JUSTC

Article

http://justc.ustc.edu.cn

Received: May 20, 2024; Accepted: September 01, 2024

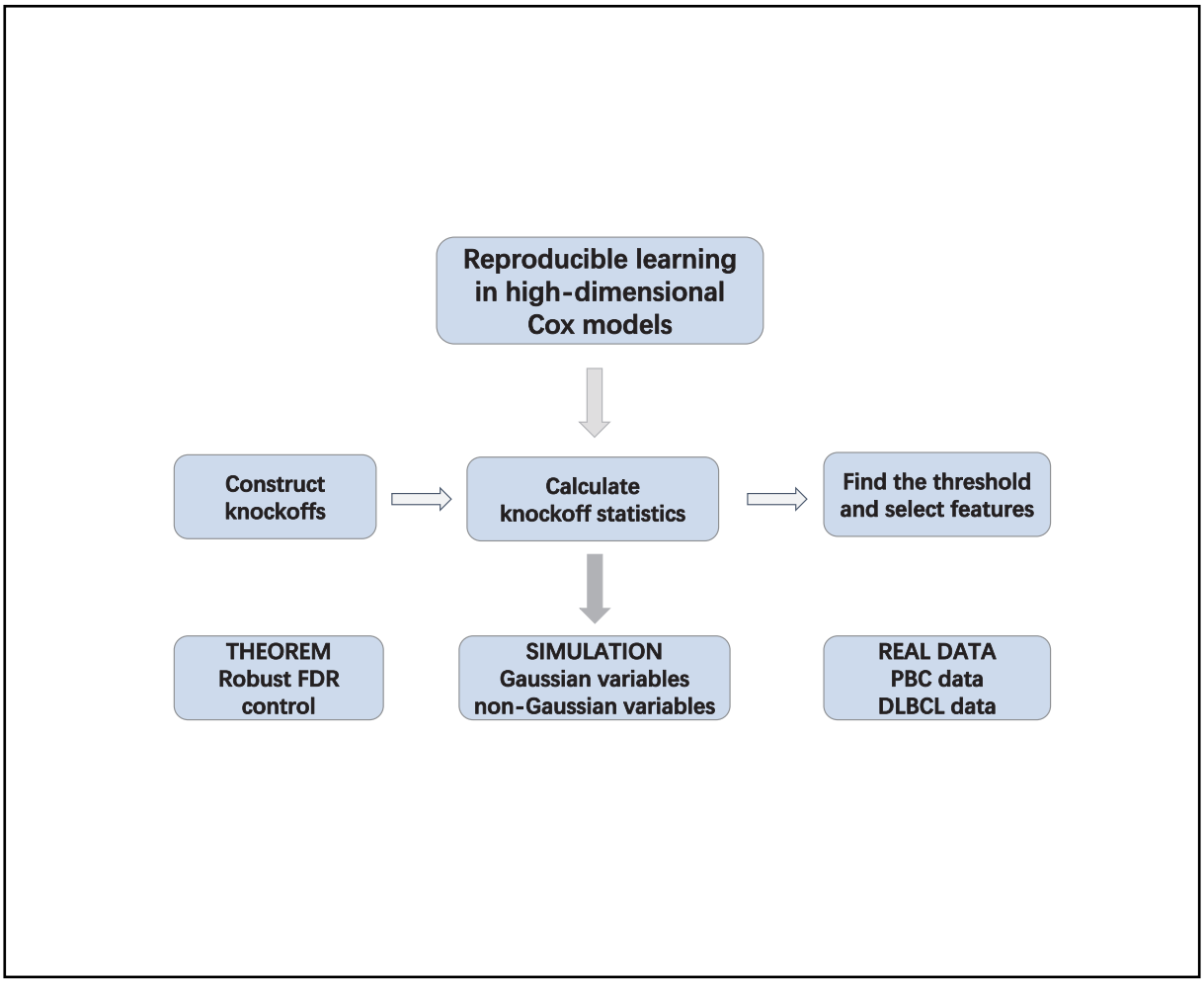
Reproducible learning in high-dimensional Cox models

Xin Zhou, Yan Dong, and Yang Li ✉

International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230601, China

✉Correspondence: Yang Li, E-mail: tjly@mail.ustc.edu.cn

Graphical abstract



Controlling the FDR in the high-dimensional Cox models.

Public summary

- We propose a robust knockoff method to control the FDR in high-dimensional Cox models.
- We demonstrate the robustness of FDR control for our method when the population precision matrix of predictors is unknown and replaced by some consistent estimates.
- Our numerical studies and real-data applications validate the effectiveness of the proposed knockoff procedure.

Citation: Zhou X, Dong Y, Li Y. Reproducible learning in high-dimensional Cox models. *JUSTC*, **2025**, 55(0): . CSTR: 32290.14.JUSTC-2024-0075. DOI: 10.52396/JUSTC-2024-0075

JUSTC

Article

http://justc.ustc.edu.cn

Received: May 20, 2024; Accepted: September 01, 2024

Reproducible learning in high-dimensional Cox models

Xin Zhou, Yan Dong, and Yang Li ✉

International Institute of Finance, School of Management, University of Science and Technology of China, Hefei 230601, China

✉Correspondence: Yang Li, E-mail: tjly@mail.ustc.edu.cn

Cite This: *JUSTC*, **2025**, 55(X): (11pp)

Read Online

Abstract: In survival analysis, the Cox proportional hazards model is one of the most popular models widely used in biomedical sciences and epidemiology. Despite significant advancements, how to ensure the reproducibility and robustness of feature selection in high-dimensional Cox models remains largely unexplored, especially when the population precision matrix of features is unknown. In this paper, we introduce a new feature selection procedure using knockoffs to facilitate false discovery rate (FDR) control in high-dimensional Cox models. We establish the robustness of FDR control for the proposed method when the population precision matrix is unknown and replaced by some consistent estimate. Through numerical studies and real-data applications, we demonstrate the effectiveness of our method in controlling FDR while ensuring statistical power.

Keywords: high dimensionality; Cox model; false discovery rate; robustness; feature selection

CLC number: O212 **Document code:** A

2020 Mathematics Subject Classification: Primary 62F05; Secondary 62N03

1 Introduction

Survival analysis is ubiquitous across various fields, including biomedical sciences, epidemiology, and finance[1–3]. Among survival analysis models, the Cox proportional hazards model[1] has gained increasing popularity because of its flexibility in accommodating diverse risk and censoring patterns. With the increasing size and complexity of datasets in areas such as genomics, effective feature selection has become crucial for identifying significant variables in high-dimensional Cox models. Over the past few decades, various regularization methods for feature selection have been extended from linear regression to Cox proportional hazards models; examples include Lasso[4], SCAD[5], and Dantzig selector[6], among others. Despite rapid progress, most existing studies on Cox models have mainly focused on statistical power. However, it is crucial to consider another key perspective in feature selection: reproducibility, ensuring that the results are consistent across different samples.

To ensure the reproducibility of statistical inference, Benjamini and Hochberg[7] proposed the Benjamini–Hochberg procedure, a method based on multiple testing, to control the false discovery rate (FDR). This procedure selects a feature only if its associated p-value is smaller than a certain threshold, where the FDR is defined as the expectation of the fraction of false discoveries among all discoveries. Due to the importance of FDR control, the Benjamini–Hochberg procedure has been generalized to various settings[8,9].

However, the aforementioned p-value based FDR control procedures can be invalid in some high-dimensional settings[10]. Therefore, Barber and Candès[11] suggested a new procedure called the knockoff filter to control the FDR of feature selection through control variables. This procedure was later extended to high-dimensional settings with the aid of feature screening and data splitting[12]. Moreover, Candès et al.[13] introduced the model-X knockoffs for FDR control under arbitrary dimensions and any dependence structure when the distribution of the covariates is given. The robustness of model-X knockoffs has been further demonstrated in Refs. [14–16]. Furthermore, the model-X knockoffs framework has been applied to some survival models to control the FDR[17,18]. However, both Dong et al.[17] and Li et al.[18] assumed that the feature distribution is known. When the population precision matrix of predictors is unknown and needs to be estimated, how to achieve robust FDR control for the high-dimensional Cox model is still largely unexplored.

Notably, existing robust methods for FDR control in the knockoff framework mainly focused on scenarios where the response is perfectly observed[14,15,19]. However, to explore the robust FDR control for high-dimensional Cox models, we need to address the challenges posed by censored data and the unknown precision matrix simultaneously. Even if we employ some consistent estimate of the precision matrix, the constructed knockoff variables will not satisfy the so-called pairwise exchangeability and will depend on the censored data, invalidating the key properties required to guarantee FDR control in Ref. [13]. To address this issue, by employing an independent and consistent estimate of the precision matrix, we can characterize the FDR as a function of covariates and their knockoff copies. This ensures that the FDR function depends on the precision matrix only through the covariates and their knockoff counterparts, demonstrating that the randomness of the censored data will not affect the robust FDR analysis. By extracting the precision matrix from the

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

knockoff matrix, we can directly characterize the relationship between the FDR and the precision matrix. This motivates us to utilize the stability argument to ensure that our knockoff procedure is stable under small perturbations of the precision matrix estimation.

In this study, we propose a robust knockoff method to control the FDR in high-dimensional Cox models. The new method combines the strengths of model-X knockoffs and penalized partial log-likelihood estimation. We demonstrate the robustness of FDR control for our method when the population precision matrix of predictors is unknown and replaced by some consistent estimates. Our numerical studies and real-data applications further validate the effectiveness of the proposed knockoff procedure.

The rest of this paper is organized as follows. Section 2 outlines the model setting, reviews the model-X knockoffs framework, and describes our proposed knockoff procedure. Theoretical properties are established in Section 3. Numerical studies and real-data applications, which illustrate the efficacy of our approach, are presented in Sections 4 and 5, respectively. We discuss some extensions and possible future work in Section 6. All proofs and technical details are relegated to Appendix.

2 Methodology

2.1 Model setting

We consider a sequence of independent and identically distributed (i.i.d.) samples $\{(\boldsymbol{x}_i, T_i)\}_{i=1}^n$ drawn from a population $(\boldsymbol{x}, T)$. Each sample consists of a p-dimensional covariate vector $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ip})^\top$, corresponding to the ith individual. In survival analysis, situations in which not all survival times are fully observable are common. In such cases, there is a sequence of right censoring times $\{C_i\}_{i=1}^n$. Consequently, the observed data are $(y_i, \boldsymbol{X}_i, \delta_i)$, where $y_i = \min(T_i, C_i)$ represents the observed event time, and $\delta_i = \mathbb{1}\{T_i \le C_i\}$ serves as the censoring indicator. The Cox proportional hazards model formalizes the hazard function for the ith individual as follows:

$$\lambda(t \mid \boldsymbol{x}_i) = \lambda_0(t)\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}_0), \tag{1}$$

where $\lambda_0(t)$ denotes the unknown baseline hazard function at time t and $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ is an unknown true coefficient vector. Then the maximum partial log-likelihood is defined as

$$L(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \delta_i \left\{ \boldsymbol{x}_i^\top \boldsymbol{\beta} - \log\Big(\frac{1}{n}\sum_{j=1}^n \mathbf{1}\{Y_j \ge Y_i\}\exp(\boldsymbol{x}_j^\top \boldsymbol{\beta})\Big)\right\}.$$

In high-dimensional settings where the number of covariates p can be significantly larger than the sample size n, it is commonly assumed that the true coefficient vector $\boldsymbol{\beta}_0$ has a sparse structure to ensure model identifiability. This implies that only a small subset of the components of $\boldsymbol{\beta}_0$ are nonzero. Let $\mathcal{S}_0 = \{j : \beta_{0,j} \ne 0\}$ denote the set of predictors that are truly relevant to the hazard function, with $\mathcal{S}_0^c$ representing its complement. To evaluate the effectiveness of a feature selection procedure that identifies a subset $\widehat{\mathcal{S}} \subset [p]$, we define the FDR of this procedure as follows:

$$\mathrm{FDR} := \mathbb{E}[\mathrm{FDP}], \quad \mathrm{FDP} := \frac{|\widehat{\mathcal{S}} \cap \mathcal{S}_0^c|}{|\widehat{\mathcal{S}}| \vee 1}, \tag{2}$$

where FDP is the false discovery proportion, $|\cdot|$ indicates the cardinality of a set, and the notation $a \vee b = \max\{a, b\}$. Our goal is to accurately identify the relevant variables within $\mathcal{S}_0$ while controlling the FDR.

2.2 A brief review of the model-X knockoffs framework

Before introducing our procedure, we briefly review the model-X knockoffs framework developed by Candès et al.[13] for completeness. The model-X knockoffs framework constructs the knockoff variables that mimic the dependency structure among the original features and are independent of the response conditional on the original features.

Definition 1.[13] Model-X knockoff features for a set of random features $\boldsymbol{x} = (x_1, \cdots, x_p)^\top$ are a new set of random features $\widetilde{\boldsymbol{x}} = (\widetilde{x}_1, \cdots, \widetilde{x}_p)^\top$ that satisfies two properties:

(i) For any subset $\mathcal{S} \subset [p]$, $[\boldsymbol{x}, \widetilde{\boldsymbol{x}}]_{\mathrm{swap}(\mathcal{S})} \overset{\mathrm{d}}{=} [\boldsymbol{x}, \widetilde{\boldsymbol{x}}]$, where swap$(\mathcal{S})$ means swapping components x_j and $\widetilde{x}_j$ for each $j \in \mathcal{S}$ and $\overset{\mathrm{d}}{=}$ denotes equal distribution;

(ii) $\widetilde{\boldsymbol{x}}$ is independent of the response conditional on $\boldsymbol{x}$.

When the covariate distribution is characterized by a Gaussian distribution $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{\Omega}_0^{-1})$ with $\boldsymbol{\Omega}_0$ the precision matrix of $\boldsymbol{x}$, the knockoff vector $\widetilde{\boldsymbol{x}}$ can be generated from the conditional distribution

$$\widetilde{\boldsymbol{x}} | \boldsymbol{x} \sim N(\mathbf{C}^{\boldsymbol{\Omega}_0}\boldsymbol{x},\ (\mathbf{B}^{\boldsymbol{\Omega}_0})^2), \tag{3}$$

where $\mathbf{C}^{\boldsymbol{\Omega}_0} = \mathbf{I}_p - \boldsymbol{\Omega}_0 \mathrm{diag}\{\boldsymbol{s}\}$, $\mathbf{B}^{\boldsymbol{\Omega}_0} = (2\mathrm{diag}\{\boldsymbol{s}\} - \mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{\Omega}_0\mathrm{diag}\{\boldsymbol{s}\})^{1/2}$, and $\mathrm{diag}\{\boldsymbol{s}\}$ is the diagonal matrix with components in the vector $\boldsymbol{s}$. The original covariate $\boldsymbol{x}$ and its knockoff copy $\widetilde{\boldsymbol{x}}$ satisfy the following joint distribution

$$\begin{pmatrix} \boldsymbol{x} \\ \widetilde{\boldsymbol{x}} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_0 - \mathrm{diag}\{\boldsymbol{s}\} \\ \boldsymbol{\Sigma}_0 - \mathrm{diag}\{\boldsymbol{s}\} & \boldsymbol{\Sigma}_0 \end{pmatrix} \right) \tag{4}$$

where $\boldsymbol{\Sigma}_0 = \boldsymbol{\Omega}_0^{-1}$ is the covariance matrix of covariates $\boldsymbol{x}$. In addition, the vector $\boldsymbol{s}$ can be selected via the procedures suggested by Candès et al.[13] to ensure that $\boldsymbol{\Sigma}_0 - 2^{-1}\mathrm{diag}\{\boldsymbol{s}\} > 0$ is positive definite.

Indeed, it has been a common practice to use the Gaussian distribution to construct knockoffs since it provides an explicit conditional distribution expression to sample the knockoff variables, as demonstrated in various studies[13, 15, 20]. Moreover, as shown in Ref. [14], even if the covariates do not follow the Gaussian distribution, the knockoff variables constructed via the methods outlined above can be good approximations of the knockoffs satisfying Definition 1.

The next step is to construct knockoff statistics $W_1, \cdots, W_p$ to evaluate the feature importance of the original variables. Denote by $\widetilde{\boldsymbol{X}} = (\widetilde{\boldsymbol{x}}_1, \cdots, \widetilde{\boldsymbol{x}}_n)^\top$ the constructed knockoff matrix of the covariate matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)^\top$. For each $j \in \{1, \cdots, p\}$, we write $W_j = w_j([\boldsymbol{X}, \widetilde{\boldsymbol{X}}], \boldsymbol{y})$ as a function of the augmented data matrix $[\boldsymbol{X}, \widetilde{\boldsymbol{X}}]$ and response $\boldsymbol{y} = (y_1, \cdots, y_n)^\top$ satisfying the sign-flip property

$$w_j([\boldsymbol{X}, \widetilde{\boldsymbol{X}}]_{\mathrm{swap}(S)}, \boldsymbol{y}) = \begin{cases} w_j([\boldsymbol{X}, \widetilde{\boldsymbol{X}}], \boldsymbol{y}), & j \notin S; \\ -w_j([\boldsymbol{X}, \widetilde{\boldsymbol{X}}], \boldsymbol{y}), & j \in S, \end{cases}$$

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

where $S \subset \{1,\cdots,p\}$ is any given subset. Intuitively, since all the knockoff features are noise features, we would expect the important features to demonstrate large positive values in their knockoff statistics, whereas the noise features should exhibit small magnitudes symmetrically distributed around zero.

The final step is to select features whose knockoff statistics exceed a data-driven threshold T, which results in the set of identified features

$$\widehat{\mathcal{S}} = \widehat{\mathcal{S}}(T) = \{1 \le j \le p : W_j \ge T\}.$$

Following Baeber and Candès[11] and Candès et al.[13], there are two thresholds for a prespecified FDR level $q \in (0,1)$. One is the knockoff threshold for controlling the modified FDR, which replaces $|\widehat{\mathcal{S}}|$ with $q^{-1}+|\widehat{\mathcal{S}}|$ in the denominator of the FDR in (2). It is defined as

$$T = \min\left\{t \in \mathcal{W} : \frac{\left|\{j : W_j \le -t\}\right|}{\left|\{j : W_j \ge t\}\right|} \le q\right\} \tag{5}$$

or $T = +\infty$ if the above set is empty, where $\mathcal{W} = \left\{\left|W_j\right| : 1 \le j \le p\right\} / \{0\}$. The other is the knockoff+ threshold

$$T_+ = \min\left\{t \in \mathcal{W} : \frac{1+\left|\{j : W_j \le -t\}\right|}{\left|\{j : W_j \ge t\}\right|} \le q\right\}. \tag{6}$$

It is clear that the knockoff+ threshold is more conservative since it aims at controlling the exact FDR. Candès et al.[13] has proved that the model-X knockoffs framework controls the FDR with the knockoff+ threshold, and controls the modified FDR with the knockoff threshold.

2.3 The robust knockoff filter in high-dimensional Cox models

We now introduce the proposed method for FDR control in high-dimensional Cox models. As discussed in Section 2.2, following Fan et al.[15] and Bai et al.[20], we assume that the rows of the design matrix $\boldsymbol{X}$ are sampled as i.i.d. copies from $N(\mathbf{0}, \boldsymbol{\Omega}_0^{-1})$.

Step 1. Constructing knockoffs. When the population precision matrix $\boldsymbol{\Omega}_0$ is known, we can generate knockoff variables by (3). However, since the population precision matrix $\boldsymbol{\Omega}_0$ is unknown in practice, we first need to construct its estimate, denoted as $\boldsymbol{\Omega}$. In fact, various methods can be employed to estimate the high-dimensional precision matrix $\boldsymbol{\Omega}_0$ such as the graphical Lasso[21], CLIME[22], and ISEE[23].

With the estimated precision matrix $\boldsymbol{\Omega}$, we propose generating the knockoff variables $\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}} = (\widetilde{\boldsymbol{x}}_1^{\boldsymbol{\Omega}},\cdots,\widetilde{\boldsymbol{x}}_n^{\boldsymbol{\Omega}})^\top \in \mathbb{R}^{n\times p}$ by

$$\widetilde{\boldsymbol{x}}_i^{\boldsymbol{\Omega}} | \boldsymbol{x}_i \sim N\left(\mathbf{C}^{\boldsymbol{\Omega}}\boldsymbol{x}_i,\ (\mathbf{B}^{\boldsymbol{\Omega}})^2\right), \tag{7}$$

where $\mathbf{C}^{\boldsymbol{\Omega}} = \mathbf{I}_p - \boldsymbol{\Omega}\mathrm{diag}\{\boldsymbol{s}\}$, and $\mathbf{B}^{\boldsymbol{\Omega}} = (2\mathrm{diag}\{\boldsymbol{s}\} - \mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{\Omega}\mathrm{diag}\{\boldsymbol{s}\})^{1/2}$. Here we use the superscript $\boldsymbol{\Omega}$ to emphasize the dependence of the knockoff variables on $\boldsymbol{\Omega}$. Formally, we can write the knockoff matrix $\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}$ as

$$\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}} = \boldsymbol{X}\mathbf{C}^{\boldsymbol{\Omega}} + \mathbf{Z}\mathbf{B}^{\boldsymbol{\Omega}}, \tag{8}$$

where $\mathbf{Z} = (\boldsymbol{z}_i,\cdots,\boldsymbol{z}_n)^\top$, and $\boldsymbol{z}_i \sim N(\mathbf{0}, \mathbf{I}_p)$ represent i.i.d. normal random vectors that are independent of $\boldsymbol{x}_i$'s. Note that when $\boldsymbol{\Omega}$ equals $\boldsymbol{\Omega}_0$, the knockoffs generated above are the exact model-X knockoffs as described in (3).

Step 2. Calculating knockoff statistics. As outlined in Section 2.2, there are many options for the knockoff statistics as long as they satisfy the so-called flip-sign property. Here we employ the Lasso coefficient difference statistics due to its high power in empirical performance[13]. Specifically, we consider the following penalized estimator from the partial log-likelihood,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{2p}}{\arg\min}\left\{-\widetilde{\ell}(\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1\right\}, \tag{9}$$

where $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1,\cdots,\widehat{\beta}_{2p})^\top$, $\lambda > 0$ is the regularization parameter and

$$\widetilde{\ell}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\delta_i\Big\{[\boldsymbol{x}_i^\top, (\widetilde{\boldsymbol{x}}_i^{\boldsymbol{\Omega}})^\top]\boldsymbol{\beta} - \log\Big(\frac{1}{n}\sum_{j=1}^{n}\mathbf{1}\{Y_j \ge Y_i\}\exp([\boldsymbol{x}_i^\top, (\widetilde{\boldsymbol{x}}_i^{\boldsymbol{\Omega}})^\top]\boldsymbol{\beta})\Big)\Big\},$$

Then the knockoff statistic W_j is defined as

$$W_j = |\widehat{\beta}_j| - |\widehat{\beta}_{j+p}|. \tag{10}$$

Step 3. Find the threshold and select features. With the knockoff statistics W_j, $j = 1,\cdots,p$, we sort them from high to low and select features whose statistics exceed the data-driven threshold T defined in (5) or (6), thus identifying the relevant features in the set

$$\widehat{\mathcal{S}} = \widehat{\mathcal{S}}(T) = \{1 \le j \le p : W_j \ge T\}. \tag{11}$$

3 Theoretical properties

In this section, we establish theoretical guarantees of the proposed method. We first establish the exact FDR control of our procedure when the population precision matrix $\boldsymbol{\Omega}_0$ is known. Next, we investigate the robustness of FDR control for our procedure.

Condition 1. The censoring time C is independent of the irrelevant features in $\mathcal{S}_0^c$.

Condition 1 is a mild condition similar to that used by Dong et al.[17] and Li et al.[18], indicating that the censoring time depends only on the relevant covariates in $\mathcal{S}_0$. In particular, this condition is satisfied when the censoring time is independent of all the covariates.

Lemma 1. Assume that Condition 1 holds and the constructed knockoffs satisfy Definition 1. For any $q \in (0,1)$, our method with the knockoff threshold T in (5) identifies the set $\widehat{\mathcal{S}} = \{1 \le j \le p : W_j \ge T\}$ satisfying the modified FDR control

$$\mathrm{mFDR} = \mathbb{E}\left[\frac{|\widehat{\mathcal{S}} \cap \mathcal{S}_0^c|}{|\widehat{\mathcal{S}}| + q^{-1}}\right] \le q,$$

and with the knockoff+ threshold T_+ in (6) identifies the set $\widehat{\mathcal{S}}_+ = \{1 \le j \le p : W_j \ge T_+\}$ satisfying the exact FDR control

$$\mathrm{FDR} = \mathbb{E}\left[\frac{|\widehat{\mathcal{S}} \cap \mathcal{S}_0^c|}{|\widehat{\mathcal{S}}| \vee 1}\right] \le q.$$

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

Lemma 1 is the same as Theorem 1 in Ref. [18], which establishes the FDR control of our method in the ideal scenario where the constructed knockoffs satisfy Definition 1. If we can obtain the population precision matrix $\boldsymbol{\Omega}_0$, the knockoffs generated from $\boldsymbol{\Omega}_0$ will inherently satisfy Definition 1. Moreover, the knockoff+ threshold is more conservative since it controls the exact FDR.

Next, we investigate the theoretical properties of our method when the precision matrix $\boldsymbol{\Omega}_0$ is unknown and replaced by an estimator $\boldsymbol{\Omega}$. The following definition characterizes the properties that an acceptable estimator $\boldsymbol{\Omega}$ should satisfy.

Definition 2 (Acceptable estimator). A $p\times p$ symmetric matrix $\boldsymbol{\Omega}$ is called an acceptable estimator of the true precision matrix $\boldsymbol{\Omega}_0$ if it satisfies:

(i) $\boldsymbol{\Omega}$ is independent of the data $(\boldsymbol{X},\boldsymbol{y})$;

(ii) $\|\boldsymbol{\Omega}-\boldsymbol{\Omega}_0\|_2 \le Ca_{np}$ holds with probability at least $1-O(p^{-c})$ for some constants $C,c>0$ and $a_{np}\to 0$ as $n\to\infty$.

A similar class of acceptable estimators was adopted in Ref. [19]. The first property in Definition 2 is imposed so that the randomness of $\widehat{\boldsymbol{\Omega}}$ would not interfere with the subsequent knockoff inference procedure. In practice, this can be achieved via the technique of data splitting[12,24,25], or by using auxiliary data of covariates, which are independent of the original data[26]. Furthermore, we mention that independence is more of a technical assumption than a practical need. The second property can be satisfied by many existing methods for high-dimensional precision matrix estimation, such as graphical Lasso[21], CLIME[22], and ISEE[23].

The FDR function depends on the precision matrix $\boldsymbol{\Omega}$ through the construction of the knockoff matrix $\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}$. To link the FDR of our procedure based on the true precision matrix with that derived from the estimated one, we reformulate the FDP and FDR functions in (2) as

$$\mathrm{FDR}_n(\boldsymbol{\Omega})=\mathbb{E}[\mathrm{FDP}_n(\boldsymbol{y},\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}})], \tag{12}$$

where the subscript n is used to stress the dependence of the FDP and FDR functions on the sample size n. In particular, with the population precision matrix $\boldsymbol{\Omega}_0$, it follows from Lemma 1 that the FDR of our proposed method based on $(\boldsymbol{y},\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}_0})$, $\mathrm{FDR}_n(\boldsymbol{\Omega}_0)$, is controlled at the target level q. To study the robustness of our method, we establish a connection between the functions $\mathrm{FDR}_n(\boldsymbol{\Omega})$ and $\mathrm{FDR}_n(\boldsymbol{\Omega}_0)$.

In light of (12), we can first study the link between the augmented knockoff matrices $[\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}]$ and $[\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}_0}]$. For any precision matrix $\boldsymbol{\Omega}$, by (8) we know that the corresponding knockoff matrix is $\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}=\boldsymbol{X}\mathbf{C}^{\boldsymbol{\Omega}}+\mathbf{Z}\mathbf{B}^{\boldsymbol{\Omega}}$, where $\mathbf{C}^{\boldsymbol{\Omega}}=\mathbf{I}_p-\boldsymbol{\Omega}\mathrm{diag}\{\boldsymbol{s}\}$, and $\mathbf{B}^{\boldsymbol{\Omega}}=(2\mathrm{diag}\{\boldsymbol{s}\}-\mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{\Omega}\mathrm{diag}\{\boldsymbol{s}\})^{1/2}$. Then by some calculations, it yields that

$$[\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}]=\boldsymbol{X}_{\mathrm{aug}}\mathbf{H}^{\boldsymbol{\Omega}}, \tag{13}$$

where $\boldsymbol{X}_{\mathrm{aug}}=[\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}_0}]=[\boldsymbol{X},\boldsymbol{X}\mathbf{C}^{\boldsymbol{\Omega}_0}+\mathbf{Z}\mathbf{B}^{\boldsymbol{\Omega}_0}]$ and

$$\mathbf{H}^{\boldsymbol{\Omega}}=\begin{pmatrix}\mathbf{I}_p & \mathbf{C}^{\boldsymbol{\Omega}}-\mathbf{C}^{\boldsymbol{\Omega}_0}(\mathbf{B}^{\boldsymbol{\Omega}_0})^{-1}\mathbf{B}^{\boldsymbol{\Omega}}\\ \mathbf{0} & (\mathbf{B}^{\boldsymbol{\Omega}_0})^{-1}\mathbf{B}^{\boldsymbol{\Omega}}\end{pmatrix} \tag{14}$$

with $\mathbf{C}^{\boldsymbol{\Omega}_0}=\mathbf{I}_p-\boldsymbol{\Omega}_0\mathrm{diag}\{\boldsymbol{s}\}$, $\mathbf{B}^{\boldsymbol{\Omega}_0}=(2\mathrm{diag}\{\boldsymbol{s}\}-\mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{\Omega}_0\mathrm{diag}\{\boldsymbol{s}\})^{1/2}$. Since $\boldsymbol{\Omega}_0$ is the population precision matrix, we can treat $\mathbf{H}^{\boldsymbol{\Omega}}$ as a function of $\boldsymbol{\Omega}$. Thus, with an estimated precision matrix $\boldsymbol{\Omega}$, we can see that $\mathbf{H}^{\boldsymbol{\Omega}}$ measures the difference between the estimated augmented knockoff matrix $[\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}]$ and the true one $[\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}_0}]$ with respect to $\boldsymbol{\Omega}$. In particular, when $\boldsymbol{\Omega}=\boldsymbol{\Omega}_0$, we have $\mathbf{H}^{\boldsymbol{\Omega}_0}=\mathbf{I}_{2p}$.

On the basis of the above discussions, since the FDR function depends on $\boldsymbol{\Omega}$ through $[\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}]$ in view of (12), we can further characterize the dependency of the FDR function on $\boldsymbol{\Omega}$ through the matrix $\mathbf{H}^{\boldsymbol{\Omega}}$.

Proposition 1. For any given symmetric positive definite matrix $\boldsymbol{\Omega}\in\mathbb{R}^{p\times p}$, we have

$$\mathrm{FDR}_n(\boldsymbol{\Omega})=\mathbb{E}\left[g_n(\boldsymbol{X}_{\mathrm{aug}}\mathbf{H}^{\boldsymbol{\Omega}})\right],$$

where $\boldsymbol{X}_{\mathrm{aug}}=[\boldsymbol{X},\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}_0}]\in\mathbb{R}^{n\times 2p}$ is the augmented knockoff matrix with the knockoff matrix $\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}_0}$ generated vua the true precision matrix $\boldsymbol{\Omega}_0$. The function $g_n(\cdot)$ is some conditional expectation of the FDP function, whose form is free of $\boldsymbol{\Omega}$, and $\mathbf{H}^{\boldsymbol{\Omega}}$ is defined in (14).

According to Proposition 1, when $\boldsymbol{\Omega}=\boldsymbol{\Omega}_0$, $\mathbf{H}^{\boldsymbol{\Omega}_0}=\mathbf{I}_{2p}$, and $\mathrm{FDR}_n(\mathbf{H}^{\boldsymbol{\Omega}_0})=\mathbb{E}[g_n(\boldsymbol{X}_{\mathrm{aug}})]\le q$ according to Lemma 1. Since the augmented knockoff matrix $\boldsymbol{X}_{\mathrm{aug}}$ is generated through the true precision matrix $\boldsymbol{\Omega}_0$, the only difference between $\mathrm{FDR}_n(\boldsymbol{\Omega})$ and $\mathrm{FDR}_n(\boldsymbol{\Omega}_0)$ lies in the matrix $\mathbf{H}^{\boldsymbol{\Omega}}$. Moreover, Proposition 1 demonstrates that the dependence of the FDR function on $\boldsymbol{\Omega}$ is fully characterized by the matrix $\mathbf{H}^{\boldsymbol{\Omega}}$ so that we can reparameterize the FDR function as $\mathrm{FDR}_n(\mathbf{H}^{\boldsymbol{\Omega}})$.

Condition 2. There exists some positive constant L such that

$$|\mathrm{FDR}_n(\mathbf{H}^{\boldsymbol{\Omega}})-\mathrm{FDR}_n(\mathbf{H}^{\boldsymbol{\Omega}_0})|\le L\|\mathbf{H}^{\boldsymbol{\Omega}}-\mathbf{H}^{\boldsymbol{\Omega}_0}\|_2$$

for any given symmetric positive definite $\boldsymbol{\Omega}\in\mathbb{R}^{p\times p}$ satisfying $\|\boldsymbol{\Omega}-\boldsymbol{\Omega}_0\|_2\le Ca_{np}$ with some constant $C>0$ and $a_{np}\to 0$.

Condition 2 is a similar continuity condition to that in Fan et al.[15] and Zhou et al.[19], which assumes the Lipschitz continuity of the FDR function with respect to $\mathbf{H}^{\boldsymbol{\Omega}}$ when the estimation loss of the precision matrix is no larger than a_{np}. According to Proposition 1, the FDR function $\mathrm{FDR}_n(\mathbf{H}^{\boldsymbol{\Omega}})$ can be expressed as $\mathbb{E}\left[g_n(\boldsymbol{X}_{\mathrm{aug}}\mathbf{H}^{\boldsymbol{\Omega}})\right]$. Note that $g_n(\cdot)$ is some conditional expectation with respect to the probability law of $\boldsymbol{X}_{\mathrm{aug}}$, which follows a matrix normal distribution with independent rows in view of (4). Furthermore, the conditional expectation is indeed a projection onto normal random variables. Motivated by the smoothness of normal random variables, we expect $g_n(\cdot)$ to be smooth. Thus, given the smoothness of $g_n(\cdot)$, we expect that Condition 2 holds.

We establish the robustness of FDR control for our procedure in the following theorem.

Theorem 1. Assume that all the eigenvalues of $\boldsymbol{\Omega}_0$ are bounded away from 0 and ∞ and that the smallest eigenvalue of $2\mathrm{diag}\{\boldsymbol{s}\}-\mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{\Omega}_0\mathrm{diag}\{\boldsymbol{s}\}$ is bounded from below by some positive constant. Then under Condition 2, it holds that

$$\sup_{\|\boldsymbol{\Omega}-\boldsymbol{\Omega}_0\|_2\le Ca_{np}}|\mathrm{FDR}_n(\mathbf{H}^{\boldsymbol{\Omega}})-\mathrm{FDR}_n(\mathbf{H}^{\boldsymbol{\Omega}_0})|=O(a_{np}).$$

Moreover, with Condition 1 and the acceptable estimator $\boldsymbol{\Omega}$ satisfying Definition 2, the FDR of the knockoff procedure can be bounded from above by $q+O(a_{np})+O(p^{-c})$ for any prespecified FDR level $q\in(0,1)$ and some positive constant c.

Theorem 1 establishes the robustness of FDR control with

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

respect to the estimated precision matrix. The terms $O(a_{np})$ and $O(p^{-c})$ reflect the price that we pay for estimating the unknown precision matrix. Specifically, the term $O(a_{np})$ represents the estimation accuracy of the precision matrix in terms of the matrix spectral norm. When $\mathbf{\Omega}_0$ is L_p-sparse with each row containing at most L_p nonzeros for some L_p that can diverge with p, it has been shown that the convergence rate a_{np} is typically on the order of $L_p\sqrt{(\log p)/n}$ under various popular methods such as graphical Lasso[21], CLIME[22], and ISEE[23]. In addition, in high-dimensional settings where p is larger than n, the term $O(p^{-c})$ is the tail probability of the consistent results, which converges to zero rapidly. Therefore, with a sufficiently large sample size, the reproducibility of the suggested procedure will be guaranteed for high-dimensional Cox models.

4 Simulation studies

In this section, we conduct simulation studies to examine the finite-sample performance of the proposed method. Since Definition 2 requires that the estimated precision matrix in Step 1 be consistent and independent of the data used for the subsequent knockoff procedure, we investigate the impact of precision matrix estimation by comparing the performance of the oracle knockoff procedure, our method with data splitting, and our method without data splitting. Specifically, the oracle knockoff procedure is carried out by generating knockoff variables via the population precision matrix $\mathbf{\Omega}_0$. For our method with data splitting, we first randomly divide the data into two parts of equal size. The first part is used to estimate the precision matrix. And with the estimated precision matrix, the subsequent knockoff procedure is applied to the second part to obtain the final selected set. In contrast, our method without data splitting uses the same data for both precision matrix estimation and the knockoff procedure.

In addition, both the knockoff thresholds defined in (5) and (6) are applied to the three aforementioned procedures. For convenience, we refer to the proposed method without data splitting as "knockoff" and "knockoff+", and the proposed method with data splitting as "split knockoff" and "split knockoff+". Furthermore, the oracle procedure with the two knockoff thresholds is denoted as "oracle knockoff" and "oracle knockoff+", respectively.

For comparison, we also report a data splitting method with FDR control similar to that proposed in Refs. [27, 28]. This method consists of four steps. First, the data are randomly split into two equal folds. Second, the first fold is used to estimate the active set of predictors via Lasso. Third, the second fold is used to fit the selected predictors in the active set with the maximum partial log-likelihood method and calculate the corresponding p-values. In addition, the remaining p-values are set to 1. Finally, for a prespecified target FDR level q, the Benjamini–Hochberg procedure[7] is applied to the obtained p-values to determine the final selected set of predictors. For simplicity, we refer to this method as BHq.

In the implementation of our knockoff procedure, we employ the ISEE method[23] to estimate the precision matrix because of its scalability and simple tuning. Similar to Fan et al.[15], we choose $s_j = 1/\Lambda_{\max}(\widehat{\mathbf{\Omega}})$ for all $1 \le j \le p$, where $\widehat{\mathbf{\Omega}}$ denotes the estimated precision matrix and $\Lambda_{\max}$ is the largest eigenvalue of a matrix. To examine the feature selection performance, we focus on the empirical FDR and the power based on 100 replications, where the power is defined as the proportion of true discoveries to all true variables. The target FDR level q is set at 0.1 for all simulation examples.

4.1 Simulation example 1

In our simulation setup, the sample size n is fixed at 400 and the dimension p varies among $\{600, 800, 1000, 1200, 1400\}$. The rows of the design matrix $\boldsymbol{X}$ are generated as i.i.d. copies from $N(\mathbf{0}, \mathbf{\Sigma})$ with covariance matrix $\mathbf{\Sigma} = (0.5^{|i-j|})_{1\le i,j\le p}$. The true coefficient vector β_0 is set as a sparse vector with $k = 20$ nonzero components, whose locations are randomly selected and the values of these nonzero elements are randomly drawn from $\{\pm 1.5\}$. We generate survival times $\{T_i\}_{i=1}^n$ on the basis of the hazard function (1) with $\lambda_0(t) = 0.3$. The censoring times $\{C_i\}_{i=1}^n$ are sampled from an exponential distribution with parameter λ_c, which is set to achieve an approximate censoring rate of 20%.

The simulation results are summarized in Table 1. It can be seen from Table 1 that the three knockoff procedures with knockoff+ threshold can control the FDR at the target level while maintaining high power across different dimensions. It is reasonable that the procedure with knockoff threshold has an empirical FDR slightly over the target level since the knockoff threshold tends to select more variables than the knockoff+ threshold by definition. Moreover, our knockoff and knockoff+ methods closely mimic split knockoff and split knockoff+, indicating that data splitting is more of a technical assumption. Furthermore, the performance of both meth-

Table 1. The empirical FDR and power of different methods over 100 replications for simulation example 1 in Section 4.1.

Method	$p = 600$		$p = 800$		$p = 1000$		$p = 1200$		$p = 1400$	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
Knockoff	0.1250	1	0.1293	1	0.1194	1	0.1092	1	0.1290	1
Knockoff+	0.1009	1	0.1063	1	0.0864	1	0.0796	1	0.0870	1
Oracle knockoff	0.1227	1	0.1319	1	0.1472	1	0.1379	1	0.1308	1
Oracle knockoff+	0.0886	1	0.0958	1	0.0888	1	0.0996	1	0.1121	1
Split knockoff	0.1191	1	0.1337	1	0.1368	1	0.1214	1	0.1521	0.9983
Split knockoff+	0.0605	1	0.1036	0.9983	0.1103	1	0.0823	1	0.1043	0.9983
BHq	0.0079	0.9750	0.0032	0.9241	0.0032	0.8917	0.0127	0.8467	0.0167	0.7524

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

ods is comparable to that of the oracle knockoff procedure, which employs the true precision matrix for knockoff generation. This further demonstrates the robustness of our procedure. In contrast, BHq is a more conservative method, as it controls the FDR at a level much lower than the target level while also suffering from a loss of power as the dimension increases. This might be because the asymptotic null p-values could be highly nonuniform in high-dimensional cases[10,13].

4.2 Simulation example 2

In example 2, the setup is similar to that in example 1 with the exception of the baseline hazard function $\lambda_0(t)$ and the censoring time. Here we model the baseline hazard function as time-dependent by setting $\lambda_0(t) = 2t$. In addition, we generate the censoring time as $\lambda_c \exp(\boldsymbol{x}_i^\top \boldsymbol{\gamma})$ to make it dependent on some important features. Specifically, after generating $\boldsymbol{\beta}_0$, we generate $\boldsymbol{\gamma}$ with 5 nonzero elements, each of which is valued at 1. These nonzero elements of $\boldsymbol{\gamma}$ are randomly allocated among the nonzero locations of $\boldsymbol{\beta}_0$. Furthermore, the constant λ_c is chosen to achieve a censoring rate of approximately 20%.

Table 2 summarizes the empirical FDR and power of different methods under the scenario where the censoring time depends on covariates and the baseline hazard function is time-dependent. Table 2 shows that our methods, both with and without data splitting, still enjoy good and similar performance in terms of empirical FDR and power. Moreover, we also observe that BHq can achieve FDR control but suffers from a loss of power.

4.3 Simulation example 3

In example 3, we consider that the distribution of the covariates $\boldsymbol{X}$ is non-Gaussian. Specifically, the rows of $\boldsymbol{X}$ are i.i.d. samples drawn from the multivariate centered t-distribution $t_{10}(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (0.5^{|i-j|})_{1\le i,j\le p}$. The remaining part of this simulation setting is the same as that in example 2.

When the covariate distribution is non-Gaussian, we can see from Table 3 that our knockoff+ and split knockoff+ methods can still control the FDR at the target level and retain high power across various dimensions. Our procedure with the knockoff threshold also has high power and achieves an FDR value slightly above the target level. For BHq, this method can control the FDR at a much lower level than the target level while suffering much more from power loss.

5 Two real data applications

In this section, we showcase the practical utility of the proposed method on two real datasets.

5.1 Primary biliary cirrhosis data

The primary biliary cirrhosis (PBC) data provided by Therneau[29], were collected from the Mayo Clinic trial in PBC of the liver, which was conducted from January 1974 to May 1984. This study included 424 patients, 312 of whom participated in a randomized trial of the drug D-penicillamine versus placebo. Among these 312 patients, we discarded those observations with missing values in the dataset, resulting in a sample of $n = 276$ patients. At the end of the follow-

Table 2. The empirical FDR and power of different methods over 100 replications for simulation example 2 in Section 4.2.

Method	$p = 600$		$p = 800$		$p = 1000$		$p = 1200$		$p = 1400$	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
Knockoff	0.1477	1	0.1536	1	0.1041	1	0.1465	1	0.0995	1
Knockoff+	0.1105	1	0.1088	1	0.0802	1	0.0902	1	0.0771	1
Oracle knockoff	0.0973	1	0.1721	1	0.1200	1	0.1197	1	0.1091	1
Oracle knockoff+	0.0695	1	0.1311	1	0.0961	1	0.0961	1	0.0643	1
Split knockoff	0.1493	1	0.1060	1	0.1155	0.9900	0.1574	0.9917	0.0915	0.9700
Split knockoff+	0.0937	1	0.0632	1	0.0768	0.9883	0.1170	0.9917	0.0657	0.9533
BHq	0.0035	0.9624	0.0000	0.8883	0.0000	0.8083	0.0049	0.6717	0.0035	0.5567

Table 3. The empirical FDR and power of different methods over 100 replications for simulation example 3 in Section 4.3.

Method	$p = 600$		$p = 800$		$p = 1000$		$p = 1200$		$p = 1400$	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
Knockoff	0.1138	1	0.0852	1	0.1014	1	0.1601	1	0.0972	1
Knockoff+	0.0729	1	0.0700	1	0.0731	1	0.0633	1	0.0546	1
Oracle knockoff	0.1150	1	0.1070	1	0.1428	1	0.0989	1	0.0985	1
Oracle knockoff+	0.0827	1	0.0770	1	0.0977	1	0.0670	1	0.0685	1
Split knockoff	0.1305	0.9983	0.1274	1	0.1178	0.9750	0.1194	0.9717	0.1246	0.9500
Split knockoff+	0.0950	0.9917	0.0934	1	0.0732	0.9617	0.0594	0.9667	0.0695	0.9450
BHq	0.0081	0.9317	0.0016	0.8250	0.0016	0.7050	0.0085	0.6183	0.0024	0.5033

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

up period, 111 of these 276 patients had died, indicating a censoring rate of approximately 60%. The dataset comprises $p = 17$ features including demographic variables (age, sex) and clinical characteristics (e.g., spiders, ascites, edema). For a comprehensive overview of these variables, see Ref. [29]. Our goal was to identify the true risk factors that influence survival time among the 17 covariates included in the dataset.

As demonstrated in Section 4, since our method without data splitting can control the FDR at the desired level and maintain high power, we implement our method without data splitting. For comparison, we also apply BHq and Lasso to this dataset. The precision matrix is estimated via the ISEE method[23] and the target FDR level is set at $q = 0.2$. The selection results are summarized in Table 4. Compared with Lasso, the knockoff+ method selects seven predictors, excluding ascites and SGOT. As evidenced by Tibshirani[4], the coefficients for these two features are notably small and not statistically significant, suggesting that they are likely noise variables. This selective capability suggests the effectiveness of the knockoff+ method in distinguishing relevant from irrelevant predictors within the dataset. In addition, BHq selects fewer variables. Moreover, the variables identified by knockoff+ but not by BHq, such as age and oedema, have been shown to be significant in Refs. [4, 30]. This reveals the potential of the knockoff+ method to uncover more clinically relevant variables that might be missed by BHq, thereby aiding in the development of more accurate treatment strategies.

5.2 Diffuse large-B-cell lymphoma data

We further demonstrated the usefulness of our method by analyzing the microarray diffuse large-B-cell lymphoma (DLBCL) data originally studied by Rosenwald et al.[31] This dataset consists of 240 patients after chemotherapy, with $p = 7399$ cDNA microarray expressions of each individual patient as predictors. Among 240 patients, five patients had survival times close to 0; thus, we removed them and obtained a sample of $n = 235$ patients. The median follow-up time was 2.8 years and 138 patients died during the study period. The aim of this study was to formulate a molecular predictor of survival after chemotherapy for this disease.

Since the dimensionality of this dataset is ultrahigh, we implement a screening procedure as described by Liu et al.[32] to obtain a reduced model comprising $[n/\log(n)] = 43$ features. Similar to Section 5.1, we apply our method without data splitting, as well as BHq and Lasso to the reduced model obtained from the screening stage. The target level of the FDR is set at $q = 0.2$.

The selected genes are listed in Table 5. Lasso tends to select more genes than the knockoff procedure and BHq does, whereas the BHq approach is more conservative. Table 6 lists the 12 genes selected by means of knockoff+ method. We can see from Table 6 that the genes identified by the knockoff+ method belong to both the proliferation signature group and the lymph node signature group, whereas the BHq method identifies genes from only the proliferation signature group. As evidenced by Hong et al.[33], both the proliferation and lymph node signature groups are relevant to patient survival from the time of chemotherapy. In addition, the lymph node signature, which reflects a host response characterized by the abundant expression of extracellular matrix components and infiltration of tumors with immune cells other than T cells, is important for understanding the tumor microenvironment and its interaction with the immune system[34]. Thus, compared with BHq, the ability of the knockoff method to identify genes from the lymph node signature group underscores its capacity to capture more critical genes for patient survival.

6 Conclusions

In this paper, we propose a new method for feature selection in high-dimensional Cox models that achieves asymptotic FDR control, facilitating reproducible learning in survival analysis. Both the established theoretical properties and empirical performance demonstrate the efficiency of our pro-

Table 4. Variable selection results of different methods for the real data application in Section 5.1.

Method	Selected variables
Lasso	Age, ascites, oedema, serum bilirubin, albumin, urine copper, alkaline phosphatase, SGOT, prothrombine time
BHq	Serum bilirubin, urine copper, albumin, prothrombine time
Knockoff	Age, oedema, serum bilirubin, albumin, urine copper, alkaline phosphatase, SGOT, prothrombine time
Knockoff+	Age, oedema, serum bilirubin, albumin, urine copper, alkaline phosphatase, prothrombine time

Table 5. IDs of genes selected by different methods for the real data application in Section 5.2.

Method	Gene IDs								
Lasso	31981	29652	30157	16292	26163	16309	34771	30282	31037
	17765	27267	17236	32787	27774	32679	31242	16937	24933
	29652	32238	24376	22162	17765	27267			
BHq	31981	16292	16309	34771	31037				
Knockoff	31981	29652	30157	16292	26163	16309	34771	30282	31037
	17765	27267	17236	32787	27774	32679	31242	16937	
Knockoff+	31981	29652	30157	16292	26163	16309	34771	30282	31037
	17765	27267	17236						

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

Table 6. The list of 12 selected genes by knockoff+ for the real data application in Section 5.2.

ID	Signature	Description
17236	Proliferation	Eukaryotic translation elongation factor 1 alpha 1
31981	Proliferation	Septin 1
29652	Proliferation	BUB3 budding uninhibited by benzimidazoles 3 homolog
30157	Proliferation	Centromere protein F
16292	Proliferation	Centromere protein F
26163	Proliferation	Ring-box 1
16309	Proliferation	Guanine monphosphate synthetase
34771	Proliferation	Tubulin, alpha, ubiquitous
30282	Proliferation	Tubulin, alpha, ubiquitous
31037	Proliferation	Protein phosphatase 1G, magnesium-dependent, gamma isoform
17765	Lymph node	Integrin, alpha 5 (fibronectin receptor, alpha polypeptide)
27267	Lymph node	Actinin, alpha 1

posed method. It would be interesting to extend our method to more general models in survival analysis and develop the robustness of power analysis in our method. The main difficulty lies in how to establish the consistency of the coefficients used to generate the knockoff statistics, particularly when the knockoff matrix is generated on the basis of an estimated feature distribution. These extensions are beyond the scope of this study and are interesting topics for future research.

The independence between the estimated precision matrix and the data plays an important role in our technical analysis and can be achieved via data splitting. Our numerical examples, however, suggest that this independence is more of a technical assumption than a practical necessity. It would be interesting to develop theoretical guarantees for our procedure without the independence assumption. Moreover, the robustness of FDR control for our procedure utilizes the stability argument in Condition 2 that the knockoff procedure exhibits stability properties under small perturbations of the precision matrix. Similar stability arguments and technical conditions are also used in Barber et al.[14] and Fan et al.[16]. Nevertheless, it would be interesting to provide more justifications for these arguments. The main challenge is that the FDR function is data-driven and it depends on the data through unknown and complex forms. Thus, these extensions are important and interesting topics for future research.

Acknowledgements

This work was supported by National Key R&D Program of China (2022YFA1008000), National Natural Science Foundation of China (12101584), Fundamental Research Funds for the Central Universities (WK2040000114, WK2040000079), Hefei Postdoctoral Research Project Funds in 2021, and Anhui Postdoctoral Research Project Funds in 2021.

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Xin Zhou is a Ph.D. candidate at the School of Management, University of Science and Technology of China. His research mainly focuses on ensemble learning, high-dimensional variable selection, and statistical inference.

Yang Li is a Special Associate Researcher at the School of Management, University of Science and Technology of China (USTC). He obtained his Ph.D. degree from USTC. His research mainly focuses on the broad areas of machine learning and distributed learning.

References

[1] Cox D R. Regression models and life-tables. *Journal of the Royal Statistical Society Series B*, **1972**, *34*: 187–220.
[2] Gepp A, Kumar K. The role of survival analysis in financial distress prediction. *International Research Journal of Finance and Economics*, **2008**, *16*: 13–34.
[3] Kvamme H, Borgan O, Scheel I. Time-to-event prediction with neural networks and Cox regression. *Journal of Machine Learning Research*, **2019**, *20*: 1–30.
[4] Tibshirani R. The lasso method for variable selection in the Cox model. *Statistics in Medicine*, **1997**, *16*: 385–395.
[5] Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. *The Annals of Statistics*, **2002**, *30*: 74–99.
[6] Antoniadis A, Fryzlewicz P, Letue F. The Dantzig selector in Cox's proportional hazards model. *Scandinavian Journal of Statistics*, **2010**, *37*: 531–552.
[7] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **1995**, *57*: 289–300.
[8] Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **2001**, *29*: 1165–1188.
[9] Chudik A, Kapetanios G, Pesaran H. A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica*, **2018**, *86*: 1479–1512.
[10] Fan Y, Demirkaya E, Lv J. Nonuniformity of *p*-values can occur early in diverging dimensions. *Journal of Machine Learning Research*, **2019**, *20*: 2849–2881.
[11] Barber R F, Candès E J. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, **2015**, *43*: 2055–2085.
[12] Barber R F, Candès E J. A knockoff filter for high-dimensional selective inference. *The Annals of Statistics*, **2019**, *47*: 2504–2537.
[13] Candès E J, Fan Y, Janson L, et al. Panning for gold: 'model-X' knockoffs for high-dimensional controlled variable selection. *Journal of the Royal Statistical Society Series B*, **2018**, *80*: 551–577.
[14] Barber R F, Candès E J, Samworth R J. Robust inference with knockoffs. *Annals of Statistics*, **2020**, *48*: 1409–1431.

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

[15] Fan Y, Demirkaya E, Li G, et al. RANK: large-scale inference with graphical nonlinear knockoffs. *Journal of the American Statistical Association*, **2020**, *115*: 362–379.
[16] Fan Y, Lv J, Sharifvaghefi M, et al. IPAD: stable interpretable forecasting with knockoffs inference. *Journal of the American Statistical Association*, **2020**, *115*: 1822–1834.
[17] Dong Y, Li D, Zheng Z, et al. Reproducible feature selection in high-dimensional accelerated failure time models. *Statistics & Probability Letters*, **2022**, *181*: 109275.
[18] Li D, Yu J, Zhao H. CoxKnockoff: Controlled feature selection for the Cox model using knockoffs. *Stat*, **2023**, *12*: e607.
[19] Zhou J, Li Y, Zheng Z, et al. Reproducible learning in large-scale graphical models. *Journal of Multivariate Analysis*, **2022**, *189*: 104934.
[20] Bai X, Ren J, Fan Y, et al. KIMI: Knockoff inference for motif identification from molecular sequences with controlled false discovery rate. *Bioinformatics*, **2021**, *37*: 759–766.
[21] Ravikumar P, Wainwright M J, Raskutti G, et al. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics*, **2011**, *5*: 935–980.
[22] Cai T, Liu W, Luo X. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, **2011**, *106*: 594–607.
[23] Fan Y, Lv J. Innovated scalable efficient estimation in ultra-large Gaussian graphical models. *The Annals of Statistics*, **2016**, *44*: 2098–2126.
[24] Dai C, Lin B, Xing X, et al. A scale-free approach for false discovery rate control in generalized linear models. *Journal of the American Statistical Association*, **2023**, *118*: 1551–1565.
[25] Dai C, Lin B, Xing X, et al. False discovery rate control via data splitting. *Journal of the American Statistical Association*, **2023**, *118*: 2503–2520.
[26] Loh P L, Wainwright M J. High-dimensional regression with noisy and missing data: Provable guarantees with nonconvexity. *The Annals of Statistics*, **2012**, *40*: 1637–1664.
[27] Meinshausen N, Meier L, Bühlmann P. P-values for high-dimensional regression. *Journal of the American Statistical Association*, **2009**, *104*: 1671–1681.
[28] Wasserman L, Roeder K. High-dimensional variable selection. *The Annals of Statistics*, **2009**, *37*: 2178–2201.
[29] Therneau T M. Extending the Cox model. In: Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis. New York: Springer, **1997**: 51–84.
[30] Leng C, Ma S. Path consistent model selection in additive risk model via Lasso. *Statistics in Medicine*, **2007**, *26*: 3753–3770.
[31] Rosenwald A, Wright G, Chan W, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *The New England Journal of Medicine*, **2002**, *346*: 1937–1947.
[32] Liu Y, Chen X, Li G. A new joint screening method for right-censored time-to-event data with ultra-high dimensional covariates. *Statistical Methods in Medical Research*, **2020**, *29*: 1499–1513.
[33] Hong H G, Kang J, Li Y. Conditional screening for ultra-high dimensional covariates with survival outcomes. *Lifetime Data Analysis*, **2018**, *24*: 45–71.
[34] Bea S, Zettl A, G. Wright, et al. Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood*, **2005**, *106*: 3183–3190.
[35] Fan Y, Kong Y, Li D, et al. Innovated interaction screening for high-dimensional nonlinear classification. *The Annals of Statistics*, **2015**, *43*: 1243–1272.
[36] Schmitt B A. Perturbation bounds for matrix square roots and Pythagorean sums. *Linear Algebra and Its Applications*, **1992**, *174*: 215–227.

Appendix

This section contains the proofs of main results. Throughout the proofs, let C denote some generic positive constant whose value may change from line to line. For ease of presentation, for the population precision matrix $\mathbf{\Omega}_0$, denote by $\widetilde{\boldsymbol{X}}_0 = \boldsymbol{X}\mathbf{C}_0 + \mathbf{Z}\mathbf{B}_0$, where $\mathbf{C}_0 = \mathbf{I}_p - \mathbf{\Omega}_0\mathrm{diag}\{\boldsymbol{s}\}$, and $\mathbf{B}_0 = (2\mathrm{diag}\{\boldsymbol{s}\} - \mathrm{diag}\{\boldsymbol{s}\}\mathbf{\Omega}_0\mathrm{diag}\{\boldsymbol{s}\})^{1/2}$.

A.1 Proof of Proposition 1

In this proof, we consider the precision matrix $\mathbf{\Omega}$ is given. Recall that $\mathbf{C}^{\mathbf{\Omega}} = \mathbf{I}_p - \mathbf{\Omega}\mathrm{diag}\{\boldsymbol{s}\}$, and $\mathbf{B}^{\mathbf{\Omega}} = (2\mathrm{diag}\{\boldsymbol{s}\} - \mathrm{diag}\{\boldsymbol{s}\}\mathbf{\Omega}\mathrm{diag}\{\boldsymbol{s}\})^{1/2}$. By (7), we can generate the knockoffs matrix $\widetilde{\boldsymbol{X}}^{\mathbf{\Omega}} = [\widetilde{\boldsymbol{x}}_1^{\mathbf{\Omega}}, \cdots, \widetilde{\boldsymbol{x}}_n^{\mathbf{\Omega}}]^\top$ as

$$\widetilde{\boldsymbol{X}}^{\mathbf{\Omega}} = \boldsymbol{X}\mathbf{C}^{\mathbf{\Omega}} + \mathbf{Z}\mathbf{B}^{\mathbf{\Omega}},$$

where we use the superscript $\mathbf{\Omega}$ to emphasize the dependence of knockoffs matrix on $\mathbf{\Omega}$, $\mathbf{Z} = (z_i, \cdots, z_n)^\top$, and $z_i \sim N(\mathbf{0}, \mathbf{I}_p)$ represent i.i.d. normal random vectors that are independent of $\boldsymbol{x}_i$'s.

By the same argument, for the knockoffs matrix $\widetilde{\boldsymbol{X}}_0$ generating by the precision matrix $\mathbf{\Omega}_0$, we can get that

$$\widetilde{\boldsymbol{X}}_0 = \boldsymbol{X}\mathbf{C}_0 + \mathbf{Z}\mathbf{B}_0.$$

Recall that $\boldsymbol{X}_{\mathrm{aug}} = [\boldsymbol{X}, \widetilde{\boldsymbol{X}}_0] = [\boldsymbol{X}, \boldsymbol{X}\mathbf{C}_0 + \mathbf{Z}\mathbf{B}_0]$ is the augmented matrix consisting of the columns of $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}_0$, and

$$\mathbf{H}^{\mathbf{\Omega}} = \begin{pmatrix} \mathbf{I}_p & \mathbf{C}^{\mathbf{\Omega}} - \mathbf{C}_0\mathbf{B}_0^{-1}\mathbf{B}^{\mathbf{\Omega}} \\ \mathbf{0} & \mathbf{B}_0^{-1}\mathbf{B}^{\mathbf{\Omega}} \end{pmatrix}.$$

By some calculations, it follows that

$$[\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\mathbf{\Omega}}] = [\boldsymbol{X}_{\mathrm{aug}}\mathbf{H}^{\mathbf{\Omega}}] \tag{A1}$$

and have identical joint distribution.

In view of our knockoffs procedure, we calculate the knockoff statistics W_j's using $(\widetilde{\boldsymbol{y}}, \boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\mathbf{\Omega}})$. Thus, the FDR function can be written as

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

$$\mathrm{FDR}_n(\boldsymbol{\Omega}) = \mathbb{E}[\mathrm{FDP}_n(\boldsymbol{y}, \boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}})] = \mathbb{E}[g_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}})], \tag{A2}$$

where $g_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}) = \mathbb{E}[\mathrm{FDP}_n(\boldsymbol{y}, \boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}})|\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}]$. It can be seen that the function g_n is the conditional FDP when knockoff matrix $\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}$ is generated using $\boldsymbol{\Omega}$. Since the response $\boldsymbol{y}$ is independent of $\widetilde{\boldsymbol{X}}^{\boldsymbol{\Omega}}$ given $\boldsymbol{X}$, it is clear that the function form of g_n is free of the matrix $\boldsymbol{\Omega}$ used to generate knockoff variables.

Thus, combining Eqs. (A1) and (A2), we can rewrite the FDR function as

$$\mathrm{FDR}_n(\boldsymbol{\Omega}) = \mathbb{E}[g_n(\boldsymbol{X}_{\mathrm{aug}}\mathbf{H}^{\boldsymbol{\Omega}})],$$

which completes the proof of Proposition 1.

A.2 Lemma 2 and its proof

Lemma 2. Assume that $\boldsymbol{\Omega}$ is an acceptable estimator satisfying Definition 2. If $\Lambda_{\max}(\boldsymbol{\Sigma}_0) \le c_0^{-1}$ and $\Lambda_{\min}(2\mathrm{diag}\{\boldsymbol{s}\} - \mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{\Omega}_0\mathrm{diag}\{\boldsymbol{s}\}) \ge c_0$ for some constant $c_0 > 0$, it holds that

$$\|\mathbf{B}_0^{-1}\mathbf{B}^{\boldsymbol{\Omega}} - \mathbf{I}_p\|_2 \le c_1\|\boldsymbol{\Omega} - \boldsymbol{\Omega}_0\|_2 = O(a_{np}),$$

where c_1 is some positive constant.

Proof. By the definition of $\mathbf{B}_0$ and $\mathbf{B}^{\boldsymbol{\Omega}}$, we have

$$(\mathbf{B}^{\boldsymbol{\Omega}})^2 - \mathbf{B}_0^2 = -\Big(\mathrm{diag}(\boldsymbol{s})(\boldsymbol{\Omega} - \boldsymbol{\Omega}_0)\mathrm{diag}(\boldsymbol{s})\Big). \tag{A3}$$

By the definition of $\boldsymbol{s}$ that $\boldsymbol{\Sigma}_0 - 2^{-1}\mathrm{diag}(\boldsymbol{s})$ is positive definite and $\Lambda_{\max}(\boldsymbol{\Sigma}_0) \le c_0$, it holds that

$$\|\boldsymbol{s}\|_{\max} \le 2\Lambda_{\max}(\boldsymbol{\Sigma}_0) \le C. \tag{A4}$$

In addition, since $\boldsymbol{\Omega}$ is an acceptable estimator satisfying Definition 2, we have $\|\boldsymbol{\Omega} - \boldsymbol{\Omega}_0\|_2 = O(a_{np})$. Combining the above results yields that

$$\|(\mathbf{B}^{\boldsymbol{\Omega}})^2 - \mathbf{B}_0^2\|_2 \le C\|\boldsymbol{\Omega} - \boldsymbol{\Omega}_0\|_2 = O(a_{np}). \tag{A5}$$

On the other hand, by the triangle inequality, for sufficiently large n, it follows that

$$\begin{aligned}\Lambda_{\min}((\mathbf{B}^{\boldsymbol{\Omega}})^2) \ge& \Lambda_{\min}(\mathbf{B}_0^2) + \Lambda_{\min}((\mathbf{B}^{\boldsymbol{\Omega}})^2 - \mathbf{B}_0^2) \ge \Lambda_{\min}(\mathbf{B}_0^2) - O(a_{np}) = \\ & \Lambda_{\min}(2\mathrm{diag}(\boldsymbol{s}) - \mathrm{diag}(\boldsymbol{s})\boldsymbol{\Omega}_0\mathrm{diag}(\boldsymbol{s})) - O(a_{np}) \ge c_0/2.\end{aligned}$$

In addition, it holds that $\Lambda_{\min}(\mathbf{B}_0^2) = \Lambda_{\min}(2\mathrm{diag}(\boldsymbol{s}) - \mathrm{diag}(\boldsymbol{s})\boldsymbol{\Omega}_0\mathrm{diag}(\boldsymbol{s})) \ge c_0/2$. Based on the two inequalities, by Lemma 2.2 in Ref. [36], we get that

$$\|(\mathbf{B}^{\boldsymbol{\Omega}}) - \mathbf{B}_0\|_2 \le (\sqrt{c_0/2} + \sqrt{c_0/2})^{-1}\|(\mathbf{B}^{\boldsymbol{\Omega}})^2 - \mathbf{B}_0^2\|_2 \le C\|\boldsymbol{\Omega} - \boldsymbol{\Omega}_0\|_2 = O(a_{np}),$$

where the last step is due to (A5). Thus it concludes that

$$\|\mathbf{B}_0^{-1}\mathbf{B}^{\boldsymbol{\Omega}} - \mathbf{I}_p\|_2 \le \|(\mathbf{B}^{\boldsymbol{\Omega}}) - \mathbf{B}_0\|_2\|\mathbf{B}_0^{-1}\|_2 \le C\|\boldsymbol{\Omega} - \boldsymbol{\Omega}_0\|_2\Lambda_{\min}^{-1/2}(\mathbf{B}_0^2) \le C\|\boldsymbol{\Omega} - \boldsymbol{\Omega}_0\|_2,$$

where the last step uses assumption $\Lambda_{\min}(\mathbf{B}_0^2) = \Lambda_{\min}\{2\mathrm{diag}\{\boldsymbol{s}\} - \mathrm{diag}\{\boldsymbol{s}\}\boldsymbol{\Omega}_0\mathrm{diag}\{\boldsymbol{s}\}\} \ge c_0$. This completes the proof of Lemma 2.

A.3 Proof of Theorem 1

We use the same notation as in the proof of Proposition 1. By the definition of $\mathbf{H}^{\boldsymbol{\Omega}}$ and $\mathbf{H}^{\boldsymbol{\Omega}_0}$, we have

$$\mathbf{H}^{\boldsymbol{\Omega}} - \mathbf{H}^{\boldsymbol{\Omega}_0} = \mathbf{H}^{\boldsymbol{\Omega}} - \mathbf{I}_{2p} = \begin{pmatrix} \mathbf{0} & \mathbf{C}^{\boldsymbol{\Omega}} - \mathbf{C}^{\boldsymbol{\Omega}_0}(\mathbf{B}_0)^{-1}\mathbf{B}^{\boldsymbol{\Omega}} \\ \mathbf{0} & (\mathbf{B}_0)^{-1}\mathbf{B}^{\boldsymbol{\Omega}} - \mathbf{I}_p \end{pmatrix}.$$

Then it follows from the matrix norm inequality that

$$\begin{aligned}\|\mathbf{H}^{\boldsymbol{\Omega}} - \mathbf{H}^{\boldsymbol{\Omega}_0}\|_2 \le& \|\mathbf{C}^{\boldsymbol{\Omega}} - \mathbf{C}_0\mathbf{B}_0^{-1}\mathbf{B}^{\boldsymbol{\Omega}}\|_2 + \|\mathbf{B}_0^{-1}\mathbf{B}^{\boldsymbol{\Omega}} - \mathbf{I}_p\|_2 \le \\ & \|\mathbf{C}^{\boldsymbol{\Omega}} - \mathbf{C}_0\|_2 + \|\mathbf{C}_0(\mathbf{B}_0^{-1}\mathbf{B}^{\boldsymbol{\Omega}} - \mathbf{I}_p)\|_2 + \|\mathbf{B}_0^{-1}\mathbf{B}^{\boldsymbol{\Omega}} - \mathbf{I}_p\|_2 \le \\ & \|\mathbf{C}^{\boldsymbol{\Omega}} - \mathbf{C}_0\|_2 + (\|\mathbf{C}_0\|_2 + 1)\|\mathbf{B}_0^{-1}\mathbf{B}^{\boldsymbol{\Omega}} - \mathbf{I}_p\|_2.\end{aligned} \tag{A6}$$

Since all the eigenvalues of $\boldsymbol{\Omega}_0$ are bounded away from 0 and ∞, we have $\|\boldsymbol{\Omega}_0\|_2 \le C$. This together with (A4) leads to

$$\|\mathbf{C}_0\|_2 = \|\mathbf{I} - \boldsymbol{\Omega}_0\mathrm{diag}(\boldsymbol{s})\|_2 \le 1 + \|\boldsymbol{\Omega}_0\|_2\|\mathrm{diag}(\boldsymbol{s})\|_2 \le C.$$

In addition, since $\boldsymbol{\Omega}$ is an acceptable estimator satisfying Definition 2, we have $\|\boldsymbol{\Omega} - \boldsymbol{\Omega}_0\|_2 = O(a_{np})$. Thus, we have

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):

$$\|\mathbf{C}^{\mathbf{\Omega}} - \mathbf{C}_0\|_2 = \|(\mathbf{\Omega} - \mathbf{\Omega}_0)\mathrm{diag}(\boldsymbol{s})\|_2 \le Ca_{np}.$$

These together with Lemma 2 entail that

$$\|\mathbf{H}^{\mathbf{\Omega}} - \mathbf{H}^{\mathbf{\Omega}_0}\|_2 \le Ca_{np}.$$

In light of Condition 2, we have

$$|\mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}}) - \mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}_0})| \le L\|\mathbf{H}^{\mathbf{\Omega}} - \mathbf{H}^{\mathbf{\Omega}_0}\|_2. \tag{A7}$$

Thus, combining the above results leads to

$$\sup_{\|\mathbf{\Omega}-\mathbf{\Omega}_0\|_2 \le Ca_{np}} |\mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}}) - \mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}_0})| \le O(a_{np}). \tag{A8}$$

Next we establish the FDR control of our procedure. Define the event $\mathcal{E}_0 = \{\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}_0\|_2 \le Ca_{np}\}$. By Definition 2, the acceptable estimator $\mathbf{\Omega}$ is independent of $(\boldsymbol{X}, \boldsymbol{y})$ and the event $\mathcal{E}_0$ holds with probability at least $1 - O(p^{-c})$. Then it follows from (A.8) that

$$\left|\mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\mathbf{\Omega}})\middle|\mathcal{E}_0\right] - \mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}_0)\middle|\mathcal{E}_0\right]\right| \le \sup_{\|\mathbf{\Omega}-\mathbf{\Omega}_0\|_2 \le Ca_{np}} \left|\mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\mathbf{\Omega}})\right] - \mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}_0)\right]\right| =$$
$$\sup_{\|\mathbf{\Omega}-\mathbf{\Omega}_0\|_2 \le Ca_{np}} |\mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}}) - \mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}_0})| \le O(a_{np}). \tag{A9}$$

By the property of conditional expectation, we have

$$\mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}}) - \mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}_0}) =$$
$$\left(\mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\mathbf{\Omega}})\middle|\mathcal{E}_0\right] - \mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}_0)\middle|\mathcal{E}_0\right]\right) \cdot \mathbb{P}(\mathcal{E}_0) + \left(\mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\mathbf{\Omega}})\middle|\mathcal{E}_0^c\right] - \mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}_0)\middle|\mathcal{E}_0^c\right]\right) \cdot \mathbb{P}(\mathcal{E}_0^c) = A_1 + A_2.$$

For the first term A_1 in above, by (A9), it holds that

$$|A_1| \le \left|\mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}^{\mathbf{\Omega}})\middle|\mathcal{E}_0\right] - \mathbb{E}\left[\mathrm{FDP}_n(\boldsymbol{X}, \widetilde{\boldsymbol{X}}_0)\middle|\mathcal{E}_0\right]\right| \le O(a_{np}).$$

For term A_2, since FDP is always bounded between 0 and 1, we have

$$|A_2| \le 2\mathbb{P}(\mathcal{E}_0^c) \le O(p^{-c}).$$

Combining the above results leads to

$$\left|\mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}}) - \mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}_0})\right| \le O(a_{np}) + O(p^{-c}).$$

For any pre-specified FDR level $q \in (0,1)$, the above result together with Lemma 1 yields that $\mathrm{FDR}_n(\mathbf{H}^{\mathbf{\Omega}}) \le q + O(a_{np}) + O(p^{-c})$. It concludes the proof of Theorem 1.

CSTR: 32290.14.JUSTC-2024-0075
DOI: 10.52396/JUSTC-2024-0075
JUSTC, **2025**, 55(X):