

Article

High-Dimensional Consistencies of KOO Methods for the Selection of Variables in Multivariate Linear Regression Models with Covariance Structures

Yasunori Fujikoshi ^{1,*} and Tetsuro Sakurai ²¹ Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-2 Kagamiyama, Hiroshima 739-8626, Japan² School of General and Management Studies, Suwa University of Science, 5000-1 Toyohira, Chino 391-0292, Japan

* Correspondence: fujikoshi_y@yahoo.co.jp

Abstract: In this paper, we consider the high-dimensional consistencies of KOO methods for selecting response variables in multivariate linear regression with covariance structures. Here, the covariance structures are considered as (1) independent covariance structure with the same variance, (2) independent covariance structure with different variances, and (3) uniform covariance structure. A sufficient condition for model selection consistency is obtained using a KOO method under a high-dimensional asymptotic framework, such that sample size n , the number p of response variables, and the number k of explanatory variables are large, as in $p/n \rightarrow c_1 \in (0, 1)$ and $k/n \rightarrow c_2 \in [0, 1)$, where $c_1 + c_2 < 1$.

Keywords: consistency property; covariance structures; high-dimensional asymptotic framework; KOO methods; multivariate linear regression

MSC: 62H12; 62H10



Citation: Fujikoshi, Y.; Sakurai, T. High-Dimensional Consistencies of KOO Methods for the Selection of Variables in Multivariate Linear Regression Models with Covariance Structures. *Mathematics* **2023**, *11*, 671. <https://doi.org/10.3390/math11030671>

Academic Editors: Alexander Tikhomirov and Liangxiao Jiang

Received: 18 November 2022

Revised: 28 December 2022

Accepted: 17 January 2023

Published: 28 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

We focus on a multivariate linear regression model of p response variables y_1, \dots, y_p on a subset of k explanatory variables x_1, \dots, x_k . Suppose that there are n observations on a p -dimensional response vector $\mathbf{y} = (y_1, \dots, y_p)'$ and a k -dimensional explanatory vector $\mathbf{x} = (x_1, \dots, x_k)'$, and let $\mathbf{Y} : n \times p$ and $\mathbf{X} : n \times k$ be the observation matrices of \mathbf{y} and \mathbf{x} with sample size n , respectively. The multivariate linear regression model including all the explanatory variables under normality is written as follows:

$$\mathbf{Y} \sim N_{n \times p}(\mathbf{X}\Theta, \Sigma \otimes \mathbf{I}_n), \quad (1)$$

where Θ is a $k \times p$ unknown matrix of regression coefficients, and Σ is a $p \times p$ unknown covariance matrix that is positive definite. $N_{n \times p}(\cdot, \cdot)$ is the normal matrix distribution, such that the mean of \mathbf{Y} is $\mathbf{X}\Theta$, and the covariance matrix of $\text{vec}(\mathbf{Y})$ is $\Sigma \otimes \mathbf{I}_n$; equivalently, the rows of \mathbf{Y} are independently normal with the same covariance matrix Σ . Here, $\text{vec}(\mathbf{Y})$ is the $np \times 1$ column vector that is obtained by stacking the columns of \mathbf{Y} on top of one another. We assumed that $\text{rank}(\mathbf{X}) = k$.

In multivariate linear regression, the selection of variables for the model is an important concern. One of the approaches is to first consider variable selection models and then apply model selection criteria such as AIC and BIC. Such a criterion for Full Model (1) is expressed as follows:

$$\text{GIC} = -2 \log L(\hat{\Xi}) + dg, \quad (2)$$

where $L(\hat{\Xi})$ is the maximal likelihood, $\Xi = \{\Theta, \Sigma\}$, $d > 0$ is the penalty term, and g is the number of unknown parameters given by $\{kp + \frac{1}{2}p(p+1)\}$. For AIC and BIC, d is

defined as 2 and $\log n$, respectively. In the selection of k variables x_1, \dots, x_k , we identified $\{x_1, \dots, x_k\}$ with the index set $\{1, \dots, k\} \equiv \omega$, and denote GIC for subset $j \subset \omega$ by GIC_j . Then, the model selection based on GIC chooses the following model:

$$\tilde{j} = \arg \min_j \text{GIC}_j. \quad (3)$$

Here the minimum is usually taken for all combinations of response variables. There are computational problems for the methods based on GIC, including AIC and BIC methods, since we need to compute $2^k - 1$ statistics for the selection of k explanatory variables. To avoid this computational problem, [1] proposed a method that was essentially thanks to [2]. The method, which was named the knock-one-out (KOO) method by [3], determines “selection” or “no selection” for each variable by comparing the model removing that variable and the full model. More precisely, the KOO method chooses the model or the set of variables given by

$$\hat{j} = \{j \in \omega \mid \text{GIC}_{\omega \setminus j} > \text{GIC}_\omega\}, \quad (4)$$

where $\omega \setminus j$ is a short expression for $\omega \setminus \{j\}$, which is the set obtained by removing element j from the set ω . In general, the KOO method can be applied to a method or criterion, not only AIC, a general variable selection criterion or method.

In the literature on multivariate linear regression, numerous papers have dealt with the variable selection problem, as it relates to selecting explanatory variables. When Σ is unknown positive definite, [4–6], for example, indicated that, in a high-dimensional case, AIC and C_p have consistency properties, but BIC is not necessarily consistent. KOO methods in the multivariate regression model were studied by [3] and [7,8]. The KOO method in discriminant analysis; see [9], and [10]. For a review, see [11].

In this paper, we assume that the covariance structure was one of three covariance structures: (1) an independent covariance structure with the same variance, (2) an independent covariance structure with different variances, and (3) a uniform covariance structure. The numbers of unknown parameters in covariance structures (1)–(3) were 1, p , and 2, respectively. Sufficient conditions for the KOO method given by (4) to be consistent were derived under a high-dimensional asymptotic framework, such that sample size n , the number p of response variables, and the number k of explanatory variables were large, as in $p/n \rightarrow c_1 \in (0, 1)$ and $k/n \rightarrow c_2 \in [0, 1)$, where $c_1 + c_2 < 1$. Ref. [12] considered similar problems under covariance structures (1), (3), and (4), an autoregressive covariance structure, but did not consider them under (2). Moreover, in the study of asymptotic consistencies, they assumed that k was fixed, but in this paper, k may tend to infinity, such that $k/n \rightarrow c_2 \in [0, 1)$. From the numerical experiments in [12], we know that the probability of choosing the true model in Cases (1) and (3) results from the following table (Table 1). In variable selection for multivariate linear regression using the KOO method, the probability of selecting the true model is shown in the following table. Here, we examine Cases (1), an independent covariance structure with the same variance, and (3), a uniform covariance structure.

Table 1. KOO Based on AIC.

$k = 3$ (n, p)	KOO Based on AIC		KOO Based on AIC	
	(20, 10)	(200, 100)	(20, 10)	(200, 100)
(1)	0.74	1.00	0.77	1.00
(3)	0.47	1.00	0.22	1.00

In this table (Table 1), k is the number of nonzero true explanatory variables, and the true parameter values were omitted. In [12], k was treated as finite. In this paper, k may tend to infinity, such that $k/n \rightarrow c_2 \in [0, 1)$.

The present paper is organized as follows. In Section 2, we present notations and preliminaries. In Section 3, we state KOO methods with Covariance Structures (1)–(3) in

terms of key statistics. Further, an approach for their consistencies is stated in Section 3. In Sections 4–6, we discuss consistency properties of KOO methods under Covariance Structures (1)–(3). In Section 7, our conclusions are discussed.

2. Notations and Preliminaries

Suppose that j denotes a subset of $\omega = \{1, \dots, k\}$ containing k_j elements, and \mathbf{X}_j denotes the $n \times k_j$ matrix comprising the columns of \mathbf{X} indexed by the elements of j . Then, $\mathbf{X}_\omega = \mathbf{X}$. Further, we assumed that covariance matrix Σ had a covariance structure Σ_c . Then, we have a generic candidate model:

$$M_{c,j} : \mathbf{Y} \sim N_{n \times p}(\mathbf{X}_j \Theta_j, \Sigma_{c,j} \otimes \mathbf{I}_n), \quad (5)$$

where Θ_j is a $k_j \times p$ unknown matrix of regression coefficients. We assumed that $\text{rank}(\mathbf{X}) = k$.

When $\Sigma_{c,j}$ is a $p \times p$ unknown covariance matrix, we could write the GIC in (2) as follows:

$$\text{GIC}_{c,j} = n \log |\hat{\Sigma}_j| + np(\log 2\pi + 1) + d \left\{ k_j p + \frac{1}{2} p(p+1) \right\}, \quad (6)$$

where $n\hat{\Sigma}_j = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_j)\mathbf{Y}$ and $\mathbf{P}_j = \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'$. When $j = \omega$, model $M_{c,\omega}$ is called the full model. $\hat{\Sigma}_{c,\omega}$ and \mathbf{P}_ω are defined from $\hat{\Sigma}_{c,j}$ and \mathbf{P}_j as $j = \omega$, $k_\omega = k$ and $\mathbf{X}_\omega = \mathbf{X}$.

In this paper, we considered the cases in which the covariance matrix Σ_c belonged to each of the following three structures:

- (1) Independent covariance structure with the same variance (ICSS).

$$\Sigma_v = \sigma_v^2 \mathbf{I}_p,$$

- (2) Independent covariance structure with different variances (ICSD).

$$\Sigma_b = \text{diag}(\sigma_1^2, \dots, \sigma_p^2),$$

- (3) Uniform covariance structure (UCS).

$$\Sigma_u = \sigma_u^2 (\rho_u^{1-\delta_{ij}})_{1 \leq i,j \leq p}.$$

The models considered in this paper can be expressed as in (5) with $\Sigma_{v,j}$, $\Sigma_{b,j}$, and $\Sigma_{u,j}$ for $\Sigma_{c,j}$. Let $f(\mathbf{Y}; \Theta_j, \Sigma_{c,j})$ be the density of \mathbf{Y} in (5) with $\Sigma = \Sigma_{c,j}$. In the derivation of the GIC, under the covariance structure $\Sigma = \Sigma_{c,j}$, we use the following equality:

$$\begin{aligned} -2 \log \max_{\Theta_j, \Sigma_{c,j}} f(\mathbf{Y}; \Theta_j, \Sigma_{c,j}) &= np \log(2\pi) \\ &+ \min_{\Sigma_{c,j}} \left\{ np \log |\Sigma_{c,j}| + \text{tr} \Sigma_{c,j}^{-1} \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_j) \mathbf{Y} \right\}. \end{aligned} \quad (7)$$

Let $\hat{\Sigma}_{c,j}$ be the quantity minimizing the right-hand side of (7). Then, in our model, it satisfies $\text{tr} \hat{\Sigma}_{c,j}^{-1} \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_j) \mathbf{Y} = np$, and we obtain

$$\begin{aligned} \text{GIC}_{c,j} &= -2 \log f(\mathbf{Y}; \hat{\Theta}_j, \hat{\Sigma}_c) + dm_{c,j} \\ &= np \log |\hat{\Sigma}_{c,j}| + np(\log 2\pi + 1) + dm_{c,j}, \end{aligned} \quad (8)$$

where $m_{c,j}$ is the number of independent unknown parameters under $M_{c,j}$, and d is a positive constant that may depend on n . For AIC and BIC, d is defined by 2 ([13]) and $\log n$ ([14]), respectively.

3. Approach to Consistencies of KOO Methods

Our KOO method is based on

$$T_{c,j;d} = \text{GIC}_{c,\omega \setminus j} - \text{GIC}_{c,\omega}. \quad (9)$$

In fact, the KOO method chooses the following model:

$$\hat{j}_{c;d} = \{j \mid T_{c,j;d} > 0\}. \quad (10)$$

Its consistency can be proven by showing the following two properties:

$$\text{Q1 : } [\text{F1}] \equiv \sum_{j \in j_*} \Pr(T_{c,j;d} \leq 0) \rightarrow 0, \quad (11)$$

$$\text{Q2 : } [\text{F2}] \equiv \sum_{j \notin j_*} \Pr(T_{c,j;d} \geq 0) \rightarrow 0, \quad (12)$$

as in [11]. The result can be shown by using the following inequality:

$$\begin{aligned} \Pr(\hat{j}_{c;d} = j_*) &= \Pr\left(\bigcap_{j \in j_*} "T_{c,j;d} > 0" \bigcap_{j \notin j_*} "T_{c,j;d} < 0"\right) \\ &= 1 - \Pr\left(\bigcup_{j \in j_*} "T_{c,j;d} \leq 0" \bigcup_{j \notin j_*} "T_{c,j;d} \geq 0"\right) \\ &\geq 1 - \sum_{j \in j_*} \Pr(T_{c,j;d} \leq 0) - \sum_{j \notin j_*} \Pr(T_{c,j;d} \geq 0). \end{aligned}$$

Here, [F1] denotes the probability that true variables are not selected, and [F2] denotes the probability that nontrue variables are selected. Such notations are used for other variable selection methods. x_j is included in the true set of variables if $\theta_j \neq 0$.

Here, we list some of our main assumptions:

A1: The set j_* of the true explanatory variables is included in the full subset, i.e., $j_* \subset \omega$. and the set j_* is finite.

A2: The high-dimensional asymptotic framework:

$p \rightarrow \infty$, $n \rightarrow \infty$, $k \rightarrow \infty$, $p/n \rightarrow c_1 \in (0, 1)$, $k/n \rightarrow c_2 \in [0, 1)$,
where $0 < c_1 + c_2 < 1$.

A general model selection criterion $\hat{j}_{c;d}$ is high-dimensionally consistent if

$$\lim \Pr(\hat{j}_{c;d} = j_*) = 1,$$

under a high-dimensional asymptotic framework. Here, "lim" means the limit under A2.

4. Asymptotic Consistency under an Independent Covariance Structure

In this section, we show an asymptotic consistency of the KOO method on the basis of a general information criterion under an independent covariance structure. A generic candidate model when the set of explanatory variables is j can be expressed as follows:

$$M_{v,j} : \mathbf{Y} \sim N_{n \times p}(\mathbf{X}_j \Theta_j, \Sigma_{v,j} \otimes \mathbf{I}_n), \quad (13)$$

where $\Sigma_{v,j} = \sigma_{v,j}^2 \mathbf{I}_p$ and $\sigma_{v,j}^2 > 0$. Let us denote the density of \mathbf{Y} under (13) with $f(\mathbf{Y}; \Theta_j, \sigma_{v,j})$. Then, we have

$$\begin{aligned} -2 \log f(\mathbf{Y}; \Theta_j, \sigma_{v,j}^2) &= np \log(2\pi) + np \log \sigma_{v,j}^2 \\ &\quad + \frac{1}{\sigma_{v,j}^2} \text{tr}(\mathbf{Y} - \mathbf{X}_j \Theta_j)' (\mathbf{Y} - \mathbf{X}_j \Theta_j). \end{aligned}$$

Therefore, the maximal estimators of Θ_j and $\sigma_{v,j}^2$ under $M_{v,j}$ are given as follows:

$$\hat{\Theta}_j = (\mathbf{X}_j' \mathbf{X}_j)^{-1} \mathbf{X}_j' \mathbf{Y}, \quad \hat{\sigma}_{v,j}^2 = \frac{1}{np} \text{tr} \mathbf{Y}' (\mathbf{I}_n - \mathbf{P}_j) \mathbf{Y}. \quad (14)$$

General Information Criterion (8) is given by

$$\text{GIC}_{v,j} = np \log \hat{\sigma}_{v,j}^2 + np(\log 2\pi + 1) + dm_{v,j}, \quad (15)$$

where d is a positive constant, and $m_{v,j} = k_j p + 1$.

Using (9) and (15), we have

$$\begin{aligned} T_{v,j;d} &\equiv \text{GIC}_{v,\omega \setminus j} - \text{GIC}_{v,\omega} \\ &= np \log \left(1 + U_{2j} U_1^{-1} \right) - dp, \end{aligned} \quad (16)$$

where

$$\begin{aligned} U_1 &= \text{tr} \mathbf{Y}' (\mathbf{I}_n - \mathbf{P}_\omega) \mathbf{Y} = \sum_{\ell=1}^p \mathbf{y}_\ell' (\mathbf{I}_n - \mathbf{P}_\omega) \mathbf{y}_\ell, \\ U_{2j} &= \text{tr} \mathbf{Y}' (\mathbf{P}_\omega - \mathbf{P}_{\omega \setminus j}) \mathbf{Y} = \sum_{\ell=1}^p \mathbf{y}_\ell' (\mathbf{P}_\omega - \mathbf{P}_{\omega \setminus j}) \mathbf{y}_\ell. \end{aligned}$$

$U_1/\sigma_{v,j_*}^2$ and $U_{2j}/\sigma_{v,j_*}^2$ are independently distributed as a central and a noncentral chi-squared distribution, respectively. More precisely, assume that

$$\mathbf{E}(\mathbf{Y}) = \mathbf{X}_{j_*} \Theta_{j_*}, \quad (17)$$

and let $\sigma_{v,*}^2 = \sigma_{v,j_*}^2$. Then, using basic distributional properties (see, [15]) on quadratic forms of normal variates and Wishart matrices, we have the following results:

$$\begin{aligned} (1) \quad &U_1/\sigma_{v,*}^2 \sim \chi_{(n-k)p}^2, \\ (2) \quad &U_{2j}/\sigma_{v,*}^2 \sim \chi_p^2(\delta_{v,j}^2), \\ (3) \quad &U_1 \perp U_{2j}, \end{aligned} \quad (18)$$

where noncentrality parameter $\tau_{v,j}^2$ is defined by

$$\delta_{v,j}^2 = \frac{1}{\sigma_{v,*}^2} \text{tr} (\mathbf{X}_{j_*} \Theta_{j_*})' (\mathbf{P}_\omega - \mathbf{P}_{\omega \setminus j}) \mathbf{X}_{j_*} \Theta_{j_*}.$$

If $j \notin j_*$, $\delta_{v,j}^2 = 0$, and if $j \in j_*$, in general, $\tau_{v,j}^2 \neq 0$. For a sufficient condition for the consistency of the KOO method based on $\text{GIC}_{v,j}$, we assumed

$$\text{A3v} : \text{For any } j \in j_*, \delta_{v,j}^2 = O(np), \text{ and } \lim_{p/n \rightarrow c_1} \frac{1}{np} \delta_{v,j}^2 = \eta_{v,j}^2 > 0. \quad (19)$$

Now, we consider the high-dimensional asymptotic consistency of the KOO method based on $\text{GIC}_{v,j}$ in (15), whose selection method is given by $\hat{j}_{v,j;d} = \{j \mid T_{v,j;d} > 0\}$. When $j \notin j_*$, from (16), we can write

$$T_{v,j;d} = np \log \left\{ 1 + \chi_p^2 / \chi_m^2 \right\} - dp, \quad m = (n-k)p.$$

Therefore, we have

$$\begin{aligned}
[F2] &= \sum_{j \notin j_*} \Pr(np \log \{1 + \chi_p^2 / \chi_m^2\} \geq dp) \\
&= (k - k_{j_*}) \Pr(U \geq h) \\
&\leq (k - k_{j_*}) \Pr(U \geq h_0),
\end{aligned} \tag{20}$$

where

$$\begin{aligned}
U &= \frac{\chi_p^2}{\chi_m^2} - \frac{p}{m-2}, \\
h &= e^{d/n} - 1 - \frac{p}{m-2}, \quad h_0 = \frac{d}{n} - \frac{p}{m-2}.
\end{aligned} \tag{21}$$

Note that $h_0 < h$. Then, under the assumption $h_0 > 0$, we have

$$[F2] \leq (k - k_{j_*}) h^{-2\ell} \mathbb{E}[U^{2\ell}] \leq (k - k_{j_*}) h_0^{-2\ell} \mathbb{E}[U^{2\ell}]. \tag{22}$$

Related to the assumption $h_0 > 0$, we assumed

$$A4v : d > \frac{np}{m-2} \rightarrow \frac{1}{1-c_2}, \text{ and } d = O(n^a), \quad 0 < a < 1. \tag{23}$$

The first part in A4v implies $h_0 > 0$. It is easy to see that

$$\mathbb{E}[U^2] = \frac{2p(m+p-2)}{(m-2)^2(m-4)} = O((n^2p)^{-1}).$$

Here, for the first equality, assumption $m > 4$ is required. Further, $h_0^{-2} = O(n^{2(1-a)})$. Therefore, from (22), we have that $[F2] \rightarrow 0$.

When $j \in j_*$, we can write $T_{v,j;d} = np \log \{1 + \chi_p^2(\delta_{v,j}^2) / \chi_m^2\} - dp$. Therefore, we can express [F1] as

$$[F1] = \sum_{j \in j_*} \Pr(\tilde{T}_{v,j;d} \leq 0),$$

where

$$\tilde{T}_{v,j;d} = \frac{p}{n} \log \left\{ 1 + \frac{\chi_p^2(\delta_{v,j}^2)}{\chi_m^2} \right\} - \frac{d}{n}.$$

Assumptions A3v and A4v easily show that

$$\tilde{T}_{v,j;d} \rightarrow c_1 \log(1 + \eta_{v,j}^2) > 0.$$

This implies that $\Pr(\tilde{T}_{v,j;d} \leq 0) \rightarrow 0$.

These imply the following theorem.

Theorem 1. Suppose that Assumptions A1, A2 A3v, and A4v are satisfied. Then, the KOO method based on general information criteria $GIC_{v,j}$ defined by (15) is asymptotically consistent.

An alternative approach for “[F1] $\rightarrow 0$ ”. When $j \in j_*$, we can write

$$T_{v,j;d} = np \log \{1 + \chi_p^2(\delta_{v,j}^2) / \chi_m^2\} - dp.$$

Therefore, we have

$$\begin{aligned}
[F1] &= \sum_{j \in j_*} \Pr(np \log \left\{ 1 + \chi_p^2(\delta_{v,j}^2) / \chi_m^2 \right\} \leq dp) \\
&= \sum_{j \in j_*} \Pr(\tilde{U}_j \leq \tilde{h}_j),
\end{aligned}$$

where, for $j \in j_*$,

$$\tilde{U}_j = \frac{\chi_p^2(\delta_{v,j}^2)}{\chi_m^2} - \frac{p + \delta_{v,j}^2}{m - 2}, \quad \tilde{h}_j = e^{d/n} - 1 - \frac{p + \delta_{v,j}^2}{m - 2} = h - \frac{\delta_{v,j}^2}{m - 2}.$$

Then, under $d = O(n^a)$ ($0 < a < 1$), A3v in (19) and the assumption $\tilde{h}_j < 0$ (or equivalently $h < \delta_j^2 / (m - 2)$), we have

$$[F1] \leq k_{j_*} \max_j |\tilde{h}_j|^{-2\ell} E[\tilde{U}^{2\ell}].$$

It is easily seen that

$$E[\tilde{U}_j^2] = \frac{2(p + 2\delta_{v,j}^2)(m + p - 2 + \delta_{v,j}^2)}{(m - 2)^2(m - 4)} = O((n^2 p)^{-1}),$$

where $m > 4$ and under $d = n^a$ ($0 < a < 1$) and A3v,

$$|\tilde{h}_j|^2 \rightarrow \frac{\eta_{v,j}^2}{c_1(1 - c_2)}.$$

These imply that $[F1] \rightarrow 0$. In this approach, it was assumed that $\tilde{h}_j < 0$ (or equivalently $h < \delta_j^2 / (m - 2)$).

5. Asymptotic Consistency under an Independent Covariance Structure with Different Variances

In this section, we assumed that covariance matrix Σ had an independent covariance matrix with different variances, i.e., $\Sigma = \Sigma_b = \text{diag}(\sigma_{b1}^2, \dots, \sigma_{bp}^2)$. First, let us consider deriving a key statistic $T_{b,j;d} = \text{GIC}_{b,\omega \setminus j} - \text{GIC}_{b,\omega}$. Consider a candidate model with $E(\mathbf{Y}) = \mathbf{X}\Theta$,

$$M_{b,\omega} : \mathbf{Y} \sim N_{n \times p}(\mathbf{X}\Theta, \Sigma_b \otimes \mathbf{I}_n). \quad (24)$$

Let the density in the full model be expressed as $f(\mathbf{Y}; \Theta, \Sigma_b)$. Then, we have

$$\begin{aligned}
-2 \log f(\mathbf{Y}; \Theta, \Sigma_b) &= np \log(2\pi) \\
&+ \sum_{\ell=1}^p \left\{ n \log \sigma_{b\ell}^2 + \frac{1}{\sigma_{b\ell}^2} (\mathbf{y}_\ell - \mathbf{X}\boldsymbol{\theta}_\ell)' (\mathbf{y}_\ell - \mathbf{X}\boldsymbol{\theta}_\ell) \right\}.
\end{aligned}$$

It holds that

$$\begin{aligned}
-2 \log \max_{\Theta, \Sigma_b} f(\mathbf{Y}; \Theta, \Sigma_b) &= np(\log 2\pi + 1) \\
&+ \sum_{\ell=1}^p n \log \frac{1}{n} \mathbf{y}_\ell' (\mathbf{I}_n - \mathbf{P}\omega) \mathbf{y}_\ell.
\end{aligned} \quad (25)$$

Next, consider the model removing the j th explanatory variable from the full model $M_{b,\omega}$, which is denoted by $M_{b,\omega \setminus j}$ or $M_{b,\omega \setminus j}$. Similarly,

$$\begin{aligned} -2 \log \max_{M_{b,\omega \setminus j}} f(\mathbf{Y}; \Theta, \Sigma_b) &= np(\log 2\pi + 1) \\ &+ \sum_{\ell=1}^p n \log \frac{1}{n} \mathbf{y}'_{\ell} (\mathbf{I}_n - \mathbf{P}_{\omega \setminus j}) \mathbf{y}_{\ell}. \end{aligned} \quad (26)$$

Using (25) and (26), we can obtain a general information criterion (8) for two models, $M_{b,\omega}$ and $M_{b,\omega \setminus j}$, and we have

$$\begin{aligned} T_{b,j;d} &\equiv \text{GIC}_{b,\omega \setminus j} - \text{GIC}_{b,\omega} \\ &= \sum_{\ell=1}^p n \log \left(1 + U_{2\ell} U_{1\ell}^{-1} \right) - dp, \end{aligned} \quad (27)$$

where

$$\begin{aligned} U_{1\ell} &= \mathbf{y}'_{\ell} (\mathbf{I}_n - \mathbf{P}_{\omega}) \mathbf{y}_{\ell}, \quad \ell = 1, \dots, p, \\ U_{2\ell} &= \mathbf{y}'_{\ell} (\mathbf{P}_{\omega} - \mathbf{P}_{\omega \setminus j}) \mathbf{y}_{\ell}, \quad \ell = 1, \dots, p. \end{aligned}$$

Let us assume that

$$\mathbf{E}(\mathbf{Y}) = \mathbf{X}_{j*} \Theta_{j*} \text{ and } \sigma_{b,*}^2 = \sigma_{b,j*}^2 \quad (28)$$

Then, as in (18), we have the following results:

$$\begin{aligned} (1) \quad &U_{1\ell} / \sigma_{b,*}^2 \sim \chi_{n-k}^2, \quad \ell = 1, \dots, p, \\ (2) \quad &U_{2\ell} / \sigma_{b,*}^2 \sim \chi_1^2(\delta_{b,j;\ell}^2), \quad \ell = 1, \dots, p, \\ (3) \quad &U_{1\ell}, U_{2\ell}, (\ell = 1, \dots, p) \text{ are independent,} \end{aligned} \quad (29)$$

where noncentral parameters $\delta_{b,j;\ell}^2$ are defined by

$$\delta_{b,j;\ell}^2 = \frac{1}{\sigma_{b,*}^2} (\mathbf{X}_{j*} \boldsymbol{\theta}_*^{(\ell)})' (\mathbf{P}_{\omega} - \mathbf{P}_{\omega \setminus j}) (\mathbf{X}_{j*} \boldsymbol{\theta}_*^{(\ell)}),$$

with $\Theta_* = (\boldsymbol{\theta}_*^{(1)}, \dots, \boldsymbol{\theta}_*^{(p)})$. If $j \notin j_*$, $\delta_{b,j;\ell}^2 = 0$, and if $j \in j_*$, $\delta_{b,j;\ell}^2 \neq 0$. For a sufficient condition for consistency of the KOO method based on $\text{GIC}_{b,j}$, we assumed

$$\begin{aligned} \text{A3b : For any } j \in j_*, \lim_{n \rightarrow \infty} (n-k)^{-1} \delta_{b,j;\ell}^2 &= \eta_{b,j;\ell}^2 > 0, \text{ and} \\ \lim_{n \rightarrow \infty} \frac{1}{p} \sum_{\ell=1}^p \log \left\{ 1 + \frac{1}{n-k} \delta_{b,j;\ell}^2 \right\} &\rightarrow \eta_{b,j}^2 > 0. \end{aligned} \quad (30)$$

Now, we consider the high-dimensional asymptotic consistency of the KOO method based on $T_{b,j;d}$ in (9), whose selection method is given by $\hat{j}_{v,j;d} = \{j \mid T_{b,j;d} > 0\}$. When $j \notin j_*$, we have

$$\begin{aligned} [\text{F2}] &= \sum_{j \notin j_*} \Pr \left(\sum_{\ell=1}^p n \log \left\{ 1 + U_{2\ell} U_{1\ell}^{-1} \right\} \geq d \right) \\ &\leq \sum_{j \notin j_*} \sum_{\ell=1}^p \Pr(n \log \left\{ 1 + U_{2\ell} U_{1\ell}^{-1} \right\} \geq d). \end{aligned}$$

This implies that

$$\begin{aligned} [\text{F2}] &\leq p(k - k_{j_*}) \Pr(n \log \{1 + \chi_1^2 / \chi_{n-k}^2\} \geq d) \\ &= p(k - k_{j_*}) \Pr(V \geq r), \end{aligned} \quad (31)$$

where

$$\begin{aligned} V &= \frac{\chi_1^2}{\chi_{n-k}^2} - \frac{1}{n-k-2}, \\ r &= e^{d/n} - 1 - \frac{1}{n-k-2}, \quad r_0 = \frac{d}{n} - \frac{1}{n-k-2}. \end{aligned} \quad (32)$$

Note that $r_0 < r$. Then, under the assumption $r_0 > 0$, we have

$$[\text{F2}] \leq p(k - k_{j_*}) r^{-2\ell} \mathbb{E}[V^{2\ell}] \leq p(k - k_{j_*}) r_0^{-2\ell} \mathbb{E}[V^{2\ell}]. \quad (33)$$

Related to the assumption $r_0 > 0$, we assumed

$$\text{A4b} : d > \frac{n}{n-k-2} \rightarrow \frac{1}{1-c_2}, \text{ and } d = O(n^a), \quad 0 < a < 1. \quad (34)$$

The first part in A4b implies $r_0 > 0$. It is easy to see that

$$\mathbb{E}[V^2] = \frac{2(n-k-1)}{(n-k-2)^2(n-k-4)} = O((n^2)^{-1}).$$

Further, $r_0^{-2} = O(n^{2(1-a)})$. Therefore, from (33), we have that $[\text{F2}] \rightarrow 0$.

When $j \in j_*$, we can write $T_{b,j;d} = n \sum_{\ell=1}^p \log \{1 + U_{2\ell} U_{1\ell}^{-1}\} - dp$. Therefore, we can express [F1] as follows:

$$[\text{F1}] = \sum_{j \in j_*} \Pr(\tilde{T}_{b,j;d} \leq 0),$$

where

$$\tilde{T}_{b,j;d} = \frac{1}{p} \sum_{\ell=1}^p \log \left\{ 1 + \frac{\chi_{1;\ell}^2 (\delta_{b,j;\ell}^2)}{\chi_{n-k;\ell}^2} \right\} - \frac{d}{n}.$$

Assumptions A3b and A4b easily show that

$$\tilde{T}_{b,j;d} \rightarrow \eta_{b,j}^2 > 0.$$

This implies that $\Pr(\tilde{T}_{b,j;d} \leq 0) \rightarrow 0$.

These imply the following theorem.

Theorem 2. Suppose that Assumptions A1, A2, A3b and A4b are satisfied. Then, the KOO method based on $T_{b,j;d}$ in (27) is asymptotically consistent.

Let us consider an alternative approach for "[F1] $\rightarrow 0$ " as in the case of independent covariance structure. When $j \in j_*$, we can write

$$\begin{aligned} [\text{F1}] &= \sum_{j \in j_*} \Pr \left(\sum_{\ell=1}^p \left\{ n \log \left(1 + \frac{\chi_{1;\ell}^2(\delta_{b,j;\ell}^2)}{\chi_{n-k;\ell}^2} \right) - d \right\} \leq 0 \right) \\ &\leq \sum_{j \in j_*} \sum_{\ell=1}^p \Pr \left(n \log \left(1 + \frac{\chi_{1;\ell}^2(\delta_{b,j;\ell}^2)}{\chi_{n-k;\ell}^2} \right) - d \leq 0 \right) \\ &= \sum_{j \in j_*} \sum_{\ell=1}^p \Pr(\tilde{V}_{j,\ell} \leq \tilde{r}_{j,\ell}). \end{aligned}$$

Here, for $j \in j_*$,

$$\begin{aligned} \tilde{V}_{j,\ell} &= \frac{\chi_{1;\ell}^2(\delta_{b,j;\ell}^2)}{\chi_{n-k;\ell}^2} - \frac{1 + \delta_{b,j;\ell}^2}{n-k-2}, \quad \ell = 1, \dots, p, \\ \tilde{r}_{j,\ell} &= e^{d/n} - 1 - \frac{1 + \delta_{b,j;\ell}^2}{n-k-2} = r - \frac{\delta_{b,j}^2}{n-k-2}, \quad \ell = 1, \dots, p, \end{aligned}$$

where r is the same one as in (32). Note that $\chi_{1;\ell}^2(\delta_{b,j;\ell}^2)$, $\ell = 1, \dots, p$ are distributed as a noncentral distribution $\chi_1^2(\delta_{b,j;\ell}^2)$, and they are independent. Then, under the assumption $\tilde{r}_j < 0$ (or equivalently $r < \delta_{b,j}^2 / (n-k-2)$), we have

$$[\text{F1}] \leq k_{j_*} \sum_{\ell=1}^p |\tilde{r}_{j,\ell}|^{-2s} \mathbb{E}[\tilde{V}_{j,\ell}^{2s}], \quad s = 1, 2, \dots \quad (35)$$

In the above upper bound, it holds that

$$|\tilde{r}_{j,\ell}| \sim \delta_{b,j;\ell}^2 / (n-k) \rightarrow \eta_{b,j;\ell}^2. \quad (36)$$

Useful bounds are obtained by giving the first few moments of $\tilde{V}_{j,\ell}$. For example,

$$\begin{aligned} \mathbb{E}[\tilde{V}_{j,\ell}^2] &= \frac{2(1 + 2\delta_{v,j;\ell}^2)(n-k-1 + \delta_{v,j;\ell}^2)}{(n-k-2)^2(n-k-4)} = O(n^{-1}), \\ \mathbb{E}[\tilde{V}_{j,\ell}^4] &= O(n^{-2}). \end{aligned}$$

Then, Bound (35) with $s = 2$ can be asymptotically expressed as follows:

$$k_{j_*} \sum_{\ell=1}^p \eta_{b,j;\ell}^{-4} \mathbb{E}[\tilde{V}_{j,\ell}^4] = k_{j_*} p \left(\frac{1}{p} \sum_{\ell=1}^p \eta_{b,j;\ell}^{-4} \right) \times O(n^{-2}).$$

The above expression is $O(n^{-1})$ under the assumption that $\frac{1}{p} \sum_{\ell=1}^p \eta_{b,j;\ell}^{-4}$ tends to a quantity.

6. Asymptotic Consistency under a Uniform Covariance Structure

In this section, we show an asymptotic consistency of KOO method based on a general information criterion under a uniform covariance structure. First, following [12], we derive a $\text{GIC}_{u,j}$ as in (6), and a key statistic $T_{u,j;d}$ as in (9). A uniform covariance structure is given by

$$\Sigma_u = \sigma_u^2(\rho_u^{1-\delta_{ij}}) = \sigma_u^2\{(1-\rho_u)\mathbf{I}_p + \rho_u \mathbf{1}_p \mathbf{1}_p'\}, \quad (37)$$

with Kronecker delta δ_{ij} . The covariance structure is expressed as follows:

$$\Sigma_u = \alpha \left(\mathbf{I}_p - \frac{1}{p} \mathbf{G}_p \right) + \beta \frac{1}{p} \mathbf{G}_p,$$

where

$$\alpha = \sigma_u^2(1 - \rho_u), \quad \beta = \sigma_u^2\{1 + (p-1)\rho_u\}, \quad \mathbf{G}_p = \mathbf{1}_p \mathbf{1}_p',$$

and $\mathbf{1}_p = (1, \dots, 1)'$. Matrices $\mathbf{I}_p - \frac{1}{p} \mathbf{G}_p$ and $\frac{1}{p} \mathbf{G}_p$ are orthogonal idempotent matrices, so we have

$$|\Sigma_u| = \beta \alpha^{p-1}, \quad \Sigma_u^{-1} = \frac{1}{\alpha} \left(\mathbf{I}_p - \frac{1}{p} \mathbf{G}_p \right) + \frac{1}{\beta} \cdot \frac{1}{p} \mathbf{G}_p.$$

Now, we consider the multivariate regression model $M_{u,j}$ given by

$$M_{u,j} : \mathbf{Y} \sim N_{n \times p}(\mathbf{X}_j \Theta_j, \Sigma_{u,j} \otimes \mathbf{I}_n), \quad (38)$$

where $\Sigma_{u,j} = \alpha_j (\mathbf{I}_p - p^{-1} \mathbf{G}_p) + \beta_j p^{-1} \mathbf{G}_p$. Let $\mathbf{H} = (\mathbf{h}_1, \mathbf{H}_2)$ be an orthogonal matrix where $\mathbf{h}_1 = p^{-1/2} \mathbf{1}_p$, and let

$$\mathbf{W}_j = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_j)\mathbf{Y} \text{ and } \mathbf{U}_j = \mathbf{H}'\mathbf{W}_j\mathbf{H}.$$

Here, \mathbf{h}_1 is a characteristic vector of $\Sigma_{u,j}$, and each column vector of \mathbf{H}_2 is a characteristic vector of $\Sigma_{u,j}$. Let the density function of \mathbf{Y} under $M_{u,j}$ be denoted by $f(\mathbf{Y}; \Theta_j, \alpha_j, \beta_j)$. Then, we have

$$\begin{aligned} g(\alpha_j, \beta_j) &= -2 \log \max_{\Theta_j} f(\mathbf{Y}; \Theta_j, \alpha_j, \beta_j) \\ &= np \log(2\pi) + n(p-1) \log \alpha_j + n \log \beta_j + \text{tr} \Psi_j^{-1} \mathbf{U}_j, \end{aligned}$$

where $\Psi_j = \text{diag}(\beta_j, \alpha_j, \dots, \alpha_j)$. Therefore, the maximum likelihood estimators of α_j and β_j under $M_{u,j}$ are given by

$$\begin{aligned} \hat{\alpha}_j &= \frac{1}{n(p-1)} \text{tr} \mathbf{H}_2' \mathbf{Y}' (\mathbf{I}_n - \mathbf{P}_j) \mathbf{Y} \mathbf{H}_2, \\ \hat{\beta}_j &= \frac{1}{n} \mathbf{h}_1' \mathbf{Y}' (\mathbf{I}_n - \mathbf{P}_j) \mathbf{Y} \mathbf{h}_1. \end{aligned}$$

The number of independent parameters under $M_{u,j}$ is $m_j = k_j p + 2$. Noting that Ψ_j is diagonal, we can obtain the general information criterion (GIC) in (8) for \mathbf{Y} in (38) as follows:

$$\text{GIC}_{u,j} = n(p-1) \log \hat{\alpha}_j + n \log \hat{\beta}_j + np(\log 2\pi + 1) + d(k_j p + 2). \quad (39)$$

Therefore, we have

$$\begin{aligned} T_{u,j;d} &\equiv \text{GIC}_{u,\omega \setminus j} - \text{GIC}_{u,\omega} \\ &= n(p-1) \log \left\{ \hat{\alpha}_{\omega \setminus j} (\hat{\alpha}_{\omega})^{-1} \right\} + n \log \left\{ \hat{\beta}_{\omega \setminus j} (\hat{\beta}_{\omega})^{-1} \right\} - dp \\ &= Z_{1j} + Z_{2j}. \end{aligned} \quad (40)$$

Here, Z_{1j} and Z_{2j} are defined as follows:

$$\begin{aligned} Z_{1j} &= n(p-1) \log \left\{ 1 + V_{2j}^{(1)} \left(V_1^{(1)} \right)^{-1} \right\} - d(p-1), \\ Z_{2j} &= n \log \left\{ 1 + V_{2j}^{(2)} \left(V_1^{(2)} \right)^{-1} \right\} - d, \end{aligned} \quad (41)$$

using the following $V_1^{(i)}, V_{2j}^{(i)}, i = 1, 2$:

$$\begin{aligned} V_1^{(1)} &= \text{tr} \mathbf{H}_2' \mathbf{Y} (\mathbf{I}_n - \mathbf{P}_\omega) \mathbf{Y} \mathbf{H}_2, & V_{2j}^{(1)} &= \text{tr} \mathbf{H}_2' \mathbf{Y}' (\mathbf{P}_\omega - \mathbf{P}_{\omega \setminus j}) \mathbf{Y} \mathbf{H}_2, \\ V_1^{(2)} &= \mathbf{h}_1' \mathbf{Y}' (\mathbf{I}_n - \mathbf{P}_\omega) \mathbf{Y} \mathbf{h}_1, & V_{2j}^{(2)} &= \mathbf{h}_1' \mathbf{Y}' (\mathbf{P}_\omega - \mathbf{P}_{\omega \setminus j}) \mathbf{Y} \mathbf{h}_1. \end{aligned}$$

Related to the distributional reductions of $Z_{1j}, Z_{2j}, j = 1, \dots, k$, we use the following Lemma frequently.

Lemma 1. Let \mathbf{W} have a noncentral Whishart distribution $W_p(m, \Sigma; \Omega)$. Let the covariance matrix Σ be decomposed into characteristic roots and vectors as follows:

$$\begin{aligned} \Sigma &= \mathbf{H} \Lambda \mathbf{H}' \\ &= (\mathbf{H}_1, \dots, \mathbf{H}_h) \text{diag}(\lambda_1 \mathbf{I}_{q_1}, \dots, \lambda_h \mathbf{I}_{q_h}) (\mathbf{H}_1, \dots, \mathbf{H}_h)', \end{aligned}$$

where $\lambda_1 > \dots > \lambda_h > 0$ and \mathbf{H} is an orthogonal matrix. Then, $\text{tr} \mathbf{H}_j' \mathbf{W}_j \mathbf{H}_j, i = 1, \dots, h$ are independently distributed to noncentral chi-squared distributions with mk_j degrees of freedom and noncentrality parameters $\delta_j^2 = \text{tr} \mathbf{H}_j' \Omega \mathbf{H}_j$.

Proof. The result may be proven by considering the characteristic function of $(\text{tr} \mathbf{H}_1' \mathbf{W} \mathbf{H}_1, \dots, \text{tr} \mathbf{H}_q' \mathbf{W} \mathbf{H}_q)$ which is expressed as follows (see Theorem 2.1.2 in [15]):

$$\begin{aligned} & \mathbb{E} \left[e^{it_1 \text{tr} \mathbf{H}_1' \mathbf{W} \mathbf{H}_1 + \dots + it_h \text{tr} \mathbf{H}_h' \mathbf{W} \mathbf{H}_h} \right] \\ &= \mathbb{E} [\text{etr}(\mathbf{K})] \\ &= |\mathbf{I}_p - 2\Sigma \mathbf{K}|^{-m/2} \text{etr} \left\{ \Omega \mathbf{K} (\mathbf{I}_p - 2\Sigma \mathbf{K})^{-1} \right\}, \end{aligned}$$

where $\mathbf{K} = it_1 \mathbf{H}_1 \mathbf{H}_1' + \dots + it_h \mathbf{H}_h \mathbf{H}_h'$. The result can be easily obtained by checking that the above last expression equals

$$\prod_{j=1}^q (1 - 2it_j)^{-nk_j/2} \exp \left\{ \frac{it_j}{1 - 2it_j} \text{tr} \mathbf{H}_j' \Omega \mathbf{H}_j \right\}.$$

□

Assume that the true model is expressed as

$$M_{u,j*} : \mathbf{Y} \sim N_{n \times p}(\mathbf{X}_{j*} \Theta_{j*}, \Sigma_{u,*} \otimes \mathbf{I}_n), \quad (42)$$

where $\Sigma_{u,*} = \alpha_* (\mathbf{I}_p - p^{-1} \mathbf{G}_p) + \beta_* p^{-1} \mathbf{G}_p$. Using Lemma 1, we have the following lemma.

Lemma 2. Under True Model (42), it holds that

- (1) $V_1^{(1)} / \alpha_*$ and $V_{2j}^{(1)} / \alpha_*$ are independently distributed to a central chi-squared distribution $\chi_{(p-1)(n-k)}^2$ and a noncentral chi-squared distribution $\chi_{p-1}^2(\delta_{1j}^2)$, respectively.
- (2) $V_1^{(2)} / \beta_*$ and $V_{2j}^{(2)} / \beta_*$ are independently distributed to a central chi-squared distribution χ_{n-k}^2 and a noncentral chi-squared distribution $\chi_1^2(\delta_{2j}^2)$, respectively.
- (3) Noncentrality parameters δ_{1j}^2 and δ_{2j}^2 are defined as follows:

$$\begin{aligned} \delta_{1j}^2 &= \frac{1}{\alpha_*} \text{tr} \mathbf{H}_2' (\mathbf{X}_{j*} \Theta_{j*})' (\mathbf{P}_\omega - \mathbf{P}_{\omega \setminus j}) (\mathbf{X}_{j*} \Theta_{j*}) \mathbf{H}_2 \\ \delta_{2j}^2 &= \frac{1}{\beta_*} \mathbf{h}_1' (\mathbf{X}_{j*} \Theta_{j*})' (\mathbf{P}_\omega - \mathbf{P}_{\omega \setminus j}) (\mathbf{X}_{j*} \Theta_{j*}) \mathbf{h}_1. \end{aligned}$$

Here, if $j \notin j_*$, then $\delta_{1j}^2 = 0$ and $\delta_{2j}^2 = 0$.

Now, we consider the high-dimensional asymptotic consistency of the KOO method based on $T_{b,j;d}$ in (27), whose selection method is given by $\hat{j}_{v,j;d} = \{j \mid T_{b,j;d} > 0\}$. For a sufficient condition for the consistency of $\hat{j}_{v,j;d}$, we assumed

A3u: For any $j \in j_*$, $\delta_{1j}^2 = O(np)$, $\delta_{2j}^2 = O(n)$ and

$$\lim_{n,p} \frac{1}{np} \delta_{1j}^2 = \eta_{1j}^2 > 0, \quad \lim_{n,p} \frac{1}{n} \delta_{2j}^2 = \eta_{2j}^2 > 0, \quad (43)$$

When $j \notin j_*$, we have

$$\begin{aligned} [F2] &= \sum_{j \notin j_*} \{\Pr(Z_{1j} + Z_{2j} \geq 0)\} \\ &\leq \sum_{j \notin j_*} \{\Pr(Z_{1j} \geq 0) + \Pr(Z_{2j} \geq 0)\} \\ &= (k - k_{j_*}) \left\{ \Pr(Z^{(1)} \geq s_0^{(1)}) + \Pr(Z^{(2)} \geq s_0^{(2)}) \right\}. \end{aligned}$$

Here,

$$\begin{aligned} Z^{(1)} &= \frac{\chi_{p-1}^2}{\chi_{(p-1)(n-k)}^2} - \frac{p-1}{(p-1)(n-k)-2}, \\ s^{(1)} &= e^{d/n} - 1 - \frac{p-1}{(p-1)(n-k)-2}, \quad s_0^{(1)} = \frac{d}{n} - \frac{p-1}{(p-1)(n-k)-2}, \\ Z^{(2)} &= \frac{\chi_1^2}{\chi_{n-k}^2} - \frac{1}{n-k-2}, \\ s^{(2)} &= e^{d/n} - 1 - \frac{1}{n-k-2}, \quad s_0^{(2)} = \frac{d}{n} - \frac{1}{n-k-2}. \end{aligned}$$

Note that $s_0^{(1)} < s^{(1)}$ and $s_0^{(2)} < s^{(2)}$. Then, under the assumption that $s_0^{(1)} > 0$ and $s_0^{(2)} > 0$, we have

$$[F2] \leq (k - k_{j_*}) \left[\left(s_0^{(1)} \right)^{-2\ell} \mathbb{E} \left[(Z^{(1)})^{2\ell} \right] + \left(s_0^{(2)} \right)^{-2\ell} \mathbb{E} \left[(Z^{(2)})^{2\ell} \right] \right]. \quad (44)$$

Related to assumptions $s_0^{(1)} > 0$ and $s_0^{(2)} > 0$, we assumed

$$\begin{aligned} \text{A4u} : d &> \frac{n(p-1)}{(p-1)(n-k)-2} \rightarrow \frac{1}{1-c_2}, \quad d > \frac{n}{n-k-2} \rightarrow \frac{1}{1-c_2}, \\ \text{and } d &= O(n^a), \quad 0 < a < 1. \end{aligned} \quad (45)$$

The first part in A4u implies $s_0^{(1)} > 0$ and $s_0^{(2)} > 0$. It is easy to see that

$$\begin{aligned} \mathbb{E}[(Z^{(1)})^2] &= \frac{2(p-1)^2(n-k+1)}{\{(p-1)(n-k)-2\}^2 \{(p-1)(n-k)-4\}} = O((n^3)^{-1}), \\ \mathbb{E}[(Z^{(2)})^2] &= \frac{2(n-k-1)}{(n-k-2)^2(n-k-4)} = O((n^2)^{-1}). \end{aligned}$$

Further, $(s_0^{(1)})^{-2} = O(n^{2(1-a)})$ and $(s_0^{(2)})^{-2} = O(n^{2(1-a)})$. Therefore, from (44), we have that $[F2] \rightarrow 0$.

When $j \in j_*$, we can write $T_{b,j;d} = n \sum_{\ell=1}^p \log\{1 + U_{2\ell} U_{1\ell}^{-1}\} - dp$. Therefore, we can express $[F1]$ as follows:

$$[F1] = \sum_{j \in j_*} \Pr(\tilde{T}_{b,j;d} \leq 0),$$

where

$$\tilde{T}_{b,j;d} = \frac{1}{p} \sum_{\ell=1}^p \log \left\{ 1 + \frac{\chi_{1,\ell}^2 (\delta_{b,j;\ell}^2)}{\chi_{n-k;\ell}^2} \right\} - \frac{d}{n}.$$

Assumptions A3b and A4b easily show that

$$\tilde{T}_{v,j;d} \rightarrow \log(1 + \gamma_{v,j}^2) > 0.$$

This implies that $\Pr(\tilde{T}_{v,j;d} \leq 0) \rightarrow 0$, and $[F1] \rightarrow 0$.

These imply the following theorem.

Theorem 3. Suppose that Assumptions A1, A2, A3u and A4u are satisfied. Then, the KOO method based on $T_{u,j;d}$ in (40) is asymptotically consistent.

7. Concluding Remarks

In this paper, we considered selecting regression variables in a p variate regression model with one of three covariance structures: (1) ICSS (an independent covariance structure with the same variance), (2) ICSD (an independent covariance structure with different variances), and (3) UCS (a uniform covariance structure). It was proposed to use a KOO method on the basis of a general information criterion with a penalty term d . We indicated high-dimensional consistencies of the KOO methods with $d = O(n^a)$, $0 < a < 1$. Ref. [12] studied the asymptotic consistencies of KOO methods in (1) and (3). However, in their approach, the number of explanatory variables was fixed; in this paper, the number of explanatory variables may have tended to infinity. KOO methods may be feasible in computation. The idea goes back to [1], and [2]. However, high-dimensional properties were recently studied in [7–9,11].

A high-dimensional study of the KOO method under an autoregressive covariance structure (AUTO), and extending our results to the case of non-normality remain as future work.

Author Contributions: Conceptualization, Y.F.; Methodology, Y.F. and T.S.; Software, T.S.; Writing—original draft, Y.F. and T.S.; Writing—review & editing, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to express their gratitude to Vladimir V. Ulyanov and the three referees for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nishii, R.; Bai, Z.D.; Krishnaia, P.R. Strong consistency of the information criterion for model selection in multivariate analysis. *Hiroshima Math. J.* **1988**, *18*, 451–462. [\[CrossRef\]](#)
2. Zhao, L.C.; Krishnaia, P.R.; Bai, Z.D. On detection of the number of signals in presence of white noise. *J. Multivar. Anal.* **1986**, *20*, 1–25. [\[CrossRef\]](#)
3. Bai, Z.; Fujikoshi, Y.; Hu, J. *Strong Consistency of the AIC, BIC, Cp and KOO Methods in High-Dimensional Multivariate Linear Regression*; Hiroshima Statistical Research Group: Hiroshima, Japan, 2018; TR; 18–09.
4. 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. [\[CrossRef\]](#)
5. 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. [\[CrossRef\]](#)
6. Fujikoshi, Y.; Sakurai, T. High-dimensional consistency of rank estimation criteria in multivariate linear Model. *J. Multivar. Anal.* **2016**, *149*, 199–212. [\[CrossRef\]](#)

7. 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. [[CrossRef](#)]
8. Oda, R.; Yanagihara, H. A consistent likelihood-based variable selection method in normal multivariate linear regression. In *Intelligent Decision Technologies*; Czarnowski, I., Ed.; Springer: Singapore, 2021; Volume 238, pp. 391–401.
9. Fujikoshi, Y.; Sakurai, T. Consistency of test-based method for selection of variables in high-dimensional two group-discriminant analysis. *Jpn. J. Stat. Data Sci.* **2019**, *2*, 155–171. [[CrossRef](#)]
10. Oda, R.; Suzuki, Y.; Yanagihara, H.; Fujikoshi, Y. A consistent variable selection method in high-dimensional canonical discriminant analysis. *J. Multivar. Anal.* **2020**, *175*, 1–13. [[CrossRef](#)]
11. Fujikoshi, Y. High-dimensional consistencies of KOO methods in multivariate regression model and discriminant analysis. *J. Multivar. Anal.* **2022**, *188*, 104860. [[CrossRef](#)]
12. Sakurai, T.; Fujikoshi, Y. Exploring consistencies of information criterion and test-based criterion for high-dimensional multivariate regression models under three covariance structures. In *Festschrift in Honor of Professor Dietrich von Rosen's 65th Birthday*; Holgerson, T., Singnull, M., Eds.; Springer: Berlin, Germany, 2020; pp. 313–334.
13. Akaike, H. Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*; Petrov, B.N., Csáki, F., Eds.; Akadémiai Kiadó: Budapest, Hungary, 1973; pp. 267–281.
14. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
15. Fujikoshi, Y.; Ulyanov, V.V.; Shimizu, R. *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*; Wiley: Hoboken, NJ, USA, 2010.

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.