.000	1.000	1.000	1.000
100	30	1.001	1.153	1.003	1.008	1.007	30	1.002	1.156	1.003	1.008	1.008
300	30	0.999	1.002	1.000	0.999	0.999	90	1.000	1.011	1.000	1.000	1.001
500	30	1.000	1.001	1.000	0.999	1.000	150	0.999	1.002	0.999	0.999	0.999
1000	30	1.000	1.001	1.001	1.000	1.000	300	1.000	1.000	1.000	1.000	1.000
100	50	1.002	2.820	1.003	1.006	1.027	50	1.003	2.822	1.004	1.007	1.028
300	50	1.000	1.002	1.000	1.000	1.000	150	1.000	2.563	1.000	1.000	1.001
500	50	1.000	1.001	1.000	1.000	1.001	250	1.000	2.455	1.000	1.000	1.000
1000	50	1.001	1.001	1.001	1.001	1.001	500	1.000	2.327	1.000	1.000	1.000
100	70	1.020	3.000	1.020	2.751	1.571	70	1.018	2.997	1.018	2.750	1.562
300	70	1.000	1.004	1.000	1.000	1.000	210	1.000	3.000	1.000	1.005	1.004
500	70	1.000	1.000	1.000	1.000	1.000	350	1.000	3.000	1.000	1.000	1.000
1000	70	1.000	1.000	1.000	1.000	1.000	700	1.000	3.000	1.000	1.000	1.000
300	90	1.000	1.010	1.000	1.000	1.000	270	1.010	3.000	1.011	3.000	1.753
500	90	1.000	1.000	1.000	1.000	1.000	450	1.001	3.000	1.001	3.000	1.040
1000	90	1.000	1.000	1.000	1.000	1.000	900	1.000	3.000	1.000	3.000	1.002
Ave.		1.004	1.191	1.015	1.080	1.027		1.003	1.891	1.006	1.341	1.062
Case of $\tau = 1.2$												
100	5	1.134	1.129	1.113	1.428	1.185	5	1.134	1.127	1.111	1.429	1.184
300	5	1.016	1.073	1.067	1.040	1.007	15	1.002	1.012	1.008	1.144	1.002
500	5	1.011	1.064	1.060	1.000	1.002	25	1.001	1.003	1.002	1.002	1.001
1000	5	1.011	1.065	1.063	1.001	1.002	50	1.000	1.000	1.000	1.000	1.000
100	10	1.098	1.073	1.064	1.383	1.099	10	1.099	1.072	1.064	1.384	1.100
300	10	1.003	1.021	1.016	1.098	1.001	30	1.000	1.004	1.001	1.177	1.001
500	10	1.003	1.020	1.017	1.001	1.001	50	1.000	1.000	1.000	1.000	1.000
1000	10	1.003	1.019	1.018	1.001	1.001	100	1.000	1.000	1.000	1.000	1.000
100	30	1.057	1.154	1.038	1.160	1.028	30	1.057	1.154	1.038	1.160	1.028
300	30	1.001	1.005	1.002	1.177	1.002	90	1.000	1.010	1.000	1.014	1.000
500	30	1.000	1.001	1.000	1.002	1.000	150	1.000	1.002	1.000	1.000	1.000
1000	30	1.001	1.001	1.001	1.001	1.001	300	1.000	1.000	1.000	1.000	1.000
100	50	1.048	2.826	1.038	1.029	1.034	50	1.050	2.822	1.039	1.029	1.035
300	50	1.000	1.002	1.000	1.121	1.000	150	1.000	2.558	1.000	1.000	1.001
500	50	1.001	1.001	1.001	1.001	1.001	250	1.000	2.456	1.000	1.000	1.000
1000	50	0.999	0.999	0.999	0.999	0.999	500	1.000	2.335	1.000	1.000	1.000
100	70	1.070	3.000	1.067	2.751	1.569	70	1.071	2.998	1.067	2.750	1.571
300	70	1.000	1.004	1.000	1.052	1.001	210	1.000	3.001	1.000	1.005	1.005
500	70	1.001	1.001	1.001	1.001	1.001	350	1.000	3.000	1.000	1.000	1.001
1000	70	1.000	1.000	1.000	1.000	1.000	700	1.000	3.000	1.000	1.000	1.000
300	90	1.010	1.094	1.012	1.004	1.033	270	1.033	4.286	1.034	4.286	2.350
500	90	1.012	1.036	1.016	1.182	1.032	450	1.011	4.240	1.012	4.240	1.167
1000	90	1.010	1.001	1.003	1.738	1.000	900	1.002	4.198	1.002	4.198	1.071
Ave.		1.021	1.200	1.026	1.181	1.043		1.020	2.056	1.016	1.557	1.109

Table 5. Relative expected loss function selected by each of the five criteria when the true variance–covariance matrix is a uniform correlation matrix.

Relative Expected Loss							Relative Expected Loss					
n	p	$HEMC_p$	C_p	MC_p	BC_p	$HCGC_p$	p	$HEMC_p$	C_p	MC_p	BC_p	$HCGC_p$
Case of $\tau = 0.4$												
100	5	1.023	1.094	1.067	1.068	1.019	5	1.023	1.094	1.068	1.068	1.020
300	5	1.013	1.071	1.065	1.000	1.003	15	1.001	1.010	1.006	1.004	1.000
500	5	1.012	1.066	1.062	1.000	1.002	25	0.999	1.002	1.000	1.000	0.999
1000	5	1.010	1.061	1.060	0.999	1.000	50	0.999	0.999	0.999	0.999	0.999
100	10	1.029	1.052	1.028	1.221	1.029	10	1.027	1.051	1.027	1.220	1.027
300	10	1.003	1.022	1.017	1.000	1.001	30	1.002	1.005	1.002	1.299	1.002
500	10	1.002	1.019	1.016	0.999	1.000	50	0.999	1.000	0.999	1.434	0.999
1000	10	1.001	1.016	1.014	0.998	0.999	100	1.000	1.000	1.000	1.634	1.000
100	30	1.057	1.154	1.039	1.149	1.030	30	1.059	1.159	1.040	1.150	1.030
300	30	1.001	1.004	1.001	1.296	1.001	90	1.024	1.011	1.013	1.256	1.002
500	30	1.000	1.002	1.001	1.014	1.000	150	1.014	1.002	1.006	1.301	1.000
1000	30	1.001	1.001	1.001	1.001	1.001	300	1.002	1.000	1.001	1.467	1.000
100	50	1.045	2.970	1.044	1.045	1.066	50	1.045	2.971	1.044	1.045	1.065
300	50	1.001	1.003	1.000	1.315	1.000	150	1.013	2.639	1.014	1.068	1.027
500	50	1.000	1.000	1.000	1.433	1.000	250	1.007	2.518	1.010	1.105	1.026
1000	50	1.001	1.001	1.001	1.001	1.001	500	1.005	2.375	1.007	1.194	1.018
100	70	1.028	3.353	1.032	3.077	1.721	70	1.029	3.354	1.033	3.074	1.720
300	70	1.011	1.004	1.004	1.273	1.001	210	1.028	3.251	1.027	1.060	1.059
500	70	1.000	1.000	1.000	1.413	1.000	350	1.015	3.245	1.014	1.028	1.061
1000	70	1.000	1.000	1.000	1.067	1.000	700	1.002	3.225	1.002	1.003	1.071
300	90	1.024	1.011	1.013	1.256	1.002	270	1.023	3.349	1.024	3.349	1.943
500	90	1.001	1.001	1.001	1.289	1.001	450	1.023	3.346	1.023	3.346	1.108
1000	90	1.000	1.000	1.000	1.537	1.000	900	1.018	3.328	1.018	3.328	1.058
Ave.		1.011	1.213	1.020	1.237	1.038		1.016	1.997	1.016	1.541	1.097
Case of $\tau = 1.2$												
100	5	1.165	1.140	1.127	1.458	1.223	5	1.165	1.140	1.127	1.458	1.223
300	5	1.015	1.069	1.063	1.111	1.011	15	1.025	1.014	1.013	1.465	1.029
500	5	1.011	1.065	1.062	1.003	1.002	25	1.006	1.003	1.002	1.433	1.004
1000	5	1.010	1.061	1.060	0.999	1.000	50	0.999	1.000	0.999	1.429	0.999
100	10	1.113	1.082	1.078	1.238	1.114	10	1.113	1.080	1.078	1.238	1.114
300	10	1.013	1.024	1.019	1.412	1.020	30	1.048	1.017	1.027	1.457	1.032
500	10	1.004	1.020	1.017	1.160	1.002	50	1.036	1.006	1.016	1.546	1.012
1000	10	1.001	1.017	1.016	0.999	0.999	100	1.017	1.001	1.005	1.630	1.001
100	30	1.025	1.317	1.042	1.004	1.064	30	1.025	1.318	1.041	1.004	1.064
300	30	1.047	1.016	1.026	1.457	1.031	90	1.009	1.093	1.012	1.004	1.033
500	30	1.013	1.003	1.004	1.489	1.007	150	1.019	1.086	1.014	1.028	1.040
1000	30	1.000	1.001	1.001	1.630	1.000	300	1.015	1.090	1.008	1.047	1.070
100	50	1.019	3.833	1.027	1.044	1.119	50	1.019	3.826	1.028	1.045	1.119
300	50	1.029	1.030	1.024	1.171	1.025	150	1.006	3.284	1.010	1.000	1.079
500	50	1.036	1.006	1.017	1.545	1.011	250	1.003	3.091	1.007	1.000	1.087
1000	50	1.000	1.000	1.000	1.429	1.000	500	1.002	2.891	1.004	1.000	1.099
100	70	1.037	4.297	1.042	3.933	2.028	70	1.037	4.298	1.042	3.930	2.027
300	70	1.036	1.064	1.030	1.067	1.038	210	1.005	4.118	1.007	1.087	1.085
500	70	1.012	1.005	1.010	1.288	1.007	350	1.002	4.064	1.002	1.021	1.082
1000	70	1.001	1.000	1.000	1.781	1.000	700	1.000	4.014	1.001	1.001	1.109
300	90	1.000	1.011	1.000	1.012	1.000	270	1.014	2.999	1.014	2.999	1.761
500	90	1.000	1.000	1.000	1.000	1.000	450	1.001	3.000	1.001	3.000	1.040
1000	90	1.000	1.000	1.000	1.000	1.000	900	1.000	3.000	1.000	3.000	1.002
Ave.		1.026	1.307	1.029	1.358	1.074		1.025	2.236	1.020	1.557	1.135

Under the asymptotic framework in (3), it is generally expected that the accuracy of asymptotic approximations deteriorates as c_0 approaches 1. However, in the present simulation study, no clear deterioration in the performance of the variable selection method based on the $HEMC_p$ criterion was observed when c_0 approaches 1.

From the results of these simulation studies, we confirmed that the variable selection method based on the $HEMC_p$ criterion consistently exhibited superior performance across all simulation settings considered here, regardless of the size of p . Therefore, we anticipate that the $HEMC_p$ criterion will perform well in practical applications.

Funding: This research was funded by JSPS KAKENHI Grant Number 23H00809, and the ISM General Cooperative Research 2 Grant Number 2026-ISMCRP-2010.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: The author is grateful to Hirofumi Wakaki, Hiroshima University, for his valuable comments on the proof of Lemma A2. The author also wishes to thank two reviewers for their helpful comments.

Conflicts of Interest: The author declares no conflicts of interest.

Appendix A. Proofs of Lemmas and Theorems

Appendix A.1. Proof of Lemma 1

Let $\Theta_j^{*\top}$ for $j \subseteq j_*$ be a $p \times k_j$ matrix of which the i th column vector consists of $\theta_{a_i}^*$, where a_i is the i th element of j , and Φ_j^* for $j \subseteq j_*$ be a $k_j \times k_j$ matrix defined by $\Phi_j^* = \Theta_j^* \Sigma_*^{-1} \Theta_j^{*\top}$. By using this notation, for any $j \in \mathcal{J}_-$, δ_j is rewritten as

$$\delta_j = \frac{1}{n} \text{tr} \left\{ \Phi_{j^c \cap j_*}^* \mathbf{X}_{j^c \cap j_*}^\top (\mathbf{I}_n - \mathbf{P}_j) \mathbf{X}_{j^c \cap j_*} \right\}. \quad (\text{A1})$$

To assess which elements of Φ_j^* are convergent or divergent for $j \subseteq j_*$, we prepare the following lemma.

Lemma A1. Let $\mathbf{a}_p = (a_1, \dots, a_p)^\top$ and $\mathbf{b}_p = (b_1, \dots, b_p)^\top$ be p -dimensional real vectors and let $\mathbf{S}_p$ be a $p \times p$ positive definite real matrix. Moreover, we define sequences $\{c_p\}$, $\{d_p\}$, and $\{h_p\}$ as

$$c_p = \mathbf{a}_p^\top \mathbf{S}_p^{-1} \mathbf{a}_p, \quad d_p = \mathbf{b}_p^\top \mathbf{S}_p^{-1} \mathbf{b}_p, \quad h_p = \mathbf{a}_p^\top \mathbf{S}_p^{-1} \mathbf{b}_p.$$

Then, if $c_p = O(1)$ and $d_p = O(1)$ hold as $p \rightarrow \infty$, sequences $\{c_p\}$, $\{d_p\}$ and $\{h_p\}$ are convergent, and sequences $\{c_p\}$ and $\{d_p\}$ converge to positive values.

Proof of Lemma A1. From the definitions of c_p and d_p , we can observe that $\{c_p\}$ and $\{d_p\}$ are positive sequences. Let

$$\mathbf{a}_{p+1} = \begin{pmatrix} \mathbf{a}_p \\ a_{p+1} \end{pmatrix}, \quad \mathbf{b}_{p+1} = \begin{pmatrix} \mathbf{b}_p \\ b_{p+1} \end{pmatrix}, \quad \mathbf{S}_{p+1} = \begin{pmatrix} \mathbf{S}_p & \mathbf{r}_{p+1} \\ \mathbf{r}_{p+1}^\top & s_{p+1} \end{pmatrix}. \quad (\text{A2})$$

It follows from the general formula for the inverse of a block matrix, e.g., th. 8.5.11 [29], that

$$\mathbf{S}_{p+1}^{-1} = \begin{pmatrix} \mathbf{S}_p^{-1} & \mathbf{0}_p \\ \mathbf{0}_p^\top & 0 \end{pmatrix} + \frac{1}{w_{p+1}} \mathbf{u}_{p+1} \mathbf{u}_{p+1}^\top, \quad (\text{A3})$$

where $\mathbf{0}_p$ is the p -dimensional vector of zeros, and w_{p+1} and $\mathbf{u}_{p+1}$ are given by

$$w_{p+1} = s_{p+1} - \mathbf{r}_{p+1}^\top \mathbf{S}_p^{-1} \mathbf{r}_{p+1}, \quad \mathbf{u}_{p+1} = \begin{pmatrix} \mathbf{S}_p^{-1} \mathbf{r}_{p+1} \\ -1 \end{pmatrix}.$$

It should be kept in mind that $w_{p+1} > 0$ because we assume that $\mathbf{S}_{p+1}$ is also a positive definite real matrix. By using (A2) and (A3), we derive

$$\begin{aligned} c_{p+1} &= \mathbf{a}_{p+1}^\top \mathbf{S}_{p+1}^{-1} \mathbf{a}_{p+1} = \mathbf{a}_p^\top \mathbf{S}_p^{-1} \mathbf{a}_p + \frac{1}{w_{p+1}} (\mathbf{a}_{p+1}^\top \mathbf{u}_{p+1})^2 \\ &= c_p + \frac{1}{w_{p+1}} (\mathbf{a}_{p+1}^\top \mathbf{u}_{p+1})^2. \end{aligned} \quad (\text{A4})$$

This means that $\{c_p\}$ is a monotonically increasing sequence. On the other hand, the assumption indicates that sequence $\{c_p\}$ is bounded from above. Hence, $\{c_p\}$ converges to a positive value. By a similar method, $\{d_p\}$ also converges to a positive value. Let us define $\mathbf{u}_1 = -1$ and $w_1 = s_1$. From (A4), we can see that sequences $\{c_p\}$, $\{d_p\}$, and $\{h_p\}$ are expressed as

$$c_p = \sum_{i=1}^p \left(\frac{\mathbf{a}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right)^2, \quad d_p = \sum_{i=1}^p \left(\frac{\mathbf{b}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right)^2, \quad h_p = \sum_{i=1}^p \left(\frac{\mathbf{a}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right) \left(\frac{\mathbf{b}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right).$$

It follows from the relationship between the arithmetic mean and geometric mean that

$$\begin{aligned} h_p^+ &= \sum_{i=1}^p \left| \left(\frac{\mathbf{a}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right) \left(\frac{\mathbf{b}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right) \right| = \sum_{i=1}^p \sqrt{\left(\frac{\mathbf{a}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right)^2 \left(\frac{\mathbf{b}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right)^2} \\ &\leq \frac{1}{2} \sum_{i=1}^p \left\{ \left(\frac{\mathbf{a}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right)^2 + \left(\frac{\mathbf{b}_i^\top \mathbf{u}_i}{\sqrt{w_i}} \right)^2 \right\} = \frac{1}{2} (c_p + d_p). \end{aligned}$$

It is straightforward to see that $\{h_p^+\}$ is a monotonically increasing sequence. Since $\{c_p\}$ and $\{d_p\}$ are bounded from above, sequence $\{h_p^+\}$ is also bounded from above. This implies that $\{h_p^+\}$ converges to a positive value. Since h_p^+ is the absolute convergent series, h_p is the convergent series. Hence, h_p is also convergent. $\square$

It follows from Lemma A1 that $\Phi_{\zeta_*}^*$ converges to some positive semidefinite matrix, where ζ_* is given in (17). Since $\mathbf{X}^\top \mathbf{X}/n$ is convergent, $\mathbf{X}_{j^c \cap j_*}^\top (\mathbf{I}_n - \mathbf{P}_j) \mathbf{X}_{j^c \cap j_*}/n$ is also convergent. Therefore, δ_j is convergent for any $j \in \mathcal{S}_-$. When $j \in \mathcal{S}_-^c \cap \mathcal{J}_-$, we have

$$\exists a \in \zeta_* \cap j_* \text{ s.t. } a \notin j.$$

Then, by using the above result and $\theta_a^{*\top} \Sigma_*^{-1} \theta_a^* \rightarrow \infty$ as $p \rightarrow \infty$, δ_j in (A1) is rewritten as

$$\delta_j \geq \frac{1}{n} \text{tr} \left\{ \Phi_{j^c \cap j_*}^* \mathbf{X}_{j^c \cap j_*}^\top (\mathbf{I}_n - \mathbf{P}_{\omega \setminus \{a\}}) \mathbf{X}_{j^c \cap j_*} \right\} = \frac{1}{n} \theta_a^{*\top} \Sigma_*^{-1} \theta_a^* \mathbf{X}_{\{a\}}^\top (\mathbf{I}_n - \mathbf{P}_{\omega \setminus \{a\}}) \mathbf{X}_{\{a\}} \rightarrow \infty.$$

Hence, δ_j is divergent for any $j \in \mathcal{S}_-^c \cap \mathcal{J}_-$. Consequently, Lemma 1 is proved.

Appendix A.2. Proof of Lemma 2

When $c_0 \neq 0$, $p \rightarrow \infty$ holds. Notice that $\text{tr}(\mathcal{E}^\top \mathbf{P}_j \mathcal{E})/p \xrightarrow{p} k_j$ as $p \rightarrow \infty$ because $\text{tr}(\mathcal{E}^\top \mathbf{P}_j \mathcal{E})$ is distributed according to a chi-square distribution with pk_j degrees of freedom. From the result and Lemma 1, Lemma 2 is proved.

Appendix A.3. Proof of Lemma 3

Suppose that $j_1 \subset j_2$ for $j_1, j_2 \in \mathcal{J}_+$. Then, $P(\text{tr}(\mathcal{E}^\top P_{j_2} \mathcal{E}) - \text{tr}(\mathcal{E}^\top P_{j_1} \mathcal{E}) > 0) = 1$ holds because $P_{j_2} - P_{j_1}$ is positive semidefinite and the distribution of $\mathcal{E}$ is continuous. This implies that $P(\hat{j}_{LE} \in \mathcal{J}_+ \setminus \{j_*\}) = 0$. Since $(\mathcal{J}_+ \setminus \{j_*\})^c = \{j_*\} \cup \mathcal{J}_-$, Lemma 3 (1) is proved.

Next, we show the proof of Lemma 3 (2). Notice that

$$\frac{1}{n} \text{tr}(\mathcal{E}^\top P_j \mathcal{E}) = \frac{p}{n} \cdot \frac{1}{p} \text{tr}(\mathcal{E}^\top P_j \mathcal{E}) \xrightarrow{p} k_j c_0,$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. By using this result, we show the following basic results used for proving Lemma 3 (2):

$$\begin{aligned} \forall j \in \mathcal{J}_-, \quad \frac{1}{n\delta_j} \{\mathcal{L}(j) - \mathcal{L}(j_*)\} &= 1 + \frac{1}{\delta_j} \left\{ \frac{1}{n} \text{tr}(\mathcal{E}^\top P_j \mathcal{E}) - \frac{1}{n} \text{tr}(\mathcal{E}^\top P_{j_*} \mathcal{E}) \right\} \\ &= 1 + \frac{1}{\delta_j} (k_j - k_*) c_0 + o_p(1), \end{aligned} \quad (\text{A5})$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Recall that $P(\hat{j}_{LE} \in \mathcal{J}_+ \setminus \{j_*\}) = 0$. Hence, to prove Lemma 3 (2), it is sufficient to show that the following equation is satisfied:

$$\forall j \in \{\{j_*\} \cup \mathcal{J}_-\} \setminus \{j_0\}, \quad \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{h_j} \{\mathcal{L}(j) - \mathcal{L}(j_0)\} = \tau_j > 0, \quad (\text{A6})$$

(see, e.g., the result in [30]), where j_0 is the asymptotically loss optimal model given by (19), and h_j is some positive constant depending on model j , which does not converge to 0. First, we give the proof when $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$. From Lemma 2 (1) and (19), we have

$$\forall j \in \{\{j_*\} \cup \mathcal{S}_-\} \setminus \{j_0\}, \quad \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n} \{\mathcal{L}(j) - \mathcal{L}(j_0)\} = \lambda(j) - \lambda(j_0) > 0, \quad (\text{A7})$$

where $\lambda(j)$ is given by (18). On the other hand, from Lemma 2, for any $j \in \mathcal{S}_-^c \cap \mathcal{J}_-$, δ_j goes to ∞ as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Then, from (A7), we can see that $\{\mathcal{L}(j_*) - \mathcal{L}(j_0)\} / (n\delta_j)$ is 0 or converges to 0 in probability as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. It follows from this result, the divergence of δ_j and (A5) that $\forall j \in \mathcal{S}_-^c \cap \mathcal{J}_-$,

$$\begin{aligned} &\text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n\delta_j} \{\mathcal{L}(j) - \mathcal{L}(j_0)\} \\ &= \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n\delta_j} \{\mathcal{L}(j) - \mathcal{L}(j_*)\} + \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n\delta_j} \{\mathcal{L}(j_*) - \mathcal{L}(j_0)\} = 1 > 0. \end{aligned}$$

Therefore, (A6) is proved when $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$. Next, we give the proof when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$. Recall that $\xi_* = \emptyset \Leftrightarrow \delta_j \rightarrow \infty$ for any $j \in \mathcal{J}_-$. Hence, substituting $\delta_j \rightarrow \infty$ or $c_0 = 0$ into (A5) yields

$$\forall j \in \mathcal{J}_-, \quad \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n\delta_j} \{\mathcal{L}(j) - \mathcal{L}(j_*)\} = 1 > 0.$$

Recall that $j_0 = j_*$ when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$. Hence, (A6) is proved when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$. Consequently, Lemma 3 (2) is proved.

Appendix A.4. Proof of Lemma 4

We first give the lemma required for proving Lemma 4.

Lemma A2. Let X be a Bernoulli random variable with probability of success θ , and let h be a positive value such that

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} h = \infty.$$

Suppose that

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} \theta = 0.$$

Then, without depending on the order of divergence of h , $hX \xrightarrow{p} 0$ as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$.

Proof of Lemma A2. Notice that hX is a non-negative random variable. For any $\eta > 0$, we have

$$P(hX > \eta) = \begin{cases} \theta & (\text{when } \eta \leq h) \\ 0 & (\text{when } \eta > h) \end{cases}.$$

From the above result and the θ assumption, it follows that

$$P(hX > \eta) \leq \theta \rightarrow 0,$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. This means that $hX \xrightarrow{p} 0$ as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$. Hence, Lemma A2 is proved. $\square$

To permit a unified description of proofs for $\mathcal{L}(\hat{j}_{LE})$ and $\mathcal{L}(\hat{j}_\alpha)$, we write the optimal model and the selected model as $\hat{j}$. Notice that $I(\hat{j} = j)$ is a Bernoulli random variable. It follows from the assumption that

$$\forall j \in \mathcal{J} \setminus \{\ell\}, \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} P(I(\hat{j} = j) = 1) = 0. \quad (\text{A8})$$

This indicates that for any $j \in \mathcal{J} \setminus \{\ell\}$, $I(\hat{j} = j) = o_p(1)$ as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. From (A8) and Lemma A2, we have

$$\forall j \in \mathcal{J} \setminus \{\ell\}, \quad p \cdot I(\hat{j} = j) \xrightarrow{p} 0, \quad \forall j \in \mathcal{J} \setminus \{\ell\}, \quad n\delta_j \cdot I(\hat{j} = j) \xrightarrow{p} 0, \quad (\text{A9})$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Notice that $\text{tr}(\mathcal{E}^\top \mathbf{P}_j \mathcal{E})/p = O_p(1)$ as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. It follows from this equation and (A9) that for any $j \in \mathcal{J} \setminus \{\ell\}$

$$\text{tr}(\mathcal{E}^\top \mathbf{P}_j \mathcal{E}) \cdot I(\hat{j} = j) = \frac{1}{p} \text{tr}(\mathcal{E}^\top \mathbf{P}_j \mathcal{E}) \cdot \{p \cdot I(\hat{j} = j)\} \xrightarrow{p} 0, \quad (\text{A10})$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. By using (A9) and (A10), we have

$$\forall j \in \mathcal{J} \setminus \{\ell\}, \quad \mathcal{L}(j)I(\hat{j} = j) \xrightarrow{p} 0,$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Substituting the above equation into (15) yields

$$\mathcal{L}(\hat{j}) = \mathcal{L}(\ell)I(\hat{j} = \ell) + o_p(1) = \mathcal{L}(\ell) - \mathcal{L}(\ell)I(\hat{j} \neq \ell) + o_p(1), \quad (\text{A11})$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Notice that $I(\hat{j} \neq \ell)$ is also a Bernoulli random variable having

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} P(I(\hat{j} \neq \ell) = 1) = 0.$$

Therefore, from Lemma A2, we can see that

$$\mathcal{L}(\ell)I(\hat{j} \neq \ell) \xrightarrow{p} 0,$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Substituting the above equations into (A11) yields

$$\mathcal{L}(\hat{j}) = \mathcal{L}(\ell) + o_p(1),$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. It follows from the above equation that $\mathcal{L}(\hat{j}_{LE}) = \mathcal{L}(\ell) + o_p(1)$ and $\mathcal{L}(\hat{j}_\alpha) = \mathcal{L}(\ell) + o_p(1)$ as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. These equations imply that $\mathcal{L}(\hat{j}_{LE})/\mathcal{L}(\hat{j}_\alpha) = 1 + o_p(1)$ as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Hence, Lemma 4 is proved.

Appendix A.5. Proof of Lemma 5

We first give lemmas required for proving Lemma 5 (see [18], for the proof).

Lemma A3. Suppose that Assumptions A1 and A2 hold. The GC_p criterion is consistent when $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$ if α satisfies Conditions C1 and C2.

Lemma A4. Suppose that Assumptions A1 and A2 hold. When $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$, statistics based on transformed $\text{tr}(\mathbf{S}_j \mathbf{S}_\omega^{-1})$ converge to positive values in probability as $n \rightarrow \infty$ and $p/n \rightarrow c_0 \in [0, 1)$, i.e.,

$$\begin{aligned} \forall j \in \mathcal{S}_- \cup \mathcal{J}_+, \quad \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n} \left\{ (n - k_j) \text{tr}(\mathbf{S}_j \mathbf{S}_\omega^{-1}) - (n - k_j)p \right\} &= \frac{\delta_{j,0} + (k - k_j)c_0}{1 - c_0}, \\ \forall j \in \mathcal{S}_-^c \cap \mathcal{J}_-, \quad \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n\delta_j} \left\{ (n - k_j) \text{tr}(\mathbf{S}_j \mathbf{S}_\omega^{-1}) - (n - k_j)p \right\} &= 1, \end{aligned}$$

where $\mathbf{S}_j$ is given by (5).

To prove Lemma 5, it suffices to show that the following equation is satisfied:

$$\forall j \in \{\{j_*\} \cup \mathcal{J}_-\} \setminus \{j_0\}, \quad \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{h_j} \{GC_p(j|\alpha) - GC_p(j_0|\alpha)\} = \tau_j > 0, \quad (\text{A12})$$

(see, e.g., the result in [30], where h_j is some positive value depending on the model j , which does not converge to 0. Recall that $j_0 = j_*$ when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$, $j_0 = j_*$. Hence, when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$, the equations in (A12) hold if α satisfies the conditions for consistency given in Lemma A3, i.e., Conditions C1 and C2.

Next, we consider the case of $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$. Although it is stated in (19) that $\lambda(j_0)$ is the minimum among all $j \in \{j_*\} \cup \mathcal{S}_-$, we know that

$$\forall j \in \{\mathcal{S}_- \cup \mathcal{J}_+\} \setminus \{j_0\}, \quad \lambda(j_0) < \lambda(j), \quad (\text{A13})$$

where $\lambda(j)$ is given by (18). This is because $\lambda(j) = k_j c_0$ and $k_* < k_j$ holds for any $j \in \mathcal{J}_+ \setminus \{j_*\}$. It follows from the above result, (A13), Lemma A4, and Condition C3 that

$$\begin{aligned} \forall j \in \{\mathcal{S}_- \cup \mathcal{J}_+\} \setminus \{j_0\}, \quad \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n} \{GC_p(j|\alpha) - GC_p(j_0|\alpha)\} &= \frac{\lambda(j) - \lambda(j_0)}{1 - c_0} > 0, \\ \forall j \in \mathcal{S}_-^c \cap \mathcal{J}_-, \quad \text{plim}_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{1}{n\delta_j} \{GC_p(j|\alpha) - GC_p(j_0|\alpha)\} &= 1 > 0. \end{aligned}$$

Consequently, Lemma 5 is proved.

Appendix A.6. Proof of Lemma 6

We first give a lemma required for proving Lemmas 6 and 8.

Lemma A5. Let X be a Bernoulli random variable with probability of success θ satisfying

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} \theta = 0.$$

Then, for any $j \in \mathcal{J}$, $E[\mathcal{L}(j)X]/p$ can be expanded as

$$\frac{1}{p}E[\mathcal{L}(j)X] = \frac{n}{p}\delta_j\theta + o(1),$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$.

Proof of Lemma A5. Notice that

$$\begin{aligned} E[p^{-1}\mathcal{L}(j)] &= p^{-1}n\delta_j + k_j, & E[X] &= \theta, \\ \text{Var}[p^{-1}\mathcal{L}(j)] &= \text{Var}[p^{-1}\text{tr}(\mathbf{E}^\top \mathbf{P}_j \mathbf{E})] = 2p^{-1}k_j, & \text{Var}[X] &= \theta(1 - \theta). \end{aligned}$$

From the above equations and the θ assumption, it follows that

$$\begin{aligned} &|E[p^{-1}\mathcal{L}(j)X] - E[p^{-1}\mathcal{L}(j)]E[X]| = |\text{Cov}[p^{-1}\mathcal{L}(j), X]| \\ &\leq \sqrt{\text{Var}[p^{-1}\mathcal{L}(j)]\text{Var}[X]} = \sqrt{2p^{-1}k_j\theta(1 - \theta)} \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Hence, we derive

$$\frac{1}{p}E[\mathcal{L}(j)X] = \frac{1}{p}E[\mathcal{L}(j)]E[X] + o(1) = \left(\frac{n}{p}\delta_j + k_j\right)\theta + o(1) = \frac{n}{p}\delta_j\theta + o(1),$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Consequently, Lemma A5 is proved. $\square$

Since $\mathcal{R}(j_{\text{ME}})$ is the minimum value of $\mathcal{R}(j)$ among $j \in \mathcal{J}$, $\mathcal{R}(j_{\text{ME}}) \leq \mathcal{R}(j_0)$ holds. On the other hand, $\mathcal{L}(\hat{j}_{\text{LE}})$ is the minimum value of $\mathcal{L}(j)$ among $j \in \mathcal{J}$, $\mathcal{L}(\hat{j}_{\text{LE}}) \leq \mathcal{L}(j_{\text{ME}})$ holds. This implies that $E[\mathcal{L}(\hat{j}_{\text{LE}})] \leq E[\mathcal{L}(j_{\text{ME}})] = \mathcal{R}(j_{\text{ME}})$. Therefore, we have

$$E[\mathcal{L}(\hat{j}_{\text{LE}})] \leq \mathcal{R}(j_{\text{ME}}) \leq \mathcal{R}(j_0). \quad (\text{A14})$$

By using the equation in (15) and j_0 in (19), we derive

$$\mathcal{L}(j_0) \geq \mathcal{L}(\hat{j}_{\text{LE}}) = \sum_{j \in \mathcal{J}} \mathcal{L}(j)I(\hat{j}_{\text{LE}} = j).$$

This implies that

$$\begin{aligned} \mathcal{L}(j_0)\{1 - I(\hat{j}_{\text{LE}} = j_0)\} &\geq \sum_{j \in \mathcal{J} \setminus \{j_0\}} \mathcal{L}(j)I(\hat{j}_{\text{LE}} = j) \\ \iff \mathcal{L}(j_0)I(\hat{j}_{\text{LE}} \neq j_0) &\geq \sum_{j \in \mathcal{J} \setminus \{j_0\}} \mathcal{L}(j)I(\hat{j}_{\text{LE}} = j). \end{aligned} \quad (\text{A15})$$

It follows from Theorem 3 that $P(\hat{j}_{\text{LE}} \neq j_0) \rightarrow 0$ as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Since $I(\hat{j}_{\text{LE}} \neq j_0) \sim B(1, P(\hat{j}_{\text{LE}} \neq j_0))$, by using Lemma A5, we have

$$\frac{1}{p}E[\mathcal{L}(j_0)I(\hat{j}_{\text{LE}} \neq j_0)] = \frac{n}{p}\delta_{j_0}P(\hat{j}_{\text{LE}} \neq j_0) + o(1),$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Notice that

$$\frac{n}{p}\delta_{j_0} = \begin{cases} c_0^{-1}\delta_{j_0,0} + o(1) & (\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}) \\ 0 & (\{\xi_* = \emptyset\} \cup \{c_0 = 0\}) \end{cases}. \quad (\text{A16})$$

Since $0 < c_0^{-1}\delta_{j_0,0} < \infty$, we have

$$\frac{1}{p}E[\mathcal{L}(j_0)I(\hat{j}_{\text{LE}} \neq j_0)] \rightarrow 0, \quad (\text{A17})$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. It follows from (A15) and (A17) that

$$\frac{1}{p} \sum_{j \in \mathcal{J} \setminus \{j_0\}} E[\mathcal{L}(j)I(\hat{j}_{\text{LE}} = j)] \rightarrow 0,$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. This implies that

$$\begin{aligned} \frac{1}{p}E[\mathcal{L}(\hat{j}_{\text{LE}})] &= \frac{1}{p}E[\mathcal{L}(j_0)I(\hat{j}_{\text{LE}} = j_0)] + o(1) \\ &= \frac{1}{p}E[\mathcal{L}(j_0)] - \frac{1}{p}E[\mathcal{L}(j_0)I(\hat{j}_{\text{LE}} \neq j_0)] + o(1) \\ &= \frac{1}{p}\mathcal{R}(j_0) + o(1), \end{aligned}$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. By using the above result and (A14), Lemma 6 is proved.

Appendix A.7. Proof of Lemma 7

We first give a lemma required for proving Lemma 7.

Lemma A6. Let $\mathcal{W}_-$ be a set of underspecified models which satisfy Equation (16), i.e.,

$$\mathcal{W}_- = \left\{ j \in \mathcal{J}_- \mid \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{p}{n\delta_j} \left(-\frac{n}{n-p} + \alpha \right) = 0 \right\}.$$

Suppose that Assumptions A1 and A2 hold. The order of the selection probability of $j \in \mathcal{W}_-$ by the GC_p is given by

$$P(\hat{j}_\alpha = j) = O(n^{-r}),$$

where r is any positive integer.

Proof of Lemma A6. Let $q_j = k_{j^c \cap j_*}$ for any $j \in \mathcal{J}_-$, and let $V, U_1, \dots, U_{q_j}$ be mutually independent random variables defined by $V \sim \chi_m^2$ and $U_i \sim \chi_{n\delta_{j,i}}^2$, where $\delta_{j,i}$ is a positive value satisfying $\delta_{j,1} + \dots + \delta_{j,q_j} = \delta_j$ and $\delta_{i,j} \geq \delta_j/q_j^2$ for all $i \in \{1, \dots, q_j\}$. Here, m is given in (8) and δ_j is given by (6). Suppose that Assumption A1 holds. Then, from [31] or [32], the probability of selecting $j \in \mathcal{J}_-$ by a GC_p is bounded as

$$P(\hat{j}_\alpha = j) \leq \sum_{i=1}^{q_j} P(U_i/V \leq p\alpha/(n-k)). \quad (\text{A18})$$

Let $\gamma_{j,i}$ be a constant defined by

$$\gamma_{j,i} = \frac{1}{\delta_{j,i}} \left(\frac{p + n\delta_{j,i}}{m-2} - \frac{p\alpha}{n-k} \right) \quad (j \in \mathcal{J}_-, i = 1, \dots, q_j).$$

From a simple calculation, we have

$$\gamma_{j,i} = \frac{n}{m-2} + \frac{p(n+kp)}{(m-2)(n-k)(n-p)\delta_{j,i}} - \frac{p}{(n-k)\delta_{j,i}} \left(-\frac{n}{n-p} + \alpha \right). \quad (\text{A19})$$

It follows from the relations $\delta_{j,i} \geq \delta_j/q_j^2$ and $q_j \leq k_*$, the boundedness of $1/\delta_j$, and the definition of $\mathcal{W}_-$ that

$$\begin{aligned} \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{p(n+kp)}{(m-2)(n-k)(n-p)\delta_{j,i}} &= 0, \\ \lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{p}{(n-k)\delta_{j,i}} \left(-\frac{n}{n-p} + \alpha \right) &= 0. \end{aligned} \quad (\text{A20})$$

Using Equations (A19) and (A20) yields

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} \gamma_{j,i} = \frac{1}{1-c_0} > 0.$$

Hence, $\gamma_{j,i}$ is positive for sufficiently large n or large n and p . This implies that for sufficiently large n or large n and p

$$\begin{aligned} P\left(\frac{U_i}{V} \leq \frac{p\alpha}{n-k}\right) &= P\left(\frac{1}{\delta_{j,i}} \left(\frac{U_i}{V} - \frac{p+n\delta_{j,i}}{m-2}\right) \leq -\gamma_{j,i}\right) \\ &\leq P\left(\frac{1}{\delta_{j,i}} \left|\frac{U_i}{V} - \frac{p+n\delta_{j,i}}{m-2}\right| \geq \gamma_{j,i}\right) \\ &\leq \frac{1}{(\gamma_{j,i}\delta_{j,i})^{2r}} E\left[\left(\frac{U_i}{V} - \frac{p+n\delta_{j,i}}{m-2}\right)^{2r}\right], \end{aligned} \quad (\text{A21})$$

where r is any positive integer. The results in [31] indicate that

$$E\left[\left(\frac{U_i}{V} - \frac{p+n\delta_{j,i}}{m-2}\right)^{2r}\right] = O(n^{-r}\delta_{j,i}^{2r}). \quad (\text{A22})$$

Substituting (A21) and (A22) into (A18) yields $P(\hat{j}_\alpha = j) = O(n^{-r})$. Consequently, Lemma A6 is proved. $\square$

If α satisfies Condition C2, $\mathcal{W}_- = \mathcal{J}_-$ holds. On the other hand, when α satisfies Condition C3, for Equation (16) to hold, it must be the case that $\delta_j = \infty$. Hence, if α satisfies Condition C3, $\mathcal{W}_- = \mathcal{S}_-^c \cap \mathcal{J}_-$ holds. Recall that r in Lemma A6 is any positive integer. From Lemma A6, the selection probability of $j \in \mathcal{W}_-$ can be expressed as

$$P(\hat{j}_\alpha = j) = \frac{M_j}{n^{\epsilon+1}},$$

where M_j is a constant independent of n and p , and ϵ is the positive integer defined in Assumption A3. This implies that

$$\frac{n}{p}\delta_j P(\hat{j}_\alpha = j) = \frac{1}{p} \frac{\delta_j}{n^\epsilon} M_j = 0,$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Consequently, Lemma 7 is proved.

Appendix A.8. Proof of Lemma 8

From Lemma 5, if α satisfies Conditions C1 and C2 when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$, or Condition C3 when $\{\xi_* \neq \emptyset\} \cap \{c_0 \neq 0\}$, the following equations are satisfied:

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} P(\hat{j}_\alpha \neq j_0) = 0, \quad \lim_{n \rightarrow \infty, p/n \rightarrow c_0} P(\hat{j}_\alpha = j) = 0 \quad (\forall j \in \mathcal{J} \setminus \{j_0\}),$$

where j_0 is the asymptotically loss optimal model given by (19). Notice that

$$I(\hat{j}_\alpha \neq j_0) \sim B(1, P(\hat{j}_\alpha \neq j_0)), \quad I(\hat{j}_\alpha = j) \sim B(1, P(\hat{j}_\alpha = j)), \quad (\forall j \in \mathcal{J} \setminus \{j_0\}).$$

By using Lemma A5 and Equation (A16), we have

$$\frac{1}{p} E[\mathcal{L}(j_0) I(\hat{j}_\alpha \neq j_0)] = \frac{n}{p} \delta_{j_0} P(\hat{j}_\alpha \neq j_0) + o(1) = o(1) \text{ as } n \rightarrow \infty, p/n \rightarrow c_0. \quad (\text{A23})$$

Recall that $\delta_j = 0$ holds for any $j \in \mathcal{J}_+$, or $np^{-1}\delta_j = O(1)$ as $n \rightarrow \infty$ and $p/n \rightarrow c_0$ for any $j \in \mathcal{S}_-$ when $c_0 \neq 0$. From these equations and Lemma 7, we can see that

$$\begin{aligned} \frac{n}{p} \delta_j P(\hat{j}_\alpha = j) &= \begin{cases} 0 & (\forall j \in \mathcal{J}_+ \setminus \{j_0\}) \\ o(1) & \begin{cases} \forall j \in \mathcal{S}_- \setminus \{j_0\} & \text{when } c_0 \neq 0 \\ \forall j \in \mathcal{S}_-^c \cap \mathcal{J}_- & \text{when } c_0 \neq 0 \\ \forall j \in \mathcal{J}_- & \text{when } c_0 = 0 \end{cases} \end{cases} \\ &= \begin{cases} 0 & (\forall j \in \mathcal{J}_+ \setminus \{j_0\}) \\ o(1) & (\forall j \in \mathcal{J}_- \setminus \{j_0\}) \end{cases}, \end{aligned}$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. It follows from the equation and Lemma A5 that for any $j \in \mathcal{J} \setminus \{j_0\}$

$$\frac{1}{p} E[\mathcal{L}(j) I(\hat{j}_\alpha = j)] = \frac{n}{p} \delta_j P(\hat{j}_\alpha = j) + o(1) = o(1), \quad (\text{A24})$$

as $n \rightarrow \infty$ and $p/n \rightarrow c_0$. Notice that

$$\begin{aligned} \frac{1}{p} E[\mathcal{L}(\hat{j}_\alpha)] &= \frac{1}{p} \sum_{j \in \mathcal{J}} E[\mathcal{L}(j) I(\hat{j}_\alpha = j)] \\ &= \frac{1}{p} \mathcal{R}(j_0) - \frac{1}{p} E[\mathcal{L}(j_0) I(\hat{j}_\alpha \neq j_0)] + \frac{1}{p} \sum_{j \in \mathcal{J} \setminus \{j_0\}} E[\mathcal{L}(j) I(\hat{j}_\alpha = j)]. \end{aligned}$$

By substituting (A23) and (A24) into the above equation, Lemma 8 is proved.

Appendix A.9. Proof of Corollary 1

To prove Corollary 1, it is sufficient to show that α satisfying Condition C3 satisfies Condition C2 when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$. Recall that $\xi_* = \emptyset \Leftrightarrow \delta_j \rightarrow \infty$ for any $j \in \mathcal{J}_-$. Notice that $p/(n\delta_j)$ and $p\alpha/(n\delta_j)$ converge to 0 when $\delta_j \rightarrow \infty$ or $c_0 = 0$. These convergences to 0 and $n/(n-p) \rightarrow 1/(1-c_0)$ imply that

$$\lim_{n \rightarrow \infty, p/n \rightarrow c_0} \frac{p}{n\delta_j} \left(-\frac{n}{n-p} + \alpha \right) = 0.$$

This indicates that Condition C2 holds when $\{\xi_* = \emptyset\} \cup \{c_0 = 0\}$. Consequently, Corollary 1 is proved.

References

1. Srivastava, M.S. *Methods of Multivariate Statistics*; John Wiley & Sons, New York, NY, USA, 2002.
2. Timm, N.H. *Applied Multivariate Analysis*; Springer-Verlag: New York, NY, USA, 2002.
3. Darroudi, S.; Eslamiyeh, M.; Mansoori, A.; Nazar, E.; Alinezhad-Namaghi, M.; Rezvani, R.; Ghayour-Mobarhan, M.; Esmaily, H.; Moohebat, M. Multivariate linear regression to predict association of non-invasive arterial stiffness with cardiovascular events. *ESC Heart Fail.* **2025**, *12*, 1141–1150.
4. Friston, K.J.; Holmes, A.P.; Worsley, K.J.; Poline, J.-P.; Frith, C.D.; Frackowiak, R.S.J. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **1994**, *2*, 189–210.
5. Mountford, A.; Uhlig, H. What are the effects of fiscal policy shocks? *J. Appl. Econom.* **2009**, *24*, 960–992.
6. Naylor, M.G.; Cardenas, V.A.; Tosun, D.; Schuff, N.; Weiner, M.; Schwartzman, A.; Voxelwise multivariate analysis of multimodality magnetic resonance imaging. *Hum. Brain Mapp.* **2014**, *35*, 831–846.
7. Classical Multivariate Regression. Available online: <https://cran.r-project.org/web/packages/rrr/vignettes/rrr.html> (accessed on 25 April 2026).
8. Fox, J.; Weisberg, S. Multivariate linear models in R. In *An R Companion to Applied Regression*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2019. Available online: <https://www.john-fox.ca/Companion/appendices.html> (accessed on 23 April 2026).
9. Atkinson, A.C. A note on the generalized information criterion for choice of a model. *Biometrika* **1980**, *67*, 413–418.
10. Nagai, I.; Yanagihara, H.; Satoh, K. Optimization of ridge parameters in multivariate generalized ridge regression by plug-in methods. *Hiroshima Math. J.* **2012**, *42*, 301–324.
11. Sparks, R.S.; Coutsourides, D.; Troskie, L. The multivariate C_p . *Comm. Statist. A—Theory Methods* **1983**, *12*, 1775–1793.
12. Mallows, C.L. Some comments on C_p . *Technometrics* **1973**, *15*, 661–675.
13. Mallows, C.L. More comments on C_p . *Technometrics* **1995**, *37*, 362–372.
14. Fujikoshi, Y.; Satoh, K. Modified AIC and C_p in multivariate linear regression. *Biometrika* **1997**, *84*, 707–716.
15. Shao, J. An asymptotic theory for linear model selection. *Stat. Sin.* **1997**, *7*, 221–264.
16. Shibata, R. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **1980**, *8*, 147–164.
17. Shibata, R. An optimal selection of regression variables. *Biometrika* **1981**, *68*, 45–54.
18. Yanagihara, H. A high-dimensionality-adjusted consistent C_p -type statistic for selecting variables in a normality-assumed linear regression with multiple responses. *Procedia Comput. Sci.* **2016**, *96*, 1096–1105.
19. Imori, S. Asymptotic optimality of C_p -type criteria in high-dimensional multivariate linear regression models. *Statist. Sin.* **2023**, *33*, 1233–1248.
20. Akaike, H. Information theory and an extension of the maximum likelihood principle. In *International Symposium on Information Theory*, 2nd ed.; Petrov, B.N., Csáki, F., Eds.; Akadémiai Kiadó: Budapest, Hungary, 1973; pp. 267–281.
21. Schwarz, G. Estimating the dimension of a model. *Ann. Statist.* **1978**, *6*, 461–464.
22. Nishii, R. Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.* **1984**, *12*, 758–765.
23. Yanagihara, H.; Wakaki, H.; Fujikoshi, Y. A consistency property of the AIC for multivariate linear models when the dimension and the sample size are large. *Electron. J. Stat.* **2015**, *9*, 869–897.
24. Fujikoshi, Y.; Sakurai, T.; Yanagihara, H. Consistency of high-dimensional AIC-type and C_p -type criteria in multivariate linear regression. *J. Multivar. Anal.* **2014**, *123*, 184–200.
25. Rao, C.R.; Wu, Y. A strongly consistent procedure for model selection in a regression problem. *Biometrika* **1989**, *76*, 369–374.
26. Bai, Z.D.; Rao, C.R.; Wu, Y. Model selection with data-oriented penalty. *J. Stat. Plann. Inference* **1999**, *77*, 103–117.
27. Bozdogan, H. Model selection and Akaike’s information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **1987**, *52*, 345–370.
28. Hannan, E.J.; Quinn, B.G. The determination of the order of an autoregression. *J. Roy. Stat. Soc. Ser. B* **1979**, *41*, 190–195.
29. Harville, D.A. *Matrix Algebra from a Statistician’s Perspective*; Springer-Verlag: New York, NY, USA, 1997.
30. Yanagihara, H. Conditions for consistency of a log-likelihood-based information criterion in normal multivariate linear regression models under the violation of normality assumption. *J. Jpn. Stat. Soc.* **2015**, *45*, 21–56.
31. Oda, R.; Yanagihara, H. A fast and consistent variable selection method for high-dimensional multivariate linear regression with a large number of explanatory variables. *Electron. J. Stat.* **2020**, *14*, 1386–1412.
32. Yanagihara, H. Convergence order of the probability of selecting the true set of variables by consistent C_p -type criterion in normal multivariate linear regression models. *Sūrikaiseikikenkyūsho Kōkyūroku* **2016**, *1999*, 72–85. (In Japanese)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.