KOO Approach for Scalable Variable Selection Problem in Large-dimensional Regression

Zhidong Bai
Northeast Normal University, China
baizd@nenu.edu.cn

Kwok Pui Choi
National University of Singapore,
Singapore
stackp@nus.edu.sg

Yasunori Fujikoshi
Hiroshima University, Japan
fujikoshi_y@yahoo.co.jp

Jiang Hu
Northeast Normal University, China
huj156@nenu.edu.cn

An important issue in many multivariate regression problems is to eliminate candidate predictors with null predictor vectors. In large-dimensional (LD) setting where the numbers of responses and predictors are large, model selection encounters the scalability challenge. Knock-one-out (KOO) statistics hold promise to meet this challenge. In this paper, the almost sure limits and the central limit theorem of the KOO statistics are derived under the LD setting and mild distributional assumptions (finite fourth moments) of the errors. These theoretical results guarantee the strong consistency of a subset selection rule based on the KOO statistics with a general threshold. For enhancing the robustness of the selection rule, we also propose a bootstrap threshold for the KOO approach. Simulation results support our conclusions and demonstrate the selection probabilities by the KOO approach with the bootstrap threshold outperform the methods using Akaike information threshold, Bayesian information threshold and Mallow's C_p threshold. We compare the proposed KOO approach with those based on information threshold to a chemometrics dataset and a yeast cell-cycle dataset, which suggests our proposed method identifies useful models.

1. INTRODUCTION

In multivariate statistical analysis, linear regression is a basic and commonly used type of approach. The overall idea of regression is to examine which variables in particular are significant predictors of the outcome variables, and in what way do they indicated by the magnitude and sign of the outcome variables. Specifically,

$$\mathbf{Y} = \mathbf{X}\mathbf{\Theta} + \mathbf{E}\mathbf{\Sigma}^{1/2}, \tag{1}$$

where the $n \times p$ response matrix $\mathbf{Y} = (y_{ij}) = (\mathbf{y}_1, \ldots, \mathbf{y}_n)'$, the $n \times k$ predictor matrix $\mathbf{X} = (\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_n)' = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$, the $k \times p$ regression coefficient matrix $\mathbf{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k)'$, the $n \times p$ random errors matrix $\mathbf{E} = (\mathbf{e}_1, \ldots, \mathbf{e}_p) = (e_{ij})$ and the $p \times p$ covariance matrix $\mathbf{\Sigma}$ with full rank. A main goal in multivariate linear regression (MLR) is to estimate the regression coefficients Θ. The estimates should be such that the estimated regression plane explains the variation in the values of the responses with great accuracy.

Model (1) (referred to hereinafter as the full model), however, is not always satisfactory because some of the predictors may be uncorrelated with the responses. We take a simple example to illustrate this fact. Let $\mathbf{j}$ be a subset of $[k] = \{1, 2, \ldots, k\}$, $\mathbf{X_j} = (\mathbf{x}_j, j \in \mathbf{j})$ and $\mathbf{\Theta_j} = (\boldsymbol{\theta}_j, j \in \mathbf{j})'$. Denote model $\mathbf{j}$ by

$$M_{\mathbf{j}}: \quad \mathbf{Y} = \mathbf{X_j}\mathbf{\Theta_j} + \mathbf{E}\mathbf{\Sigma}^{1/2}. \tag{2}$$

The classical linear least-squares solution is to estimate the matrix of regression coefficients $\widehat{\boldsymbol{\Theta}}$ of the full model (1) by

$$\widehat{\boldsymbol{\Theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y},$$

which minimizes the sum of the squares of errors, i.e.,

$$\widehat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} \operatorname{tr}(\mathbf{Y} - \mathbf{X}\boldsymbol{\Theta})(\mathbf{Y} - \mathbf{X}\boldsymbol{\Theta})'.$$

If there exists a predictor vector $\boldsymbol{\theta}_j = \mathbf{0}$, then the least-squares estimator of the regression coefficients of model $M_{[k]\setminus j}$ is

$$\widehat{\boldsymbol{\Theta}}_{[k]\setminus j} = (\mathbf{X}'_{[k]\setminus j}\mathbf{X}_{[k]\setminus j})^{-1}\mathbf{X}'_{[k]\setminus j}\mathbf{Y},$$

It is known that in this case the mean squared error (MSE) of the predictions from $\widehat{\boldsymbol{\Theta}}_{[k]\setminus j}$ is smaller than that from $\widehat{\boldsymbol{\Theta}}$ under some mild conditions. Moreover, even though the elements of $\boldsymbol{\theta}_j$ are not equal to zero but small enough, the MSE of the predictions from $\widehat{\boldsymbol{\Theta}}_{[k]\setminus j}$ is also smaller than that from $\widehat{\boldsymbol{\Theta}}$ (e.g., Fujikoshi et al. (2010)). Therefore, removing these "non-significant" predictors from the full model improves the model. How to determine the significance of each predictor for the response and to select the true model from the full model are important problems in multiple regression model. Here, the true model is the data-generating model and is denoted by

(3) $$M_{\mathbf{j}_*} : \quad \mathbf{Y} = \mathbf{X}_{\mathbf{j}_*}\boldsymbol{\Theta}_{\mathbf{j}_*} + \mathbf{E}\boldsymbol{\Sigma}^{1/2},$$

where for all $j \in [k]\setminus\mathbf{j}_*$, $\boldsymbol{\theta}_j = \mathbf{0}$.

To measure the significance of the predictors for the response, one can make use of the regression coefficients, the partial correlation or the multiple correlation coefficient between each predictor and the responses. However, these direct measures are unstable under high-dimensional regression because they all highly depend on the values of each predictor. Instead, we consider removing one predictor vector from the full model and measuring how much "information" we lose. Hence, we refer to this kind of statistics KOO (knock-one-out or kick-one-out) statistics in the technical report (Bai et al., 2018a). This KOO idea can be traced back to Nishii et al. (1988), who investigated the discriminant analysis and canonical correlation analysis under fixed dimensions. In this paper, we study the KOO statistics in high-dimensional responses and predictors.The KOO method was motivated to address the issue of computational complexity in traditional AIC and BIC methods. Moreover, we find that the KOO method exhibits excellent stability, particularly in high-dimensional response settings.

There has been a lot of recent interest in variable selection problems for high-dimensional linear regression models because of the increasingly frequent and important in diverse fields of economics, finance and machine learning. For univariate (or single) response case (i.e., $p = 1$), a variety of methods have been developed. This includes the penalty-based methods such as the least angle and shrinkage selection operator (LASSO, Tibshirani (1996)), the adaptive LASSO (Zou, 2006), the smoothly clipped absolute deviation (SCAD Fan & Li (2001)), the minimax convex penalty (MCP, Zhang (2010)); the screening-based methods such as the sure independence screening (SIS, Fan & Lv (2008)), the covariate assisted screening estimates (CASE, Ke et al. (2014)); the testing based methods such as the multiple testing approach by the false discovery rate (FDR) (Liu & Luo, 2014; Xia et al., 2018) and many other related methods. We refer to some recent review papers (Shao, 1997; Fan & Lv, 2010; Huang et al., 2012; Anzanello & Fogliatto, 2014; Heinze et al., 2018; Desboulets, 2018; Lee et al., 2019; Cai et al., 2023) for more details. However, there is comparatively less literature available for multiple responses (i.e. $p > 1$). Xia (2017) proposed a row-wise multiple testing procedure when p is fixed; Kong et al. (2017) suggested a screening method via the distance correlations of the responses

and each covariate for high-dimensional multi-response interaction models. For $p \to \infty$, following Bai et al. (2014), Bai et al. (2022) investigated the asymptotic properties of the classical AIC, BIC and C_p criteria; and Sakurai & Fujikoshi (2020); Oda & Yanagihara (2020) established the consistencies of the KOO methods with AIC, BIC and C_p thresholds under normality errors.

Main contributions of this paper are: (1) We obtain the asymptotic distributions of the KOO statistics $\mathcal{K}_j$ for any $j = 1, \ldots, k$ under some mild moment conditions and 3L asymptotic framework: large-response ($p \to \infty$), large-model ($k \to \infty$) and large-sample ($n \to \infty$). These theoretical results are applicable to many other model selection rules, such as growth curve model, multiple discriminant analysis, principal component analysis, canonical correlation analysis, and graphical model (e.g., Fujikoshi & Sakurai (2019); Oda et al. (2020); Fujikoshi et al. (2023)). (2) A scalable model selection method based on the KOO statistics is proposed. In practice, we use a multiplier bootstrap procedure to estimate the asymptotic thresholds. Simulation studies and real data analyses suggest the proposed model selection method performs favorably against the existing KOO methods with AIC, BIC and C_p thresholds.

The remainder of this paper is organized as follows. In Section 2, we state the main results of this paper, which include the almost sure limit and central limit theorem (CLT) of the KOO statistics. In Section 3, we propose a model selection method for the high-dimensional linear regression model based on the KOO statistics and information criteria. In Sections 4 and 5, we conduct some simulation studies and real data analysis, respectively. Proofs of the main theorems under normality are given in Section 6 since they are less technical and of independent interests. Proofs for general error distributions using random matrix theory are provided in the Appendix for interested readers.

2. KOO STATISTICS

2.1. **Notation and preliminary.** We begin this section with some basic notation and definitions. In this paper, matrices and vectors are denoted by boldface uppercase and lowercase letters, respectively. Let $\mathbf{I}_n$ denote the identity matrix of order n,

$$(4) \qquad \widehat{\boldsymbol{\Sigma}}_{\mathbf{j}} = \frac{1}{n}\mathbf{Y}'\mathbf{Q}_{\mathbf{j}}\mathbf{Y}, \ \mathbf{Q}_{\mathbf{j}} = \mathbf{I}_n - \mathbf{P}_{\mathbf{j}}, \ \mathbf{P}_{\mathbf{j}} = \mathbf{X}_{\mathbf{j}}(\mathbf{X}_{\mathbf{j}}'\mathbf{X}_{\mathbf{j}})^{-1}\mathbf{X}_{\mathbf{j}}',$$

$|\mathbf{j}|$ the cardinality of subset $\mathbf{j}$, and $|\widehat{\boldsymbol{\Sigma}}_{\mathbf{j}}|$ the determinant $\widehat{\boldsymbol{\Sigma}}_{\mathbf{j}}$. Note that $\mathbf{P}_{\mathbf{j}}$ is an orthogonal projection of rank $|\mathbf{j}|$ onto the subspace spanned by $\mathbf{X}_{\mathbf{j}}$, and $\mathbf{Q}_{\mathbf{j}}$ is the orthogonal projection of rank $n - |\mathbf{j}|$ onto the orthogonal complement subspace spanned by $\mathbf{X}_{\mathbf{j}}$. For brevity, we suppress the subscript $[k]$ for full model, and denote the true model subscript by $*$ and the subscript of model $[k]\backslash j$ by j (e.g., $\mathbf{Q} := \mathbf{Q}_{[k]}$, $\mathbf{Q}_{\mathbf{j}_*} := \mathbf{Q}_*$ and $\mathbf{Q}_j := \mathbf{Q}_{[k]\backslash j}$). The identity matrix, all-zero matrix, all-one vector and all-zero vector, whose orders are often clear from the context and thus will not be indicated, are denoted by $\mathbf{I}$, $\mathbf{O}$, $\mathbf{1}$, and $\mathbf{0}$, respectively. We call j (or variable $\mathbf{x}_j$) true if $j \in \mathbf{j}_*$, and j (or variable $\mathbf{x}_j$) is spurious if $j \notin \mathbf{j}_*$. For a matrix $\mathbf{A}$, its spectral norm and maximum norm are denoted by $\|\mathbf{A}\|$ and $\|\mathbf{A}\|_\infty$, respectively. The largest and smallest eigenvalues of $\mathbf{A}$ are denoted by $\lambda_{\max}^{\mathbf{A}}$ and $\lambda_{\min}^{\mathbf{A}}$, respectively. For two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimension, $\mathbf{A} \circ \mathbf{B}$ stands for the Hadamard product of $\mathbf{A}$ and $\mathbf{B}$. We denote the probability by $\mathbb{P}$, the expectation by $\mathbb{E}$, and the trace by tr. Define $c_n := p/n$ and $\alpha_n := k/n$. Throughout this paper, we use $o(1)$ (respectively, $o_p(1)$, $o_{a.s.}(1)$) to denote (respectively, in probability, almost surely) scalar negligible entries. And the notations $O(1)$, $O_p(1)$ and $O_{a.s.}(1)$ are used in a similar way.

We now introduce the KOO statistics

$$\mathcal{K}_j = \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Sigma}}_j) - p.$$

It is known that for testing $\boldsymbol{\theta}_j = \mathbf{0}$ under normality, the Lawley-Hotelling trace statistic can be expressed as $(n - k)(\mathcal{K}_j + p)$. Next we will investigate the statistical properties of $\mathcal{K}_j$ under the 3L asymptotic framework: large-model (k), large-sample (n) and large-dimensional response (p). Before presenting our main theoretical results, we briefly analyze the statistic $\mathcal{K}_j$. Let

$$\mathbf{a}_j = \mathbf{Q}_j \mathbf{x}_j / \|\mathbf{Q}_j \mathbf{x}_j\|.$$

By Sylvester's determinant theorem, we have that

(5)
$$n\widehat{\boldsymbol{\Sigma}}_j = n\widehat{\boldsymbol{\Sigma}} + \mathbf{Y}'\mathbf{a}_j\mathbf{a}_j'\mathbf{Y}$$

which implies

(6)
$$\mathcal{K}_j = n^{-1}\mathbf{a}_j'\mathbf{Y}\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{Y}'\mathbf{a}_j.$$

If we plug the model (1) into the jth KOO statistic, we have

$$\mathcal{K}_j = (\mathbf{a}_j'\mathbf{X}_*\boldsymbol{\Theta}_*\boldsymbol{\Sigma}^{-1/2} + \mathbf{a}_j'\mathbf{E})(\mathbf{E}'\mathbf{Q}\mathbf{E})^{-1}(\mathbf{E}'\mathbf{a}_j + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Theta}_*'\mathbf{X}_*'\mathbf{a}_j).$$

When j is spurious (i.e., $j \notin \mathbf{j}_*$), $\mathbf{a}_j$ and $\mathbf{X}_*$ are orthogonal. Thus, in this case,

$$\mathcal{K}_j = \mathbf{a}_j'\mathbf{E}(\mathbf{E}'\mathbf{Q}\mathbf{E})^{-1}\mathbf{E}'\mathbf{a}_j.$$

On the other hand, when j is true (i.e., $j \in \mathbf{j}_*$), then

$$\mathcal{K}_j \asymp \mathbf{a}_j'\mathbf{E}(\mathbf{E}'\mathbf{Q}\mathbf{E})^{-1}\mathbf{E}'\mathbf{a}_j + \mathbf{x}_j'\mathbf{Q}_j\mathbf{x}_j\boldsymbol{\theta}_j'\boldsymbol{\Sigma}^{-1/2}(\mathbf{E}'\mathbf{Q}\mathbf{E})^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}_j.$$

We emphasize that, for spurious j, the KOO statistics $\mathcal{K}_j$ are independent of the population covariance matrix $\boldsymbol{\Sigma}$. This property is highly desirable as it eliminates the involvement of unknown parameters. Furthermore, the term $\mathbf{x}_j'\mathbf{Q}_j\mathbf{x}_j\boldsymbol{\theta}_j'\boldsymbol{\Sigma}^{-1/2}(\mathbf{E}'\mathbf{Q}\mathbf{E})^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\theta}_j > 0$ becomes a key indicator to distinguish between spurious and true variables, with its value serving as a crucial factor in the determination process. The detailed discussion is stated in the next subsection.

2.2. **Asymptotical properties of the KOO statistics.** In this subsection, we state the asymptotics of the KOO statistics and illustrate how the KOO statistics of true variables behave differently from that of the KOO statistics of spurious variables under some mild conditions. Before stating these results, we collect the needed conditions below.

(C1) As $\min\{k, p, n\} \to \infty$, $c_n \to c \in (0, 1)$ and $\alpha_n \to \alpha \in [0, 1)$ satisfying $\alpha + c < 1$.
(C2) The true model $\mathbf{j}_* \subset [k]$, and $|\mathbf{j}_*|$ is allowed to diverge as $k \to \infty$.
(C3) The entries e_{ij} of $\mathbf{E}$ are independent and identically distributed (i.i.d.) with zero means, unit variances, and finite fourth moments, i.e., $\tau = \mathbb{E}e_{ij}^4 - 3 \in (-\infty, \infty)$.
(C4) Matrix $\mathbf{X}'\mathbf{X}$ is positive definite for all $n > k + p$.

Our main results of this paper are stated below. The proofs, under normality of errors, will be given in Section 6; and the general proofs without assuming normality of errors will be given in the Appendix. Let

(7)
$$\delta_j := \delta_{nj} = p^{-1}\mathbf{x}_j'\mathbf{Q}_j\mathbf{x}_j\boldsymbol{\theta}_j'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_j.$$

The following theorem identifies the strong limits of the KOO statistics $\mathcal{K}_j$ for all $j \in [k]$.

Theorem 2.1. *Under conditions* (C1) – (C4), *we have uniformly in* $j \in [k]$,

$$\mathcal{K}_j = \begin{cases} \frac{c_n}{1 - c_n - \alpha_n} + o_{a.s.}(1), & \text{if } j \notin \mathbf{j}_*, \\ (1 + \delta_j)\left[\frac{c_n}{1 - c_n - \alpha_n} + o_{a.s.}(1)\right], & \text{if } j \in \mathbf{j}_*. \end{cases}$$

As $\boldsymbol{\theta}_j$ and $\boldsymbol{\Sigma}$ are typically unknown in practice, the limits of $\mathcal{K}_j$'s for $j \in \mathbf{j}_*$ are unknown. However, the fluctuations of the $\mathcal{K}_j$'s for spurious variables are pretty simple, which is described in the following theorem.

Theorem 2.2. *Under conditions* (C1) – (C4), *for any fixed integer $q > 0$ and $\{j_1, \ldots, j_q\} \subset \mathbf{j} \backslash \mathbf{j}_*$, the random vector*

$$\sqrt{p} \mathbf{G}_q^{-1/2} \left[(\mathcal{K}_{j_1}, \ldots, \mathcal{K}_{j_q})' - \frac{c_n}{1 - c_n - \alpha_n} \mathbf{1}_q \right]$$

converges weakly to the standard q-dimensional Gaussian random vector, where

$$\mathbf{G}_q = \frac{c_n^2}{(1 - \alpha_n - c_n)^2} \left[\frac{2(1 - \alpha_n)}{(1 - \alpha_n - c_n)} (\mathcal{A}_q' \mathcal{A}_q)^2 + \tau (\mathcal{A}_q \circ \mathcal{A}_q)'(\mathcal{A}_q \circ \mathcal{A}_q) \right],$$

and $\mathcal{A}_q = (\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_q})$ is an $n \times q$ non-random matrix.

Theorem 2.2 is of independent interest: As $\mathcal{K}_j$'s are the basic statistics for testing the hypothesis that $\boldsymbol{\theta}_j = \mathbf{0}$, this theorem can be used to obtain the CLTs of these statistics under the null hypothesis. Moreover, if $\tau = 0$ (e.g., $\{e_{ij}\}$ come from a standard normal distribution), then the second term in $\mathbf{G}_q$ vanishes; or if $\max_{j \in \mathbf{j} \backslash \mathbf{j}_*} \|\mathbf{a}_j\|_\infty = o(1)$, then the second term in the covariance matrix $\mathbf{G}_q$ tends to 0 as $n \to \infty$.

When $\tau \neq 0$, we propose an estimator of τ,

$$\hat{\tau} = \left\{ p^{-1} \text{tr}[(\mathbf{Y}'\mathbf{Q}\mathbf{Y} - (n-k)\mathbf{I}) \circ (\mathbf{Y}'\mathbf{Q}\mathbf{Y} - (n-k)\mathbf{I})] - 2(n-k) \right\} / \text{tr}(\mathbf{Q} \circ \mathbf{Q}),$$

which is shown to be unbiased and weakly consistent in Theorem 2.3 below.

Theorem 2.3. *Under the conditions* (C1) – (C4), *$\hat{\tau}$ is an unbiased and weakly consistent estimator of τ.*

Combining Theorems 2.2 and 2.3, the rejection region of the KOO statistics for testing whether some variables are spurious can be constructed. However, in order to know the power, we also need to know the fluctuations for the statistics of the true variables. The following theorem states that under some additional assumptions, the KOO statistic of the true variable is comparable to that of the spurious variables.

Theorem 2.4. *In addition to the conditions* (C1) – (C4), *for $j \in \mathbf{j}_*$, we assume that*

(C5): $\mathbb{E}e_{11}^3 = 0$.
(C6): As $\min\{p, n, k\} \to \infty$, $\|\mathbf{a}_j\|_\infty = o(1)$, $\mathbf{x}_j'\mathbf{Q}_j\mathbf{x}_j \|\boldsymbol{\theta}_j' \boldsymbol{\Sigma}^{-1/2}\|_\infty^2 = o(p)$.
(C7): As $\min\{p, n, k\} \to \infty$, δ_j tends to a constant.
Then,

$$\sqrt{p} \left(\mathcal{K}_j - \frac{c_n(1 + \delta_j)}{1 - c_n - \alpha_n} \right) / \sigma_{nj} \xrightarrow{D} N(0, 1),$$

where $\sigma_{nj}^2 = 2c_n^2[(1 - \alpha_n)(1 + 2\delta_j) + c_n \delta_j^2]/(1 - \alpha_n - c_n)^3$.

2.3. Some remarks on the theorems.

Remark 2.5. *The condition, $c > 0$, in* (C1) *is due to technical reasons:* our main tools are from random matrix theory (RMT) and RMT generally assumes the limit p/n exists and is positive. Note further that we make no explicit use of the unknown limits α and c in all the theorems below. Rather, we used α_n and c_n, which are always positive, in our results.

Remark 2.6. *If the model size k is greater than the sample size n but the true model size k_* is fixed, one can first apply screening methods (such as the sure independence screening method based on the distance correlation (Li et al., 2012), and interaction pursuit via distance correlation (Kong et al., 2017)) to ensure condition (C1) holds. For further details on the screening methods, see (Fan & Lv, 2008, 2010).*

Remark 2.7. *If the entries e_{ij} of $\mathbf{E}$ are independent but not necessarily identically distributed, our results in this paper continue to hold provided an additional Lindeberg-type condition:*

$$\frac{1}{\eta^4 n^2} \sum_{i,j} \mathbb{E}\left[|e_{ij}|^4 \mathbb{1}\left\{|e_{ij}| \geq \eta\sqrt{n}\right\}\right] = o(1),$$

for any $\eta > 0$. Here, $\mathbb{1}\{\cdot\}$ stands for the indicator function. The proofs are analogous but slightly more tedious, and we do not pursue this extension in this paper.

Remark 2.8. *From Theorems 2.2 and 2.4, we can theoretically investigate the asymptotic power of whether a variable is spurious. However, for testing whether a variable is true, the asymptotic distribution of the true KOO statistic (i.e., Theorem 2.4) cannot be applied directly since δ_j is unknown when j is a true variable. Variable selection problem will be discussed in the next section in detail.*

3. SELECTION CRITERIA BASED ON THE KOO STATISTICS

Theorem 2.1 highlights the crucial role of δ_j in differentiating the true variables from the spurious ones. For spurious variables, $\mathcal{K}_j$'s should be close to the point $c_n/(1 - c_n - \alpha_n)$ when n, p, k are large. Since δ_j is always positive for $j \in \mathbf{j}_*$, the true variables would be separated from $c_n/(1 - c_n - \alpha_n)$ and thus can be identified by the largest $\mathcal{K}_j$'s. Moreover, we can deduce a strongly consistent estimator for the true variables from this theorem. Let

$$\hat{\mathbf{j}}_\vartheta = \left\{ j \in [k] | \mathcal{K}_j > \frac{c_n(1 + \vartheta)}{1 - \alpha_n - c_n} \right\}, \ \vartheta > 0.$$

Then, we have the following corollary of Theorem 2.1.

Corollary 3.1. *Assume that conditions (C1) – (C4) hold and $\lim \delta_j > 0$ for all $j \in \mathbf{j}_*$. Then, for any fixed value $\vartheta \in (0, \min_{j \in \mathbf{j}_*}\{\lim \delta_j\})$,*

$$\lim_{n,p \to \infty} \hat{\mathbf{j}}_\vartheta \overset{a.s.}{\to} \mathbf{j}_*.$$

Remark 3.2. *This corollary implies the strong consistency for the KOO methods with AIC, BIC and C_p thresholds if δ_j satisfies the conditions.*

In practice, however, choosing a suitable ϑ is important but very challenging because (1) the largest spurious KOO statistic may converge to its limit slowly; (2) the spurious KOO statistics are correlated; and (3) the limits of the true KOO statistics are unknown. Hence, we propose a high-dimensional multiplier bootstrap procedure to approximate the distribution of the largest spurious KOO statistic $\mathcal{K}_j$, from which a selection criterion for the linear regression model (1) under the 3L framework is formulated.

Denote the estimator of the true model be

$$\hat{\mathbf{j}}_* = \{j \in [k] : \mathcal{K}_j > K_\nu\},$$

where K_ν is the critical value with at significance level ν, which is estimated by Algorithm 1.

From Theorem 2.2, the critical value K_ν may depend on $\|\mathbf{a}_i\|_\infty$ or the excess kurtosis but not on the exact distribution of the errors. The boxplots of the spurious KOO statistics $\mathcal{K}_j$'s for different distributions presented in Fig. 1 support this claim. In this simulation, we set $\mathbf{\Theta} = \mathbf{O}$, $\mathbf{\Sigma} = \mathbf{I}$ and

---

Algorithm 1: Estimation of K_ν

Input: ν, $\mathbf{Y}$, $\mathbf{X}$ and estimator $\hat{\tau}$ based on $\{\mathbf{Y}, \mathbf{X}\}$

Output: Estimator $\hat{K}_\nu$

1 Compute $\mathcal{A}_k = (\mathbf{a}_1, \ldots, \mathbf{a}_k)$.

2 Generate a random matrix $\tilde{\mathbf{E}}$ with $n \times p$ i.i.d. zero mean, unit variance and $\hat{\tau}$ excess kurtosis elements.

3 Compute $\mathbf{K} = \mathcal{A}'_k \tilde{\mathbf{E}} (\tilde{\mathbf{E}}' \mathbf{Q} \tilde{\mathbf{E}})^{-1} \tilde{\mathbf{E}}' \mathcal{A}_k$.

4 Compute the largest value of the diagonal elements of $\mathbf{K}$ and denote it by $\tilde{\mathcal{K}}^{(1)}$.

5 Repeat N times of the above procedures 2–4, and obtain $\{\tilde{\mathcal{K}}^{(1)}, \ldots, \tilde{\mathcal{K}}^{(N)}\}$.

6 Compute the $100(1-\nu)$th quantile of $\{\tilde{\mathcal{K}}^{(1)}, \ldots, \tilde{\mathcal{K}}^{(N)}\}$ and denote it by $\hat{K}_\nu$.

---

generate two predictor matrices: the first one is a 2000×600 matrix with i.i.d. entries from $U(1,5)$; and the second one is a 2000×600 diagonal matrix. As the values of the diagonal elements do not affect the result, the diagonal entries were chosen to be 1 in our simulation. We examine six different distributions of the errors: standard normal distribution $N(0,1)$, standardized uniform distribution $U(0,1)$, standardized Bernoulli distribution $B(1, \rho)$ with parameter $\rho = (6 - \sqrt{6})/12$, standardized chi-square distribution with 12 degrees of freedom $\chi^2(12)$, standardized t-distribution with 10 degrees of freedom t_{10}, standardized Poisson distribution with parameter 1 $Pois(1)$, standardized exponential distribution with rate parameter 1 $Exp(1)$ and standardized chi-square distribution with 2 degrees of freedom $\chi^2(2)$. Note that $\|\mathbf{a}_i\|_\infty \to 0$ for the random predictor matrix, $\|\mathbf{a}_i\|_\infty = 1$ for the rectangular diagonal predictor matrix, the excess kurtosis of $N(0,1)$ is 0, the excess kurtoses of $Exp(1)$ and $\chi^2(2)$ are 2, the excess kurtoses of $\chi^2(12)$, t_{10} and $Pois(1)$ are 1, and the excess kurtoses of $U(0,1)$ and $B(1, (6 - \sqrt{6})/12)$ are $-6/5$. Hence, in practice for convenience, we can use standardized χ^2 distribution with $12/\hat{\tau}$ degrees of freedom if $\hat{\tau} > 0$ and standardized Bernoulli distribution $B(1, \rho)$ with parameter ρ satisfying $\rho(1 - \rho) = 1/(6 - \hat{\tau})$ if $\hat{\tau} < 0$. Of course, if $\max_i \|\mathbf{a}_i\|_\infty \to 0$, we can use the standard normal distribution directly.

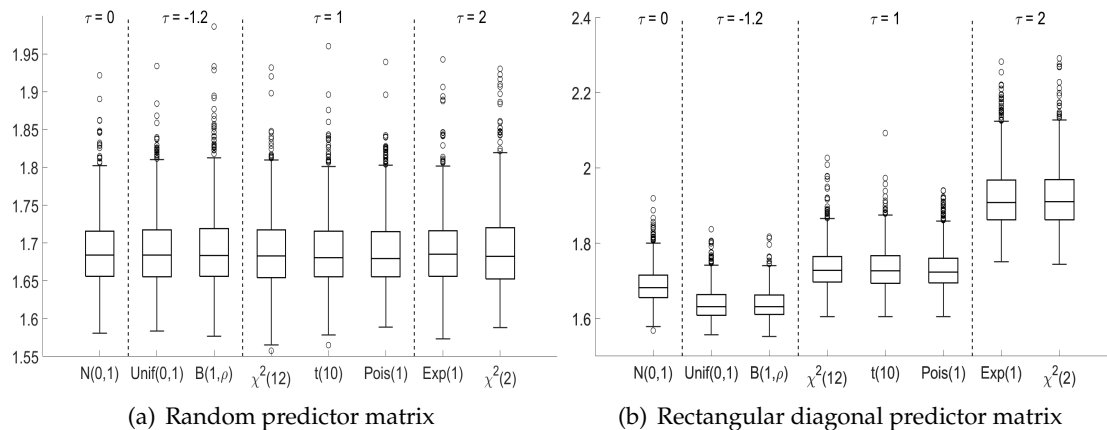(a) Random predictor matrix

(b) Rectangular diagonal predictor matrix

FIGURE 1. Boxplots of the spurious KOO statistics $\{\mathcal{K}^{(j)}, j = 1, \ldots, 1000\}$ with six different normalized distributions and two predictor matrices. The y-axis represents the values of $\mathcal{K}^{(j)}$'s.

4. SIMULATION STUDIES

In this section, we numerically examine the properties of the proposed KOO method in a 3L framework with different settings. For comparison, we also report the results of KOO methods with AIC, BIC and C_p thresholds as proposed by Nishii et al. (1988) and implemented by Fujikoshi & Sakurai (2019); Oda et al. (2020); Nakagawa et al. (2021); Fujikoshi (2022). Specifically, the KOO methods with AIC, BIC and C_p thresholds, respectively, choose the model

$$
\begin{aligned}
\hat{\mathbf{j}}_*^A &= \{j \in [k] : \log(1 + \mathcal{K}_j) > 2c_n\}, \\
\hat{\mathbf{j}}_*^B &= \{j \in [k] : \log(1 + \mathcal{K}_j) > \log(n)c_n\}, \\
\hat{\mathbf{j}}_*^C &= \{j \in [k] : (1 - \alpha_n)\mathcal{K}_j > 2c_n\}.
\end{aligned}
$$

For simplicity, we abbreviate the KOO method with our bootstrapping threshold to KBT, the KOO method with AIC threshold as KAIC. Similar abbreviations KBIC and KCp are used. We consider the following two settings.

Setting I: Fix $k_* = 5$, $p/n = \{0.2, 0.4\}$ and $k/n = \{0.2, 0.4\}$ with $n = 100, 500, 1000, 2000$. The results for $n = 2000$ are given in the Appendix. Set $\boldsymbol{\Sigma} = \mathbf{I}$, $\mathbf{X} = (x_{ij})_{n \times k}$, $\Theta_{\mathbf{j}_*} = \mathbf{1}_5\boldsymbol{\theta}_*$ and $\Theta = (\Theta_{\mathbf{j}_*}, \mathbf{0})$, where $\{x_{ij}\}$ are i.i.d. generated from the continuous uniform distributions $U(1, 5)$, $\mathbf{1}_5$ is a five-dimensional vector of ones and $\boldsymbol{\theta}_* = ((-0.5)^0, \ldots, (-0.5)^{p-1})$.

Setting II: Same as Setting I, except $\mathbf{X} = (\mathbf{I}_k, \mathbf{O}_{k \times (n-k)})'$ and $\Theta_{\mathbf{j}_*} = \sqrt{n}\mathbf{1}_5\boldsymbol{\theta}_*$.

For Setting I, we consider three cases for the distribution of $\mathbf{E}$:

  (i) Standard normal distribution, $e_{ij} \sim N(0, 1)$;
  (ii) Standardized t distribution with three degrees of freedom, i.e., $e_{ij} \sim t_3/\sqrt{5/3}$;
  (iii) Standardized chi-square distribution with two degrees of freedom, i.e., $e_{ij} \sim (\chi^2(3) - 3)/\sqrt{6}$.

Since $\|\mathbf{a}_j\|_\infty \to 0$ in Setting I, we use $\tilde{\mathbf{E}}$ with the standard normal distribution to estimate $\hat{K}_\nu$. We emphasize that the excess kurtosis of distribution t_3 is infinite.

For Setting II, we consider three cases for the distribution of $\mathbf{E}$:

  (iv) Standardized exponential distribution with rate parameter 1, i.e., $e_{ij} \sim Exp(1) - 1$;
  (v) Standardized Poisson distribution with parameter 1, i.e., $e_{ij} \sim Pois(1) - 1$;
  (vi) Standardized uniformly distribution, i.e., $e_{ij} \sim U(-\sqrt{3}, \sqrt{3})$.

Since $\|\mathbf{a}_j\|_\infty = 1$ in Setting II, we use $\tilde{\mathbf{E}}$ with standardized χ^2 distribution and standardized Bernoulli distribution, respectively, to estimate $\hat{K}_\nu$ with some suitably chosen parameter values.

In all the simulation studies, we choose two critical points in the KOO methods:.

$$
\hat{\mathbf{j}}_*^{(0)} = \{j \in [k] : \mathcal{K}_j > \hat{K}_0\} \quad \text{and} \quad \hat{\mathbf{j}}_*^{(5)} = \{j \in [k] : \mathcal{K}_j > \hat{K}_{0.05}\},
$$

where $\hat{K}_0$ and $\hat{K}_{0.05}$ are the largest and the 95th percentile of 1,000 bootstrap values, respectively.

We first explain our choices of the settings and the distributions. Since the KOO criteria depend on the values $\delta_j = p^{-1}\mathbf{x}_j'\mathbf{Q}_j\mathbf{x}_j\theta_j'\boldsymbol{\Sigma}^{-1}\theta_j$, it suffices to set $\boldsymbol{\Sigma} = \mathbf{I}$ and vary $\boldsymbol{\Theta}_*$ and $\mathbf{X}$ in conducting our simulation studies. Settings I and II both ensure δ_j are bounded above. For the case $\delta_j \to \infty$, the KOO statistics for the true variables and spurious variables are well separated, and all the compared selection methods will not show significant differences. The selection of distributions comprises five continuous distributions and one discrete distribution. The distribution described in (ii) only has finite second moment. This selection was made to investigate the implications of not satisfying the condition of finite fourth moment. To measure in greater detail the performance of these selection rules, the numbers of times, in 1000 repetitions, a selection rule under-specifies the true model, exactly identifies it and over-specifies it were tabulated. When the selection rule over-specifies the true

model, we also report the average number of spurious variables selected in the last row of each sub-table. Due to space consideration, we present selected results, but typical, of Setting I (i) and Setting II (iv) in Tables 1 and 2, respectively. Full set of results, including those for $n = 2000$, can be found in the Appendix.

	$\alpha = 0.2, c = 0.4$														
	$n = 100$					$n = 500$					$n = 1000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	938	0	640	19	0	1000	0	0	0	0	1000	0	0	0
T-S	35	62	0	360	940	2	0	0	1000	953	23	0	0	998	957
O-S	965	0	1000	0	41	998	0	1000	0	47	977	0	1000	2	43
A-S	3.69	–	7.06	–	1.05	6.86	–	46.30	–	1.04	3.92	–	95.28	1	1.05
	$\alpha = 0.4, c = 0.2$														
	$n = 100$					$n = 500$					$n = 1000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	42	0	828	41	0	129	0	0	0	0	729	0	0	0
T-S	0	923	3	172	919	0	871	0	998	965	0	271	41	1000	954
O-S	1000	35	997	0	40	1000	0	1000	2	35	1000	0	959	0	46
A-S	16.50	1.09	6.67	–	1.12	100.87	–	8	1	1	213.52	–	3.28	–	1.02

TABLE 1. Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (I) and (i) based on 1,000 replications. Here U-S, T-S, O-S and A-S stand for number of times a selection method under-specified the true model, number of times a selection method identified the true model exactly, number of times a selection method over-specified the true model, and the average number of spurious variables a selection method identified when it over-specified the model, respectively.

Based on our simulation results, the following observations are made: (1) The proposed KBT are the most robust among the compared methods, especially when the sample size n is large. (2) If the sample size n is small, we recommend choosing a bigger ν in order to avoid missing the true variables. After all, admitting a small number of spurious variables is a better tradeoff than missing some true variables. (3) Choosing a bigger ν may select more spurious variables, but unlike the KAIC and KCp , the number of spurious variables selected is still under control. (4) The simulation results are very similar across different distributions of errors, which suggests these selection rules are rather robust against the distributions of errors. (5) When $\max_i \|\mathbf{a}_i\|_\infty \to 0$, our proposed methods also work well even the finite fourth moment condition does not hold, suggesting that our theorems continue to hold even under weaker conditions. Our guess is that finite second moment of the underlying error distributions is enough. (6) The performances of KAIC, KBIC and KCp are not acceptable under our settings: KAIC and KCp frequently over-specify the true models quite substantially, and KBIC frequently under-specifies the true models. Under some special cases, KBIC has good selection times, however, KBT in general outperforms KBIC.

5. REAL DATA ANALYSIS

We apply the proposed methods to two real examples. The first example is a chemometrics data taken from Skagerberg et al. (1992) (we replaced the value 19203 with 1.9203 in the 37th observation).

	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
	$\alpha = 0.2, c = 0.4$											
	\multicolumn{5}{c}{$n = 100$}	\multicolumn{5}{c}{$n = 500$}					\multicolumn{5}{c}{$n = 1000$}				
U-S	0	925	0	991	622	0	1000	0	2	0	0	1000	0	0	0
T-S	2	74	0	9	361	0	0	0	997	934	0	0	0	1000	938
O-S	998	1	1000	0	17	1000	0	1000	1	66	1000	0	1000	0	62
A-S	4.74	1	7	–	1.06	16.40	–	45.88	1	1	18.74	–	95.36	–	1.03
	$\alpha = 0.4, c = 0.2$											
	\multicolumn{5}{c}{$n = 100$}	\multicolumn{5}{c}{$n = 500$}					\multicolumn{5}{c}{$n = 1000$}				
U-S	0	4	0	999	597	0	7	0	7	0	0	27	0	0	0
T-S	0	348	0	1	386	0	993	0	993	961	0	973	0	1000	939
O-S	1000	648	1000	0	17	1000	0	1000	0	39	1000	0	1000	0	61
A-S	15.31	1.67	9.23	–	1	94.64	–	28.61	–	1	198.92	–	31.12	–	1.03

TABLE 2. Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (II) and (iv) based on 1,000 replications. Here U-S, T-S, O-S and A-S stand for number of times a selection method under-specified the true model, number of times a selection method identified the true model exactly, number of times a selection method over-specified the true model, and the average number of spurious variables a selection method identified when it over-specified the model, respectively.

The data are taken from a simulation of a low-density tubular polyethylene reactor studying the relationship between polymer properties and the process. The predictor variables consist of 20 temperatures measured at equal distances along the low-density polyethylene reactor section, together with the wall temperature of the reactor and the solvent feed rate. The responses are the output characteristics of the polymers, including two molecular weights, two branching frequencies and the contents of two groups. This data set has been studied by Breiman & Friedman (1997) and Similä & Tikka (2007). Similar to Breiman & Friedman (1997), we log-transformed the response values because they are highly skewed to the right. In total, there are $n = 56$ observations with $k = 22$ predictor variables and $p = 6$ responses.

We present the scatterplot of $\{\mathcal{K}_j\}$ in descending order in Figure 2. We also indicate the critical values of KAIC, KBIC and KCp, and $\hat{K}_0$, $\hat{K}_{0.05}$ estimated by Algorithm 1 with standard normal distribution and $N = 1,000$. Since the dimension is relatively small, we recommend using a larger significance level ν to prevent under-specifying. It seems that the variables $\{22, 3, 4\}$ are significant and variables $\{21, 11\}$ are potentially significant too. KAIC and KCp, however, select many more variables, which are likely to be spurious. The second example is a multivariate yeast cell-cycle dataset from Spellman et al. (1998), which can be found in the R package "spls". This data set contains 542 cell-cycle-related genes (i.e., $n = 542$). Each gene contains 106 binding levels of transcription factors (i.e., $k = 106$) and 18 time points covering two cell cycles (i.e., $p = 18$). The binding levels of the transcription factors play a role in determining which genes are expressed and help delineate the process behind eukaryotic cell cycles. Further explanations of the dataset can be found in (Wang et al., 2007; Chun & Keleş, 2010; Chen & Huang, 2012; Kong et al., 2017). Our results are presented in Figure 3. The transcription factors {SWI5, STE12, ACE2, NDD1}, corresponding to the four largest $\mathcal{K}_j$-values, have been confirmed to be related to the cell cycle regulation by experiment Wang et al.
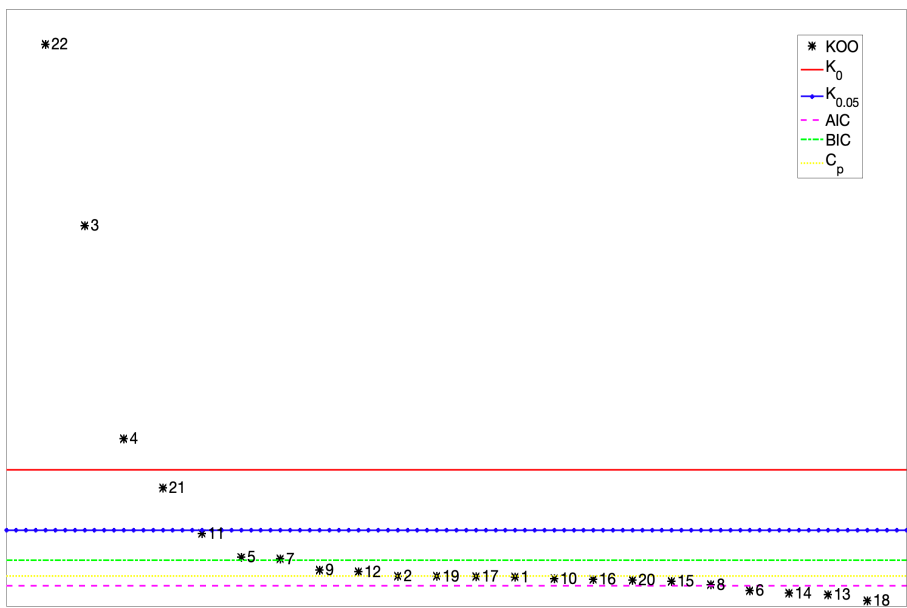
FIGURE 2. Scatterplots for the chemometrics dataset.

(2007). And the other two transcription factors {RME1, HIR2} could possibly be related to the cell cycle regulation. KBIC, however, will have missed identifying the TFs {STE12, ACE2, NDD1} in the yeast cell-cycle. On the other hand, KAIC and KCp will have identified more TFs, many of which may not be related to the yeast cell-cycle.
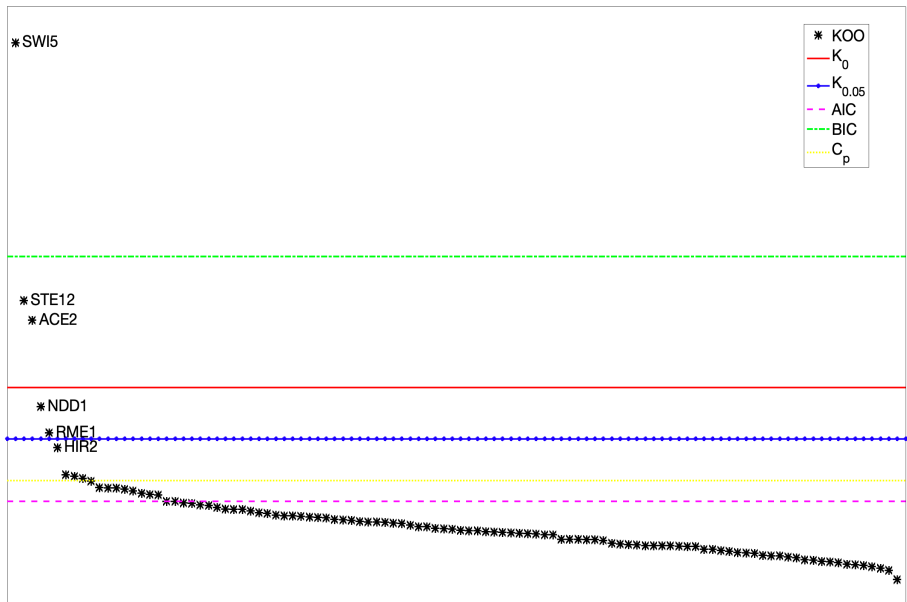

FIGURE 3. Scatterplots for the yeast cell-cycle dataset.

6. Proofs of Theorems 2.1, 2.2 and 2.4 under normality

If the errors follow the standard normal distribution, the KOO statistic can be written as the quotient of two independent chi-squared random variables. As a result, the proofs of Theorems 2.1, 2.2 and 2.4 are easier to present. The proofs may also be of independent interest. Hence, we prove these results under normality in this section. The proofs of these theorems for general error distributions via random matrix theory are postponed in the Appendix for interested readers. Note that there is no need to prove Theorem 2.3 when the errors follow the standard normal distribution.

Recall the KOO statistic

$$\mathcal{K}_j = \mathbf{v}_j' \mathbf{W}^{-1} \mathbf{v}_j,$$

where

$$\mathbf{v}_j = \boldsymbol{\Sigma}^{-1/2} \mathbf{Y}' \mathbf{a}_j \text{ and } \mathbf{W} = \mathbf{E}' \mathbf{Q} \mathbf{E}.$$

When $\mathbf{E}$ has the standard normal distribution, it follows that

$$\mathbf{W} \sim W_p(n-k, \mathbf{I}_p), \ \mathbf{v}_j \sim N_p(\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta}_* \mathbf{X}_*' \mathbf{a}_j, \mathbf{I}_p), \quad j = 1, \dots, k,$$

and $\mathbf{W}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ are independent. Note that $\mathbf{v}_1, \dots, \mathbf{v}_k$ are not necessarily independent. If $j \notin \mathbf{j}_*$, $\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta}_* \mathbf{X}_*' \mathbf{a}_j = \mathbf{0}$, on the other hand if $j \in \mathbf{j}_*$, $\boldsymbol{\Sigma}^{-1/2} \boldsymbol{\Theta}_* \mathbf{X}_*' \mathbf{a}_j \neq \mathbf{0}$. Moreover, under the assumption of normality, $\tau = 0$ in assumption (C3). Next, we state a preliminary lemma.

Lemma 6.1. *Let* $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_q)$ *be a* $p \times q$ *random matrix with* $q \leq p$, *and let* $\mathbf{W}$ *be a* $p \times p$ *random matrix which is distributed as Wishart distribution* $W_p(m, \mathbf{I}_p)$. *Assume that* $\mathbf{V}$ *and* $\mathbf{W}$ *are independent. Let* $\mathbf{H}$ *be a* $p \times p$ *random orthogonal matrix such that the first* q *columns are* $\mathbf{V}(\mathbf{V}'\mathbf{V})^{-1/2}$, *that is,* $\mathbf{H} = (\mathbf{V}(\mathbf{V}'\mathbf{V})^{-1/2}, \cdot)$. *Let*

(8) $$\mathbf{Z} = \mathbf{H}' \mathbf{W} \mathbf{H} = \begin{pmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{12} \\ \mathbf{Z}_{21} & \mathbf{Z}_{22} \end{pmatrix}, \quad \mathbf{Z}_{11 \cdot 2} = \mathbf{Z}_{11} - \mathbf{Z}_{12} \mathbf{Z}_{22}^{-1} \mathbf{Z}_{21},$$

and $\mathbf{Z}_{21}$ *is a* $(p-q) \times q$ *matrix. Then,*

(9) $$\mathbf{Z} \sim W_p(m, \mathbf{I}_p),$$

(10) $$\mathbf{V}' \mathbf{W}^{-1} \mathbf{V} = (\mathbf{V}'\mathbf{V})^{1/2} \mathbf{Z}_{11 \cdot 2}^{-1} (\mathbf{V}'\mathbf{V})^{1/2},$$

(11) $$\mathbf{Z}_{11 \cdot 2} \sim W_q(m - (p-q), \mathbf{I}_q).$$

When $q = 1$,

$$\mathbf{v}_1' \mathbf{W}^{-1} \mathbf{v}_1 = \frac{\mathbf{v}_1' \mathbf{v}_1}{\mathbf{Z}_{11 \cdot 2}},$$

where $\mathbf{Z}_{11 \cdot 2} \sim \chi^2(m - (p-1))$ *is a chi-square variate with* $m - (p-1)$ *degrees of freedom, and* $\mathbf{v}_1' \mathbf{v}_1$ *and* $\mathbf{Z}_{11 \cdot 2}$ *are independent. Further, if* $\mathbf{v}_1 \sim N_p(\boldsymbol{\mu}, \mathbf{I}_p)$,

$$\mathbf{v}_1' \mathbf{W}^{-1} \mathbf{v}_1 = \frac{\mathbf{v}_1' \mathbf{v}_1}{\mathbf{Z}_{11 \cdot 2}} \sim \frac{\chi^2(p; \boldsymbol{\mu}' \boldsymbol{\mu})}{\chi^2(m - (p-1))}.$$

Here, $\chi^2(p; \delta^2)$ *denotes a noncentral chi-square variate with* p *degrees of freedom and noncetrality parameter* $\boldsymbol{\mu}' \boldsymbol{\mu}$, *and* $\chi^2(m - (p-1))$ *denotes a chi-square variate with* $m - (p-1)$ *degrees of freedom, and they are independent.*

Proof. The result (9) is straightforward by considering the conditional distribution of $\mathbf{Z}$ given $\mathbf{H}$, and noting that the obtained result does not depend on $\mathbf{H}$. Next we consider the result (10). Noting that

$\mathbf{H}$ is orthogonal, we have

$$\mathbf{V}'\mathbf{W}^{-1}\mathbf{V} = \mathbf{V}'\mathbf{H}(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1}\mathbf{H}'\mathbf{V}$$
$$= \left[(\mathbf{V}'\mathbf{V})^{1/2},\, \mathbf{O}\right]' \mathbf{Z}^{-1} \left[(\mathbf{V}'\mathbf{V})^{1/2},\, \mathbf{O}\right],$$

which implies (10). (11) is a well known result on Wishart distribution (e.g., Theorem 2.2.3 in (Fujikoshi et al., 2010)).

Next we consider the case of $q = 1$. Note that the $(1,1)$ element of $\mathbf{Z}^{-1}$ is $\mathbf{Z}_{11\cdot 2}^{-1}$. Then

$$\mathbf{v}_1'\mathbf{W}^{-1}\mathbf{v}_1 = \mathbf{v}_1\mathbf{H}'(\mathbf{H}'\mathbf{W}\mathbf{H})^{-1}\mathbf{H}'\mathbf{v}_1$$
$$= ((\mathbf{v}_1'\mathbf{v}_1)^{1/2}, 0, \ldots, 0)\mathbf{Z}^{-1}((\mathbf{v}_1'\mathbf{v}_1)^{1/2}, 0, \ldots, 0)'$$
$$= \mathbf{v}_1'\mathbf{v}_1\mathbf{Z}_{11\cdot 2}^{-1}.$$

The required result follows from the fact that $\mathbf{v}_1'\mathbf{v}_1 \sim \chi^2(p; \delta^2)$ and $\mathbf{Z}_{11\cdot 2} \sim \chi^2(m - (p-1))$. Then we complete the proof of this lemma. $\qquad\square$

6.1. **Proof of Theorem 2.1.** By Lemma 6.1, we can express $\mathcal{K}_j$ as a ratio of two independent chi-square variates as

$$\mathcal{K}_j = \begin{cases} \dfrac{\chi^2(p)}{\chi^2(\widetilde{m})} & \text{if } j \notin \mathbf{j}_* \\ \dfrac{\chi^2(p; p\delta_j)}{\chi^2(\widetilde{m})} & \text{if } j \in \mathbf{j}_* \end{cases},$$

where $\delta_j = p^{-1}\mathbf{x}_j'\mathbf{Q}_j\mathbf{x}_j\theta_j'\mathbf{\Sigma}^{-1}\theta_j$ and $\widetilde{m} = n - k - p + 1$. For $j \notin \mathbf{j}_*$, let

$$Z_1 = \frac{\chi^2(p) - p}{\sqrt{2p}} \text{ and } Z_2 = \frac{\chi^2(\widetilde{m}) - \widetilde{m}}{\sqrt{2\widetilde{m}}}.$$

Then, it is clear that

$$\mathcal{K}_j = \frac{p + \sqrt{2p}Z_1}{\widetilde{m} + \sqrt{\widetilde{m}}Z_2} = \frac{p/n + \sqrt{2p}Z_1/n}{\widetilde{m}/n + \sqrt{\widetilde{m}}Z_2/n}$$
$$= \frac{c_n}{1 - c_n - \alpha_n} + o_{a.s.}(1).$$

For $j \in \mathbf{j}_*$, let

$$\widetilde{Z}_1 = \frac{\chi^2(p; p\delta_j) - p(1 + \delta_j)}{\sqrt{2p(1 + 2\delta_j)}}.$$

Note that $\widetilde{Z}_1 \overset{D}{\to} N(0,1)$ as $p \to \infty$ or $p\delta_j^2 \to \infty$. Thus we can find that

$$\mathcal{K}_j = \left\{p(1 + \delta_j) + \sqrt{2p(1 + 2\delta_j)}\widetilde{Z}_1\right\}\left\{\widetilde{m} + \sqrt{2\widetilde{m}}Z_2\right\}^{-1}$$
$$= \frac{p}{\widetilde{m}}\left\{1 + \delta_j + \sqrt{2(1 + 2\delta_j)p^{-1}}\widetilde{Z}_1\right\}\left\{1 - \sqrt{2\widetilde{m}^{-1}}Z_2\right\}^{-1},$$

which implies

$$\mathcal{K}_j - \frac{c_n(1 + \delta_j)}{1 - c_n - \alpha_n} = o_{a.s.}(1 + \delta_j).$$

Then we complete the proof of Theorem 2.1.

6.2. **Proof of Theorem 2.2.** For simplicity, we consider the case $q = 2$, and assume that $\{1, 2\} \subset [k]\setminus\mathbf{j}_*$. To prove Theorem 2.2, it is sufficient to show that for any non-null vector $\mathbf{h} = (h_1, h_2)'$, $\sqrt{p}[h_1\mathcal{K}_1 + h_2\mathcal{K}_q - \frac{c_n}{1-c_n-\alpha_n}(h_1 + h_2)]$ converges weakly to a normal distribution with mean zero and variance $\frac{2c^2(1-\alpha)}{(1-\alpha-c)^3}\mathbf{h}'(\mathcal{A}_2'\mathcal{A}_2)^2\mathbf{h}$.

Under the normality assumption, we can express $\mathcal{K}_1$ and $\mathcal{K}_2$ as follows:

(12) $$\mathcal{K}_1 = \mathbf{v}_1'\mathbf{W}^{-1}\mathbf{v}_1, \quad \mathcal{K}_2 = \mathbf{v}_2'\mathbf{W}^{-1}\mathbf{v}_2.$$

Here, $\mathbf{v}_i \sim N_p(\mathbf{0}, \mathbf{I}_p), i = 1, 2$, $\mathbf{W} \sim W_p(m, \mathbf{I}_p)$, $m = n - k$, $\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\mathbf{W}$ are independent, but $\mathbf{v}_1$ and $\mathbf{v}_2$ are not independent. Let $\mathcal{K}_0 = \mathbf{v}_1'\mathbf{W}^{-1}\mathbf{v}_2$ and

$$\mathcal{K} = \begin{pmatrix} \mathcal{K}_1 & \mathcal{K}_0 \\ \mathcal{K}_0' & \mathcal{K}_2 \end{pmatrix}.$$

Note that

$$h_1\mathcal{K}_1 + h_2\mathcal{K}_2 = \text{tr}D_h\mathcal{K},$$

where $D_h = \begin{pmatrix} h_1 & 0 \\ 0 & h_2 \end{pmatrix}$.

Let $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2)$. Using Lemma 6.1, we can write $\mathcal{K}$ as

$$\mathcal{K} = \mathbf{V}'\mathbf{W}^{-1}\mathbf{V}$$

(13) $$= \left(\frac{1}{p}\mathbf{V}'\mathbf{V}\right)^{1/2}\left(\frac{1}{\widetilde{m}}\mathbf{Z}_{11\cdot2}\right)^{-1}\left(\frac{1}{p}\mathbf{V}'\mathbf{V}\right)^{1/2}\frac{p}{\widetilde{m}},$$

where $\mathbf{Z}_{11\cdot2} \sim W_2(\widetilde{m}, \mathbf{I}_p)$ is defined in (8) and $\widetilde{m} = m - (p - 2)$. Note that

$$\mathbf{V}'\mathbf{V} \sim W_2(p, \mathbf{\Lambda}), \quad \mathbf{\Lambda} = \begin{pmatrix} 1 & \lambda \\ \lambda & 1 \end{pmatrix},$$

where $\lambda = \mathbf{a}_1'\mathbf{a}_2$. Let

(14) $$\mathbf{F} = \begin{pmatrix} f_1 & f_3 \\ f_3 & f_2 \end{pmatrix} = \sqrt{p}\left(\frac{1}{p}\mathbf{V}'\mathbf{V} - \mathbf{\Lambda}\right),$$

(15) $$\mathbf{G} = \begin{pmatrix} g_1 & g_3 \\ g_3 & g_2 \end{pmatrix} = \sqrt{\widetilde{m}}\left(\frac{1}{\widetilde{m}}\mathbf{Z}_{11\cdot2} - \mathbf{I}_2\right).$$

It follows from the asymptotic distribution of a Wishart matrix (e.g., Theorem 2.5.1 in (Fujikoshi et al., 2010)) that the limiting distribution of $(f_1, f_2, f_3)'$ (respectively, $(g_1, g_2, g_3)'$) is a 3-variate normal distribution with mean zero and covariance matrix

$$\begin{pmatrix} 2 & 2\lambda^2 & 2\lambda \\ 2\lambda^2 & 2 & 2\lambda \\ 2\lambda & 2\lambda & 1 + \lambda^2 \end{pmatrix}, \quad \left(\text{respectively, } \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right).$$

Consequently, it is straightforward to show that

(16) $$\text{tr}D_h\mathbf{F} \xrightarrow{D} N_2(0, 2\text{tr}(D_h\mathbf{\Lambda})^2)$$

and

(17) $$\text{tr}\mathbf{\Lambda}^{1/2}D_h\mathbf{\Lambda}^{1/2}\mathbf{G} \xrightarrow{D} N_2(0, 2\text{tr}(D_h\mathbf{\Lambda})^2).$$

Then, by substituting

$$\frac{1}{p}\mathbf{V}'\mathbf{V} = \mathbf{\Lambda} + \frac{1}{\sqrt{p}}\mathbf{F}$$

and

$$\left(\frac{1}{\widetilde{m}}\mathbf{Z}_{11\cdot 2}\right)^{-1} = \left(\mathbf{I}_2 + \frac{1}{\sqrt{\widetilde{m}}}\mathbf{G}\right)^{-1} = \mathbf{I}_2 - \frac{1}{\sqrt{\widetilde{m}}}\mathbf{G} + \frac{1}{\widetilde{m}}\left(\mathbf{I}_2 + \frac{1}{\sqrt{\widetilde{m}}}\mathbf{G}\right)^{-1}\mathbf{G}^2$$

into (13), we can expand $\mathcal{K}$ as

(18) $$\mathcal{K} = \left\{\Lambda + \frac{1}{\sqrt{p}}\mathbf{F} - \frac{1}{\sqrt{\widetilde{m}}}(\Lambda + \frac{1}{\sqrt{p}}\mathbf{F})^{1/2}\mathbf{G}(\Lambda + \frac{1}{\sqrt{p}}\mathbf{F})^{1/2} + \varnothing\right\}\frac{c_n}{1 - c_n - \alpha_n},$$

where $\varnothing$ denotes the terms of order $O_p(n^{-1})$. Using (18), we have

$$\sqrt{p}\left\{h_1\mathcal{K}_1 + h_2\mathcal{K}_2 - \frac{c_n}{1 - c_n - \alpha_n}(h_1 + h_2)\right\}$$

(19) $$= \frac{c_n}{1 - c_n - \alpha_n}\left\{\mathrm{tr}D_h\mathbf{F} - \left(\frac{c_n}{1 - c_n - \alpha_n}\right)^{1/2}\mathrm{tr}\Lambda^{1/2}D_h\Lambda^{1/2}\mathbf{G}\right\} + O_p(n^{-1/2}).$$

By (16) and (17), we can see that the limiting distribution of (19) is normal with mean zero and variance

$$\frac{2c^2}{(1 - c - \alpha)^2}\left(1 + \frac{c}{1 - c - \alpha}\right)\mathrm{tr}(D_h\Lambda)^2 = \frac{2c^2(1 - \alpha)}{(1 - \alpha - c)^3}\mathbf{h}'(\mathcal{A}_2'\mathcal{A}_2)^2\mathbf{h}.$$

This completes the proof of Theorem 2.2 .

6.3. **Proof of Theorem 2.4.** In the proof of Theorem 2.1, recall that for $j \in \mathbf{j}_*$, $\mathcal{K}_j$ can be expressed as a ratio of two independent chi-square variates:

$$\mathcal{K}_j = \frac{\chi^2(p; p\delta_j^2)}{\chi^2(\tilde{m})},$$

where $\chi^2(p; p\delta_j^2)$ denotes a noncentral chi-square variate with p degrees of freedom and noncentrality parameter $p\delta_j^2$, and $\chi^2(\tilde{m})$ denotes a chi-square variate with $\tilde{m} = n - k - p + 1$ degrees of freedom, and they are independent. Let

$$\widetilde{Z}_1 = \frac{\chi^2(p; p\delta_j^2) - p - p\delta_j^2}{\sqrt{2(p + 2p\delta_j^2)}}, \quad \widetilde{Z}_2 = \frac{\chi^2(\tilde{m}) - \tilde{m}}{\sqrt{2\tilde{m}}}.$$

Then, it is checked that $\widetilde{Z}_1$ and $\widetilde{Z}_2$ converge to the standard normal distribution. Note that

$$\mathcal{K}_j = \left\{(p + p\delta_j^2) + \sqrt{2(p + 2p\delta_j^2)}\widetilde{Z}_1\right\}\left\{\tilde{m} + \sqrt{2\tilde{m}}\widetilde{Z}_2\right\}^{-1}$$

$$= \frac{p}{\tilde{m}}\left\{1 + \delta_j^2 + \sqrt{2p^{-1}(1 + 2\delta_j^2)}\widetilde{Z}_1\right\}\left\{1 + \sqrt{2\tilde{m}^{-1}}\widetilde{Z}_2\right\}^{-1}.$$

This implies that

$$\sqrt{p}\left\{\mathcal{K}_j - \frac{p}{\tilde{m}}(1 + \delta_j^2)\right\}$$

$$= \frac{p}{\tilde{m}}\left\{\sqrt{2(1 + 2\delta_j^2)}\widetilde{Z}_1 - (1 + \delta_j^2)\left(\frac{2p}{\tilde{m}}\right)^{1/2}\widetilde{Z}_2\right\} + O_p(n^{-1/2}).$$

Theorem 2.4 follows from noting that $\widetilde{Z}_1$ and $\widetilde{Z}_2$ independently converge to the standard normal distribution.

APPENDIX A. SIMULATION RESULTS

The simulation results for Settings I and II, and six cases of distribution of $\mathbf{E}$ are tabulated in Tables 3–8.

APPENDIX B. PROOFS OF THEOREMS 2.1–2.4

In this appendix, we present the proofs of Theorems 2.1–2.4 under general distributions by random matrix theory. Before that, we first give some notation and preliminary results which will be used in the sequel frequently. For simplicity, we denote $\mathbf{M} = p^{-1}\mathbf{E}'\mathbf{Q}\mathbf{E}$ and $\mathbf{M}_l = \frac{1}{p}\mathbf{E}'_l\mathbf{Q}\mathbf{E}_l$, where $\mathbf{E}_l$ is the $n \times (p-1)$ submatrix of $\mathbf{E}$ with the l-th column removed. Denote by $\mathbb{E}_l$ the conditional expectation given $\{\mathbf{e}_1, \ldots, \mathbf{e}_l\}$ and by $\mathbb{E}_0 = \mathbb{E}$ the unconditional expectation, where $\mathbf{e}_i$ is the n-vector of the i-th column of $\mathbf{E}$. Let $\mathbf{b} = \mathbf{\Sigma}^{-1/2}\mathbf{\Theta}_*\mathbf{X}'_*\mathbf{a}_1$ and $\mathbf{b}_l$ be the $p-1$ sub-vector of $\mathbf{b}$ with the l-th entry b_l removed. Then we have

$$\mathbf{a}'_1\mathbf{Y}\mathbf{\Sigma}^{-1/2}(\mathbf{E}'\mathbf{Q}\mathbf{E})^{-1}\mathbf{\Sigma}^{-1/2}\mathbf{Y}'\mathbf{a}_1 = p^{-1}(\mathbf{b}' + \mathbf{a}'_1\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}).$$

Modifying the truncation argument of Bai et al. (2018b), we can assume that the variables $\{e_{ij}, i = 1 \ldots n, j = 1 \ldots p\}$ satisfy the following additional condition:

(20) $$|e_{ij}| < \eta_n\sqrt{n}, \quad \text{for all } i, j,$$

where $\eta_n \to 0$ slowly enough. By the theorem in the appendix of Bai & Silverstein (2004), we know for any positive constant $d < (1 - \sqrt{c})^2$ and any given $t > 0$, $\lambda_{\min}^{\frac{1}{n}\mathbf{E}'\mathbf{E}} \overset{a.s.}{\to} (1 - \sqrt{c})^2$ and

$$\mathbb{P}(\lambda_{\min}^{\frac{1}{n}\mathbf{E}'\mathbf{E}} < d) = o(n^{-t}).$$

Moreover, by Theorem 1.2 in Bai & Silverstein (1999), we conclude that for any positive constant $d < (1 - \sqrt{c/(1-\alpha)})^2$ and any given $t > 0$, $\lambda_{\min}^{\frac{1}{n}\mathbf{E}'\mathbf{Q}\mathbf{E}} \overset{a.s.}{\to} (1 - \sqrt{c/(1-\alpha)})^2$ and

$$\mathbb{P}(\lambda_{\min}^{\frac{1}{n}\mathbf{E}'\mathbf{Q}\mathbf{E}} < d) = o(n^{-t}).$$

Denote

$$\beta_l = \frac{1}{p}\mathbf{e}'_l\mathbf{Q}\mathbf{e}_l - \frac{1}{p^2}\mathbf{e}'_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}'_l\mathbf{Q}\mathbf{e}_l$$

and

$$\beta_1^{tr} = \text{tr}[\frac{1}{p}\mathbf{Q} - \frac{1}{p^2}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}'_l\mathbf{Q}] = \frac{n - k - p + 1}{p}.$$

It follows that

(21) $$\frac{1}{\beta_l} = \frac{1}{\beta_1^{tr}} - \frac{\xi_l}{\beta_l\beta_1^{tr}},$$

where $\xi_l = \beta_l - \beta_1^{tr}$. By Lemma 7.2 in Bai & Yao (2005) (see Lemma B.4), we have that for any $2 \le \ell \le \log(n)$,

(22) $$\mathbb{E}|\xi_l|^\ell = O(p^{-1}\eta_n^{2\ell-4}),$$

$\alpha = 0.2, c = 0.2$									
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	79	0	15	0	0	589	0	0	0
T-S	198	921	228	983	966	570	411	655	999	954
O-S	802	0	772	2	34	430	0	345	1	46
A-S	**2.05**	–	1.97	1	1	1.37	–	1.27	1	1.04
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	993	0	0	0	0	1000	0	0	0
T-S	956	7	972	1000	958	1000	0	1000	999	951
O-S	44	0	28	0	42	0	0	0	1	49
A-S	1.02	–	1	–	1.02	–	–	–	1	1.04

$\alpha = 0.2, c = 0.4$									
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	938	0	640	19	0	1000	0	0	0
T-S	35	62	0	360	940	2	0	0	1000	953
O-S	965	0	1000	0	41	998	0	1000	0	47
A-S	3.69	–	7.06	–	1.05	6.86	–	46.30	–	1.04
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	23	0	0	998	957	505	0	0	1000	961
O-S	977	0	1000	2	43	495	0	1000	0	39
A-S	3.92	–	95.28	1	1.05	1.47	–	194.82	–	1

$\alpha = 0.4, c = 0.2$									
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	42	0	828	41	0	129	0	0	0
T-S	0	923	3	172	919	0	871	0	998	965
O-S	1000	35	997	0	40	1000	0	1000	2	35
A-S	16.50	1.09	6.67	–	1.12	100.87	–	8	1	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	729	0	0	0	0	999	0	0	0
T-S	0	271	41	1000	954	0	1	748	995	940
O-S	1000	0	959	0	46	1000	0	252	5	60
A-S	213.52	–	3.28	–	1.02	450.55	–	1.16	1	1

$\alpha = 0.4, c = 0.4$									
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	623	0	999	889	0	1000	0	10	0
T-S	0	294	0	1	103	0	0	0	990	963
O-S	1000	83	1000	0	8	1000	0	1000	0	37
A-S	31.05	1.51	29.35	–	1.25	194.59	–	193.17	–	1.05
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	998	952	0	0	0	996	930
O-S	1000	0	1000	2	48	1000	0	1000	4	70
A-S	394.99	–	394.85	1	1.06	795	–	795	1	1.01

TABLE 3. Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (I) and (i) based on 1,000 replications. Here U-S , T-S, O-S and A-S stand for number of times a selection method under-specified the true

	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
$\alpha = 0.2, c = 0.2$										
	$n = 100$					$n = 500$				
U-S	1	109	1	45	10	0	307	0	0	0
T-S	198	891	219	955	958	563	693	646	1000	961
O-S	801	0	780	0	32	437	0	354	0	39
A-S	2.09	–	1.96	–	1.03	1.35	–	1.25	–	1.05
	$n = 1000$					$n = 2000$				
U-S	0	780	0	0	0	0	1000	0	0	0
T-S	963	220	978	999	946	1000	0	1000	999	953
O-S	37	0	22	1	54	0	0	0	1	47
A-S	1.03	–	1	1	1.02	–	–	–	1	1.04
$\alpha = 0.2, c = 0.4$										
	$n = 100$					$n = 500$				
U-S	2	734	0	487	41	0	1000	0	0	0
T-S	40	266	0	513	937	1	0	0	1000	945
O-S	958	0	1000	0	22	999	0	1000	0	55
A-S	3.63	–	6.99	–	1.05	6.59	–	45.85	–	1.02
	$n = 1000$					$n = 2000$				
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	26	0	0	1000	962	490	0	0	999	943
O-S	974	0	1000	0	38	510	0	1000	1	57
A-S	3.95	–	95.28	–	1	1.46	–	194.80	1	1.02
$\alpha = 0.4, c = 0.2$										
	$n = 100$					$n = 500$				
U-S	0	56	2	245	53	0	113	0	0	0
T-S	0	917	2	755	913	0	887	0	999	958
O-S	1000	27	996	0	34	1000	0	1000	1	42
A-S	16.56	1.04	6.79	–	1.03	102.01	–	8.19	1	1
	$n = 1000$					$n = 2000$				
U-S	0	465	0	0	0	0	902	0	0	0
T-S	0	535	54	1000	940	0	98	780	999	959
O-S	1000	0	946	0	60	1000	0	220	1	41
A-S	213.34	–	3.27	–	1.02	449.91	–	1.11	1	1.05
$\alpha = 0.4, c = 0.4$										
	$n = 100$					$n = 500$				
U-S	0	465	1	984	744	0	996	0	17	3
T-S	0	420	0	16	240	0	4	0	983	937
O-S	1000	115	999	0	16	1000	0	1000	0	60
A-S	30.98	1.39	29.18	–	1.06	194.59	–	193.12	–	1
	$n = 1000$					$n = 2000$				
U-S	0	1000	0	3	1	0	1000	0	0	0
T-S	0	0	0	996	949	0	0	0	999	950
O-S	1000	0	1000	1	50	1000	0	1000	1	50
A-S	394.99	–	394.86	1	1	795	–	795	1	1

TABLE 4. Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (I) and (ii) based on 1,000 replications. Here U-S , T-S, O-S and A-S stand for number of times a selection method under-specified the true

	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
$\alpha = 0.2, c = 0.2$										
			$n = 100$					$n = 500$		
U-S	0	120	0	25	0	0	543	0	0	0
T-S	182	880	204	974	969	572	457	663	999	955
O-S	818	0	796	1	31	428	0	337	1	45
A-S	2.10	–	2.02	1	1.10	1.35	–	1.23	1	1.04
			$n = 1000$					$n = 2000$		
U-S	0	975	0	0	0	0	1000	0	0	0
T-S	963	25	981	1000	963	999	0	1000	999	953
O-S	37	0	19	0	37	1	0	0	1	47
A-S	1	–	1	–	1	1	–	–	1	1.02
$\alpha = 0.2, c = 0.4$										
			$n = 100$					$n = 500$		
U-S	0	895	0	219	24	0	1000	0	0	0
T-S	37	105	1	778	932	4	0	0	999	941
O-S	963	0	999	3	44	996	0	1000	1	59
A-S	3.68	–	7.11	1	1.07	6.60	–	46.14	1	1.02
			$n = 1000$					$n = 2000$		
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	35	0	0	1000	937	454	0	0	1000	941
O-S	965	0	1000	0	63	546	0	1000	0	59
A-S	3.90	–	95.25	–	1.05	1.46	–	194.82	–	1
$\alpha = 0.4, c = 0.2$										
			$n = 100$					$n = 500$		
U-S	0	61	0	331	43	0	153	0	0	0
T-S	0	894	5	666	897	0	847	0	999	969
O-S	1000	45	995	3	60	1000	0	1000	1	31
A-S	16.59	1.02	6.91	1	1.08	101.70	–	8.23	1	1
			$n = 1000$					$n = 2000$		
U-S	0	715	0	0	0	0	1000	0	0	0
T-S	0	285	44	1000	947	0	0	788	999	963
O-S	1000	0	956	0	53	1000	0	212	1	37
A-S	213.18	–	3.22	–	1.04	449.63	–	1.13	1	1
$\alpha = 0.4, c = 0.4$										
			$n = 100$					$n = 500$		
U-S	0	589	0	994	837	0	1000	0	2	0
T-S	0	319	0	6	150	0	0	0	995	950
O-S	1000	92	1000	0	13	1000	0	1000	3	50
A-S	30.97	1.34	29.24	–	1.38	194.60	–	193.15	1	1.02
			$n = 1000$					$n = 2000$		
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	1000	950	0	0	0	994	945
O-S	1000	0	1000	0	50	1000	0	1000	6	55
A-S	394.99	–	394.83	–	1.06	795	–	795	1	1.05

TABLE 5. Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (I) and (iii) based on 1,000 replications. Here U-S , T-S, O-S and A-S stand for number of times a selection method under-specified the true

$\alpha = 0.2, c = 0.2$									
	$n = 100$					$n = 500$			
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	77	0	971	215	0	648	0	0	0
T-S	41	847	46	29	753	0	352	1	1000	967
O-S	959	76	954	0	32	1000	0	999	0	33
A-S	3.11	1.04	3.01	–	1	7.40	–	6.68	–	1.03
	$n = 1000$					$n = 2000$			
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	993	0	0	0	0	1000	0	0	0
T-S	4	7	6	1000	944	177	0	267	997	942
O-S	996	0	994	0	56	823	0	733	3	58
A-S	5.51	–	4.65	–	1.04	2.11	–	1.77	1	1.02

$\alpha = 0.2, c = 0.4$									
	$n = 100$					$n = 500$			
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	925	0	991	622	0	1000	0	2	0
T-S	2	74	0	9	361	0	0	0	997	934
O-S	998	1	1000	0	17	1000	0	1000	1	66
A-S	4.74	1	7	–	1.06	16.40	–	45.88	1	1
	$n = 1000$					$n = 2000$			
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	1000	938	0	0	0	999	924
O-S	1000	0	1000	0	62	1000	0	1000	1	76
A-S	18.74	–	95.36	–	1.03	13.77	–	194.65	1	1.01

$\alpha = 0.4, c = 0.2$									
	$n = 100$					$n = 500$			
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	4	0	999	597	0	7	0	7	0
T-S	0	348	0	1	386	0	993	0	993	961
O-S	1000	648	1000	0	17	1000	0	1000	0	39
A-S	15.31	1.67	9.23	–	1	94.64	–	28.61	–	1
	$n = 1000$					$n = 2000$			
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	27	0	0	0	0	354	0	0	0
T-S	0	973	0	1000	939	0	646	0	1000	956
O-S	1000	0	1000	0	61	1000	0	1000	0	44
A-S	198.92	–	31.12	–	1.03	417.94	–	21.32	–	1.02

$\alpha = 0.4, c = 0.4$									
	$n = 100$					$n = 500$			
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	243	0	1000	898	0	988	0	188	2
T-S	0	233	0	0	97	0	12	0	812	965
O-S	1000	524	1000	0	5	1000	0	1000	0	33
A-S	28.58	1.99	26.89	–	1.20	191.27	–	186.32	–	1.03
	$n = 1000$					$n = 2000$			
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	1000	942	0	0	0	1000	928
O-S	1000	0	1000	0	58	1000	0	1000	0	72
A-S	394.33	–	391.74	–	1.05	794.99	–	794.74	–	1.04

TABLE 6. Selection times of the KOO methods with AIC, BIC, C_p thresholds and boot-strap methods under Settings (II) and (iv) based on 1,000 replications. Here U-S, T-S, O-S and A-S stand for number of times a selection method under-specified the true

$\alpha = 0.2, c = 0.2$									
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	63	0	66	0	0	732	0	0	0
T-S	127	936	144	933	963	169	268	245	998	969
O-S	873	1	856	1	37	831	0	755	2	31
A-S	2.38	1	2.27	1	1	2.16	–	1.92	1	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	688	0	779	994	949	994	0	998	1000	937
O-S	312	0	221	6	51	6	0	2	0	63
A-S	1.16	–	1.11	1	1.02	1	–	1	–	1.05

$\alpha = 0.2, c = 0.4$									
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	973	0	615	100	0	1000	0	0	0
T-S	21	27	2	385	863	3	0	0	999	948
O-S	979	0	998	0	37	997	0	1000	1	52
A-S	3.86	–	6.98	–	1.05	8.99	–	46.12	1	1.02
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	2	0	0	999	952	132	0	0	998	947
O-S	998	0	1000	1	48	868	0	1000	2	53
A-S	6.64	–	95.24	1	1.02	2.30	–	193.88	1	1.06

$\alpha = 0.4, c = 0.2$									
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1	0	384	17	0	1	0	0	0
T-S	0	841	1	615	939	0	999	0	1000	944
O-S	1000	158	999	1	44	1000	0	1000	0	56
A-S	15.86	1.09	7.40	1	1.07	98.76	–	13.10	–	1.02
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	6	0	0	0	0	374	0	0	0
T-S	0	994	0	1000	954	0	626	197	999	954
O-S	1000	0	1000	0	46	1000	0	803	1	46
A-S	208.66	–	7.73	–	1.02	438.75	–	2.01	1	1.02

$\alpha = 0.4, c = 0.4$									
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	263	0	969	684	0	994	0	1	0
T-S	0	540	0	31	302	0	6	0	999	948
O-S	1000	197	1000	0	14	1000	0	1000	0	52
A-S	30.52	1.38	28.73	–	1.07	194.24	–	192.21	–	1.04
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	999	953	0	0	0	999	950
O-S	1000	0	1000	1	47	1000	0	1000	1	50
A-S	394.98	–	394.64	1	1.02	795	–	795	1	1.02

TABLE 7. Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (II) and (v) based on 1,000 replications. Here U-S, T-S, O-S and A-S stand for number of times a selection method under-specified the true

$\alpha = 0.2, c = 0.2$

	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
			$n = 100$					$n = 500$		
U-S	0	61	0	0	0	0	788	0	0	0
T-S	483	939	534	999	974	960	212	977	1000	956
O-S	517	0	466	1	26	40	0	23	0	44
A-S	1.49	–	1.44	1	1.08	1	–	1	–	1
			$n = 1000$					$n = 2000$		
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	1000	0	1000	1000	937	1000	0	1000	999	951
O-S	0	0	0	0	63	0	0	0	1	49
A-S	–	–	–	–	1.10	–	–	–	1	1.02

$\alpha = 0.2, c = 0.4$

	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
			$n = 100$					$n = 500$		
U-S	0	987	0	161	13	0	1000	0	0	0
T-S	60	13	0	838	951	28	0	0	1000	958
O-S	940	0	1000	1	36	972	0	1000	0	42
A-S	3.01	–	6.83	1	1.06	3.94	–	45.52	–	1.05
			$n = 1000$					$n = 2000$		
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	237	0	0	999	949	870	0	0	999	954
O-S	763	0	1000	1	51	130	0	1000	1	46
A-S	1.92	–	94.99	1	1	1.05	–	193.56	1	1

$\alpha = 0.4, c = 0.2$

	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
			$n = 100$					$n = 500$		
U-S	0	0	0	3	0	0	0	0	0	0
T-S	0	995	17	996	941	0	1000	82	998	943
O-S	1000	5	983	1	59	1000	0	918	2	57
A-S	16.54	1	4.84	1	1.07	102.68	–	2.96	1	1.05
			$n = 1000$					$n = 2000$		
U-S	0	3	0	0	0	0	345	0	0	0
T-S	0	997	655	1000	946	0	655	992	1000	951
O-S	1000	0	345	0	54	1000	0	8	0	49
A-S	217.91	–	1.17	–	1.07	465.89	–	1	–	1.06

$\alpha = 0.4, c = 0.4$

	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
			$n = 100$					$n = 500$		
U-S	0	237	0	961	464	0	997	0	0	0
T-S	0	658	0	39	505	0	3	0	1000	952
O-S	1000	105	1000	0	31	1000	0	1000	0	48
A-S	32.02	1.32	30.24	–	1.13	194.90	–	194.26	–	1.02
			$n = 1000$					$n = 2000$		
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	999	937	0	0	0	999	939
O-S	1000	0	1000	1	63	1000	0	1000	1	61
A-S	395	–	394.98	1	1.02	795	–	795	1	1.05

TABLE 8. Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (II) and (vi) based on 1,000 replications. Here U-S, T-S, O-S and A-S stand for number of times a selection method under-specified the true

which indicate that ξ_l tends to 0 in probability with order of $o(n^{-t})$ for any $t > 0$. Analogously, for application later, together with the condition that $p^{-1/2}\mathbf{b}$ is bounded in Euclidean norm, we conclude that for $2 \le \ell \le \log(n)$,

(23) $$\max\{\mathbb{E}|p^{-1}\mathbf{b}_l'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l|^\ell, \mathbb{E}|p^{-1}\mathbf{a}_l\mathbf{E}_l'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l|^\ell\} = O(p^{\ell/2-1}\eta_n^{\ell-2})$$

and

(24) $$\max\{\mathbb{E}|p^{-2}\mathbf{b}_l'\mathbf{\Psi}_l\mathbf{b}_l|^\ell, \mathbb{E}|p^{-2}\mathbf{a}_1\mathbf{E}_l'\mathbf{\Psi}_l\mathbf{b}_l|^\ell, \mathbb{E}|p^{-2}\mathbf{a}_1\mathbf{E}_l'\mathbf{\Psi}_l\mathbf{E}_l\mathbf{a}_1|^\ell\} = O(p^{\ell-2}\eta_n^{2\ell-4})$$

where

$$\mathbf{\Psi}_l = \mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1} - \mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}.$$

As we only need to prove the weak convergence conclusion and $\beta_1^{tr} \to (1 - \alpha - c)/c > 0$, thus throughout the proofs, we can safely assume $\|\mathbf{M}^{-1}\|$, $\|\mathbf{M}_l^{-1}\|$ and $|1/\beta_l|$ are all bounded for large n.

B.1. **Proof of Theorem 2.1.** Theorem 2.1 can be obtained from Proposition 3.1 in Bai et al. (2022) with letting $z \downarrow 0$ directly. That is, for any non-random vectors $\mathbf{r}_1$, $\mathbf{r}_2$ $\mathbf{r}_3$ and $\mathbf{r}_4$ with suitable dimensions and bounded in Euclidean norm, under conditions in Theorem 2.1, we have that for any $t > 0$ and $\varepsilon > 0$,

(25) $$\mathbb{P}\left(\left|\mathbf{r}_1'\mathbf{M}^{-1}\mathbf{r}_2 - \frac{c_n\mathbf{r}_1'\mathbf{r}_2}{1 - c_n - \alpha_n}\right| \ge \varepsilon\right) = o(n^{-t}),$$

(26) $$\mathbb{P}\left(\left|\frac{1}{\sqrt{p}}\mathbf{r}_1'\mathbf{M}^{-1}\mathbf{E}'\mathbf{r}_3\right| \ge \varepsilon\right) = o(n^{-t}),$$

and

(27) $$\mathbb{P}\left(\left|\frac{1}{p}\mathbf{r}_3'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{r}_4 - \frac{c_n\mathbf{r}_3'\mathbf{r}_4}{1 - c_n - \alpha_n} + \frac{c_n^2\mathbf{r}_3'\mathbf{Q}\mathbf{r}_4}{(1 - c_n - \alpha_n)(1 - \alpha_n)}\right| \ge \varepsilon\right) = o(n^{-t}).$$

Then the proof of Theorem 2.1 is complete.

B.2. **Proof of Theorem 2.2.** For simple presentation, in the following we assume $\{1, \ldots, q\} \subset [k]\backslash\mathbf{j}_*$ and $j_i = i$. To prove Theorem 2.2, it is sufficient to show that for any non-null vector $\mathbf{h} = (h_1, \ldots, h_q)'$, $\sqrt{p}[(\mathcal{K}_1, \ldots, \mathcal{K}_q)\mathbf{h} - \frac{c_n}{1 - c_n - \alpha_n}\mathbf{1}_q'\mathbf{h}]$ converges weakly to a normal distribution with mean zero and variance $\frac{c^2}{(1 - \alpha_n - c_n)^2}[\frac{2(1 - \alpha_n)}{(1 - \alpha_n - c_n)}\mathbf{h}'(\mathcal{A}_q'\mathcal{A}_q)^2\mathbf{h} + \tau\mathbf{h}'(\mathcal{A}_q \circ \mathcal{A}_q)'(\mathcal{A}_q \circ \mathcal{A}_q)\mathbf{h}]$, where $\mathcal{A}_q = (\mathbf{a}_1, \ldots, \mathbf{a}_q)$.

We split the proof of this theorem into two parts. First, we show the asymptotic normality of the sequence of random variables

$$\mathcal{M}_1^{(n)} := \sqrt{p}[(\mathcal{K}_1, \ldots, \mathcal{K}_q)\mathbf{h} - \mathbb{E}(\mathcal{K}_1, \ldots, \mathcal{K}_q)\mathbf{h}].$$

Second, we prove the non-random sequence

$$\mathcal{M}_2^{(n)} = \sqrt{p}[\mathbb{E}(\mathcal{K}_1, \ldots, \mathcal{K}_q)\mathbf{h} - \frac{c_n\mathbf{1}_q'\mathbf{h}}{1 - c_n - \alpha_n}]$$

tends to zero. Note that for notational simplicity the superscript $^{(n)}$ in $\mathcal{M}_1^{(n)}$ and $\mathcal{M}_2^{(n)}$ are suppressed in the sequel.

We start to consider $\mathcal{M}_1$. Let $\mathcal{H} = Diag(h_1, \ldots, h_q)$. It follows that

$$\mathcal{M}_1 = p^{-1/2} \sum_{l=1}^{p} (\mathbb{E}_l - \mathbb{E}_{l-1}) \text{tr}(\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathcal{A}_q\mathcal{H}\mathcal{A}_q')$$

$$= p^{-1/2} \sum_{l=1}^{p} (\mathbb{E}_l - \mathbb{E}_{l-1}) \text{tr}[(\mathbf{E}\mathbf{M}^{-1}\mathbf{E}' - \mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l')\mathcal{A}_q\mathcal{H}\mathcal{A}_q'].$$

By the inversion formula of block matrix, we obtain

$$\mathbf{E}\mathbf{M}^{-1}\mathbf{E}' - \mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l' = \frac{1}{\beta_l p^2}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'$$

(28)
$$- \frac{1}{\beta_l p}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l' - \frac{1}{\beta_l p}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l' + \frac{\mathbf{e}_l\mathbf{e}_l'}{\beta_l}.$$

Then, by the equation (21), we can rewrite $\mathcal{M}_1$ as

$$\mathcal{M}_1 = \frac{1}{p^{1/2}\beta_1^{tr}} \sum_{l=1}^{p} \mathbb{E}_l(\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \text{tr}\mathbf{\Gamma}_l) - \frac{1}{p^{1/2}(\beta_1^{tr})^2} \sum_{l=1}^{p} \mathbb{E}_l(\xi_l\text{tr}\mathbf{\Gamma}_l) + \mathcal{M}_{10},$$

where

$$\mathbf{\Gamma}_l = p^{-2}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathcal{A}_q\mathcal{H}\mathcal{A}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q} - p^{-1}\mathcal{A}_q\mathcal{H}\mathcal{A}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}$$

$$- p^{-1}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l\mathcal{A}_q\mathcal{H}\mathcal{A}_q' + \mathcal{A}_q\mathcal{H}\mathcal{A}_q'$$

and

$$\mathcal{M}_{10} = -\sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l(\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \text{tr}\mathbf{\Gamma}_l)}{p^{1/2}(\beta_1^{tr})^2} + \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l^2\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l}{p^{1/2}\beta_l(\beta_1^{tr})^2}.$$

It follows from (22) that

$$\mathbb{E}|\frac{1}{p^{1/2}} \sum_{l=1}^{p} \mathbb{E}_l(\xi_l\text{tr}\mathbf{\Gamma}_l)|^2 = \frac{1}{p} \sum_{l=1}^{p} \mathbb{E}|\mathbb{E}_l(\xi_l\text{tr}\mathbf{\Gamma}_l)|^2 = O(p^{-1}).$$

By (23), (24) and the BurkHölder's inequality (see Lemma B.2) we have that $\mathcal{M}_{10} = o_p(1)$. Applying Lemma 2.7 in Bai & Silverstein (1998) (see Lemma B.3), we have that

$$\mathbb{E}|\mathbb{E}_l(\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \text{tr}\mathbf{\Gamma}_l)|^4 \le \mathbb{E}|\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \text{tr}\mathbf{\Gamma}_l|^4 = O(p\eta_n^4)$$

which verifies the condition (ii) in Lemma B.1. Thus, what we need is to obtain the limit of

$$\frac{1}{p(\beta_1^{tr})^2} \sum_{l=1}^{p} \mathbb{E}_{l-1}\{\mathbb{E}_l[\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \text{tr}\mathbf{\Gamma}_l]\}^2.$$

By Lemma B.5, we have that

$$\mathbb{E}_{l-1}\{\mathbb{E}_l[\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \text{tr}\mathbf{\Gamma}_l]\}^2 = 2\mathbb{E}_{l-1}\text{tr}(\mathbb{E}_l\mathbf{\Gamma}_l\mathbb{E}_l\mathbf{\Gamma}_l) + \tau\mathbb{E}_{l-1}\text{tr}(\mathbb{E}_l\mathbf{\Gamma}_l \circ \mathbb{E}_l\mathbf{\Gamma}_l),$$

where $\circ$ stands for the Hadamard product. Notice that

(29)
$$\text{tr}(\mathbb{E}_l\mathbf{\Gamma}_l\mathbb{E}_l\mathbf{\Gamma}_l) = p^{-4}\text{tr}(\mathbb{E}_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathcal{A}_q\mathcal{H}\mathcal{A}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2$$

(30)
$$+ p^{-2}2\text{tr}[\mathbb{E}_l(\mathcal{A}_q\mathcal{H}\mathcal{A}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l\mathcal{A}_q\mathcal{H}\mathcal{A}_q')]$$

$$+ \text{tr}(\mathcal{A}_q\mathcal{H}\mathcal{A}_q')^2.$$

Let $\widetilde{\mathbf{E}}_l$ be $\mathbf{E}_l$ by replacing $\{\mathbf{e}_{l+1}, \ldots, \mathbf{e}_p\}$ with $\{\tilde{\mathbf{e}}_{l+1}, \ldots, \tilde{\mathbf{e}}_p\}$, where $\{\tilde{\mathbf{e}}_i\}$ are i.i.d. copies of $\mathbf{e}_1$. We define $\widetilde{\mathbf{M}}_l = \frac{1}{p}\widetilde{\mathbf{E}}_l'\mathbf{Q}\widetilde{\mathbf{E}}_l$, correspondingly. As $\mathcal{H}$ is a diagonal matrix, thus we have that

$$\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathcal{A}_q\mathcal{H}\mathcal{A}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2$$
$$=\mathbb{E}_l\mathrm{tr}[\mathcal{H}\mathcal{A}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\widetilde{\Xi}_l\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathcal{A}_q]$$
$$=\mathbb{E}_l\sum_{i=1}^{q}h_i\mathbf{a}_i'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\widetilde{\Xi}_l\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_i,$$

where $\widetilde{\Xi}_l = \mathbf{Q}\widetilde{\mathbf{E}}_l\widetilde{\mathbf{M}}_l^{-1}\widetilde{\mathbf{E}}_l'\mathcal{A}_q\mathcal{H}\mathcal{A}_q'\widetilde{\mathbf{E}}_l\widetilde{\mathbf{M}}_l^{-1}\widetilde{\mathbf{E}}_l'\mathbf{Q}$. By applying the inversion formula of block matrix to $\mathbf{M}_l$, similar to (28), we have that

$$\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l' - \mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}' = \frac{1}{\beta_{lp}p^2}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\mathbf{Q}\mathbf{e}_p\mathbf{e}_p'\mathbf{Q}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'$$

(31)
$$- \frac{1}{\beta_{lp}p}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\mathbf{Q}\mathbf{e}_p\mathbf{e}_p' - \frac{1}{\beta_{lp}p}\mathbf{e}_p\mathbf{e}_p'\mathbf{Q}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}' + \frac{\mathbf{e}_p\mathbf{e}_p'}{\beta_{lp}},$$

where $\mathbf{E}_{li}$ is the $n \times (i-2)$ submatrix of $\mathbf{E}$ with the columns $\{\mathbf{e}_l, \mathbf{e}_i, \ldots, \mathbf{e}_p\}$ removed, $\mathbf{M}_{li} = \frac{1}{p}\mathbf{E}_{li}'\mathbf{Q}\mathbf{E}_{li}$ and

(32)
$$\beta_{li} = \frac{1}{p}\mathbf{e}_i'\mathbf{Q}\mathbf{e}_i - \frac{1}{p^2}\mathbf{e}_i'\mathbf{Q}\mathbf{E}_{li}\mathbf{M}_{li}^{-1}\mathbf{E}_{li}'\mathbf{Q}\mathbf{e}_i.$$

Denote

$$\beta_i^{tr} = \mathrm{tr}[\frac{1}{p}\mathbf{Q} - \frac{1}{p^2}\mathbf{Q}\mathbf{E}_{li}\mathbf{M}_{li}^{-1}\mathbf{E}_{li}'\mathbf{Q}] = \frac{n-k-p+i}{p}$$

and

$$\xi_{li} = \beta_{li} - \beta_i^{tr}.$$

We can easily check that the orders of (22)–(24) hold for replacing the subscripts l by li. Thus, analogous to the above discussion, we have that

$$\mathbb{E}_l\mathbf{a}_i'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\widetilde{\Xi}_l\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_i$$
$$=\mathbb{E}_l[\mathbf{a}_i'\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\widetilde{\Xi}_l\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\mathbf{a}_i] + o_p(p^3)$$
$$=\mathbb{E}_l[\mathbf{a}_i'\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\widetilde{\Xi}_{lp}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\mathbf{a}_i] + o_p(p^3),$$

where $\widetilde{\Xi}_{lp}$ is defined by removing b_p and $\tilde{\mathbf{e}}_p$ from $\widetilde{\Xi}_l$. We then repeat the procedure that remove b_i, $\mathbf{e}_i$ and $\tilde{\mathbf{e}}_i$, $i = l+1, \ldots, p-1$ from Ξ_{lp} and $\widetilde{\Xi}_{lp}$, respectively. Then applying Proposition 3.1 in (Bai et al., 2022), we finally obtain that

$$(37) = \sum_{i,j}^{q}h_ih_jp^{-2}[\mathbf{a}_i'\mathbf{E}_{l(l-1)}\mathbf{M}_{l(l-1)}^{-1}\mathbf{E}_{l(l-1)}'\mathbf{a}_j)]^2 + o_p(1)$$

(33)
$$=\frac{(l-1)^2}{(n-k-l+1)^2}\mathbf{h}'(\mathcal{A}_q'\mathcal{A}_q)^2\mathbf{h} + o_p(1).$$

Analogously, we have that

$$(30) = p^{-2} \sum_{i=1}^{q} h_i h_j \mathbb{E}_l(\mathbf{a}_i' \mathbf{E}_l \mathbf{M}_l^{-1} \mathbf{E}_l' \mathbf{Q}) \mathbb{E}_l(\mathbf{Q} \mathbf{E}_l \mathbf{M}_l^{-1} \mathbf{E}_l \mathbf{a}_i)$$

$$= \frac{(l-1)\mathbf{h}'(\mathcal{A}_q' \mathcal{A}_q)^2 \mathbf{h}}{n-k-l+1} + o_p(1),$$

which together with (33) and the fact that $\text{tr}(\mathcal{A}_q \mathcal{H} \mathcal{A}_q')^2 = \mathbf{h}'(\mathcal{A}_q' \mathcal{A}_q)^2 \mathbf{h}$ implies

$$\frac{1}{p} \sum_{l=1}^{p} \mathbb{E}_{l-1} \text{tr}(\mathbb{E}_l \mathbf{\Gamma}_l \mathbb{E}_l \mathbf{\Gamma}_l) = \frac{\mathbf{h}'(\mathcal{A}_q' \mathcal{A}_q)^2 \mathbf{h}}{p} \sum_{l=1}^{p} \left(\frac{l-1}{n-k-l+1} + 1 \right)^2 + o_p(1)$$

$$= \mathbf{h}'(\mathcal{A}_q' \mathcal{A}_q)^2 \mathbf{h} \frac{1-\alpha_n}{1-\alpha_n-c_n} + o_p(1).$$

We now turn to prove the term $\mathbb{E}_{l-1} \text{tr}(\mathbb{E}_l \mathbf{\Gamma}_l \circ \mathbb{E}_l \mathbf{\Gamma}_l) = o_p(1)$. Let $\mathbf{u}_j$ be an n-dimensional column vector with the j-th element being 1 and 0 otherwise. Then we have that

$$\mathbb{E}(\mathbb{E}_{l-1} \text{tr}(\mathbb{E}_l \mathbf{\Gamma}_l \circ \mathbb{E}_l \mathbf{\Gamma}_l)) = \sum_{j=1}^{n} \mathbb{E}(\mathbf{u}_j' \mathbb{E}_l \mathbf{\Gamma}_l \mathbf{u}_j)^2$$

$$= \sum_{j=1}^{n} (\mathbb{E}\mathbf{u}_j' \mathbf{\Gamma}_l \mathbf{u}_j)^2 + \sum_{j=1}^{n} \mathbb{E}(\mathbb{E}_l \mathbf{u}_j' \mathbf{\Gamma}_l \mathbf{u}_j - \mathbb{E}\mathbf{u}_j' \mathbf{\Gamma}_l \mathbf{u}_j)^2.$$

By BurkHölder's inequality, we have that

$$\mathbb{E}(\mathbb{E}_l \mathbf{u}_j' \mathbf{\Gamma}_l \mathbf{u}_j - \mathbb{E}\mathbf{u}_j' \mathbf{\Gamma}_l \mathbf{u}_j)^2 \leq \sum_{s \neq l}^{p} \mathbb{E}(\mathbf{u}_j' \mathbf{\Gamma}_l \mathbf{u}_j - \mathbf{u}_j' \mathbf{\Gamma}_{l \cdot s} \mathbf{u}_j)^2,$$

where $\mathbf{\Gamma}_{l \cdot s}$ is the submatrix of $\mathbf{\Gamma}_l$ with $\mathbf{e}_s$ removed. Applying the inversion formula of block matrix (31) again, we have that

$$(34) \qquad \mathbf{Q} \mathbf{E}_l \mathbf{M}_l^{-1} \mathbf{E}_l' = \mathbf{Q} \mathbf{E}_{l \cdot s} \mathbf{M}_{l \cdot s}^{-1} \mathbf{E}_{l \cdot s}' + \frac{\mathbf{Q} \mathbf{e}_s \mathbf{e}_s'}{\beta_{l \cdot s}}$$

$$+ \frac{1}{\beta_{l \cdot s} p^2} \mathbf{Q} \mathbf{E}_{l \cdot s} \mathbf{M}_{l \cdot s}^{-1} \mathbf{E}_{l \cdot s}' \mathbf{Q} \mathbf{e}_s \mathbf{e}_s' \mathbf{Q} \mathbf{E}_{l \cdot s} \mathbf{M}_{l \cdot s}^{-1} \mathbf{E}_{l \cdot s}'$$

$$- \frac{1}{\beta_{l \cdot s} p} \mathbf{Q} \mathbf{E}_{l \cdot s} \mathbf{M}_{l \cdot s}^{-1} \mathbf{E}_{l \cdot s}' \mathbf{Q} \mathbf{e}_s \mathbf{e}_s' - \frac{1}{\beta_{l \cdot s} p} \mathbf{Q} \mathbf{e}_s \mathbf{e}_s' \mathbf{Q} \mathbf{E}_{l \cdot s} \mathbf{M}_{l \cdot s}^{-1} \mathbf{E}_{l \cdot s}'$$

$$:= \mathcal{U}_{ls0} + \mathcal{U}_{ls1} + \mathcal{U}_{ls2} - \mathcal{U}_{ls3} - \mathcal{U}_{ls4}$$

and

$$\mathbb{E}(\mathbf{u}_j' \mathbf{\Gamma}_l \mathbf{u}_j - \mathbf{u}_j' \mathbf{\Gamma}_{ls} \mathbf{u}_j)^2$$

$$= \mathbb{E}\{p^{-4} \sum_{i=1}^{q} h_i [\mathbf{u}_j'(\mathcal{U}_{ls1} + \mathcal{U}_{ls2} - \mathcal{U}_{ls3} - \mathcal{U}_{ls4}) \mathbf{a}_i]^2$$

$$+ 2p^{-2} \sum_{i=1}^{q} h_i \mathbf{u}_j' \mathcal{U}_{ls0} \mathbf{a}_i \mathbf{u}_j'(\mathcal{U}_{ls1} + \mathcal{U}_{ls2} - \mathcal{U}_{ls3} - \mathcal{U}_{ls4}) \mathbf{a}_i$$

$$- 2p^{-1} \sum_{i=1}^{q} h_i \mathbf{u}_j' \mathbf{a}_i \mathbf{u}_j'(\mathcal{U}_{ls1} + \mathcal{U}_{ls2} - \mathcal{U}_{ls3} - \mathcal{U}_{ls4}) \mathbf{a}_i\}^2.$$

We first consider $\mathbb{E}(\mathbf{u}_j'\mathcal{U}_{ls1}\mathbf{a}_i\mathbf{a}_i'\mathcal{U}_{ls1}'\mathbf{u}_j)^2$. Notice that

$$
\begin{aligned}
&\mathbb{E}(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s)^2 \\
=&\mathbb{E}[(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s - \mathbf{u}_j'\mathbf{Q}\mathbf{u}_j + \mathbf{u}_j'\mathbf{Q}\mathbf{u}_j)(\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s - 1 + 1)]^2 \\
=&\mathbb{E}[(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s - \mathbf{u}_j'\mathbf{Q}\mathbf{u}_j)(\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s - 1) + \mathbf{u}_j'\mathbf{Q}\mathbf{u}_j(\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s - 1) \\
&+ (\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s - \mathbf{u}_j'\mathbf{Q}\mathbf{u}_j) + \mathbf{u}_j'\mathbf{Q}\mathbf{u}_j]^2.
\end{aligned}
$$

From Lemma B.4 we have that for $\ell \geq 2$,

$$
\mathbb{E}(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s - \mathbf{u}_j'\mathbf{Q}\mathbf{u}_j)^\ell = O(n^{\ell-1}\eta_n^{2\ell-4})
$$

and

$$
\mathbb{E}(\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s - 1)^\ell = O(n^{\ell-1}\eta_n^{2\ell-4}).
$$

Then, together with the fact that $\frac{1}{\beta_{l\cdot s}}$ and $\mathbf{u}_j'\mathbf{Q}\mathbf{u}_j$ are both bounded, and the c_r-inequality, we obtain

$$
\mathbb{E}(\mathbf{u}_j'\mathcal{U}_{ls1}\mathbf{a}_i\mathbf{a}_i'\mathcal{U}_{ls1}'\mathbf{u}_j)^2 = O(n^3\eta_n^{-4}).
$$

Next, we consider the term $\mathbb{E}(\mathbf{u}_j'\mathcal{U}_{ls0}\mathbf{a}_i\mathbf{a}_i'\mathcal{U}_{ls1}'\mathbf{u}_j)^2$. It follows $\mathbf{a}_i'\mathbf{Q} = \mathbf{0}$ that

$$
\begin{aligned}
&\mathbb{E}(\mathbf{u}_j'\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j)^2 \\
=&\mathbb{E}(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s - \mathbf{a}_i'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\mathbf{a}_i)^2 = O(1).
\end{aligned}
$$

As other terms are analogous, thus by combining the above argument, we conclude that $\sum_{j=1}^n \mathbb{E}(\mathbb{E}_l\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j - \mathbb{E}\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j)^2 = o(1)$.

For $\sum_{j=1}^n (\mathbb{E}\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j)^2$, it follows from the assumption that $\{e_{ij}\}$ are i.i.d.,

$$
\begin{aligned}
\sum_{j=1}^n (\mathbb{E}\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j)^2 &= \sum_{j=1}^n (n^{-1}p^{-1}\mathbb{E}\text{tr}\mathcal{H}\mathcal{A}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathcal{A}_q + \mathbf{u}_j'\mathcal{A}_q\mathcal{H}\mathcal{A}_q'\mathbf{u}_j)^2 \\
&= \sum_{j=1}^n (\mathbf{u}_j'\mathcal{A}_q\mathcal{H}\mathcal{A}_q'\mathbf{u}_j)^2 + O(n^{-1}) \\
&= \mathbf{h}'(\mathcal{A}_q \circ \mathcal{A}_q)'(\mathcal{A}_q \circ \mathcal{A}_q)\mathbf{h} + O(n^{-1}).
\end{aligned}
$$

Here we use a result similar to (27), that is

$$
\mathbb{E}\frac{1}{p}\mathbf{r}_3'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{r}_4 - \frac{c_n\mathbf{r}_3'\mathbf{r}_4}{1 - c_n - \alpha_n} + \frac{c_n^2\mathbf{r}_3'\mathbf{Q}\mathbf{r}_4}{(1 - c_n - \alpha_n)(1 - \alpha_n)} \to 0,
$$

and the proof can be found in the proof of Proposition 3.1 in Bai et al. (2022). Then we conclude that

$$
\sum_{l=1}^p \mathbb{E}_{l-1}\text{tr}(\mathbb{E}_l\mathbf{\Gamma}_l \circ \mathbb{E}_l\mathbf{\Gamma}_l) = \mathbf{h}'(\mathcal{A}_q \circ \mathcal{A}_q)'(\mathcal{A}_q \circ \mathcal{A}_q)\mathbf{h} + o_p(1).
$$

Next, we will prove that the non-random sequence

$$
\mathcal{M}_2 = \mathcal{M}_2^{(n)} = o(1).
$$

Write $\mathbf{M}^{-1} = (M^{ij})$. Without loss of generality, we only need to prove $p^{-1}\mathbb{E}\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{a}_1 - \frac{c_n}{1-c_n-\alpha_n} = o(p^{-1/2})$. Because the entries of $\mathbf{E}$ are i.i.d., we have

$$p^{-1}\mathbb{E}\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{a}_1 = p^{-1}\sum_{i,j=1}^{p}\mathbb{E}\mathbf{a}_1'\mathbf{e}_i M^{ij}\mathbf{e}_j'\mathbf{a}_1$$

(35)
$$=\mathbb{E}\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1 M^{11} + (p-1)\mathbb{E}\mathbf{a}_1'\mathbf{e}_1 M^{12}\mathbf{e}_2'\mathbf{a}_1.$$

From the inverse matrix formula, we know that

$$M^{11} = \frac{1}{\beta_1} = \frac{1}{\beta_1^{tr}} - \frac{\xi_1}{(\beta_1^{tr})^2} + \frac{\xi_1^2}{\beta_1(\beta_1^{tr})^2}.$$

and

(36)
$$M^{12} = \frac{\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1^{tr}} - \frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p(\beta_1^{tr})^2} + \frac{\xi_1^2\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1(\beta_1^{tr})^2}.$$

Then it follows from (21), (22) and the Hölder's inequality that

$$\mathbb{E}\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1 M^{11} - \frac{c_n}{1-c_n-\alpha_n} = \mathbb{E}\frac{\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1\xi_1^2}{\beta_1(\beta_1^{tr})^2} = o(p^{-1/2}).$$

Moreover, substituting (36) into the second term of (35), we have three terms. The first one is

$$\mathbb{E}\frac{\mathbf{e}_2'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1^{tr}} = \mathbb{E}\frac{\mathbf{e}_2'\mathbf{a}_1\mathbf{a}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1^{tr}} = 0,$$

because of $\mathbf{a}_1'\mathbf{Q} = \mathbf{0}$. Applying the inversion formula to $\mathbf{M}_1^{-1}$ again, we obtain that

$$\mathbb{E}\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{p}$$
$$=\mathbb{E}\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{e}_2\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{\beta_{1\cdot2}p} - \mathbb{E}\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_{1\cdot2}\mathbf{M}_{1\cdot2}^{-1}\mathbf{E}_{1\cdot2}'\mathbf{Q}\mathbf{e}_2\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{\beta_{1\cdot2}p^2}$$
$$=\mathbb{E}\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{e}_2\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{\beta_{1\cdot2}p}.$$

Rewrite $1/\beta_{1\cdot2}$ as

$$\frac{1}{\beta_{1\cdot2}} = \frac{1}{\beta_2^{tr}} - \frac{\xi_{1\cdot2}}{\beta_{1\cdot2}\beta_2^{tr}}$$

and by the the fact that $\mathbf{Q}\mathbf{a}_1 = \mathbf{0}$, we have that

$$\mathbb{E}\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{e}_2\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{\beta_{1\cdot2}} = -\mathbb{E}\frac{\xi_{1\cdot2}\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{e}_2\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{\beta_2^{tr}\beta_{1\cdot2}} = o(p^{-1/2}).$$

Therefore, by combining the above results, we conclude that

$$\mathcal{M}_2 = o(1),$$

and we complete the proof of the theorem.

B.3. **Proof of Theorem 2.3.** Note that

$$\text{tr}[(\mathbf{Y}'\mathbf{Q}\mathbf{Y} - (n-k)\mathbf{I}) \circ (\mathbf{Y}'\mathbf{Q}\mathbf{Y} - (n-k)\mathbf{I})] = \sum_{i=1}^{p}(\mathbf{e}_i'\mathbf{Q}\mathbf{e}_i - (n-k))^2$$

and

$$\mathbb{E}(\mathbf{e}_i'\mathbf{Q}\mathbf{e}_i - (n-k))^2 = 2(n-k) + \tau\text{tr}(\mathbf{Q} \circ \mathbf{Q}).$$

Thus by the definition of $\hat{\tau}$ and $\{\mathbf{e}_i\}$ are i.i.d., we have $\mathbb{E}\hat{\tau} = \tau$. Next we will show that

$$\mathbb{E}(\hat{\tau} - \tau)^2 \to 0.$$

It follows from (20) and Lemma B.3 that

$$\mathbb{E}(\hat{\tau} - \tau)^2 = \mathbb{E}(\hat{\tau} - \mathbb{E}\hat{\tau})^2$$

$$=p^{-1}\mathbb{E}[(\mathbf{e}_1'\mathbf{Q}\mathbf{e}_1 - (n-k))^2 - \mathbb{E}(\mathbf{e}_1'\mathbf{Q}\mathbf{e}_1 - (n-k))]^2/\text{tr}^2(\mathbf{Q} \circ \mathbf{Q})$$

$$=\{p^{-1}\mathbb{E}(\mathbf{e}_1'\mathbf{Q}\mathbf{e}_1 - (n-k))^4 - p^{-1}[\mathbb{E}(\mathbf{e}_1'\mathbf{Q}\mathbf{e}_1 - (n-k))]^2\}/\text{tr}^2(\mathbf{Q} \circ \mathbf{Q})$$

$$\leq Kp^{-1}[((n-k)^2 + n^2(n-k)\eta_n^4 + (n-k)^2 + \tau^2\text{tr}^2(\mathbf{Q} \circ \mathbf{Q})]/\text{tr}^2(\mathbf{Q} \circ \mathbf{Q}),$$

where K is a positive constant. By c_r-inequality, we have that $\text{tr}(\mathbf{Q} \circ \mathbf{Q}) \geq n^{-1}(n-k)^2$, which together with condition (C1) implies $\mathbb{E}(\hat{\tau} - \tau)^2 \to 0$. Then we complete the proof of this theorem.

B.4. **Proof of Theorem 2.4.** For simple presentation, in the following we assume $\{1\} \subset \mathbf{j}_*$ and let $j = 1$. Then by the notation $\mathbf{b} = \mathbf{\Sigma}^{-1/2}\mathbf{\Theta}_*\mathbf{X}_*'\mathbf{a}_1$, $\mathcal{K}_1 = p^{-1}(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b})$. Note that the proof procedure of Theorem 2.4 is the same as that of Theorem 2.2. And the difference is that Theorem 2.4 requires the consideration of linear combinations of three different forms of random variables, namely $\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{a}_1$, $\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{b}$ and $\mathbf{b}'\mathbf{M}^{-1}\mathbf{b}$. As the asymptotic normality of $\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{a}_1$ is proved in last subsection, in the sequel we only focus on the other two terms and their correlations.

Analogously, we split the proof of this theorem into two parts. It is worthy noting that next we may use the same notation as in the proof of Theorem 2.2, but they represent a little different content. First, we show the asymptotic normality of the sequence of random variables

$$\mathcal{M}_3 := \sqrt{p}[p^{-1}(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}) - \mathbb{E}p^{-1}(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b})].$$

Second, we prove the non-random sequence

$$\mathcal{M}_4 = \sqrt{p}[\mathbb{E}p^{-1}(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}) - \frac{c_n(1 + \delta_1)}{1 - c_n - \alpha_n}]$$

tends to zero. It follows that

$$\mathcal{M}_1 = p^{-1/2}\sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b})$$

$$= p^{-1/2}\sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})[(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}) - (\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)].$$

By the inversion formula of block matrix, we obtain

$$(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}) - (\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)$$

$$= \frac{1}{\beta_l p^2}(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)$$

$$- \frac{2}{\beta_l p}(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l(\mathbf{e}_l'\mathbf{a}_1 + b_l) + \frac{(\mathbf{e}_l'\mathbf{a}_1 + b_l)^2}{\beta_l}.$$

Then, by the equation (21), we can rewrite $\mathcal{M}_3$ as

$$\mathcal{M}_3 = \frac{1}{p^{1/2}\beta_1^{tr}} \sum_{l=1}^{p} \mathbb{E}_l[\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \text{tr}\mathbf{\Gamma}_l + 2\mathbf{e}_l'\boldsymbol{\gamma}_l] + \mathcal{M}_{30},$$

where

$$\begin{aligned}
\mathbf{\Gamma}_l =& p^{-2}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q} \\
& - p^{-1}\mathbf{a}_1(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q} - p^{-1}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)\mathbf{a}_1' + \mathbf{a}_1\mathbf{a}_1',
\end{aligned}$$

$$\boldsymbol{\gamma}_l = -p^{-1}b_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1) + \mathbf{a}_1 b_l$$

and

$$\begin{aligned}
\mathcal{M}_{30} =& -\sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)}{p^{5/2}\beta_l\beta_1^{tr}} \\
& + 2\sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l(\mathbf{e}_l'\mathbf{a}_1 + b_l)}{p^{3/2}\beta_l\beta_1^{tr}} \\
& - \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l(\mathbf{e}_l'\mathbf{a}_1 + b_l)^2}{p^{1/2}\beta_l\beta_1^{tr}} \\
:=& -\mathcal{M}_{301} + 2\mathcal{M}_{302} - \mathcal{M}_{303}.
\end{aligned}$$

Next we will prove $\mathcal{M}_{10} = o_p(1)$. Substitute (21) into $\mathcal{M}_{101}$, $\mathcal{M}_{102}$ and $\mathcal{M}_{103}$ respectively, we then have that

$$\begin{aligned}
\mathcal{M}_{301} =& \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{\Psi}_l(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)}{p^{5/2}(\beta_1^{tr})^2} + \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)}{p^{3/2}(\beta_1^{tr})^2} \\
& - \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l^2(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)}{p^{5/2}\beta_l(\beta_1^{tr})^2}, \\
\mathcal{M}_{302} =& \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l(\mathbf{e}_l'\mathbf{a}_1 + b_l)}{p^{3/2}(\beta_1^{tr})^2} \\
& - \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l^2(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l(\mathbf{e}_l'\mathbf{a}_1 + b_l)}{p^{3/2}\beta_l(\beta_1^{tr})^2}, \\
\mathcal{M}_{303} =& \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l(\mathbf{e}_l'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_l - 1 + 2b_l\mathbf{a}_1'\mathbf{e}_l + b_l^2 + 1)}{p^{1/2}(\beta_1^{tr})^2} - \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l^2(\mathbf{e}_l'\mathbf{a}_1 + b_l)^2}{p^{1/2}\beta_l(\beta_1^{tr})^2}.
\end{aligned}$$

These together with (23), (24) and the BurkHölder's inequality (see Lemma B.2) implies that $\mathcal{M}_{10} = o_p(1)$. Note that here we used the fact $\mathbf{Q}\mathbf{a}_1 = \mathbf{0}$.

Applying Lemma B.3, we have that

$$\mathbb{E}|\mathbb{E}_l(\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \text{tr}\mathbf{\Gamma}_l)|^4 \le \mathbb{E}|\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \text{tr}\mathbf{\Gamma}_l|^4 = O(p\eta_n^4)$$

and

$$\mathbb{E}|\mathbb{E}_l(\mathbf{e}_l'\boldsymbol{\gamma}_l)|^4 \le \mathbb{E}|\mathbf{e}_l'\boldsymbol{\gamma}_l\boldsymbol{\gamma}_l^*\mathbf{e}_l|^2 = O(p^{-2}b_l^4),$$

which verify the condition (ii) in Lemma B.1. Thus, what we need is to obtain the limit of

$$\frac{1}{p(\beta_1^{tr})^2} \sum_{l=1}^{p} \mathbb{E}_{l-1}\{\mathbb{E}_l[\mathbf{e}_l' \mathbf{\Gamma}_l \mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l + 2\mathbf{e}_l'\boldsymbol{\gamma}_l]\}^2.$$

By Lemma B.5, we have that

$$\mathbb{E}_{l-1}\{\mathbb{E}_l[\mathbf{e}_l' \mathbf{\Gamma}_l \mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l + 2\mathbf{e}_l'\boldsymbol{\gamma}_l]\}^2$$
$$=2\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l\mathbb{E}_l\mathbf{\Gamma}_l) + 4\mathbb{E}_{l-1}(\mathbb{E}_l\boldsymbol{\gamma}_l\mathbb{E}_l\boldsymbol{\gamma}_l')$$
$$+ \tau\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l \circ \mathbb{E}_l\mathbf{\Gamma}_l) + 4\mathbb{E}e_{11}^3\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l \circ \mathbb{E}_l\boldsymbol{\gamma}_l\mathbf{1}').$$

Notice that

(37) $$\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l\mathbb{E}_l\mathbf{\Gamma}_l) = p^{-4}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2$$
$$+ p^{-2}2\mathbb{E}_l((\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)) + 1.$$

In the proof of Theorem 2.2, we have shown that

$$p^{-4}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_1\mathbf{a}_1\mathbb{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2$$
$$+ p^{-2}2\mathbb{E}_l(\mathbf{a}_1\mathbb{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_1) + 1$$
$$= \left(\frac{l-1}{n-k-l+1} + 1\right)^2 + o_p(1).$$

By the same procedure and the assumptions in Theorem 2.4, we can also have that

$$p^{-4}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}\mathbf{b}_l\mathbf{b}_l'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2 = \left(\frac{\sum_{i=1}^{l-1} b_i^2}{n-k-l+1}\right)^2 + o_p(1),$$

$$p^{-4}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}\mathbf{b}_l\mathbf{a}_1\mathbb{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2 = \frac{(l-1)\sum_{i=1}^{l-1} b_i^2}{n-k-l+1} + o_p(1),$$

$$p^{-2}\mathbb{E}_l(\mathbf{b}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_1) = o_p(1),$$

and

$$p^{-2}\mathbb{E}_l(\mathbf{b}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}\mathbf{b}_l) = \frac{\sum_{i=1}^{l-1} b_i^2}{n-k-l+1} + o_p(1),$$

which together with (37) and (33) implies

$$\frac{1}{p}\sum_{l=1}^{p}\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l\mathbb{E}_l\mathbf{\Gamma}_l) = \frac{1}{p}\sum_{l=1}^{p}\left(\frac{(l-1+\sum_{i=1}^{l-1} b_i^2)}{n-k-l+1} + 1\right)^2 + o_p(1).$$

For $\mathbb{E}_l\boldsymbol{\gamma}_l\mathbb{E}_l\boldsymbol{\gamma}_l'$, by the notation $\widetilde{\mathbf{M}}_l = \frac{1}{p}\widetilde{\mathbf{E}}_l'\mathbf{Q}\widetilde{\mathbf{E}}_l$, we have that

$$\mathbb{E}_l\boldsymbol{\gamma}_l'\mathbb{E}_l\boldsymbol{\gamma}_l$$
$$=\mathbb{E}_l[(p^{-1}b_l\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1) - \mathbf{a}_1 b_l)'((p^{-1}b_l\mathbf{Q}\mathbb{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1) - \mathbf{a}_1 b_l)]$$
$$=\mathbb{E}_l[p^{-2}b_l^2(\mathbf{b}_l + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\widetilde{\mathbf{E}}_l\widetilde{\mathbf{M}}_l^{-1}(\mathbf{b}_l + \widetilde{\mathbf{E}}_l'\mathbf{a}_1) + b_l^2].$$

Applying the inversion formula of block matrix (31) again, we obtain that

$$p^{-1} \sum_{l=1}^{p} \mathbb{E}_{l-1}(\mathbb{E}_l \boldsymbol{\gamma}_l' \mathbb{E}_l \boldsymbol{\gamma}_l) = p^{-1} \sum_{l=1}^{p} b_l^2 \left(\frac{(l-1+\sum_{i=1}^{l-1} b_i^2)}{n-k-l+1} + 1 \right) + o_p(1).$$

Then by applying Lemma B.6, we have that as $n \to \infty$,

$$\frac{1}{p} \sum_{l=1}^{p} \mathbb{E}_{l-1}(\mathrm{tr}(\mathbb{E}_l \boldsymbol{\Gamma}_l \mathbb{E}_l \boldsymbol{\Gamma}_l) + 2\mathbb{E}_l \boldsymbol{\gamma}_l \mathbb{E}_l \boldsymbol{\gamma}_l')$$

$$= \frac{1}{p} \sum_{l=1}^{p} \left(\left(\frac{l-1+\sum_{i=1}^{l-1} b_i^2}{n-k-l+1} + 1 \right)^2 + 2b_l^2 \left(\frac{l-1+\sum_{i=1}^{l-1} b_i^2}{n-k-l+1} + 1 \right) \right) + o_p(1)$$

$$= \frac{n}{p} \int_0^{c_n} \left[\left(\frac{t(1+\delta_1)}{1-\alpha_n - t} + 1 \right)^2 + 2\delta_1 \left(\frac{t(1+\delta_1)}{1-\alpha_n - t} + 1 \right) \right] dt + o_p(1)$$

$$= \frac{(1-\alpha_n)(1+2\delta_1) + c_n \delta_1^2}{1-\alpha_n - c_n} + o_p(1).$$

We now turn to the term $\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l \boldsymbol{\Gamma}_l \circ \mathbb{E}_l \boldsymbol{\Gamma}_l)$. By the notation that $\mathbf{u}_j$ is an n-dimensional column vector with the j-th element being 1 and 0 otherwise and repeating the same argument in the proof of Theorem 2.2, we can obtain that

$$\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l \boldsymbol{\Gamma}_l \circ \mathbb{E}_l \boldsymbol{\Gamma}_l) = \sum_{j=1}^{n} (\mathbb{E}\mathbf{u}_j' \boldsymbol{\Gamma}_l \mathbf{u}_j)^2 + o_p(1).$$

As $\{e_{ij}\}$ are i.i.d., thus from the assumptions of this theorem, we have that

$$\sum_{j=1}^{n} (\mathbb{E}\mathbf{u}_j' \boldsymbol{\Gamma}_l \mathbf{u}_j)^2$$

$$= \sum_{j=1}^{n} (n^{-1}p^{-1}\mathbb{E}(\mathbf{b}_l + \mathbf{E}_l' \mathbf{a}_1)' \mathbf{M}_l^{-1} (\mathbf{b}_l + \mathbf{E}_l' \mathbf{a}_1)$$

$$\qquad - 2p^{-1}\mathbf{u}_j' \mathbf{a}_1 \mathbb{E}(\mathbf{b}_l + \mathbf{E}_l' \mathbf{a}_1)' \mathbf{M}_l^{-1} \mathbf{E}_l' \mathbf{Q}\mathbf{u}_j + (\mathbf{u}_j' \mathbf{a}_1)^2]^2$$

$$= o(1).$$

Then we conclude that

$$\frac{1}{p} \sum_{l=1}^{p} \mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l \boldsymbol{\Gamma}_l \circ \mathbb{E}_l \boldsymbol{\Gamma}_l) = o_p(1).$$

Next, we will prove that the non-random sequence

$$\mathcal{M}_4 = o(1).$$

By the notation $\mathbf{M}^{-1} = (M^{ij})$ and $\{e_{ij}\}$ are i.i.d., we have that

$$\mathbb{E}(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}) = \sum_{i,j=1}^{p} \mathbb{E}(b_i + \mathbf{a}_1'\mathbf{e}_i)M^{ij}(b_j + \mathbf{a}_1'\mathbf{e}_j)$$

$$=\mathbf{b}'\mathbf{b}\mathbb{E}M^{11} + p\mathbb{E}\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1 M^{11} + 2\mathbb{E}\mathbf{a}_1'\mathbf{e}_1 M^{11}\sum_{i=1}^{p} b_i$$

$$+ \sum_{i \neq j}^{p} b_i b_j \mathbb{E}M^{12} + p(p-1)\mathbb{E}\mathbf{a}_1'\mathbf{e}_1 M^{12}\mathbf{e}_2'\mathbf{a}_1 + 2(p-1)\mathbb{E}\mathbf{a}_1'\mathbf{e}_1 M^{12}\sum_{i=1}^{p} b_i.$$

From the inverse matrix formula, we know that

$$M^{11} = \frac{1}{\beta_1} = \frac{1}{\beta_1^{tr}} - \frac{\xi_1}{(\beta_1^{tr})^2} + \frac{\xi_1^2}{\beta_1(\beta_1^{tr})^2}.$$

and

$$M^{12} = \frac{\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1^{tr}} - \frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p(\beta_1^{tr})^2} + \frac{\xi_1^2\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1(\beta_1^{tr})^2}.$$

Then it follows from (21), (22) and the Hölder's inequality that

$$p^{-1}\mathbf{b}'\mathbf{b}\mathbb{E}M^{11} - \frac{c_n\delta_1}{1 - c_n - \alpha_n} = o(p^{-1/2})$$

and

$$\mathbb{E}\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1 M^{11} - \frac{c_n}{1 - c_n - \alpha_n} = o(p^{-1/2}).$$

By the facts that

$$|\mathbb{E}\mathbf{a}_1'\mathbf{e}_1\xi_1| \leq \sqrt{\mathbb{E}|\xi_1|^2} = O(p^{-1/2})$$

and

$$|\sum_{i=1}^{p} b_i| = O(p^{1/2}),$$

we can obtain that

$$\mathbb{E}\mathbf{a}_1'\mathbf{e}_1 M^{11}\sum_{i=1}^{p} b_i = O(1).$$

It follows from

$$|p^{-1}\sum_{i \neq j}^{p} b_i b_j| = O(1), \quad |\mathbb{E}\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1| = O(1)$$

and the Hölder's inequality, we have that

$$\sum_{i \neq j}^{p} b_i b_j \mathbb{E}M^{12} = o(p^{-1/2}).$$

Therefore, similar to the proof of Theorem 2.2, we conclude that

$$\mathcal{M}_2 = o(1),$$

and we complete the proof of this theorem.

B.5. **Some useful lemmas.**

Lemma B.1 (Theorem 35.12 of Billingsley (1995)). *Suppose that for each n, $Y_{n1}, Y_{n2}, \ldots, Y_{nr_n}$ is a real martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_{nj}\}$ having second moments. If as $n \to \infty$, for each $\varepsilon > 0$,*

(i) $\sum_{j=1}^{r_n} \mathbb{E}\left(Y_{nj}^2 | \mathcal{F}_{n,j-1}\right) \xrightarrow{p} \sigma^2$, *where σ^2 is a positive constant;*

(ii) $\sum_{j=1}^{r_n} \mathbb{E}\left(Y_{nj}^2 I_{(|Y_{nj}| \geq \varepsilon)}\right) \to 0$,

then we have that

$$\sum_{j=1}^{r_n} Y_{nj} \xrightarrow{\mathcal{D}} N\left(0, \sigma^2\right).$$

Lemma B.2 (Burkholder (1971)). *Let $\{Y_k\}$ be a martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $\ell > 1$,*

$$\mathbb{E}\left|\sum X_k\right|^\ell \leq K_\ell \left(\mathbb{E}\left(\sum \mathbb{E}(|X_k|^2 | \mathcal{F}_{k-1})\right)^{\ell/2} + \sum \mathbb{E}|X_k|^\ell\right).$$

Lemma B.3 (Lemma 2.7 of Bai & Silverstein (1998)). *For $\mathbf{e} = (e_1, \ldots, e_n)'$ i.i.d. standardized entries, $\mathbf{A}$ a $n \times n$ matrix, we have, for any $\ell \geq 2$*

$$\mathbb{E}\left|\mathbf{e}'\mathbf{A}\mathbf{e} - \mathrm{tr}\mathbf{A}\right|^\ell \leq K_\ell \left(\left(\mathbb{E}|e_1|^4 \mathrm{tr}\mathbf{A}\mathbf{A}'\right)^{\ell/2} + \mathbb{E}|e_1|^{2\ell} \mathrm{tr}(\mathbf{A}\mathbf{A}')^{\ell/2}\right).$$

Lemma B.4 (Lemma 7.2 of Bai & Yao (2005)). *Let $\mathbf{e} = (e_1, \ldots, e_n)'$ be a random n-vector with i.i.d. standardized entries. Suppose $E|e_i|^4 < \infty$ and $|e_i| \leq \eta_n\sqrt{n}$ with $\eta_n \to 0$ slowly. Assume that $\mathbf{A}$ is a symmetric matrix of order n bounded in norm by M. Then, for any given $2 \leqslant \ell \leqslant b\log\left(n\eta_n^2\right)$ with some $b > 1$, there exists a constant K such that*

$$\mathbb{E}\left|\mathbf{e}'\mathbf{A}\mathbf{e} - \mathrm{tr}(\mathbf{A})\right|^\ell \leq n^\ell \left(n\eta_n^4\right)^{-1} \left(MK\eta_n^2\right)^\ell.$$

Lemma B.5. *Let $\mathbf{B}$ and $\mathbf{C}$ be $n \times n$ matrices. Let $\mathbf{d}$ be a n-vector. Let $\mathbf{e} = (e_1, \ldots, e_n)'$ be a random n-vector with i.i.d. standardized entries. Let $\tau := \mathbb{E}e_i^4 - 3$. Then, we have that*

$$\mathbb{E}\left\{\left(\mathbf{e}'\mathbf{B}\mathbf{e} - \mathrm{tr}\,\mathbf{B}\right)\left(\mathbf{e}'\mathbf{C}\mathbf{e} - \mathrm{tr}\,\mathbf{C}\right)\right\} = \mathrm{tr}(\mathbf{B}\mathbf{C}) + \mathrm{tr}\left(\mathbf{B}\mathbf{C}'\right) + \tau \sum_{i=1}^n b_{ii}c_{ii}$$

and

$$\mathbb{E}(\mathbf{e}'\mathbf{B}\mathbf{e}\mathbf{e}'\mathbf{d}) = \mathbb{E}e_1^3 \sum_{i=1}^n b_{ii}d_i.$$

Lemma B.6 (Lemma 3.1 of Bai & Pan (2012)). *Let $\left\{\mathbf{d}_n = (d_1, \ldots, d_n)'\right\}$ be a sequence of unit vectors with $\max_{k \leq n}|d_k| \to 0$. There is a permutation m of $\{1, \ldots, n\}$ given by*

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ m(1) & m(2) & \cdots & m(n) \end{pmatrix}$$

such that $\mathbf{d}_f = \left(d_{f(1)}, \ldots, d_{f(n)}\right)$ and F_{nm} tends to a uniform distribution over the interval $(0,1)$, where F_{nm} is a distribution function defined by

$$F_{nm}(t) = \sum_{i \leq nt} \left|d_{f(i)}\right|^2.$$

REFERENCES

ANZANELLO, M. J. & FOGLIATTO, F. S. (2014). A review of recent variable selection methods in industrial and chemometrics applications. *European J. of Industrial Engineering* **8**, 619.

BAI, Z., CHOI, K. P., FUJIKOSHI, Y. & HU, J. (2022). Asymptotics of AIC, BIC and Cp model selection rules in high-dimensional regression. *Bernoulli* **28**, 2375–2403.

BAI, Z., FANG, Z. & LIANG, Y.-C. (2014). *Spectral Theory of Large Dimensional Random Matrices and Its Applications to Wireless Communications and Finance Statistics: Random Matrix Theory and Its Applications*. Singapore: World Scientific Publishing Co. Pte Ltd.

BAI, Z., FUJIKOSHI, Y. & HU, J. (2018a). Strong consistency of the AIC, BIC, C_p and KOO methods in high-dimensional multivariate linear regression. https://arxiv.org/abs/1810.12609v3.

BAI, Z. D., CHOI, K. P. & FUJIKOSHI, Y. (2018b). Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT. *The Annals of Statistics* **46**, 2985–3013.

BAI, Z. D. & PAN, G. M. (2012). Limiting Behavior of Eigenvectors of Large Wigner Matrices. *Journal of Statistical Physics* **146**, 519–549.

BAI, Z. D. & SILVERSTEIN, J. W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *The Annals of Probability* **26**, 316–345.

BAI, Z. D. & SILVERSTEIN, J. W. (1999). Exact separation of eigenvalues of large dimensional sample covariance matrices. *The Annals of Probability* **27**, 1536–1555.

BAI, Z. D. & SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability* **32**, 553–605.

BAI, Z. D. & YAO, J. F. (2005). On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli* **11**, 1059–1092.

BILLINGSLEY, P. (1995). *Probability and Measure*. John Wiley&Sons, New York.

BREIMAN, L. & FRIEDMAN, J. H. (1997). Predicting Multivariate Responses in Multiple Linear Regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**, 3–54.

BURKHOLDER, D. L. (1971). Distribution function inequalities for martingales. *The Annals of Probability* **1**, 19–42.

CAI, T. T., GUO, Z. & XIA, Y. (2023). Statistical Inference and Large-scale Multiple Testing for High-dimensional Regression Models.

CHEN, L. & HUANG, J. Z. (2012). Sparse Reduced-Rank Regression for Simultaneous Dimension Reduction and Variable Selection. *Journal of the American Statistical Association* **107**, 1533–1545.

CHUN, H. & KELEŞ, S. (2010). Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Variable Selection. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **72**, 3–25.

DESBOULETS, L. D. D. (2018). A Review on Variable Selection in Regression Analysis. *Econometrics* **6**, 45.

FAN, J. & LI, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association* **96**, 1348–1360.

FAN, J. & LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 849–911.

FAN, J. & LV, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica* **20**, 101–148.

FUJIKOSHI, Y. (2022). High-dimensional consistencies of KOO methods in multivariate regression model and discriminant analysis. *Journal of Multivariate Analysis* **188**, 104860.

FUJIKOSHI, Y. & SAKURAI, T. (2019). Consistency of test-based method for selection of variables in high-dimensional two-group discriminant analysis. *Japanese Journal of Statistics and Data Science* **2**,

155–171.

FUJIKOSHI, Y., SAKURAI, T. & YAMADA, T. (2023). High-dimensional consistencies of KOO methods for selecting graphical models , Hiroshima Statistical Research Group, TR–22–6.

FUJIKOSHI, Y., ULYANOV, V. V. & SHIMIZU, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations*. Wiley Series in Probability and Statistics. Wiley: Hoboken, N.J.

HEINZE, G., WALLISCH, C. & DUNKLER, D. (2018). Variable selection - A review and recommendations for the practicing statistician. *Biometrical Journal* **60**, 431–449.

HUANG, J., BREHENY, P. & MA, S. (2012). A Selective Review of Group Selection in High-Dimensional Models. *Statistical Science* **27**, 481–499.

KE, Z. T., JIN, J. & FAN, J. (2014). Covariate Assisted Screening and Estimation. *The Annals of Statistics* **42**, 2202–2242.

KONG, Y., LI, D., FAN, Y. & LV, J. (2017). Interaction pursuit in high-dimensional multi-response regression via distance correlation. *The Annals of Statistics* **45**, 897–922.

LEE, E. R., CHO, J. & YU, K. (2019). A systematic review on model selection in high-dimensional regression. *Journal of the Korean Statistical Society* **48**, 1–12.

LI, R., ZHONG, W. & ZHU, L. (2012). Feature Screening via Distance Correlation Learning. *Journal of the American Statistical Association* **107**, 1129–1139.

LIU, W. & LUO, S. (2014). Hypothesis Testing for High-dimensional Regression Models , Technical report.

NAKAGAWA, T., WATANABE, H. & HYODO, M. (2021). Kick-one-out-based variable selection method for Euclidean distance-based classifier in high-dimensional settings. *Journal of Multivariate Analysis* **184**, 104756.

NISHII, R., BAI, Z. & KRISHNAIAH, P. R. (1988). Strong consistency of the information criterion for model selection in multivariate analysis. *Hiroshima Mathematical Journal* **18**, 451–462.

ODA, R., SUZUKI, Y., YANAGIHARA, H. & FUJIKOSHI, Y. (2020). A consistent variable selection method in high-dimensional canonical discriminant analysis. *Journal of Multivariate Analysis* **175**, 104561.

ODA, R. & YANAGIHARA, H. (2020). A fast and consistent variable selection method for high-dimensional multivariate linear regression with a large number of explanatory variables. *Electronic Journal of Statistics* **14**, 1386–1412.

SAKURAI, T. & FUJIKOSHI, Y. (2020). Exploring Consistencies of Information Criterion and Test-Based Criterion for High-Dimensional Multivariate Regression Models Under Three Covariance Structures. In *Recent Developments in Multivariate and Random Matrix Analysis: Festschrift in Honour of Dietrich von Rosen*, T. Holgersson & M. Singull, eds. Cham: Springer International Publishing, pp. 313–334.

SHAO, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica* **7**, 221–242.

SIMILÄ, T. & TIKKA, J. (2007). Input selection and shrinkage in multiresponse linear regression. *Computational Statistics & Data Analysis* **52**, 406–422.

SKAGERBERG, B., MACGREGOR, J. F. & KIPARISSIDES, C. (1992). Multivariate data analysis applied to low-density polyethylene reactors. *Chemometrics and Intelligent Laboratory Systems* **14**, 341–356.

SPELLMAN, P. T., SHERLOCK, G., ZHANG, M. Q., IYER, V. R., ANDERS, K., EISEN, M. B., BROWN, P. O., BOTSTEIN, D. & FUTCHER, B. (1998). Comprehensive Identification of Cell Cycle–regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. *Molecular Biology of the Cell* **9**, 3273–3297.

TIBSHIRANI, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267–288.

WANG, L., CHEN, G. & LI, H. (2007). Group SCAD regression analysis for microarray time course gene expression data. *Bioinformatics* **23**, 1486–1494.

XIA, Y. (2017). Testing and support recovery of multiple high-dimensional covariance matrices with false discovery rate control. *TEST* **26**, 782–801.

XIA, Y., CAI, T. & CAI, T. T. (2018). Two-Sample Tests for High-Dimensional Linear Regression with an Application to Detecting Interactions. *Statistica Sinica* **28**, 63–92.

ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **38**, 894–942.

ZOU, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* **101**, 1418–1429.