Functional knockoffs selection with applications to functional data analysis in high dimensions

Qizhai Li[1], Mingya Long[1,2], and Xinghao Qiao[2]

[1]*Academy of Mathematics and Systems Science CAS and University of Chinese Academy of Sciences, Beijing, China*

[2]*Department of Statistics, London School of Economics, London, U.K.*

Abstract

The knockoffs is a recently proposed powerful framework that effectively controls the false discovery rate (FDR) for variable selection. However, none of the existing knockoff solutions are directly suited to handle multivariate or high-dimensional functional data, which has become increasingly prevalent in various scientific applications. In this paper, we propose a novel functional model-X knockoffs selection framework tailored to sparse high-dimensional functional models, and show that our proposal can achieve the effective FDR control for any sample size. Furthermore, we illustrate the proposed functional model-X knockoffs selection procedure along with the associated theoretical guarantees for both FDR control and asymptotic power using examples of commonly adopted functional linear additive regression models and the functional graphical model. In the construction of functional knockoffs, we integrate essential components including the correlation operator matrix, the Karhunen-Loève expansion, and semidefinite programming, and develop executable algorithms. We demonstrate the superiority of our proposed methods over the competitors through both extensive simulations and the analysis of two brain imaging datasets.

Keywords: Correlation operator matrix; False discovery rate (FDR); Functional graphical model; Functional linear additive regression; Karhunen-Loève expansion; Power.

1 Introduction

Selecting important covariates associated with a response from a pool of potential candidates holds paramount importance across various scientific fields. At the same time, controlling the false discovery rate (FDR) offers an effective means to control error rates, ensuring replicable discoveries. A large body of literature on FDR control focuses on multiple testing approaches based on the p-values for assessing the significance of individual covariates; see, e.g., the seminal papers of Benjamini and Hochberg (1995); Benjamini and Yekutieli (2001). Yet, the high-dimensionality in covariates often renders many traditional approaches for p-value calculations inapplicable. Furthermore, none of the existing works along this vein directly address the problem of variable selection while simultaneously controlling the FDR.

The recent paper of Barber and Candès (2015) introduced a fixed-X knockoffs inference framework that effectively controls the FDR for variable selection in Gaussian linear model with the dimensionality p no larger than the sample size n under fixed designs. The key idea is to construct knockoff variables that mimic the dependence structure of original covariates while maintaining independence from the response conditional on the original covariates. It then compares the importance statistics (e.g., lasso coefficients) between the original covariates and knockoffs for variable selection. The fixed-X knockoffs inference has been extended to many settings, such as group-variable selection and multitask learning (Dai and Barber, 2016), high-dimensional linear model using data-splitting and feature screening (Barber and Candès, 2019) and Gaussian graphical model through a node-based local and a graph-based global procedure (Li and Maathuis, 2021).

More recently, Candès et al. (2018) proposed a model-X knockoffs extension that accommodates random design and allows for arbitrary and unknown conditional distribution of the response given the covariates, and for arbitrarily large p compared to n. The model-X knockoffs framework has witnessed a plethora of advancements. For instance, Fan et al. (2020a) developed a graphical nonlinear knockoffs method to handle the unknown covariate distribution, providing theoretical guarantees on the power and robustness. Fan et al. (2020b) applied knockoffs inference to high-dimensional latent factor models, enabling stable and intepretable forecasting. Dai et al. (2023) introduced a kernel knockoffs procedure for nonparametric additive models, employing subsampling and random feature mapping. See also Romano et al. (2020); Ren et al. (2023) and the references therein. Nevertheless, the existing efforts are primarily devoted to addressing scalar data. As a result, it remains unclear whether the model-X knockoffs framework is applicable to functional data.

The rapid development of data collection technology has led to an increased availability of multivariate or high-dimensional functional data datasets. Examples include time-course gene expression data and different types of neuroimaging data, where brain signals are measured over time at a multitude of regions of interest (ROIs). Building upon recent proposals (e.g., Zhu et al., 2016; Li and Solea, 2018; Fang et al., 2023), these signals are modelled as multivariate random functions with each ROI represented by a random function, where, under high-dimensional scaling, the number of functional variables p can be comparable to, or even exceed, the number of subjects n. To overcome the difficulties caused by high-dimensionality, various functional sparsity assumptions are commonly imposed on the model parameter space. E.g., scalar-on-function linear additive regression (SFLR) (Fan et al., 2015; Kong et al., 2016; Xue and Yao, 2021), function-on-function linear additive regression (FFLR) (Fan et al., 2014; Luo and Qi, 2017; Chang et al., 2023), functional linear discriminant analysis (Xue et al., 2023) as well as functional Gaussian graphical model (FGGM) (Qiao et al., 2019) and its various extensions (Solea and Li, 2022; Zapata et al., 2022; Lee et al., 2023). These models involve the development of sparse function-valued estimates in the sense of selecting important functional variables. However, none of the existing work has achieved the essential task of FDR control.

The major goal of our paper is to establish a methodological and theoretical foundation of model-X knockoffs selection for functional data and apply it to concrete examples of sparse high-dimensional functional models, thereby bridging an important gap in the respective fields. Specifically, we propose a functional model-X knockoffs selection framework that begins with dimension reduction via functional principal components analysis (FPCA), thus effectively converting infinite-dimensional curves into vector-valued FPC scores. We then compare group-lasso-based importance statistics between the estimated FPC scores of original and knockoff variables for variable selection. We demonstrate that our proposal is guaranteed to control the FDR below the nominal level regardless of n. We then showcase the proposed framework through three useful examples, i.e., SFLR, FFLR and FGGM, and, additionally, establish that the power for each model asymptotically approaches one as n goes to infinity. In constructing functional knockoffs, we integrate key ingredients: the correlation operator matrix, the Karhunen-Loève expansion, and semidefinite programming. We also develop executable algorithms using a coordinate representation system within finite-dimensional Hilbert spaces.

The main contribution of our paper is threefold. First, we propose a general functional model-X knockoffs selection framework. Arising from the initial dimension reduction, we have to deal with estimated FPC scores and truncation errors, whereas the conventional knockoffs is applied directly to observed data. Accounting for both estimation and truncation errors is a major undertaking in theoretical analysis. We show that the estimated vector-valued FPC scores possess two crucial properties: exchangeability and coin-flipping, which are pivotal in ensuring the effective FDR control of our proposal. By comparison, Dai and Barber (2016) developed a "truncation first" strategy that constructs group-knockoffs based on truncated FPC scores followed by group-lasso for variable selection. However, their FDR is limited to staying below the target level within the fixed-X rather than model-X knockoffs framework. Regarding the statistical power, our "knockoffs first" proposal constructs functional knockoffs before dimension reduction and group-variable selection, thus capturing more feature information and leading to improved power, as evidenced by simulation results in Section 5.

Second, we apply our proposal to two sparse functional linear additive regression models, i.e., SFLR and FFLR, effectively achieving the FDR controls. We also delve into the associated power properties, which pose greater challenges compared to non-functional models due to the inclusion of aforementioned additional errors. Furthermore, we integrate the fixed-X GGM knockoff filter (Li and Maathuis, 2021) into our functional model-X knockoffs framework to accommodate FGGM. Specifically, our proposal begins by locally constructing functional knockoffs and group-lasso coefficients for each node, and then solves a global optimization problem to determine nodewise thresholds for graph estimation. Compared to Qiao et al. (2019), we circumvent the estimation of unbounded inverse covariance functions by reformulating the graph estimation through penalized functional regressions on each functional variable against the remaining ones. Unlike Li and Maathuis (2021) which lacks theoretical power analysis, we establish the power guarantee for the more challenging task of functional model-X selection for FGGM in a high-dimensional regime.

Third, constructing knockoff variables is a pivotal step in implementing the knockoffs procedure, and our functional extension is far from incremental. One challenge lies in characterizing the dependence across infinite-dimensional objects and choosing suitable functional norm to quantify the strength of dependence. The natural functional extension, which seeks to minimize the average covariance operators in certain norms between the functional original and knockoff variables, is inappropriate, since the minimum eigenvalues of covariance operators converge to zero, thus making it exceedingly difficult to distinguish the original variables from knockoff counterparts. Motivated by the result that, under mild conditions, all eigenvalues of associated correlation operators remain bounded away from zero and infinity, we consider minimizing the average correlation operators in operator norm as opposed to other unbounded norms, which largely enhances the distinguishability and ensures good power in signal detection. The other challenges are to solve semidefinite programming problems and specify conditional distributions at the operator level with the aid of Karhunen-Loève expansions. To develop executable algorithms, we reformulate their sample versions by representing operators as matrices using the coordinate mapping.

Our paper is set out as follows. Section 2 proposes the functional model-X knockoffs selection framework. Section 3 applies the proposal to three examples, i.e., SFLR, FFLR and FGGM, and establishes the associated theoretical guarantees on the FDR and power. Section 4 presents the construction of functional knockoffs with executable algorithms. We demonstrate the superior finite-sample performance of our methods through simulations in Section 5 and the analysis of two brain imaging datasets in Section 6.

Notation. For a positive integer p, denote $[p] = \{1, \ldots, p\}$ and $\mathbf{I}_p$ as $p \times p$ identity matrix. For any vector $\mathbf{b} = (b_1, \ldots, b_p)^{\mathrm{T}}$, define $\|\mathbf{b}\| = (\sum_j b_j^2)^{1/2}$. For any matrix $\mathbf{B} = (B_{ij})_{p \times q}$, denote $\|\mathbf{B}\|_{\mathrm{F}} = (\sum_{i,j} B_{ij}^2)^{1/2}$ its Frobenius norm and $\mathbf{B}^{\dagger}$ its Moore-Penrose inverse. Let $L_2(\mathcal{U})$ be a Hilbert space of square-integrable functions on a compact interval $\mathcal{U}$ with inner product $\langle f, g \rangle = \int_{\mathcal{U}} f(u)g(u)du$ and norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$ for $f, g \in L_2(\mathcal{U})$. For $j \in [p]$, we take a separable Hilbert space $\mathcal{H}_j \subseteq L_2(\mathcal{U})$. For a compact linear operator K from $\mathcal{H}_j$ to $\mathcal{H}_k$ induced from the kernel function K with $K(f)(u) = \int_{\mathcal{U}} K(u, v)f(v)\mathrm{d}v \in \mathcal{H}_k$ for $f \in \mathcal{H}_j$, there exist two orthonormal bases $\{\phi_l\}$ and $\{\psi_l\}$ of $\mathcal{H}_j$ and $\mathcal{H}_k$,, respectively, and a sequence $\{\lambda_l\}$ in $\mathbb{R}$ tending to zero, such that K has the spectral decomposition $K = \sum_{l=1}^{\infty} \lambda_l \phi_l \otimes \psi_l$, where $\otimes$ denotes the tensor product. For notational economy, we will use K to denote both the operator and kernel function. We denote its Hilbert–Schmidt norm by $\|K\|_{\mathcal{S}} = (\sum_{l=1}^{\infty} \lambda_l^2)^{1/2} = \{\int\int K^2(u, v)\mathrm{d}u\mathrm{d}v\}^{1/2}$, nuclear norm by $\|K\|_{\mathcal{N}} = \sum_{l=1}^{\infty} |\lambda_r|$ and operator norm by $\|K\|_{\mathcal{L}} = \sup_{\|f\| \leqslant 1, f \in \mathcal{H}_j} \|K(f)\|$. Let $\boldsymbol{\mathcal{H}}$ be the Cartesian product of $\mathcal{H}_1, \ldots, \mathcal{H}_p$ with inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{j=1}^{p} \langle f_j, g_j \rangle$ for $\mathbf{f} = (f_1, \ldots, f_p)^{\mathrm{T}}$ and $\mathbf{g} = (g_1, \ldots, g_p)^{\mathrm{T}} \in \boldsymbol{\mathcal{H}}$. An operator matrix $\mathbf{K} = (K_{jk})$ is a $p \times p$ matrix of operators with its (j, k)th operator-valued entry K_{jk}, and can be thought of an operator from $\boldsymbol{\mathcal{H}}$ to $\boldsymbol{\mathcal{H}}$ with $\mathbf{K}(\mathbf{f}) = \big(\sum_{j=1}^{p} K_{1j}(f_j), \ldots, \sum_{j=1}^{p} K_{pj}(f_j)\big)^{\mathrm{T}}$ for $\mathbf{f} \in \boldsymbol{\mathcal{H}}$. We use $\mathbf{K} \geq 0$ to denote a positive semidefinite operator matrix satisfying $\langle \mathbf{K}(\mathbf{f}), \mathbf{f} \rangle \geqslant 0$ for any $\mathbf{f} \in \boldsymbol{\mathcal{H}}$. For $a, b \in \mathbb{R}$, we use $a \vee b = \max\{a, b\}$. For two sequences of positive numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if there exists some constant $c > 0$ such that $a_n \leqslant cb_n$.

2 Functional model-X knockoffs selection framework

2.1 Definition

Let $\mathbf{X}(\cdot) = \big(X_1(\cdot), \ldots, X_p(\cdot)\big)^{\mathrm{T}}$ be a random element in $\boldsymbol{\mathcal{H}}$. Before framing the variable selection in the context of sparse functional models with Y being a scalar or functional response, we define a functional covariate $X_j(\cdot)$ as null if and only if Y is independent of $X_j(\cdot)$ conditional on the remaining functional covariates $X_{-j}(\cdot) = \{X_1(\cdot), \ldots, X_p(\cdot)\} \backslash \{X_j(\cdot)\}$

and as nonnull otherwise. Let S^c denote the index set of null functional covariates, i.e.,

$$S^c = \big\{ j \in [p] : X_j(\cdot) \text{ is independent of } Y \text{ conditional on } X_{-j}(\cdot) \big\}, \tag{1}$$

and hence the index set of nonnull functional covariates is given by S, the complement of S^c. This formulation naturally establishes the equivalence between the selection of null functional covariates and functional variable selection. E.g., it follows from Lemmas A2 and A9 of the Supplementary Material that S^c is the same as the set $\{j \in [p] : \|\beta_j\| = 0\}$ in (7) for SFLR or $\{j \in [p] : \|\beta_j\|_{\mathcal{S}} = 0\}$ in (15) for FFLR. Our goal is to discover as many nonnull functional covariates as possible while controlling the FDR, defined as

$$\text{FDR} = \mathbb{E}\left[\frac{|\widehat{S} \cap S^c|}{|\widehat{S}| \vee 1} \right],$$

where $\widehat{S}$ represents the index set of functional covariates identified by the variable selection procedure, and $|\cdot|$ denotes the cardinality of a set.

The key ingredient of functional model-X knockoffs selection framework is the construction of functional model-X knockoffs, which is defined as follows.

Definition 1. *Functional model-X knockoffs for the family of random functions $\mathbf{X}(\cdot)$ are a new family of random functions $\widetilde{\mathbf{X}}(\cdot) = \big(\widetilde{X}_1(\cdot), \ldots, \widetilde{X}_p(\cdot)\big)^{\mathrm{T}} \in \boldsymbol{\mathcal{H}}$ that satisfies the following two properties: (i) $\big(\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big)_{swap(G)} \overset{D}{=} \big(\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big)$ for any subset $G \subseteq [p]$, where swap(G) means swapping components $X_j(\cdot)$ and $\widetilde{X}_j(\cdot)$ for each $j \in G$ and $\overset{D}{=}$ denotes the equality in distribution. (ii) $\widetilde{\mathbf{X}}(\cdot)$ and Y are independent conditionally on $\mathbf{X}(\cdot)$.*

Definition 1 generalizes the definition of model-X knockoffs (Candès et al., 2018) within Hilbert spaces. Property (ii) is fulfilled when constructing the knockoffs $\widetilde{\mathbf{X}}(\cdot)$ without any reference to Y, and Property (i) corresponds to the pairwise exchangability between the original and knockoff variables. Before giving an example obeying this property, we introduce some notation. For each $j, k \in [p]$, denote the mean of X_j as μ_j and the covariance operator between X_j and X_k as $\Sigma_{X_j X_k} = \text{Cov}(X_j, X_k) = \mathbb{E}\{(X_j - \mu_j) \otimes (X_k - \mu_k)\}$, which has a one-to-one correspondence with the covariance function $\Sigma_{X_j X_k}(u, v) = \text{Cov}\{X_j(u), X_k(v)\}$ for $(u, v) \in \mathcal{U}^2$. Denote the $p \times p$ covariance operator matrix of $\mathbf{X}$ as $\boldsymbol{\Sigma}_{XX}$, whose (j, k)th entry is operator $\Sigma_{X_jX_k}$. The cross-covariance operator matrix $\mathbf{\Sigma}_{X\tilde{X}}$ between $\mathbf{X}$ and $\widetilde{\mathbf{X}}$ can be defined similarly. We will utilize the example below as the way of constructing functional knockoffs in Section 4.

Example 1. *Suppose that* $\mathbf{X}$ *follows a multivariate Gaussian process (MGP) with mean zero and covariance* $\mathbf{\Sigma}_{XX}$*, denoted as MGP*$(\mathbf{0}, \mathbf{\Sigma}_{XX})$*. Then* $(\mathbf{X}^{\mathrm{T}}, \widetilde{\mathbf{X}}^{\mathrm{T}})^{\mathrm{T}} \sim$ *MGP*$(\mathbf{0}, \mathbf{\Sigma}_{(X,\tilde{X})(X,\tilde{X})})$ *satisfies Property (i), where*

$$\mathbf{\Sigma}_{XX} = \mathbf{\Sigma}_{\tilde{X}\tilde{X}}, \quad \mathbf{\Sigma}_{X\tilde{X}} = \mathbf{\Sigma}_{\tilde{X}X} = \mathbf{\Sigma}_{XX} - \mathbf{Q}_{XX}, \tag{2}$$

and $\mathbf{Q}_{XX} = \textit{diag}(Q_{X_1X_1}, \ldots, Q_{X_pX_p})$ *is selected in such a way that* $\mathbf{\Sigma}_{(X,\tilde{X})(X,\tilde{X})} \succeq 0$.

Suppose that we observe n i.i.d. realizations $\{\mathbf{X}_i(\cdot), Y_i\}_{i\in[n]}$ from the population $\{\mathbf{X}(\cdot), Y\}$. Due to the infinite-dimensionality of functional data, we adopt the standard dimension reduction approach by performing Karhunen-Loève expansions of $X_{ij}(\cdot)$ and $\widetilde{X}_{ij}(\cdot)$ for each j and truncating the expansions to the first d_j terms, which serves as the foundation of FPCA:

$$X_{ij}(\cdot) - \mu_j(\cdot) = \sum_{l=1}^{\infty} \xi_{ijl}\phi_{jl}(\cdot) \approx \boldsymbol{\xi}_{ij}^{\mathrm{T}}\boldsymbol{\phi}_j(\cdot), \quad \widetilde{X}_{ij}(\cdot) - \mu_j(\cdot) = \sum_{l=1}^{\infty} \tilde{\xi}_{ijl}\phi_{jl}(\cdot) \approx \widetilde{\boldsymbol{\xi}}_{ij}^{\mathrm{T}}\boldsymbol{\phi}_j(\cdot), \tag{3}$$

where $\boldsymbol{\phi}_j = (\phi_{j1}, \ldots, \phi_{jd_j})^{\mathrm{T}}$, $\boldsymbol{\xi}_{ij} = (\xi_{ij1}, \ldots, \xi_{ijd_j})^{\mathrm{T}}$ and $\widetilde{\boldsymbol{\xi}}_{ij} = (\tilde{\xi}_{ij1}, \ldots, \tilde{\xi}_{ijd_j})^{\mathrm{T}}$. Here $\xi_{ijl} = \langle X_{ij} - \mu_j, \phi_{jl}\rangle$ (or $\tilde{\xi}_{ijl} = \langle \widetilde{X}_{ij} - \mu_j, \phi_{jl}\rangle$), namely FPC score of original (or knockoff) variables, corresponds to a sequence of random variables with $\mathbb{E}(\xi_{ijl}) = 0$ and $\mathrm{Cov}(\xi_{ijl}, \xi_{ijl'}) = \omega_{jl}I(l = l')$, where $\omega_{j1} \geqslant \omega_{j2} \geqslant \cdots > 0$ are the eigenvalues of $\Sigma_{X_jX_j}$ and $\phi_{j1}(\cdot), \phi_{j2}(\cdot), \ldots$ are the corresponding eigenfunctions. To implement FPCA based on n observations, we compute the sample estimator of $\Sigma_{X_jX_j}$ via $\widehat{\Sigma}_{X_jX_j} = n^{-1}\sum_{i=1}^n (X_{ij} - \hat{\mu}_j) \otimes (X_{ij} - \hat{\mu}_j)$ with $\hat{\mu}_j = n^{-1}\sum_{i=1}^n X_{ij}$ and then carry out an eigenanalysis of $\widehat{\Sigma}_{X_jX_j}$ that leads to estimated eigenvalue/eigenvector pairs $\{\hat{\omega}_{jl}, \hat{\phi}_{jl}(\cdot)\}_{l\in[d_j]}$. We then obtain estimated FPC scores $\hat{\xi}_{ijl} = \langle X_{ij} - \hat{\mu}_j, \hat{\phi}_{jl}\rangle$ and $\check{\xi}_{ijl} = \langle \widetilde{X}_{ij} - \hat{\mu}_j, \hat{\phi}_{jl}\rangle$ for $l \in [d_j]$. Let $\widehat{\boldsymbol{\xi}}_{ij} = (\hat{\xi}_{ij1}, \ldots, \hat{\xi}_{ijd_j})^{\mathrm{T}}$, $\check{\boldsymbol{\xi}}_{ij} = (\check{\xi}_{ij1}, \ldots, \check{\xi}_{ijd_j})^{\mathrm{T}}$, $\widehat{\boldsymbol{\xi}}_i = (\widehat{\boldsymbol{\xi}}_{i1}^{\mathrm{T}}, \ldots, \widehat{\boldsymbol{\xi}}_{ip}^{\mathrm{T}})^{\mathrm{T}}$ and $\check{\boldsymbol{\xi}}_i = (\check{\boldsymbol{\xi}}_{i1}^{\mathrm{T}}, \ldots, \check{\boldsymbol{\xi}}_{ip}^{\mathrm{T}})^{\mathrm{T}}$. Resulting from the dimension reduction, the estimation of sparse function-valued parameters based on $\{\mathbf{X}_i(\cdot)^{\mathrm{T}}, Y_i\}_{i\in[n]}$ is transformed to the block sparse estimation of parameter vectors/matrices based on vector-valued estimated FPC scores $\{\widehat{\boldsymbol{\xi}}_i\}_{i\in[n]}$ and transformed responses $\{\widetilde{Y}_i\}_{i\in[n]}$, where, e.g., $\widetilde{Y}_i$ equals Y_i for SFLR in Section 3.1 and estimated FPC scores of $Y_i(\cdot)$ for FFLR in Section 3.2.

We next present the exchangeability condition, i.e., swapping estimated FPC scores of null functional covariates with those of corresponding functional knockoffs will not affect the joint distribution of $\widehat{\boldsymbol{\xi}}_i$ and $\check{\boldsymbol{\xi}}_i$ conditional on $\widetilde{Y}_i$.

Condition 1. *For any subset $G \subseteq S^c$ and $i \in [n]$, $\big(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \check{\boldsymbol{\xi}}_i^{\mathrm{T}}\big)|\widetilde{Y}_i \overset{D}{=} \big(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \check{\boldsymbol{\xi}}_i^{\mathrm{T}}\big)_{swap(G)}|\widetilde{Y}_i$.*

This condition can be validated across three examples we consider by applying the functional exachangeability result in Lemma A1 of the Supplementary Material, which is built upon the properties in Definition 1. It plays a crucial role in establishing the coin-flipping property in Lemma 1 below.

2.2 Feature statistics

To select the nonnull functional variables, we compute knockoff statistics $W_j = w_j(Z_j, \widetilde{Z}_j)$ for each $j \in [p]$, where w_j is antisymmetric function satisfying $w_j(\cdot, *) = -w_j(*, \cdot)$, and Z_j and $\widetilde{Z}_j$ respectively represent the feature importance measure of estimated FPC scores of $\{X_{ij}\}_{i\in[n]}$ and $\{\widetilde{X}_{ij}\}_{i\in[n]}$. For three examples we consider, we can choose Z_j and $\widetilde{Z}_j$ as the group-lasso coefficient vectors/matrices under the vector ℓ_2/matrix Frobenius norm of $\{\widehat{\boldsymbol{\xi}}_{ij}\}_{i\in[n]}$ and $\{\check{\boldsymbol{\xi}}_{ij}\}_{i\in[n]}$, respectively, and a valid knockoff statistic is $w_j(Z_j, \widetilde{Z}_j) = Z_j - \widetilde{Z}_j$. Intuitively, a large positive value of W_j suggests that $X_j(\cdot)$ is an important (i.e., nonnull) feature, while small magnitudes of W_j often correspond to unimportant (i.e., null) features. As noted in the Candès et al. (2018), the knockoffs selection can control the FDR when the feature importance measures W_j's possess the essential coin flipping property below.

Lemma 1. *Suppose that Condition 1 holds. Let $(\delta_1, \ldots, \delta_p)$ be a sequence of independent random variables such that $\delta_j = \pm 1$ with a probability of $1/2$ if $j \in S^c$, and $\delta_j = 1$ otherwise. Then $(W_1, \ldots, W_p) \overset{D}{=} (\delta_1 W_1, \ldots, \delta_p W_p)$, conditional on $(|W_1|, \ldots, |W_p|)$.*

Hence a large positive value of W_j provides evidence that $j \in S$, whereas, under $j \in S^c$, W_j is equally likely to be positive and negative. Notice that Lemma 1 is established on the sets of null/nonnull functional covariates, and fails to hold for the corresponding truncated sets specified in (11) below. Nevertheless, this does not place a constraint to control the FDR, as justified in Theorem 1 below.

The last step of our functional knockoffs selection framework is to apply the knockoff filter (Candès et al., 2018) by ranking W_j's from large to small and selecting features whose associated W_j's are at least some threshold T_δ in (5). This results in estimated nonnull sets

$$\widehat{S}_\delta = \{j \in [p] : W_j \geqslant T_\delta\}, \quad \delta = 0 \text{ or } 1. \tag{4}$$

To select a data-driven threshold as permissive as possible while still managing the control over the FDR, we choose the threshold in the following two ways:

$$T_\delta = \min\left\{t \in \{|W_j| > 0 : j \in [p]\} : \frac{\delta + |\{j : W_j \leqslant -t\}|}{|\{j : W_j \geqslant t\}|} \leqslant q\right\}, \tag{5}$$

where T_0 is used for knockoff filter and T_1 is used for more conservative knockoff+ filter. The false discovery is measured by both FDR based on T_1 and modified FDR based on T_0, which are respectively defined as,

$$\text{FDR} = \mathbb{E}\left[\frac{|\widehat{S}_1 \cap S^c|}{|\widehat{S}_1| \vee 1}\right] \quad \text{and} \quad \text{mFDR} = \mathbb{E}\left[\frac{|\widehat{S}_0 \cap S^c|}{|\widehat{S}_0| + 1/q}\right].$$

We are now ready to present a theorem regarding the effective FDR control.

Theorem 1. *Suppose that Condition 1 holds. For any sample size n and target FDR level $q \in [0, 1]$, the selected set $\widehat{S}_1$ satisfies FDR $\leqslant q$ and the selected set $\widehat{S}_0$ satisfies mFDR $\leqslant q$..*

Remark 1. *Resulting from the dimension reduction step, we are confronted with estimated FPC scores and bias terms formed by truncation errors. However, Theorem 1 makes it evident that neither estimation errors nor truncation errors affect the effectiveness of FDR control, which remains valid regardless of truncated dimensions d_j's. This is because Condition 1, concerning estimated FPC scores, serves as the foundation for proving Theorem 1 and can be verified for any values of d_j's without regard to truncation errors.*

Remark 2. *Our proposed "knockoffs first" framework begins with constructing functional knockoffs before performing dimension reduction and knockoff filter to the group-lasso coefficients for variable selection. Theorem 1 ensures that our proposal effectively controls FDR* *when the null set S^c is defined in (1). By comparison, an alternative "truncation first" approach applies fixed-X group knockoffs selection (Dai and Barber, 2016) by constructing vector-valued knockoffs based on vector-valued estimated FPC scores, followed by adopting a group-lasso-based knockoff filter for variable selection. However, it is important to note that the presence of functional versions of the conditional independence in both the null set S^c and Definition 1's Property (ii) does not imply the corresponding truncated (sample) versions of the conditional independence, which are essential to ensure FDR control within the model-X framework. We thus formulate the "truncation first" strategy within the fixed-X framework to accommodate linear model settings with $n \geqslant \sum_{j=1}^{p} d_j$. For instance, in Section 3.1 for SFLR, we define the corresponding null set S^c as the complement of S in (8), and the truncated version of the null set $\widetilde{S}^c$ in (11) with $S^c \subseteq \widetilde{S}^c$. This allows us to employ the group knockoffs selection (Dai and Barber, 2016) that results in the effective FDR control via*

$$\underbrace{\mathbb{E}\left[\frac{|\widehat{S} \cap S^c|}{|\widehat{S}| \vee 1}\right]}_{\text{functional version of FDR}} \leqslant \underbrace{\mathbb{E}\left[\frac{|\widehat{S} \cap \widetilde{S}^c|}{|\widehat{S}| \vee 1}\right]}_{\text{truncated version of FDR}} \leqslant q, \tag{6}$$

where $\widehat{S}$ denotes the set of selected variables.

Compared to our "functional knockoffs" proposal with both FDR and power guarantees, it is evident from Remarks 2 and 4 below that the "truncation first" strategy ensures FDR control but results in empirical power with possibly slower asymptotic rate of convergence for SFLR. These findings also hold true for FFLR and FGGM, and are consistent with our simulation results in Section 5.

3 Applications

3.1 High-dimensional SFLR

Consider the high-dimensional SFLR model

$$Y_i = \sum_{j=1}^{p} \int_{\mathcal{U}} \beta_j(u) X_{ij}(u) du + \varepsilon_i, \quad i \in [n], \tag{7}$$

where $\{\varepsilon_i\}_{i\in[n]}$ are i.i.d. mean-zero random errors, independent of mean-zero $\{X_{ij}(\cdot)\}_{i\in[n],j\in[p]}$. The function-valued coefficients $\{\beta_j(\cdot)\}_{j\in[p]}$ are to be estimated and are assumed to be functional s-sparse with support

$$S = \{j \in [p] : \|\beta_j\| \neq 0\} \tag{8}$$

and cardinality $s = |S| \ll p$. Our target is to select important functional variables (i.e., recovery of support S) while simultaneously controlling FDR. To integrate this task into our functional model-X knockoffs selection framework, we give the following condition.

Condition 2. *For any $j \in [p]$ and any bivariate functions $\gamma_k \in L_2(\mathcal{U}) \otimes L_2(\mathcal{U})$ with $k \in [p]\backslash\{j\}$, $X_{ij}(u) \neq \sum_{k\neq j}\int_{\mathcal{U}} \gamma_k(u,v)X_{ik}(v)dv$ for $i \in [n]$.*

Under Condition 2, Lemma A2 of the Supplementary Material establishes the equivalence between the nonnull set (1) and the support set (8) for SFLR. For each $j \in [p]$, we expand $X_{ij}(\cdot)$ according to (3). Some specific calculations lead to the representation of (7) as

$$Y_i = \sum_{j=1}^{p} \boldsymbol{\xi}_{ij}^{\mathrm{T}} \mathbf{b}_j + \epsilon_i + \varepsilon_i, \tag{9}$$

where the jth coefficient vector $\mathbf{b}_j = \int_{\mathcal{U}} \boldsymbol{\phi}_j(u)\beta_j(u)\mathrm{d}u \in \mathbb{R}^{d_j}$ and $\epsilon_i = \sum_{j=1}^{p}\sum_{l=d_j+1}^{\infty} \xi_{ijl}b_{jl}$ is the truncation error. Hence, we can rely on the group sparsity pattern in $\{\mathbf{b}_j\}_{j\in[p]}$ to recover the functional sparsity structure in $\{\beta_j(\cdot)\}_{j\in[p]}$. Within functional knockoffs framework, we denote the augmented coefficient vectors of FPC scores by $\{\mathbf{b}_j\}_{j\in[2p]}$ (the first p coefficient vectors are for the original covariates and the last p are for the knockoffs).

We initiate by adopting FPCA on $\{X_{ij}(\cdot)\}_{i\in[n]}$ for each j and obtain vector-valued estimated FPC scores of original and knockoff covariates (i.e., $\widehat{\boldsymbol{\xi}}_{ij}$'s and $\breve{\boldsymbol{\xi}}_{ij}$'s). We then estimate $\{\mathbf{b}_j\}_{j\in[2p]}$ via the group-lasso regression on the augmented set of estimated FPC scores

$$\min_{\mathbf{b}_1,\ldots,\mathbf{b}_{2p}} \frac{1}{2n}\sum_{i=1}^{n}\Big(Y_i - \sum_{j=1}^{p}\widehat{\boldsymbol{\xi}}_{ij}^{\mathrm{T}}\mathbf{b}_j - \sum_{j=p+1}^{2p}\breve{\boldsymbol{\xi}}_{i(j-p)}^{\mathrm{T}}\mathbf{b}_j\Big)^2 + \lambda_n\sum_{j=1}^{2p}\|\mathbf{b}_j\|, \tag{10}$$

where $\lambda_n \geqslant 0$ is the regularization parameter. Denote the solution to (10) by $\{\widehat{\mathbf{b}}_j\}_{j\in[2p]}$. Complying with the knockoff selection step in Section 2.2, we choose the jth feature importance measures by $Z_j = \|\widehat{\mathbf{b}}_j\|$ and $\widetilde{Z}_j = \|\widehat{\mathbf{b}}_{j+p}\|$ and the corresponding knockoff statistics is

$W_j = \|\widehat{\mathbf{b}}_j\| - \|\widehat{\mathbf{b}}_{j+p}\|$. Hence we obtain the set of selected functional covariates $\widehat{S}_\delta$ by applying the knockoff filter to $\{W_j\}_{j\in[p]}$ in (4), where the threshold T_δ is determined by (5).

With the constructed functional model-X knockoffs satisfying Definition 1, we can show that the estimated FPC scores of original and knockoff variables fulfill Condition 1. Then an application of Theorem 1 leads to the following theorem, which achieves the valid FDR control for SFLR without any constraint on the dimensionality p relative to the sample size n. As such, our proposal works for both $p < n$ and $p > n$ scenarios.

Theorem 2. *Suppose that Condition 2 holds. Then for any sample size n and target FDR level $q \in [0, 1]$, $\widehat{S}_1$ satisfies FDR $\leqslant q$ and $\widehat{S}_0$ satisfies mFDR $\leqslant q$.*

Remark 3. *As discussed in Remark 2, we formulate the "truncation first" strategy for SFLR within the fixed-X group knockoffs framework (Dai and Barber, 2016). Referring to (9), we can represent Y_i linearly as $Y_i = \sum_{j=1}^{p} \widehat{\boldsymbol{\xi}}_{ij}^{\mathrm{T}} \mathbf{b}_j + e_i$, where the error term e_i encompasses truncation, estimation and random errors. Define the corresponding null set as*

$$\widetilde{S}^c = \{j \in [p] : \|\mathbf{b}_j\| = 0\}. \tag{11}$$

Treating $\{\widehat{\boldsymbol{\xi}}_{ij}\}_{j\in[p]}$ as original group covariates, we can follow Dai and Barber (2016) to construct group knockoffs, then choose group-lasso-based coefficient vectors under the ℓ_2 norm as feature importance measures and finally apply the knockoff filter for group-variable selection. According to (6), the selected set by the "truncation first" approach achieves the FDR control. However, it may lead to declined power compared to our proposal, as argued in Remark 4.

Before asymptotic power analysis of our approach, we impose some regularity conditions.

Condition 3. *$\{\varepsilon_i\}_{i\in[n]}$ with finite variance σ^2 are i.i.d. sub-Gaussian variables, i.e., there exists some constant c such that $\mathbb{E}[e^{x\varepsilon_i}] \leqslant e^{c^2\sigma^2x^2/2}$ for any $x \in \mathbb{R}$.*

Condition 4. *For $(\mathbf{X}, \widetilde{\mathbf{X}}) \in \boldsymbol{\mathcal{H}}^2$, we denote a diagonal operator matrix by $\mathbf{D}_{(X,\tilde{X})(X,\tilde{X})} = \mathrm{diag}(\Sigma_{X_1X_1}, \ldots, \Sigma_{X_pX_p}, \Sigma_{\tilde{X}_1\tilde{X}_1}, \ldots, \Sigma_{\tilde{X}_p\tilde{X}_p})$. The infimum*

$$\underline{\mu} = \inf_{\boldsymbol{\Phi}\in\boldsymbol{\mathcal{H}}_0^2} \frac{\langle \boldsymbol{\Phi}, \boldsymbol{\Sigma}_{(X,\tilde{X})(X,\tilde{X})}(\boldsymbol{\Phi})\rangle}{\langle \boldsymbol{\Phi}, \mathbf{D}_{(X,\tilde{X})(X,\tilde{X})}(\boldsymbol{\Phi})\rangle}$$

is bounded away from 0, where $\boldsymbol{\mathcal{H}}_0^2 = \{\boldsymbol{\Phi} \in \boldsymbol{\mathcal{H}}^2 : \langle \boldsymbol{\Phi}, \mathbf{D}_{(X,\tilde{X})(X,\tilde{X})}(\boldsymbol{\Phi})\rangle \in (0, \infty)\}$.

Condition 5. *For each* $j \in [p]$, $\omega_{j1} > \omega_{j2} > \cdots > 0$, *and* $\max_{j\in[p]} \sum_{l=1}^{\infty} \omega_{jl} = O(1)$. *There exists some constant* $\alpha > 1$ *such that* $\omega_{jl} - \omega_{j(l+1)} \gtrsim l^{-\alpha-1}$ *for* $l = 1, \ldots, \infty$.

Condition 6. *(i) For each* $j \in S$, *there exists some constant* $\tau > \alpha/2+1$ *such that* $|b_{jl}| \lesssim l^{-\tau}$ *for* $l \geqslant 1$; *(ii)* $\min_{j\in S} \|\mathbf{b}_j\| \geqslant \kappa_n d^{\alpha} \lambda_n / \underline{\mu}$ *for some slowly diverging sequence* $\kappa_n \to \infty$ *as* $n \to \infty$, *and the regularization parameter* $\lambda_n \gtrsim s[d^{\alpha+2}\{\log(pd)/n\}^{1/2} + d^{1-\tau}]$; *(iii) There exists some constant* $c_1' \in (2(qs)^{-1}, 1)$ *such that* $|S_2| \geqslant c_1' s$ *with* $S_2 = \{j \in [p] : \|\mathbf{b}_j\| \gg s^{1/2} d^{\alpha} \lambda_n\}$.

To simplify notation, we assume the same d across $j \in [p]$ in the power analysis in Sections 3.1, 3.2 and 3.3, but our theoretical results extend naturally to the more general setting where d_j's are different. Condition 4 can be interpreted as requiring the minimum eigenvalue of the correlation operator matrix of $(\mathbf{X}^{\mathrm{T}}, \widetilde{\mathbf{X}}^{\mathrm{T}})^{\mathrm{T}}$ to be bounded away from zero. See similar conditions in Fan et al. (2020a); Dai et al. (2023) on the minimal eigenvalue of the corresponding covariance matrix, whose functional extension fails to hold as the infimum of the covariance operator matrix is zero. Conditions 5 and 6(i) are standard in functional regression literature (e.g., Kong et al., 2016) with parameter α capturing the tightness of gaps between adjacent eigenvalues and parameter τ controlling the level of smoothness in nonzero coefficient functions. Condition 6(ii) requires the ℓ_2 norms of nonzero coefficient vectors exceed a certain threshold, which ensures that the selected set contains the majority of important variables. Given that our knockoffs selection is built upon the group lasso, its power is upper bounded by that of the group lasso, which approaches one as $n \to \infty$ under this condition. Specifically, in the proof of Theorem 3, we obtain that, with high probability and $\widehat{S}_{\mathrm{GL}}^c = \{j \in [p] : \|\widehat{\mathbf{b}}_j\| = 0\}$, $|\widehat{S}_{\mathrm{GL}}^c \cap S| \min_{j\in S} \|\mathbf{b}_j\| \leqslant \sum_{j\in(\widehat{S}_{\mathrm{GL}}^c \cap S)} \|\mathbf{b}_j\| = \sum_{j\in(\widehat{S}_{\mathrm{GL}}^c \cap S)} \|\mathbf{b}_j - \widehat{\mathbf{b}}_j\| \leqslant \sum_{j=1}^{p} \|\mathbf{b}_j - \widehat{\mathbf{b}}_j\| = O(sd^{\alpha}\lambda_n/\underline{\mu})$. Then Condition 6(ii) implies that $|\widehat{S}_{\mathrm{GL}}^c \cap S| = O(s\kappa_n^{-1})$ and hence the group lasso exhibits asymptotic power one. Condition 6(iii) requires a large enough suitable subset of S that contains relatively strong signals to attain high power. See similar conditions in Fan et al. (2020b); Dai et al. (2023).

We are now ready to characterize the power of our proposal, which is defined as

$$\mathrm{Power}(\widehat{S}_{\delta}) = \mathbb{E}\left[\frac{|\widehat{S}_{\delta} \cap S|}{|S| \vee 1}\right]. \tag{12}$$

Theorem 3. *Suppose that Conditions 2–6 hold and $sd^{\alpha}\lambda_n \to 0$. Then the selected sets S_δ's satisfy Power$(\widehat{S}_\delta) \to 1$ as $n \to \infty$.*

Theorem 3 establishes the asymptotic power guarantee for both scenarios of $p < n$ and $p > n$, more specifically $n < p \lesssim e^{n/d^{4\alpha+4}}/d$ under a high-dimensional regime.

Remark 4. *In the proof of Theorem 3, we show that, with high probability,*

$$\frac{|\widehat{S}_\delta \cap S|}{|S| \vee 1} \geqslant 1 - O(\kappa_n^{-1}). \tag{13}$$

By comparison, we present a heuristic argument about the power of the "truncation first" strategy, employing the same sets $S, \widetilde{S}$ and $\widehat{S}$ as specified in Remarks 2–3. Under conditions similar to those in Fan et al. (2020a) within the fixed-X framework, it is expected that the truncated version of Power $= \mathbb{E}\left[\frac{|\widehat{S}\cap\widetilde{S}|}{|\widetilde{S}|\vee 1}\right] \to 1$, which holds in the sense of $|\widehat{S} \cap \widetilde{S}| \geqslant |\widetilde{S}|\{1 - O(\kappa_n^{-1})\}$ with high probability. This, together with the fact $\widetilde{S} \to S$ as $d, n \to \infty$, yields that the functional version of Power $= \mathbb{E}\left[\frac{|\widehat{S}\cap S|}{|S|\vee 1}\right] \to 1$, which holds since, with high probability,

$$\frac{|\widehat{S} \cap S|}{|S| \vee 1} \geqslant \frac{|\widetilde{S}|}{|S| \vee 1}\{1 - O(\kappa_n^{-1})\} \geqslant 1 - O(\kappa_n^{-1}) - O\big(h(d)\big). \tag{14}$$

Here $h(d) = \{1 - |\widetilde{S}|/(|S| \vee 1)\} \to 0$ as $d, n \to \infty$. Although both methods are guaranteed with asymptotic power one, the asymptotic rate for the empirical power of our "knockoffs first" proposal in (13) is not slower than that of the "truncation first" competitor in (14).

3.2 High-dimensional FFLR

Consider the high-dimensional FFLR model

$$Y_i(v) \quad = \quad \textstyle\sum_{j=1}^{p} \int_{\mathcal{U}} X_{ij}(u)\beta_j(u,v)\mathrm{d}u + \varepsilon_i(v), \;\; i \in [n], \;\; v \in \mathcal{V}, \tag{15}$$

where $\{\varepsilon_i(\cdot)\}_{i\in[n]}$ are i.i.d. mean-zero random error functions, independent of mean-zero $\{X_{ij}(\cdot)\}_{i\in[n],j\in[p]}$, and $\{\beta_j(\cdot,\cdot)\}_{j\in[p]}$ are functional coefficients to be estimated with support

$$S = \{j \in [p] : \|\beta_j\|_{\mathcal{S}} \neq 0\} \quad \text{and} \quad s = |S| \ll p. \tag{16}$$

Our target is to identify the support S and control FDR at the same time within our proposed framework in Section 2. To achieve this, we establish in Lemma A9 of the Supplementary Material that the nonnull set (1) is equivalent to the support set (16) for FFLR.

We follow (3) to expand $X_{ij}(\cdot)$ for each i, j. We also approximate $\{Y_i(\cdot)\}$ under the Karhunen-Loève expansion truncated at $\tilde{d}$, i.e., $Y_i(\cdot) \approx \boldsymbol{\eta}_i^{\mathrm{T}}\boldsymbol{\psi}(\cdot)$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{\tilde{d}})^{\mathrm{T}}$, $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{i\tilde{d}})^{\mathrm{T}}$. Some specific calculations lead to the representation of (15) as

$$\boldsymbol{\eta}_i^{\mathrm{T}} = \sum_{j=1}^{p} \boldsymbol{\xi}_{ij}^{\mathrm{T}}\mathbf{B}_j + \boldsymbol{\epsilon}_i^{\mathrm{T}} + \boldsymbol{\varepsilon}_i^{\mathrm{T}},$$

where $\mathbf{B}_j = \int_{\mathcal{U}\times\mathcal{V}} \boldsymbol{\phi}_j(u)\beta_j(u,v)\boldsymbol{\psi}(v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v \in \mathbb{R}^{d_j\times\tilde{d}}$, and $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\varepsilon}_i$ represent truncation and random errors, respectively. Hence, we can utilize the group sparsity pattern in $\{\mathbf{B}_j\}_{j\in[p]}$ to identify the functional sparsity structure in $\{\beta_j(\cdot,\cdot)\}_{j\in[p]}$. Within functional knockoffs framework, we denote the augmented coefficient matrices of FPC scores as $\{\mathbf{B}_j\}_{j\in[2p]}$ (the first p coefficient matrices are for original covariates and the last p are for knockoffs).

We start by performing FPCA on $\{Y_i(\cdot)\}_{i\in[n]}$ and $\{X_{ij}(\cdot)\}_{i\in[n]}$ for each j, resulting in estimated FPC scores for the response, original covariates, and knockoffs, respectively, denoted as $\widehat{\boldsymbol{\eta}}_i$, $\widehat{\boldsymbol{\xi}}_{ij}$, and $\breve{\boldsymbol{\xi}}_{ij}$ for $i \in [n], j \in [p]$. We then implement group lasso to estimate $\{\mathbf{B}_j\}_{j\in[2p]}$ using the augmented set of estimated FPC scores

$$\min_{\mathbf{B}_1,\ldots,\mathbf{B}_{2p}} \frac{1}{2n}\sum_{i=1}^{n} \|\widehat{\boldsymbol{\eta}}_i^{\mathrm{T}} - \sum_{j=1}^{p}\widehat{\boldsymbol{\xi}}_{ij}^{\mathrm{T}}\mathbf{B}_j - \sum_{j=p+1}^{2p}\breve{\boldsymbol{\xi}}_{i(j-p)}^{\mathrm{T}}\mathbf{B}_j\|^2 + \lambda_n\sum_{j=1}^{2p}\|\mathbf{B}_j\|_{\mathrm{F}}, \tag{17}$$

where $\lambda_n \geqslant 0$ is the regularization parameter. Let $\{\widehat{\mathbf{B}}_j\}_{j\in[2p]}$ be the solution to (17). Following the knockoffs selection step in Section 2.2, we choose the jth feature importance measures by $Z_j = \|\widehat{\mathbf{B}}_j\|_{\mathrm{F}}$ and $\widetilde{Z}_j = \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}}$, and hence the corresponding knockoff statistic is $W_j = \|\widehat{\mathbf{B}}_j\|_{\mathrm{F}} - \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}}$. Applying the knockoff filter to $\{W_j\}_{j\in[p]}$ in (4) with the threshold T_δ determined via (5), we obtain the set of selected functional covariates denoted as $\widehat{S}_\delta$.

We can verify Condition 1 with the choice of $\widetilde{Y}_i = \widehat{\boldsymbol{\eta}}_i$. Applying Theorem 2, we then attain valid FDR control in our approach for FFLR.

Theorem 4. *Suppose that Condition 2 holds. Then for any sample size n and target FDR level $q \in [0,1]$, $\widehat{S}_1$ satisfies FDR $\leqslant q$ and $\widehat{S}_0$ satisfies mFDR $\leqslant q$.*

Before presenting the power analysis, we need the following regularity conditions, which serve as the FFLR analogs of the conditions imposed for SFLR.

Condition 7. *$\{\varepsilon_i(\cdot)\}_{i\in[n]}$ with covariance operator $\Sigma_{\varepsilon\varepsilon}$ are i.i.d. sub-Gaussian processes in $L_2(\mathcal{V})$, i.e., there exists some constant $\tilde{c}$ such that $\mathbb{E}[e^{\langle x,\varepsilon\rangle}] \leqslant e^{\tilde{c}^2\langle x,\Sigma_{\varepsilon\varepsilon}(x)\rangle/2}$ for all $x \in L_2(\mathcal{V})$.*

Condition 8. *The eigenvalues of Σ_{YY} satisfy $\tilde{\omega}_1 > \tilde{\omega}_2 > \cdots > 0$, and $\sum_{l=1}^{\infty}\tilde{\omega}_l = O(1)$. There exists some constant $\tilde{\alpha} > 1$ such that $\tilde{\omega}_l - \tilde{\omega}_{l+1} \gtrsim l^{-\tilde{\alpha}-1}$ for $l \geqslant 1$.*

Condition 9. *(i) For each $j \in S$, there exists some constant $\tau > (\alpha \vee \tilde{\alpha})/2 + 1$ s.t. $|B_{jlm}| \lesssim (l+m)^{-\tau-1/2}$ for $l,m \geqslant 1$; (ii) $\min_{j\in S}\|\mathbf{B}_j\|_F \geqslant \kappa_n d^{\alpha}\lambda_n/\underline{\mu}$ for some slowly diverging sequence $\kappa_n \to \infty$ as $n \to \infty$, and $\lambda_n \gtrsim sd^{1/2}[(d^{\alpha+3/2} \vee \tilde{d}^{\tilde{\alpha}+3/2})\{\log(pd\tilde{d})/n\}^{1/2} + d^{1/2-\tau}]$; (iii) There exists some constant $c_2' \in (2(qs)^{-1}, 1)$ s.t. $|S_2| \geqslant c_2's$ with $S_2 = \{j \in [p] : \|\mathbf{B}_j\|_F \gg s^{1/2}d^{\alpha}\lambda_n\}$.*

Define the power of the proposed approach for FFLR in the same form as (12). We are now ready to present the following theorem about the asymptotic power one of our proposal.

Theorem 5. *Suppose that Conditions 2, 4–5, 7–9 hold and $sd^{\alpha}\lambda_n \to 0$. Then the selected sets S_δ's satisfy Power$(\widehat{S}_\delta) \to 1$ as $n \to \infty$.*

3.3 High-dimensional FGGM

Consider the high-dimensional FGGM, which depicts the conditional dependence structure among p Gaussian random functions $X_1(\cdot),\ldots,X_p(\cdot)$. To be specific, let $C_{jk}(u,v) = \text{Cov}\big\{X_j(u), X_k(v)|X_{-\{j,k\}}(\cdot)\big\}$ be the covariance between $X_j(u)$ and $X_k(v)$ conditional on the remaining $p-2$ random functions. Then nodes j and k are connected by an edge if and only if $\|C_{jk}\|_{\mathcal{S}} \neq 0$. Let (V,E) be an undirected graph with vertex set $V = [p]$ and edge set

$$E = \big\{(j,k) \in [p]^2 : \|C_{jk}\|_{\mathcal{S}} \neq 0, j \neq k\big\} \quad \text{with} \quad s = |E|. \tag{18}$$

Our goal is estimate E based on n i.i.d. observations. To achieve this, Qiao et al. (2019) proposed a functional graphical lasso approach to estimate a block sparse inverse covariance matrix by treating dimensions of $X_{ij}(\cdot)$'s as approaching infinity. However, their proposal cannot handle truly infinite-dimensional objects due to the unboundedness of the inverse of $\mathbf{\Sigma}_{XX}$. Along the line of Meinshausen and Buhlmann (2006), we develop a functional neighborhood selection method to estimate E by identifying the important neighborhoods of each node in a FFLR setup,

$$X_{ij}(v) = \sum_{k \neq j} \int_{\mathcal{U}} X_{ik}(u)\beta_{jk}(u,v)\mathrm{d}u + \varepsilon_{ij}(v), \quad i \in [n], j \in [p], v \in \mathcal{U}, \tag{19}$$

where, for each j, $\{\varepsilon_{ij}(\cdot)\}_{i\in[n]}$ are i.i.d. mean-zero Gaussian random errors, independent of $\{X_{i,-j}(\cdot)\}_{i\in[n]}$. Denote the neighborhood set of node j by

$$S_j = \left\{k \in [p]\backslash\{j\} : \|\beta_{jk}\|_{\mathcal{S}} \neq 0\right\} \text{ with } s_j = |S_j|.$$

Before associating the edge set E with neighborhood sets S_j's, we give a regularity condition.

Condition 10. *$\mathbf{\Sigma}^{-1}_{X_{-\{j,k\}}X_{-\{j,k\}}}\mathbf{\Sigma}_{X_{-\{j,k\}}X_j}$ and $\mathbf{\Sigma}^{-1}_{X_{-\{j,k\}}X_{-\{j,k\}}}\mathbf{\Sigma}_{X_{-\{j,k\}}X_k}$ are bounded linear operators for each $(j,k) \in [p]^2$ with $j \neq k$.*

Despite the unboundedness of $\mathbf{\Sigma}^{-1}_{X_{-\{j,k\}}X_{-\{j,k\}}}$, it is usually associated with another operator. The composite operators in Condition 10 can be viewed as regression operators, and hence can reasonably be assumed to be bounded.

Lemma 2. *Suppose that Condition 10 holds. Then $E = \{(j,k) \in [p]^2 : k \in S_j\}$.*

Lemma 2 suggests that we can recover E by estimating S_j for each j. To achieve this with FDR control, we build upon the idea of Li and Maathuis (2021), which employs a nodewise construction of knockoffs/feature statistics, and a global procedure to obtain thresholds for different nodes, and incorporate it into our framework. With each $X_{ij}(\cdot)$ expanded according to (3), some specific calculations lead to the representation of (19) as

$$\boldsymbol{\xi}_{ij}^{\mathrm{T}} = \sum_{k \neq j} \boldsymbol{\xi}_{ik}^{\mathrm{T}}\mathbf{B}_{jk} + \boldsymbol{\epsilon}_{ij}^{\mathrm{T}} + \boldsymbol{\varepsilon}_{ij}^{\mathrm{T}},$$

where $\mathbf{B}_{jk} = \int_{\mathcal{U}^2} \boldsymbol{\phi}_k(u)\beta_j(u,v)\boldsymbol{\phi}_j(v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v \in \mathbb{R}^{d_k \times d_j}$, and $\boldsymbol{\epsilon}_{ij}$ and $\boldsymbol{\varepsilon}_{ij}$ represent truncation and random errors, respectively. As a result, we can rely on the block sparsity pattern in $\{\mathbf{B}_{jk}\}_{1\leqslant j\neq k\leqslant p}$ to identify neighbourhood sets S_j's. Within functional knockoffs framework, we denote the augmented coefficients of PFC scores as $\{\mathbf{B}_{jk}\}_{k\in[2p]\backslash\{j,j+p\}}$ (the first $p-1$ coefficient matrices are for original covariates and the last $p-1$ are for knockoffs).

After performing FPCA on $\{X_{ij}(\cdot)\}_{i\in[n]}$ for each j, we solve the following group-lasso-based minimization problem using estimated FPC scores of original and knockoff variables:

$$\min_{\mathbf{B}_{jk}} \frac{1}{2n}\sum_{i=1}^{n}\|\hat{\boldsymbol{\xi}}_{ij}^{\mathrm{T}} - \sum_{1\leqslant k\neq j\leqslant p}\hat{\boldsymbol{\xi}}_{ik}^{\mathrm{T}}\mathbf{B}_{jk} - \sum_{p<k\neq(p+j)\leqslant 2p}\breve{\boldsymbol{\xi}}_{i(k-p)}^{\mathrm{T}}\mathbf{B}_{jk}\|^2 + \lambda_{nj}\sum_{k\neq j,(p+j)}\|\mathbf{B}_{jk}\|_{\mathrm{F}}, \tag{20}$$

where $\lambda_{nj}\geqslant 0$ is the regularization parameter. Let $\{\hat{\mathbf{B}}_{jk}\}_{k\in[2p]\backslash\{j,j+p\}}$ be the solution to (20). For the jth node, we select the importance measures as $Z_{jk}=\|\hat{\mathbf{B}}_{jk}\|_{\mathrm{F}}$ and $\tilde{Z}_{jk}=\|\hat{\mathbf{B}}_{j(k+p))}\|_{\mathrm{F}}$ for $k\neq j$, which results in the corresponding knockoff statistic $W_{jk}=\|\hat{\mathbf{B}}_{jk}\|_{\mathrm{F}}-\|\hat{\mathbf{B}}_{j(k+p)}\|_{\mathrm{F}}$. This forms a $p\times p$ matrix of knockoff statistics $\mathbf{W}=(W_{jk})_{p\times p}$ with $W_{jj}=0$. Given a threshold vector $\mathbf{T}_\delta=(T_{\delta,1},\ldots,T_{\delta,p})^{\mathrm{T}}$ with $\delta=1$ or 0 for knockoff or knockoff+ filter, respectively, we obtain the estimated neighborhood set of node j as

$$\hat{S}_{\delta,j}=\{k\in[p]\backslash\{j\}: W_{jk}\geqslant T_{\delta,j}\}.$$

We estimate the edge set E by applying either the AND or OR rule to $\hat{S}_{\delta,j}$'s,

$$\hat{E}_{\mathrm{A},\delta}=\big\{(k,j): k\in\hat{S}_{\delta,j} \text{ and } j\in\hat{S}_{\delta,k}\big\},\quad \hat{E}_{\mathrm{O},\delta}=\big\{(k,j): k\in\hat{S}_{\delta,j} \text{ or } j\in\hat{S}_{\delta,k}\big\}.$$

There are two options for selecting $\mathbf{T}_\delta$, the node-based local procedure and the graph-based global procedure. For the local one, we employ the knockoff filter to each row of $\mathbf{W}$ with corresponding $T_{\delta,j}$'s determined via (5). For each j, under verified Condition 1 with $\tilde{Y}_i=\hat{\boldsymbol{\xi}}_{ij}$, we can apply Theorem 2 to attain node-based FDR control at level q/p. Whereas such local procedure can be easily verified to achieve graph-based FDR control at level q, it results in substantial power loss as discussed in Li and Maathuis (2021). Inspired by Li and Maathuis (2021), we develop a graph-based global approach by solving the following optimization problems to compute $\mathbf{T}_\delta$ under the AND and OR rules, respectively.

$$\begin{aligned}&\mathbf{T}_\delta=\arg\max_{\mathbf{T}}|\hat{E}_{\mathrm{A},\delta}|,\\ &\qquad\text{subject to } \frac{a\delta+|\hat{S}_{\delta,j}^{-}|}{|\hat{E}_{\mathrm{A},\delta}|\vee 1}\leqslant\frac{2q}{c_a p} \text{ and } \ T_{\delta,j}\in\big\{|W_{jk}|,k\in[p]\big\}\cup\{\infty\}\backslash\{0\}, j\in[p],\end{aligned} \tag{21}$$

$$\mathbf{T}_\delta = \arg\max_{\mathbf{T}} |\widehat{E}_{\mathrm{O},\delta}|,$$
$$\text{subject to } \frac{a\delta + |\widehat{S}^-_{\delta,j}|}{|\widehat{E}_{\mathrm{O},\delta}| \vee 1} \leqslant \frac{q}{c_a p} \text{ and } T_{\delta,j} \in \big\{|W_{jk}|, k \in [p]\big\} \cup \{\infty\}\backslash\{0\}, j \in [p], \tag{22}$$

where, due to the coin flipping property of each row in $\mathbf{W}$, $\widehat{S}^-_{\delta,j} = \{k \in [p]\backslash\{j\} : W_{jk} \leqslant -T_{\delta,j}\}$ is used to approximate the set of false discoveries (i.e., $\widehat{S}_{\delta,j} \cap S_j^c$), $c_a > 0$ is a constant depending on a, and $\mathbf{T}_\delta = \{+\infty, \ldots, +\infty\}$ if there is no feasible solution. We then define the corresponding FDRs and mFDRs under the AND and OR rules as

$$\mathrm{FDR}_\mathrm{A} = \mathbb{E}\left[\frac{|\widehat{E}_{\mathrm{A},1} \cap E^c|}{|\widehat{E}_{\mathrm{A},1}| \vee 1}\right], \quad \mathrm{FDR}_\mathrm{O} = \mathbb{E}\left[\frac{|\widehat{E}_{\mathrm{O},1} \cap E^c|}{|\widehat{E}_{\mathrm{O},1}| \vee 1}\right],$$
$$\mathrm{mFDR}_\mathrm{A} = \mathbb{E}\left[\frac{|\widehat{E}_{\mathrm{A},0} \cap E^c|}{|\widehat{E}_{\mathrm{A},0}| + ac_a p/(2q)}\right], \quad \mathrm{mFDR}_\mathrm{O} = \mathbb{E}\left[\frac{|\widehat{E}_{\mathrm{O},0} \cap E^c|}{|\widehat{E}_{\mathrm{O},0}| + ac_a p/q}\right].$$

Remark 5. *The global approach in (21) and (22) not only enables FDR control but also ensures good power. To see this, we use FDR_O as an illustrative example. Then $FDR_O \leqslant \sum_{j=1}^p \mathbb{E}\left[\frac{\widehat{S}_{1,j} \cap S_j^c}{a + |\widehat{S}^-_{1,j}|}\right] \cdot \frac{q}{c_a p} \leqslant \sum_{i=1}^p c_a \cdot \frac{q}{c_a p} = q$, where the first inequality follows from (22) and the second inequality follows from Li and Maathuis (2021). Furthermore, the denominators in constraints of (21) and (22) are graph-based global terms instead of node-based local terms. This results in a broader feasible domain for the threshold vector, yielding larger estimated edge sets and increased power.*

We formalize the above remark about valid FDR control in the following theorem.

Theorem 6. *For any n and $q \in [0, 1]$, $FDR_A \leqslant q$ and $mFDR_A \leqslant q$ under the AND rule, while $FDR_O \leqslant q$ and $mFDR_O \leqslant q$ under the OR rule.*

To present the power theory, we give a condition akin to Condition 6 for SFLR and Condition 9 for FFLR.

Condition 11. *(i) For each $(j, k) \in E$, there exists some constant $\tau > \alpha/2 + 1$ such that $|B_{jklm}| \lesssim (l + m)^{-\tau - 1/2}$ for $l, m \geqslant 1$; (ii) For each $j \in [p]$, $\min_{k \in S_j} \|\mathbf{B}_{jk}\|_F \geqslant \kappa_{nj} d^\alpha \lambda_{nj}/\underline{\mu}$ for some slowly diverging sequences $\kappa_{nj} \to \infty$ as $n \to \infty$, and $\lambda_{nj} \gtrsim s_j[d^{\alpha+2}\{\log(pd)/n\}^{1/2} + d^{1-\tau}]$; (iii) For each j, there exists some constant $c_j \in \big((1 + a)c_a p(qs)^{-1}, 1\big)$ such that $|S_{j2}| \geqslant c_j s_j$ with $S_{j2} = \{k \in [p]\backslash\{j\} : \|\mathbf{B}_{jk}\|_F \gg s_j^{1/2} d^\alpha \lambda_{nj}\}$.*

We define the power of our proposal by

$$\text{Power}(\widehat{E}_{\text{A},\delta}) = \mathbb{E}\left[\frac{|\widehat{E}_{\text{A},\delta} \cap E|}{|E| \vee 1}\right] \quad \text{and} \quad \text{Power}(\widehat{E}_{\text{O},\delta}) = \mathbb{E}\left[\frac{|\widehat{E}_{\text{O},\delta} \cap E|}{|E| \vee 1}\right].$$

Theorem 7. *Suppose that Conditions 4–5 and 10–11 hold, and $s_j d^{\alpha} \lambda_{nj} \to 0$ for each j. Then the selected edge sets $\widehat{E}_{O,\delta}$'s satisfy Power($\widehat{E}_{O,\delta}$) $\to 1$ as $n \to \infty$.*

Theorem 7 provides the power guarantee in high dimensions when utilizing the OR rule and we leave the power analysis of $\widehat{E}_{\text{A},\delta}$ as future research.

4 Constructing functional model-X knockoffs

4.1 Exact construction

We consider the second-order functional knockoffs construction by matching the mean and covariance functions of $\mathbf{X}$ and $\widetilde{\mathbf{X}}$. Assuming that $(\mathbf{X}^{\mathrm{T}}, \widetilde{\mathbf{X}}^{\mathrm{T}})^{\mathrm{T}}$ follows MGP as specified in Example 1, we achieve alignment of the first two moments, which implies a matching of joint distributions, so that we have an exact construction of functional knockoffs.

To ensure the positive semi-definiteness of $\mathbf{\Sigma}_{(X,\tilde{X})(X,\tilde{X})}$ and maintain good statistical power, we need to solve the following optimization problem to obtain $Q_{X_jX_j}$'s:

$$\min_{\{Q_{X_jX_j}\}_{j\in[p]}} \sum_j \|\Sigma_{X_jX_j} - Q_{X_jX_j}\|_{\text{norm}} \ \text{ subject to } \ 2\mathbf{\Sigma}_{XX} - \mathbf{Q}_{XX} \succeq 0, \tag{23}$$

where $\|\cdot\|_{\text{norm}}$ denotes some proper functional norm. However, for each $j \in [p]$, the fact $X_j \in \mathcal{H}_j$ together with the constraint in (23) imply that, $\lambda_{\min}(Q_{X_jX_j}) \leqslant 2\lambda_{\min}(\Sigma_{X_jX_j}) \to 0$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue. When the minimum and maximum eigenvalues of $Q_{X_jX_j}$ are of the same order, we have $Q_{X_jX_j} \to 0$. This makes the original variables nearly indistinguishable from the knockoff counterparts, leading to substantially declined power.

To address this issue, we leverage the correlation operators between X_j's and X_k's for $j, k \in [p]$ (Baker, 1973), i.e., $C_{X_jX_k} : \mathcal{H}_k \to \mathcal{H}_j$ such that $\|C_{X_jX_k}\|_{\mathcal{L}} \leqslant 1$ and $\Sigma_{X_jX_k} = \Sigma_{X_jX_j}^{1/2} C_{X_jX_k} \Sigma_{X_kX_k}^{1/2}$, where $\Sigma_{X_jX_j}^{1/2} = \sum_{l=1}^{\infty} \omega_{jl}^{1/2} \phi_{jl} \otimes \phi_{jl}$ is the square-root of the operator $\Sigma_{X_jX_j}$. We denote $\mathbf{C}_{XX} = (C_{X_jX_k})_{p\times p}$ as the correlation operator matrix of $\mathbf{X}$. It follows from Solea and Li (2022) that, under mild regularity conditions, $\mathbf{C}_{XX} - c^*\mathbf{I} \geq 0$ for some positive constant c^*, which implies $\lambda_{\min}(\mathbf{C}_{XX}) \geqslant c^*$. E.g., when $C_{X_j X_k} = 2^{-|j-k|} \sum_l \phi_{jl} \otimes \phi_{kl}$, it is easy to verify that $\lambda_{\min}(\mathbf{C}_{XX}) \geqslant 1/3$. By utilizing $\mathbf{C}_{XX}$ instead of $\mathbf{\Sigma}_{XX}$, the constraint in (23) becomes $2\mathbf{C}_{XX} - \mathbf{R}_{XX} \geq 0$, where $\mathbf{R}_{XX} = \text{diag}(R_{X_1X_1}, \ldots, R_{X_pX_p})$. This makes it feasible to determine $R_{X_jX_j}$ with eigenvalues being bounded away from zero (see e.g., the construction of $R_{X_jX_j}$ under E1 and E2 below), which ensures discrepancies between original and knockoff variables with enhanced power. Some simple calculations show that the covariance structure in (2) reduces to the correlation structure

$$\mathbf{C}_{XX} = \mathbf{C}_{\tilde{X}\tilde{X}}, \quad \mathbf{C}_{X\tilde{X}} = \mathbf{C}_{\tilde{X}X} = \mathbf{C}_{XX} - \mathbf{R}_{XX}, \tag{24}$$

where $Q_{X_jX_j} = \Sigma_{X_jX_j}^{1/2} R_{X_jX_j} \Sigma_{X_jX_j}^{1/2}$ for $j \in [p]$. Hence we can achieve the equivalence between $\mathbf{\Sigma}_{(X,\tilde{X})(X,\tilde{X})} \geq 0$ and $2\mathbf{C}_{XX} - \mathbf{R}_{XX} \geq 0$. As a result, we propose to obtain $R_{X_jX_j}$'s by solving the following optimization problem instead of (23)

$$\min_{\{R_{X_jX_j}\}_{j\in[p]}} \sum_j \|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{L}} \text{ subject to } 2\mathbf{C}_{XX} - \mathbf{R}_{XX} \geq 0. \tag{25}$$

Remark 6. *It is crucial to select an appropriate functional norm in the objective function of (25). Notice that, for each* j, $C_{X_jX_j} - R_{X_jX_j}$ *is neither Hilbert–Schmidt nor nuclear with possibly unbounded* $\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{S}}$ *and* $\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{N}}$. *Therefore, we opt for the operator norm, considering that* $\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{L}} < \infty$.

To solve the optimization problem (25), we rely on the expression of the correlation operator $C_{X_jX_k}$ under the Karhunen-Loève expansion (3) in the following lemma.

Lemma 3. *Suppose that* $\sum_{l=1}^{\infty} \omega_{jl}^{1/2} < \infty$ *Then, for each* $j, k \in [p]$,

$$C_{X_jX_k} = \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} Corr(\xi_{jl}, \xi_{km})(\phi_{jl} \otimes \phi_{km}).$$

It then holds that $C_{X_jX_j} = \sum_{l=1}^{\infty}(\phi_{jl} \otimes \phi_{jl})$. While Lemma 3 applies to $\mathbf{C}_{XX}$, $\mathbf{C}_{\tilde{X}\tilde{X}}$ and cross-correlation operators $C_{X_j\tilde{X}_k} = C_{X_jX_k}$ for $j \neq k$ under (24), we focus on $C_{X_j\tilde{X}_j}$ that depends on $R_{X_jX_j}$, i.e., $C_{X_j\tilde{X}_j} = C_{X_jX_j} - R_{X_jX_j}$. Combining these facts, we derive three expressions of $R_{X_jX_j}$ as follows, leading to the corresponding correlation operator $C_{X_j\tilde{X}_j}$:

E1: When $R_{X_jX_j} = \sum_{l=1}^{\infty} r(\phi_{jl} \otimes \phi_{jl})$ for $r \in [0,1]$, $C_{X_j\tilde{X}_j} = \sum_{l=1}^{\infty}(1-r)(\phi_{jl} \otimes \phi_{jl})$;

E2: When $R_{X_jX_j} = \sum_{l=1}^{\infty} r_j(\phi_{jl} \otimes \phi_{jl})$ for each $r_j \in [0,1]$, $C_{X_j\tilde{X}_j} = \sum_{l=1}^{\infty}(1-r_j)(\phi_{jl} \otimes \phi_{jl})$;

E3: When $R_{X_jX_j} = \sum_{l=1}^{\infty} r_{jl}(\phi_{jl} \otimes \phi_{jl})$ for each $r_{jl} \in [0,1]$, $C_{X_j\tilde{X}_j} = \sum_{l=1}^{\infty}(1-r_{jl})(\phi_{jl} \otimes \phi_{jl})$.

In (25), we present the optimization problem at the operator level. To facilitate this task, we establish in Section B.1 of the Supplementary Material that the objective functions of (25), under E1, E2 and E3, simplify to the corresponding equivalent forms below.

E1: $\min_r (1-r)$;

E2: $\min_{(r_1,\ldots,r_p)} \sum_j (1-r_j)$;

E3: $\min_{(\mathbf{r}_1,\ldots,\mathbf{r}_p)} \sum_j \sup_l |1-r_{jl}|$, where $\mathbf{r}_j = (r_{j1}, r_{j2}, \ldots,)^{\mathrm{T}} \in \mathbb{R}^{\infty}$.

4.2 Implementation

Based on i.i.d. observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$, we can compute the sample versions $\widehat{\mathbf{C}}_{XX}$ and $\widehat{\mathbf{R}}_{XX}$ in the constraint of (25). To do this, we replace relevant terms in Lemma 3 by their sample counterparts, resulting in the sample correlation matrix operator $\widehat{\mathbf{C}}^{\mathrm{S}}_{XX} = (\widehat{C}^{\mathrm{S}}_{X_jX_k})_{p\times p}$, where $\widehat{C}^{\mathrm{S}}_{X_jX_k} = \sum_{l=1}^{\infty}\sum_{m=1}^{\infty} \widehat{\Theta}^{\mathrm{S}}_{jklm}(\hat{\phi}_{jl} \otimes \hat{\phi}_{km})$ with $\widehat{\Theta}^{\mathrm{S}}_{jklm} = n^{-1}\sum_{i=1}^{n}(\widehat{\xi}_{ijl} - n^{-1}\sum_{i=1}\widehat{\xi}_{ijl})(\widehat{\xi}_{ikm} - n^{-1}\sum_{i=1}\widehat{\xi}_{ikm})/(\hat{\omega}_{jl}^{1/2}\hat{\omega}_{km}^{1/2})$. However, the sample correlation estimator performs poorly in high-dimensional settings. Inspired by Schäfer and Strimmer (2005), we propose a shrinkage version of the sample correlation operator matrix as

$$\widehat{\mathbf{C}}_{XX} = (1-\gamma_n)\widehat{\mathbf{C}}^{\mathrm{S}}_{XX} + \gamma_n \widehat{\mathbf{I}}_{XX}, \tag{26}$$

where $\gamma_n \in [0,1]$ is the shrinkage parameter, and $\widehat{\mathbf{I}}_{XX} = \mathrm{diag}(\hat{I}_{X_1X_1}, \ldots, \hat{I}_{X_pX_p})$ with $\hat{I}_{X_jX_j} = \sum_{l=1}^{\infty} \hat{\phi}_{jl} \otimes \hat{\phi}_{jl}$. Additionally, we can obtain sample versions of $R_{X_jX_j}$ under E1, E2, E3 as $\widehat{R}_{X_jX_j} = \sum_{l=1}^{\infty} r(\hat{\phi}_{jl} \otimes \hat{\phi}_{jl})$, $\widehat{R}_{X_jX_j} = \sum_{l=1}^{\infty} r_j(\hat{\phi}_{jl} \otimes \hat{\phi}_{jl})$, $\widehat{R}_{X_jX_j} = \sum_{l=1}^{\infty} r_{jl}(\hat{\phi}_{jl} \otimes \hat{\phi}_{jl})$, respectively.

Nevertheless, it is still difficult to solve the optimization problems at the operator level. To make them executable algorithms, we map operators as matrices using a coordinate representing system within finite-dimensional Hilbert spaces (Solea and Li, 2022). The coordinate mapping employs a finite set of functions $\mathcal{B}_j = \{b_{j1}, \ldots, b_{jk_n}\}$ to approximate $\mathcal{H}_j$ for $j \in [p]$. Hence, each X_{ij} can be expressed as a linear combination of $b_{j1}, \ldots, b_{jk_n}$, where the coefficient vector is denoted by $[X_{ij}]_{\mathcal{B}_j}$ and is called the coordinate of X_{ij} with respect to $\mathcal{B}_j$. For any operator $K : \mathcal{H}_j \to \mathcal{H}_k$, the coefficient matrix $\big([K(b_{j1})]_{\mathcal{B}_k}, \ldots, [K(b_{jk_n})]_{\mathcal{B}_k}\big)$ is denoted by ${}_{\mathcal{B}_k}[K]_{\mathcal{B}_j}$ and is called the coordinate of K with respect to $\mathcal{B}_j$ and $\mathcal{B}_k$. In this way, we map each $X_{ij} \in \mathcal{H}_j$ to a vector in $\mathbb{R}^{k_n}$ and each operator $K : \mathcal{H}_j \to \mathcal{H}_k$ to a matrix in $\mathbb{R}^{k_n \times k_n}$. Let $\mathbf{G}_j = (G_{jlm})_{k_n \times k_n}$ with $G_{jlm} = \langle b_{jl}, b_{jm} \rangle$ be the Gram matrix of $\mathcal{B}_j$. See details of coordinate mapping in Section B.2 of the Supplementary Material.

In Section B.2 of the Supplementary Material, we derive that $2\widehat{\mathbf{C}}_{XX} - \widehat{\mathbf{R}}_{XX} \geq 0$ reduces to $2\widehat{\mathbf{\Omega}}_C - \widehat{\mathbf{\Omega}}_R \geq 0$ under the finite coordinate representation, where $\widehat{\mathbf{\Omega}}_C$ and $\widehat{\mathbf{\Omega}}_R$ are $pk_n \times pk_n$ matrices obtained from the coordinates of $\widehat{\mathbf{C}}_{XX}$ and $\widehat{\mathbf{R}}_{XX}$, respectively. Specifically, by (26) and (S.44)–(S.45) of the Supplementary Material, we obtain

$$\widehat{\mathbf{\Omega}}_C = \widehat{\mathbf{A}}\widehat{\mathbf{\Theta}}_C\widehat{\mathbf{A}}^{\mathrm{T}} \quad \text{with} \quad \widehat{\mathbf{\Theta}}_C = (1 - \gamma_n)(\widehat{\Theta}^{\mathrm{S}}_{jklm})_{pk_n \times pk_n} + \gamma_n \mathbf{I}_{pk_n}, \tag{27}$$

where $\widehat{\mathbf{A}} = \mathrm{diag}(\mathbf{G}_1\widehat{\mathbf{\Phi}}_1, \ldots, \mathbf{G}_p\widehat{\mathbf{\Phi}}_p) \in \mathbb{R}^{pk_n \times pk_n}$ represents the mapping matrix from the space of FPC scores to that of coordinates, and $\widehat{\mathbf{\Phi}}_j \in \mathbb{R}^{k_n \times k_n}$ with its lth column $[\hat{\phi}_{jl}]_{\mathcal{B}_j}$ for $j \in [p]$ and $l \in [k_n]$. Additionally, we can obtain the corresponding $\widehat{\mathbf{\Omega}}_R = \widehat{\mathbf{A}}\widehat{\mathbf{\Theta}}_R\widehat{\mathbf{A}}^{\mathrm{T}}$ below under E1, E2 and E3. Combing the above facts, we propose solving three sample finite-representations of the optimization problem (25):

E1: With $\widehat{\mathbf{\Theta}}_R = \mathrm{diag}(r\mathbf{I}_{k_n}, \ldots, r\mathbf{I}_{k_n}) \in \mathbb{R}^{pk_n \times pk_n}$,

$$\min_r (1 - r) \quad \text{subject to} \quad r \in [0, 1],\ 2\widehat{\mathbf{\Theta}}_C - \widehat{\mathbf{\Theta}}_R \geq 0. \tag{28}$$

E2: With $\widehat{\mathbf{\Theta}}_R = \mathrm{diag}(r_1\mathbf{I}_{k_n}, \ldots, r_p\mathbf{I}_{k_n}) \in \mathbb{R}^{pk_n \times pk_n}$,

$$\min_{(r_1,\ldots,r_p)} \sum_{j=1}^{p} (1 - r_j) \quad \text{subject to} \quad r_j \in [0, 1],\ 2\widehat{\mathbf{\Theta}}_C - \widehat{\mathbf{\Theta}}_R \geq 0. \tag{29}$$

E3: With $\widehat{\mathbf{\Theta}}_R = \mathrm{diag}(r_{11}, \ldots, r_{1k_n}, \ldots, r_{p1}, \ldots, r_{pk_n}) \in \mathbb{R}^{pk_n \times pk_n}$ for $j \in [p]$,

$$\min_{(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)} \sum_{j=1}^{p} \sum_{l=1}^{k_n} |1 - r_{jl}| \quad \text{subject to} \quad r_{jl} \in [0, 1],\ 2\widehat{\mathbf{\Theta}}_C - \widehat{\mathbf{\Theta}}_R \geq 0, \tag{30}$$

where $\bar{\mathbf{r}}_j = (r_{j1}, \ldots, r_{jk_n})^{\mathrm{T}}$.

Note that the original objective function $\min_{(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)} \sum_{j=1}^{p} \sup_{l\in[k_n]} |1 - r_{jl}|$ corresponding to E3 is difficult to handle. To simplify the computation, we consider the objective function in (30) instead. We establish the equivalence of both optimization tasks in Remark 7 of the Supplementary Material.

4.3 Algorithms

To simplify the notation, we use $[X_{ij}]_{\mathcal{B}_j}$ to denote the centered version $[X_{ij} - \hat{\mu}_j]_{\mathcal{B}_j}$. Since constructing functional knockoffs is achieved with the aid of empirical Karhunen-Loève expansion, we firstly summarize the algorithm for Karhunen-Loève expansion using the finite coordinate representation in Algorithm 1. We then present the algorithm for constructing functional model-X knockoffs in Algorithm 2. By the fact that $(\mathbf{X}_i^{\mathrm{T}}, \widetilde{\mathbf{X}}_i^{\mathrm{T}})^{\mathrm{T}}$ is MGP and the derivations in Section B.3 of the Supplementary Material, we obtain, in Step 2 that,

$$\widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{A}}^{\mathrm{T}}\big([\widetilde{X}_{i1}]_{\mathcal{B}_1}^{\mathrm{T}}, \ldots, [\widetilde{X}_{ip}]_{\mathcal{B}_p}^{\mathrm{T}}\big)^{\mathrm{T}}\Big|\widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{A}}^{\mathrm{T}}\big([X_{i1}]_{\mathcal{B}_1}^{\mathrm{T}}, \ldots, [X_{ip}]_{\mathcal{B}_p}^{\mathrm{T}}\big)^{\mathrm{T}} \sim \mathcal{N}(\widehat{\boldsymbol{\mu}}_{\tilde{X}|X}, \widehat{\boldsymbol{\Theta}}_{\tilde{X}|X}) \quad (31)$$

for each $i \in [n]$, where the normalization matrix $\widehat{\mathbf{W}} = \mathrm{diag}(\hat{\omega}_{11}, \ldots, \hat{\omega}_{1k_n}, \ldots, \hat{\omega}_{p1}, \ldots, \hat{\omega}_{pk_n})$, $\widehat{\boldsymbol{\mu}}_{\tilde{X}|X} = \big(\mathbf{I}_{pk_n} - \widehat{\boldsymbol{\Theta}}_R\widehat{\boldsymbol{\Theta}}_C^{-1}\big)\widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{A}}^{\mathrm{T}}\big([X_{i1}]_{\mathcal{B}_1}^{\mathrm{T}}, \ldots, [X_{ip}]_{\mathcal{B}_p}^{\mathrm{T}}\big)^{\mathrm{T}}$ and $\widehat{\boldsymbol{\Theta}}_{\tilde{X}|X} = 2\widehat{\boldsymbol{\Theta}}_R - \widehat{\boldsymbol{\Theta}}_R\widehat{\boldsymbol{\Theta}}_C^{-1}\widehat{\boldsymbol{\Theta}}_R$.

Algorithm 1 Algorithm for Karhunen-Loève expansion.

1: For each $j \in [p]$, choose a set of functions $\mathcal{B}_j = \{b_{j1}, \ldots, b_{jk_n}\}$ on $\mathcal{U}$ and compute $\mathbf{G}_j$.

2: Compute the coordinates $[X_{ij}]_{\mathcal{B}_j}$ relative to the basis $\mathcal{B}_j$ by least squares.

3: Perform spectral decomposition on $n^{-1}\mathbf{G}_j^{1/2}\sum_{i=1}^{n}\big([X_{ij}]_{\mathcal{B}_j}[X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}}\big)\mathbf{G}_j^{1/2}$ to obtain eigenpairs $(\hat{\omega}_{jl}, \hat{\mathbf{v}}_{jl})$ for $l \in [k_n]$.

4: Compute $\hat{\phi}_{jl} = (b_{j1}, \ldots, b_{jk_n})\mathbf{G}_j^{\dagger 1/2}\hat{\mathbf{v}}_{jl}$ and $\hat{\xi}_{ijl} = [X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}}\mathbf{G}_j^{1/2}\hat{\mathbf{v}}_{jl}$ for $i \in [n]$ and $l \in [k_n]$.

4.4 Partially observed functional data

In this section we consider a practical scenario where each $X_{ij}(\cdot)$ is partially observed, with errors, at L_{ij} random time points $U_{ij1}, \ldots, U_{ijL_{ij}} \in \mathcal{U}$. Let W_{ijl} be the observed value of $X_{ij}(U_{ijl})$ satisfying

$$W_{ijl} = X_{ij}(U_{ijl}) + e_{ijl}, \quad l = 1, \ldots, L_{ij}, \quad (32)$$

Algorithm 2 Algorithm for constructing functional model-X knockoffs.

1: Under E1, E2 and E3, recover the expressions of $\widehat{\boldsymbol{\Omega}}_R$ by replacing r, $(r_1, \ldots, r_p)$ and $(\bar{\mathbf{r}}_1, \ldots, \bar{\mathbf{r}}_p)$ with the corresponding solutions to the optimization problems in (28), (29) and (30), respectively.

2: Sample $\mathbf{Z}_1, \ldots \mathbf{Z}_n$ independently from $\mathcal{N}(\mathbf{0}_{pk_n}, \mathbf{I}_{pk_n})$. Based on the conditional distribution (31), obtain $\widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{A}}^{\mathrm{T}}([\widetilde{X}_{i1}]_{\mathcal{B}_1}^{\mathrm{T}}, \ldots, [\widetilde{X}_{ip}]_{\mathcal{B}_p}^{\mathrm{T}})^{\mathrm{T}} = \widehat{\boldsymbol{\mu}}_{\tilde{X}|X} + \widehat{\boldsymbol{\Theta}}_{\tilde{X}|X}^{1/2}\mathbf{Z}_i$, which turns to be the coefficient-vector $(\check{\xi}_{i11}, \ldots, \check{\xi}_{i1k_n}, \ldots, \check{\xi}_{ip1}, \ldots, \check{\xi}_{ipk_n})^{\mathrm{T}}$ of p-vector of functional knockoffs under respective Karhunen-Loève expansions.

3: Construct functional knockoffs as $\check{X}_{ij}(\cdot) = \sum_{l=1}^{k_n} \check{\xi}_{ijl}\hat{\phi}_{jl}(\cdot)$ for $i \in [n]$ and $j \in [p]$.

where e_{ijl}'s are i.i.d. mean-zero errors with finite variance, independent of $X_{ij}(\cdot)$. The sampling frequencies L_{ij}'s play a crucial role when choosing the smoothing strategy. When L_{ij}'s are larger than some order of n, the conventional approach employs nonparametric smoothing on $W_{ij1}, \ldots, W_{ijL_{ij}}$ to reduce the noise, see, e.g., local linear smoothers (Kong et al., 2016). This allows the reconstruction of individual curves, which can be treated as original covariates to construct functional model-X knockoffs. When L_{ij}'s are bounded, the pre-smoothing step is no longer viable. In such cases, it is recommended to apply nonparametric smoothers for estimating the mean, marginal- and cross-covariance functions, which are essential terms within the functional model-X knockoffs framework. This can be achieved by pooling data from subjects to build strength across all observations (Fang et al., 2023).

5 Simulations

In this section, we conduct a number of simulations to assess the finite-sample performance of the proposed functional knockoffs selection methods for SFLR, FFLR and FGGM. We compare our "knockoff first" proposals, which include the construction of functional knockoffs under E1, E2 and E3 (respectively denoted as KF1, KF2 and KF3), with two competing methods. The first competitor follows a "truncation first" strategy (denoted as TF), as detailed in Remark 2. The second is a group-lasso-based variable selection method (denoted as GL), which involves initial dimension reduction followed by group-lasso for group-variable selection but does not include knockoffs.

Implementing our proposals require choosing the shrinkage parameter γ_n in (27), the dimension k_n in the coordinate mapping and truncated dimensions d_j's (and $\tilde{d}$ for FFLR). To select γ_n, we can either use cross-validation or minimize the mean squared error of $\widehat{\boldsymbol{\Theta}}_C$ (Schäfer and Strimmer, 2005), while the latter approach is adopted for its computational efficiency. In the coordinate mapping, we follow Solea and Li (2022) to use cubic spline functions with 3 interior nodes, leading to 7 free parameters. Hence each $\mathcal{H}_j$ for $j \in [p]$ is spanned by these $k_n = 7$ spline functions. To determine the truncated dimensions, we take the standard approach by selecting the largest d_j (or $\tilde{d}$ for FFLR) eigenvalues of $\widehat{\Sigma}_{X_jX_j}$ (or $\widehat{\Sigma}_{YY}$ for FFLR) such that the cumulative percentage of selected eigenvalues exceeds 90%. Inspired from Wang and Zhu (2011), we consider minimizing the following high-dimensional BIC to choose the regularization parameter in the penalized least squares (17) for FFLR

$$\text{HBIC}(\lambda_n) = n \log\left\{\text{RSS}(\lambda_n)\right\} + 2\hbar \log\left(\tilde{d}\sum_{j=1}^{2p} d_j\right) \sum_{j=1}^{2p} \left\{\frac{(\tilde{d}d_j - 1)\|\widehat{\mathbf{B}}\|_{\text{F}}}{\|\widehat{\mathbf{B}}\|_{\text{F}} + \lambda_n} + \mathbf{I}(\|\widehat{\mathbf{B}}\|_{\text{F}} > 0)\right\}, \quad (33)$$

where $\text{RSS}(\lambda_n)$ represents the residual sum of squares, and $\hbar$ is a constant in $[0.1, 3]$ to maintain comparable power levels. The criterion (33) is also applicable for selecting λ_n in (10) for SFLR with $\tilde{d} = 1$ and $\|\mathbf{B}\|_{\text{F}}$ degenerated to $\|\mathbf{b}\|$. Since our proposal for FGGM involves p FFLRs, we can select the regularization parameters λ_{nj}'s in the same fashion as for FFLR. With (33), we can also determine the corresponding regularization parameters for each comparison method within each model.

To mimic the infinite-dimensionality of random functions, we generate functional variables by $X_{ij}(u) = \widetilde{\boldsymbol{\phi}}(u)^{\text{T}}\boldsymbol{\theta}_{ij}$ for $i \in [n], j = [p]$ and $u \in \mathcal{U} = [0,1]$, where $\widetilde{\boldsymbol{\phi}}(u)$ is a 25-dimensional Fourier basis function and $(\boldsymbol{\theta}_{i1}^{\text{T}}, \ldots, \boldsymbol{\theta}_{ip}^{\text{T}})^{\text{T}} \in \mathbb{R}^{25p}$ is generated independently from a mean zero multivariate normal distribution with block covariance matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{25p \times 25p}$, whose (j,k)th block is $\boldsymbol{\Lambda}_{jk} \in \mathbb{R}^{25 \times 25}$ for $j, k \in [p]$. The functional sparsity pattern in $\boldsymbol{\Sigma}_{XX} = \left(\Sigma_{jk}(\cdot,\cdot)\right)_{p \times p}$ with its (j,k)th entry $\Sigma_{jk}(u,v) = \widetilde{\boldsymbol{\phi}}(u)^{\text{T}}\boldsymbol{\Lambda}_{jk}\widetilde{\boldsymbol{\phi}}(v)$ can be captured by the block sparsity structure in $\boldsymbol{\Lambda}$. Define $\boldsymbol{\Lambda}_{jj} = \text{diag}(1^{-2}, \ldots, 25^{-2})$ and $\boldsymbol{\Lambda}_{jk} = (\Lambda_{jklm})_{25 \times 25}$, where $\Lambda_{jklm} = 0.5\rho^{|j-k|}l^{-1}m^{-1}$ for $l \neq m$ and $\rho^{|j-k|}l^{-2}$ for $l = m$ with $\rho = 0.5$. We generate $n = 100, 200$ observations of $p = 50, 100, 150$ functional variables, and replicate each simulation 200 times. For each of the three models, the data is generated as follows.

SFLR: We generate scalar responses $\{Y_i\}_{i\in[n]}$ from model (7), where ε_i's are independent standard normal. For each $j \in S = \{1, \ldots, 10\}$, we generate $\beta_j(u) = \sum_{l=1}^{25} b_{jl}\tilde{\phi}_l(u)$ for $u \in \mathcal{U}$, where $b_{jl} = (-1)^l c_b l^{-2}$ for $l = 1, \ldots, 25$, and the strength of signals via c_b's are sampled from Unif[4, 6]. For $j \in [p]\backslash S$, we set $\beta_j(u) = 0$.

FFLR: We generate functional responses $\{Y_i(v) : v \in \mathcal{U}\}_{i\in[n]}$ from model (15), where $\varepsilon_i(v) = \sum_{l=1}^{5} g_{il}\tilde{\phi}_l(v)$ with g_{il} being i.i.d. $\mathcal{N}(0, 1)$. For $j \in S$, we generate $\beta_j(u, v) = \sum_{l,m=1}^{25} B_{jlm}\tilde{\phi}_l(u)\tilde{\phi}_m(v)$ for $(u, v) \in \mathcal{U}^2$, where $B_{jlm} = (-1)^{l+m}c_b(l+m)^{-2}$ for $l, m = 1, \ldots, 25$, and c_b's are sampled from Unif[4, 6]. For $j \in [p]\backslash S$, we set $\beta_j(u, v) = 0$.

FGGM: Different from the above data generating process, we sequentially generate $X_{i1}(\cdot), \ldots, X_{ip}(\cdot)$. We firstly generate the functional errors $\varepsilon_{ij}(u) = \widetilde{\boldsymbol{\phi}}(u)^{\mathrm{T}}\widetilde{\boldsymbol{\theta}}_{ij}$ for $i \in [n]$ and $j = [p]$, where $\widetilde{\boldsymbol{\theta}}_{ij}$ are sampled independently from $\mathcal{N}(\mathbf{0}_{25}, \boldsymbol{\Lambda}_{jj})$. We then adopt the following structural equations to establish a directed acyclic graph,

$$X_{i1}(u) = \varepsilon_{i1}(u) \quad \text{and} \quad X_{ij}(u) = \sum_{(k,j)\in E_D} \int_{\mathcal{U}} X_{ik}(v)\beta_{jk}(u, v)dv + \varepsilon_{ij}(u) \text{ for } j \in [p]\backslash\{1\},$$

where E_D represents the directed edge set. A pair $(i, j) \in E_D$ is said to be directed from node i to node j if $(j, i) \notin E_D$, then node j is a child of node i. Denote the candidate edge set as $E_c = \{(k, j) \in [p]^2 : k < j\}$. To determine E_D, we randomly select one edge from E_c for each child node $j \in \{2, \ldots, p\}$ in a sequential way, and then randomly choose $p/3$ edges from the remaining E_c. We generate $\beta_{jk}(u, v) = \sum_{l,m=1}^{25} B_{jklm}\tilde{\phi}_l(u)\tilde{\phi}_m(v)$ for $(u, v) \in \mathcal{U}^2$, where $B_{jklm} = (-1)^{l+m}c_b s_j^{-1}(l+m)^{-2}$ for $l, m = 1, \ldots, 25$, and c_b's are sampled from Unif[4, 6]. We finally moralize the directed graph to obtain the undirected graph (Cowell et al., 2007).

We present numerical summaries of all comparison methods in terms of empirical power and FDR for SFLR, FFLR and FGGM in Tables 1, 2 and 3, respectively. Given the similar performance of three "knockoff first" methods for SFLR and FFLR, we only employ KF1 for FGGM due to computational efficiency. We choose to report results under the OR rule, which

p	n	KF1		KF2		KF3		TF		GL	
		FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
50	100	0.16	0.96	0.16	0.95	0.16	0.96	0.16	0.82	0.25	1.00
	200	0.13	1.00	0.14	1.00	0.13	1.00	0.17	0.96	0.19	1.00
100	100	0.13	0.89	0.15	0.89	0.16	0.89	0.19	0.68	0.47	1.00
	200	0.18	1.00	0.19	1.00	0.19	1.00	0.16	0.89	0.28	1.00
150	100	0.08	0.99	0.08	0.99	0.07	1.00	0.10	0.79	0.59	1.00
	200	0.19	1.00	0.17	1.00	0.19	1.00	0.18	0.86	0.43	1.00

Table 1: The empirical power and FDR for SFLR.

p	n	KF1		KF2		KF3		TF		GL	
		FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
50	100	0.18	0.88	0.17	0.86	0.19	0.88	0.17	0.69	0.80	1.00
	200	0.12	0.93	0.13	0.93	0.14	0.93	0.14	0.83	0.20	0.94
100	100	0.14	0.78	0.14	0.78	0.12	0.78	0.07	0.71	0.69	0.81
	200	0.08	0.96	0.08	0.96	0.08	0.96	0.07	0.85	0.88	1.00
150	100	0.17	0.99	0.15	0.99	0.19	0.99	0.16	0.70	0.93	1.00
	200	0.16	1.00	0.17	1.00	0.18	1.00	0.14	0.82	0.93	1.00

Table 2: The empirical power and FDR for FFLR.

demonstrates superior performance compared to the AND role. Several conclusions can be drawn from Tables 1–3. First, in all three models whether $p > n$ or $p < n$, the knockoffs-based methods, including KF1, KF2, KF3 and TF, effectively control the empirical FDR below the target level of $q = 0.2$. In contrast, GL results in significantly inflated FDR, especially for SFLR and FFLR. Second, across all scenarios, our "knockoff first" methods consistently achieve higher empirical powers compared to "truncation first" competitors. These empirical findings nicely validate the heuristic arguments presented in Remarks 2 and 4. Third, among KF1, KF2 and KF3, they exhibit similar performance in terms of FDR control and power. Due to its lowest computational cost, we recommend using KF1 in practice.

p	n	KF1		TF		GL	
		FDR	Power	FDR	Power	FDR	Power
50	100	0.18	0.74	0.15	0.50	0.23	0.75
	200	0.20	0.94	0.20	0.79	0.22	0.94
100	100	0.17	0.63	0.12	0.48	0.20	0.68
	200	0.19	0.84	0.18	0.73	0.22	0.86

Table 3: The empirical power and FDR for FGGM.

We also assess the finite-sample performance of competing methods for handling partially observed functional data. We generate $X_{ij}(\cdot)$ for $i = [n]$ and $j = [p]$ following the same procedure as above. We then generate the observed values W_{ijl}'s from (32), where the observational time points and errors e_{ijl}'s are independently sampled from Unif$[0, 1]$ and $\mathcal{N}(0, 0.5^2)$, respectively. We consider the dense setting $L = 51$ since brain signals in neuroimaging data are commonly measured at a dense set of points. We employ the local-linear-based pre-smoothing approach using a Gaussian kernel with the optimal bandwidth proportional to $L^{-1/5}$. The numerical results for SFLR and FGGM are respectively presented in Tables 4 and 5 of the Supplementary Material. Similar conclusions can be drawn compared to the results obtained for fully observed functional data from Tables 1–3.

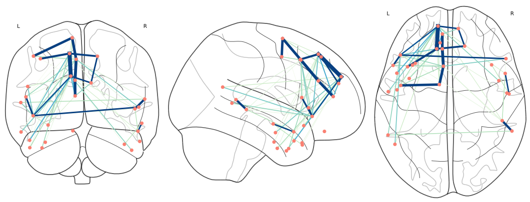
6 Real data analysis

6.1 Emotion related fMRI dataset

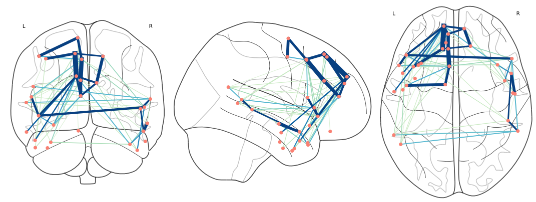
In this section, we illustrate our functional model-X knockoffs selection proposal for SFLR using a publicly available brain imaging dataset obtained from the Human Connectome Project (HCP), http://www.humanconnectome.org/. This dataset comprises $n = 848$ subjects of functional magnetic resonance imaging (fMRI) scans with Blood Oxygenation Level-Dependent (BOLD) signals in the brain. We follow recent proposals, based on HCP, to model signals as multivariate random functions, thus representing each region of interest (ROI) as one random function; see, e.g., Xue and Yao (2021); Zapata et al. (2022); Lee et al. (2023). For each subject, the BOLD signals are recorded every 0.72 seconds, totalling $L = 176$ observational time points (2.1 minutes). The response of interest is referred to as *Emotion Task Shape Acc*, which represents a continuous score measured by the Penn Emotion Recognition Test and is associated with the brain's processing of negative emotional tasks. We construct an SFLR model via (7) by treating $p = 34$ ROIs as functional covariates; see Table 6 of the Supplementary Material for details on the specific ROIs. Our goal is to identify ROIs that significantly influence the Emotion Task Shape Acc. For this purpose, we apply our proposed KF1 approach for SFLR with target FDR level of $q = 0.2$. To construct $\mathcal{H}_j$'s, we use cubic spline functions with 11 interior nodes, corresponding to $k_n = 15$. For comparison, we also implement the TF and GL methods. While TF and GL respectively select 9 and 15 ROIs, our KF1 identifies 6 ROIs, indexed by $j \in \{9, 12, 20, 22, 31, 32\}$ with sorted importance levels $W_9 > W_{31} > W_{32} > W_{20} > W_{12} > W_{22}$. Among three competitors, these six ROIs are consistently selected, and align with findings in the existing literature. Specifically, Xue and Yao (2021) identified isthmus cingulate ($j = 9$), lingual ($j = 12$) and frontal pole ($j = 31$) as important ROIs associated with Emotion Task Shape Acc. Furthermore, previous studies have found regions like pericalcarine ($j = 20$), posterior cingulate ($j = 22$) and temporal pole ($j = 32$) to be responsible for negative emotions (Sabatinelli et al., 2007; Rolls, 2019; Olson et al., 2007).

6.2 Functional connectivity analysis

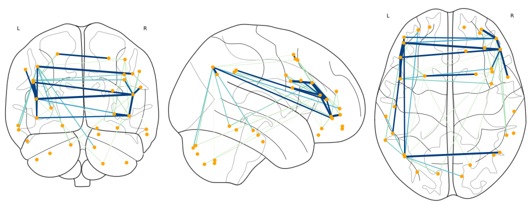
In this section, we investigate the relationship between brain functional connectivity and fluid intelligence (gF), which represents the capacity to think and reason independently of acquired knowledge. The dataset, obtained from HCP, consists of fMRI scans and the corresponding gF scores, determined based on participants' performance on the Raven's Progressive Matrices. We focus on $n_{\text{low}} = 73$ subjects with low fluid intelligence scores (gF$\leqslant 8$) and $n_{\text{high}} = 85$ subjects with high scores (gF$\geqslant 23$). In an analogy to Section 6.1,



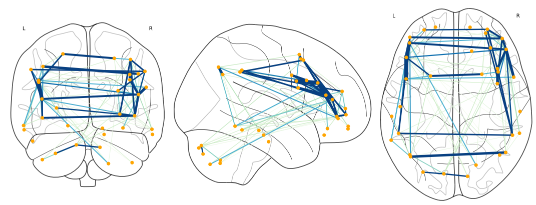
(a) gF$\leqslant$ 8: the medial frontal module



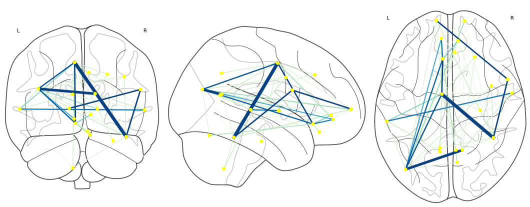
(d) gF$\geqslant$ 23: the medial frontal module



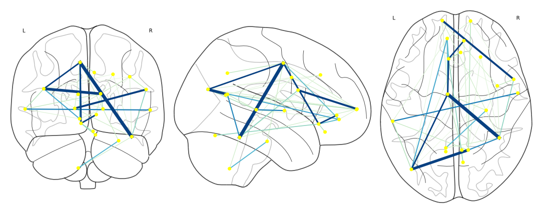
(b) gF$\leqslant$ 8: the frontoparietal module



(e) gF$\geqslant$ 23: the frontoparietal module



(c) gF$\leqslant$ 8: the default mode module



(f) gF$\geqslant$ 23: the default mode module

Figure 1: The connectivity strengths at fluid intelligence gF$\leqslant$ 8 (left column) and gF$\geqslant$ 23 (right column). Salmon, orange and yellow nodes represent the ROIs in the medial frontal, frontoparietal and default mode modules, respectively. The edge color ranging from light to dark corresponds to the value of W_{jk} from small to large.

we treat the BOLD signals at different ROIs as multivariate functional data, considering $p = 83$ ROIs across three well-established modules in neuroscience literature (Finn et al., 2015): the medial frontal module (29 ROIs), frontoparietal module (34 ROIs), and default mode module (20 ROIs). The signals for each subject at each ROI are collected every 0.72 seconds at $L = 1200$ measurement locations (14.4 minutes). To exclude unrelated frequency bands in resting-state functional connectivity, we apply ICA+FIX preprocessed pipeline and use a standard band-pass filter between 0.01 and 0.08 Hz (Glasser et al., 2016). For our analysis, we employ the proposed KF1 method for FGGM under the OR rule to construct brain functional connectivity networks, which depict the conditional dependence structures among respective ROIs within each of the three modules. We use cubic splines with 31 interior nodes, resulting in each $\mathcal{H}_j$ being spanned by $k_n = 35$ spline functions. We continue to set the target FDR level at $q = 0.2$.

Figure 1 displays the functional connectivity networks identified for subjects with gF $\leqslant 8$ and gF $\geqslant 23$. To assess the impact of gF on functional connectivity, we measure connectivity strength through W_{jk} for $j, k \in [p]$ with larger values yielding stronger connectivity. A few patterns are apparent. First, we observe increased connectivity and strength in the medial frontal and frontoparietal modules for subjects with gF$\geqslant 23$. This observation aligns with and supports the existing literature, which has reported a strong positive association between the functional connectivity and intellectual performance in these two modules (Van Den Heuvel et al., 2009). Second, it is evident that the default mode module exhibits declined connectivity and strength for subjects with higher intelligence scores, which is in line with the previous finding in neuroscience that reduced activity in the default mode module is associated with improved cognitive performance (Anticevic et al., 2012).

References

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J. and Krystal, J. H. (2012). The role of default network deactivation in cognition and disease, *Trends in Cognitive Sciences* **16**: 584–592.

Baker, C. R. (1973). Joint measures and cross-covariance operators, *Transactions of the American Mathematical Society* **186**: 273–289.

Barber, R. F. and Candès, E. J. (2015). Controlling the false discovery rate via knockoffs, *The Annals of Statistics* **43**: 2055–2085.

Barber, R. F. and Candès, E. J. (2019). A knockoff filter for high-dimensional selective inference, *The Annals of Statistics* **47**: 2504–2537.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society: Series B* **57**: 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency, *The Annals of statistics* **29**: 1165–1188.

Candès, E., Fan, Y., Janson, L. and Lv, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection, *Journal of the Royal Statistical Society: Series B* **80**: 551–577.

Chang, J., Chen, C., Qiao, X. and Yao, Q. (2023). An autocovariance-based learning framework for high-dimensional functional time series, *Journal of Econometrics, in press* .

Cowell, R. G., Dawid, P., Lauritzen, S. L. and Spiegelhalter, D. J. (2007). *Probabilistic Networks and Expert Systems*, Information Science and Statistics, Springer New York.

Dai, R. and Barber, R. F. (2016). The knockoff filter for FDR control in group-sparse and multitask regression, *International Conference on Machine Learning* **48**: 1851–1859.

Dai, X., Lyu, X. and Li, L. (2023). Kernel knockoffs selection for nonparametric additive models, *Journal of the American Statistical Association* **118**: 2158–2170.

Fan, Y., Demirkaya, E., Li, G. and Lv, J. (2020a). RANK: Large-scale inference with graphical nonlinear knockoffs, *Journal of the American Statistical Association* **115**: 362–379.

Fan, Y., Foutz, N., James, G. M. and Jank, W. (2014). Functional response additive model estimation with online virtual stock markets, *The Annals of Applied Statistics* **8**: 2435–2460.

Fan, Y., James, G. and Radehcnko, P. (2015). Functional additive regression, *The Annals of Statistics* **43**: 2296–2325.

Fan, Y., Lv, J., Sharifvaghefi, M. and Uematsu, Y. (2020b). IPAD: Stable interpretable forecasting with knockoffs inference, *Journal of the American Statistical Association* **115**: 1822–1834.

Fang, Q., Guo, S. and Qiao, X. (2023). Adaptive functional thresholding for sparse covariance function estimation in high dimensions, *Journal of the American Statistical Association, in press* .

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X. and Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity, *Nature Neuroscience* **18**: 1664–1671.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M. and Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex, *Nature* **536**: 171–178.

Kong, D., Xue, K., Yao, F. and Zhang, H. H. (2016). Partially functional linear regression in high dimensions, *Biometrika* **103**: 147–159.

Lee, K.-Y., Ji, D., Li, L., Constable, T. and Zhao, H. (2023). Conditional functional graphical models, *Journal of the American Statistical Association* **118**: 257–271.

Li, B. and Solea, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI, *Journal of the American Statistical Association* **113**: 1637–1655.

Li, J. and Maathuis, M. H. (2021). GGM knockoff filter: False discovery rate control for Gaussian graphical models, *Journal of the Royal Statistical Society: Series B* **83**: 534–558.

Luo, R. and Qi, X. (2017). Function-on-function linear regression by signal compression, *Journal of the American Statistical Association* **112**: 690–705.

Meinshausen, N. and Buhlmann, P. (2006). High dimensional graphs and variable selection with the lasso, *The Annals of Statistics* **34**: 1436–1462.

Olson, I. R., Plotzker, A. and Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing, *Brain* **130**: 1718–1731.

Qiao, X., Guo, S. and James, G. (2019). Functional graphical models, *Journal of the American Statistical Association* **114**: 211–222.

Ren, Z., Wei, Y. and Candès, E. (2023). Derandomizing knockoffs, *Journal of the American Statistical Association* **118**: 948–958.

Rolls, E. T. (2019). The cingulate cortex and limbic systems for emotion, action, and memory, *Brain Structure and Function* **224**: 3001–3018.

Romano, Y., Sesia, M. and Candès, E. (2020). Deep knockoffs, *Journal of the American Statistical Association* **115**: 1861–1872.

Sabatinelli, D., Lang, P. J., Keil, A. and Bradley, M. M. (2007). Emotional perception: Correlation of functional mri and event-related potentials, *Cerebral Cortex* **17**: 1085–1091.

Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics, *Statistical Applications in Genetics and Molecular Biology* **4**: Article32.

Solea, E. and Li, B. (2022). Copula Gaussian graphical models for functional data, *Journal of the American Statistical Association* **117**: 781–793.

Van Den Heuvel, M. P., Stam, C. J., Kahn, R. S. and Pol, H. E. H. (2009). Efficiency of functional brain networks and intellectual performance, *Journal of Neuroscience* **29**: 7619–7624.

Wang, T. and Zhu, L. (2011). Consistent tuning parameter selection in high dimensional sparse linear regression, *Journal of Multivariate Analysis* **102**: 1141–1151.

Xue, K., Yang, J. and Yao, F. (2023). Optimal linear discriminant analysis for high-dimensional functional data, *Journal of the American Statistical Association, in press* .

Xue, K. and Yao, F. (2021). Hypothesis testing in large-scale functional linear regression, *Statistica Sinica* **31**: 1101–1123.

Zapata, J., Oh, S. Y. and Petersen, A. (2022). Partial separability and functional graphical models for multivariate Gaussian processes, *Biometrika* **109**: 665–681.

Zhu, H., Strawn, N. and Dunson, D. B. (2016). Bayesian graphical models for multivariate functional data, *Journal of Machine Learning Research* **17**: 1–27.

Supplementary material to "Functional knockoffs selection with applications to functional data analysis in high dimensions"

Qizhai Li, Mingya Long and Xinghao Qiao

This supplementary material contains all technical proofs in Section A, additional methodological derivations in Section B and additional empirical results in Section C.

A Technical proofs

A.1 Proof of Theorem 2

To prove Theorem 2, we firstly present some technical lemmas with their proofs.

Lemma A1. *For any subset* $G \subseteq S^c$, $\big(\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big) \,\Big|\, Y \overset{D}{=} \big(\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big)_{swap(G)} \,\Big|\, Y.$

Proof. It is equivalent to show that $\big(\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}, Y\big) \overset{D}{=} \big(\{\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\}_{\mathrm{swap}(G)}, Y\big)$. Moreover, since $\big(\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big) \overset{D}{=} \big(\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big)_{\mathrm{swap}(G)}$ from Property (i) in Definition 1, it suffices to prove that

$$Y \,\Big|\, \big(\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big)_{\mathrm{swap}(G)} \overset{D}{=} Y \,\Big|\, \big(\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big). \tag{S.1}$$

By Properties (i) and (ii) in Definition 1, we have

$$\begin{aligned} Y \,\Big|\, \big[\{\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\}_{\mathrm{swap}(G)} = \{\mathbf{x}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{x}}(\cdot)^{\mathrm{T}}\}\big] &\overset{D}{=} Y \,\Big|\, \big[\{\mathbf{X}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\} = \{\mathbf{x}(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{x}}(\cdot)^{\mathrm{T}}\}_{\mathrm{swap}(G)}\big] \\ &\overset{D}{=} Y \,\Big|\, \big\{\mathbf{X}(\cdot)^{\mathrm{T}} = \mathbf{x}'(\cdot)^{\mathrm{T}}\big\}, \end{aligned}$$

where $\mathbf{x}(\cdot) = \big(x_1(\cdot), \ldots, x_p(\cdot)\big)^{\mathrm{T}}$, $\widetilde{\mathbf{x}}(\cdot) = \big(\widetilde{x}_1(\cdot), \ldots, \widetilde{x}_p(\cdot)\big)^{\mathrm{T}}$, and the jth element of $\mathbf{x}'(\cdot)$ is $x'_j(\cdot) = \widetilde{x}_j(\cdot)$ if $j \in G$ and $x'_j(\cdot) = x_j(\cdot)$ otherwise. Without loss of generality, assuming that $G = \{1, 2, \ldots, m\} \subseteq S^c$, we have

$$\begin{aligned} Y \,\Big|\, \big\{\mathbf{X}(\cdot)^{\mathrm{T}} = \mathbf{x}'(\cdot)^{\mathrm{T}}\big\} &\overset{D}{=} Y \,\Big|\, \big\{\mathbf{X}_{2:p}(\cdot)^{\mathrm{T}} = \mathbf{x}'_{2:p}(\cdot)^{\mathrm{T}}\big\} \\ &\overset{D}{=} Y \,\Big|\, \big\{X_1(\cdot) = x_1(\cdot), \mathbf{X}_{2:p}(\cdot)^{\mathrm{T}} = \mathbf{x}'_{2:p}(\cdot)^{\mathrm{T}}\big\}, \end{aligned} \tag{S.2}$$

where we use $\mathbf{X}_{2:p}(\cdot)$ to denote $\big(X_2(\cdot),\ldots,X_p(\cdot)\big)^{\mathrm{T}}$, $\mathbf{x}'_{2:p}(\cdot)=\big(x'_2(\cdot),\ldots,x'_p(\cdot)\big)^{\mathrm{T}}$ and the above two equalities hold since Y and $X_1(\cdot)$ are independent conditional on $\mathbf{X}_{2:p}(\cdot)$. (S.2) shows that $Y \;\Big|\; \big(\mathbf{X}(\cdot)^{\mathrm{T}},\widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big)_{\mathrm{swap}(G)} \overset{D}{=} Y \;\Big|\; \big(\mathbf{X}(\cdot)^{\mathrm{T}},\widetilde{\mathbf{X}}(\cdot)^{\mathrm{T}}\big)_{\mathrm{swap}(G\setminus\{1\})}$. Repeating this strategy until G is empty, we obtain that (S.1) holds, which completes our proof. Note that the response Y in our proof can be either scalar or functional, we use the same notation for simplicity. □

We will next demonstrate that the estimated FPC scores in SFLR satisfy Condition 1.

Corollary 1. *For any subset $G\subseteq S^c$, $\big(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}}\big) \;\Big|\; Y_i \overset{D}{=} \big(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}}\big)_{swap(G)} \;\Big|\; Y_i$.*

Proof. It is equivalent to show that $(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}},Y_i) \overset{D}{=} \big\{(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}})_{\mathrm{swap}(G)},Y_i\big\}$. Referring to the result in Lemma A1 and considering that $\big\{\big(\mathbf{X}_i(\cdot)^{\mathrm{T}},\widetilde{\mathbf{X}}_i(\cdot)^{\mathrm{T}},Y_i\big)\big\}_{i\in[n]}$ are i.i.d., we can establish that $\big(\mathbf{X}_i(\cdot)^{\mathrm{T}},\widetilde{\mathbf{X}}_i(\cdot)^{\mathrm{T}}\big) \;\Big|\; Y_i \overset{D}{=} \big(\mathbf{X}_i(\cdot)^{\mathrm{T}},\widetilde{\mathbf{X}}_i(\cdot)^{\mathrm{T}}\big)_{\mathrm{swap}(G)} \;\Big|\; Y_i$, where the response Y_i is scalar in SFLR. This implies that $\big(\mathbf{X}_i(\cdot)^{\mathrm{T}},\widetilde{\mathbf{X}}_i(\cdot)^{\mathrm{T}},Y_i\big) \overset{D}{=} \big(\{\mathbf{X}_i(\cdot)^{\mathrm{T}},\widetilde{\mathbf{X}}_i(\cdot)^{\mathrm{T}}\}_{\mathrm{swap}(G)},Y_i\big)$. This further implies that, for any $\mathbf{t}\in(\boldsymbol{\mathcal{H}}^2,\mathbb{R})$ and $\iota=\sqrt{-1}$,

$$\mathbb{E}\Big[\exp\Big\{\iota\big\langle\mathbf{t},(\mathbf{X}_i^{\mathrm{T}},\widetilde{\mathbf{X}}_i^{\mathrm{T}},Y_i)^{\mathrm{T}}\big\rangle\Big\}\Big]=\mathbb{E}\Big[\exp\Big\{\iota\big\langle\mathbf{t},\{(\mathbf{X}_i^{\mathrm{T}},\widetilde{\mathbf{X}}_i^{\mathrm{T}})_{\mathrm{swap}(G)},Y_i\}^{\mathrm{T}}\big\rangle\Big\}\Big].$$

Given $\mathbf{t}=(\mathbf{c}_1^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_1,\ldots,\mathbf{c}_p^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_p,\tilde{\mathbf{c}}_1^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_1,\ldots,\tilde{\mathbf{c}}_p^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_p,c_Y)^{\mathrm{T}}$, where $\mathbf{c}_j=(c_{j1},\ldots,c_{jd_j},0,0,\ldots)^{\mathrm{T}}$, $\tilde{\mathbf{c}}_j=(\tilde{c}_{j1},\ldots,\tilde{c}_{jd_j},0,0,\ldots)^{\mathrm{T}}$, and $\widehat{\boldsymbol{\phi}}_j=(\hat{\phi}_{j1},\hat{\phi}_{j2},\ldots)^{\mathrm{T}}$ for $j\in[p]$, we can establish that

$$\mathbb{E}\Big[\exp\Big\{\iota\big\langle\mathbf{c},(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}},Y_i)^{\mathrm{T}}\big\rangle\Big\}\Big|\mathbf{t}\Big]=\mathbb{E}\Big[\exp\Big\{\iota\big\langle\mathbf{c},\{(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}})_{\mathrm{swap}(G)},Y_i\}^{\mathrm{T}}\big\rangle\Big\}\Big|\mathbf{t}\Big],\tag{S.3}$$

for any $\mathbf{c}=(\bar{\mathbf{c}}_1^{\mathrm{T}},\ldots,\bar{\mathbf{c}}_p^{\mathrm{T}},\widetilde{\bar{\mathbf{c}}}_1^{\mathrm{T}},\ldots,\widetilde{\bar{\mathbf{c}}}_p^{\mathrm{T}},c_Y)^{\mathrm{T}}$, where $\bar{\mathbf{c}}_j=(c_{j1},\ldots,c_{jd_j})^{\mathrm{T}}$, $\widetilde{\bar{\mathbf{c}}}_j=(\tilde{c}_{j1},\ldots,\tilde{c}_{jd_j})^{\mathrm{T}}$. By (S.3) and the total expectation formula, the joint characteristic function of $(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}},Y_i)$ is equal to that of $\big\{(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}})_{\mathrm{swap}(G)},Y_i\big\}$, which implies that $(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}},Y_i) \overset{D}{=} \big\{(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}},\breve{\boldsymbol{\xi}}_i^{\mathrm{T}})_{\mathrm{swap}(G)},Y_i\big\}$, and thus completes our proof. □

Following the above corollary, we next validate Lemma 1 in the context of SFLR.

Proof of Lemma 1 in SFLR. Denote $\widehat{\boldsymbol{\Xi}}=(\widehat{\boldsymbol{\xi}}_1,\ldots,\widehat{\boldsymbol{\xi}}_n)^{\mathrm{T}}\in\mathbb{R}^{n\times\sum_j d_j}$, $\breve{\boldsymbol{\Xi}}=(\breve{\boldsymbol{\xi}}_1,\ldots,\breve{\boldsymbol{\xi}}_n)^{\mathrm{T}}\in\mathbb{R}^{n\times\sum_j d_j}$ and $\mathbf{Y}=(Y_1,\ldots,Y_n)^{\mathrm{T}}\in\mathbb{R}^n$. First, note $W_j(\widehat{\boldsymbol{\Xi}},\breve{\boldsymbol{\Xi}},\mathbf{Y})$ is a function of $\widehat{\boldsymbol{\Xi}},\breve{\boldsymbol{\Xi}}$ and $\mathbf{Y}$.

Since $W_j = \|\widehat{\mathbf{b}}_j\| - \|\widehat{\mathbf{b}}_{p+j}\|$ in SFLR for each $j \in [p]$, the flip-sign property of $W_j(\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}, \mathbf{Y})$ holds. That is, for any subset $G \subseteq [p]$,

$$W_j\big(\{\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}\}_{\mathrm{swap}(G)}, \mathbf{Y}\big) = \begin{cases} W_j\big(\{\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}\}_{\mathrm{swap}(G)}, \mathbf{Y}\big), & j \notin G \\ -W_j\big(\{\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}\}_{\mathrm{swap}(G)}, \mathbf{Y}\big), & j \in G, \end{cases}$$

where $(\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}})_{\mathrm{swap}(G)}$ is obtained from $(\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}})$ by swapping the corresponding estimated FPC scores of $X_j(\cdot)$ and $\breve{X}_j(\cdot)$ for each $j \in G$. Second, let $\boldsymbol{W} = (W_1, \ldots, W_p)^{\mathrm{T}}$, and consider any subset $G \subseteq S^c$. By swapping variables in G, we define

$$\boldsymbol{W}_{\mathrm{swap}(G)} \triangleq \Big(W_1\big(\{\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}\}_{\mathrm{swap}(G)}, \mathbf{Y}\big), \ldots, W_p\big(\{\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}\}_{\mathrm{swap}(G)}, \mathbf{Y}\big)\Big)^{\mathrm{T}}.$$

By Corollary 1, it follows that $\big((\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}), \mathbf{Y}\big) \overset{D}{=} \big((\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}})_{\mathrm{swap}(G)}, \mathbf{Y}\big)$, which consequently implies $\boldsymbol{W} \overset{D}{=} \boldsymbol{W}_{\mathrm{swap}(G)}$. Finally, consider $S^c_- = \{j \in S^c : \delta_j = -1\}$, where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^{\mathrm{T}}$ represents a sequence of independent random variables. These δ_j variables follow the Rademacher distribution if $j \in S^c$ and $\delta_j = 1$ otherwise. Through the first step, we derive $\boldsymbol{W}_{\mathrm{swap}(S^c_-)} = (\delta_1 W_1, \ldots, \delta_p W_p)^{\mathrm{T}}$. Subsequently, from the second step, we obtain $\boldsymbol{W}_{\mathrm{swap}(S^c_-)} \overset{D}{=} \boldsymbol{W}$. Combining the above results, we have $(\delta_1 W_1, \ldots, \delta_p W_p)^{\mathrm{T}} \overset{D}{=} \boldsymbol{W}$, which completes the proof. □

Lemma A2. *Suppose that Condition 2 holds. Then $\|\beta_j\| = 0$ if and only if $j \in S^c$, where S^c is defined in (1).*

Note that the proof of Lemma A2 follows the same argument as that of Lemma A9. We will provide detailed proof of Lemma A9 in Section A.3 and omit the proof of Lemma A2 here. We are now ready to prove Theorem 2.

Proof of Theorem 2. Provided that Lemma 1 holds in SFLR, it implies that the signs of the null statistics are distributed as i.i.d. coin flips. Referring to Theorem 3.4 of Candès et al. (2018), we obtain that

$$\mathbb{E}\left[\frac{|\widehat{S}_1 \cap S^c|}{|\widehat{S}_1|}\right] \leqslant q, \quad \mathbb{E}\left[\frac{|\widehat{S}_0 \cap S^c|}{|\widehat{S}_0| + 1/q}\right] \leqslant q,$$

where S^c is defined in (1). By Lemma A2, we establish that S^c in (1) is equivalent to the set $\{j : \|\beta_j\| = 0\}$, whose complement is defined in (8). This equivalence implies the effective FDR control in SFLR, which completes our proof. □

A.2 Proof of Theorem 3

Before presenting technical lemmas used in the proof of Theorem 3, we firstly introduce some notation. For $j, k \in [p]$, denote $\sigma_{jklm} = \mathbb{E}[\xi_{ijl}\xi_{ikm}]$ and its estimator $\hat{\sigma}_{jklm} = n^{-1}\sum_{i=1}^{n}\hat{\xi}_{ijl}\hat{\xi}_{ikm}$ for $l, m \in [d]$. For $j \in [p]$, $k \in [2p]\backslash[p]$, denote $\sigma_{jklm} = \mathbb{E}[\xi_{ijl}\tilde{\xi}_{i(k-p)m}]$ and its estimator $\hat{\sigma}_{jklm} = n^{-1}\sum_{i=1}^{n}\hat{\xi}_{ijl}\check{\xi}_{i(k-p)m}$. For $j, k \in [2p]\backslash[p]$, denote $\sigma_{jklm} = \mathbb{E}[\tilde{\xi}_{i(j-p)l}\tilde{\xi}_{i(k-p)m}]$ and its estimator estimator $\hat{\sigma}_{jklm} = n^{-1}\sum_{i=1}^{n}\check{\xi}_{i(j-p)l}\check{\xi}_{i(k-p)m}$. For $j \in [p]$, $l \in [d]$, denote $\sigma_{jl}^{X,Y} = \mathbb{E}[\xi_{ijl}Y_i]$ and its estimator $\hat{\sigma}_{jl}^{X,Y} = n^{-1}\sum_{i=1}^{n}\hat{\xi}_{ijl}Y_i$. For $j \in [2p]\backslash[p]$, $l \in [d]$, denote $\sigma_{jl}^{X,Y} = \mathbb{E}[\tilde{\xi}_{i(j-p)l}Y_i]$ and its estimator $\hat{\sigma}_{jl}^{X,Y} = n^{-1}\sum_{i=1}^{n}\check{\xi}_{i(j-p)l}Y_i$. Let $\mathbf{D} = \text{diag}(\mathbf{D}_1, \ldots, \mathbf{D}_p, \mathbf{D}_1, \ldots, \mathbf{D}_p) \in \mathbb{R}^{2pd\times 2pd}$ with $\mathbf{D}_j = \text{diag}(\omega_{j1}^{1/2}, \ldots, \omega_{jd}^{1/2}) \in \mathbb{R}^{d\times d}$ for $j \in [p]$ and its estimator $\widehat{\mathbf{D}} = \text{diag}(\widehat{\mathbf{D}}_1, \ldots, \widehat{\mathbf{D}}_p, \widehat{\mathbf{D}}_1, \ldots, \widehat{\mathbf{D}}_p)$ with $\widehat{\mathbf{D}}_j = \text{diag}(\hat{\omega}_{j1}^{1/2}, \ldots, \hat{\omega}_{jd}^{1/2})$. For a matrix $\mathbf{A} = (A_{jk}) \in \mathbb{R}^{p\times q}$, we denote its elementwise ℓ_∞ norm as $\|\mathbf{A}\|_{\max} = \max_{j,k}|A_{jk}|$. For a block matrix $\mathbf{B} = (\mathbf{B}_{jk}) \in \mathbb{R}^{p_1q_1\times p_2q_2}$ with its (j,k)-th block $\mathbf{B}_{jk} \in \mathbb{R}^{q_1\times q_2}$, we define its block versions of elementwise ℓ_∞ and matrix ℓ_1 norms by $\|\mathbf{B}\|_{\max}^{(q_1\times q_2)} = \max_{j,k}\|\mathbf{B}_{jk}\|_{\text{F}}$ and $\|\mathbf{B}\|_1^{(q_1\times q_2)} = \max_k\sum_j\|\mathbf{B}_{jk}\|_{\text{F}}$, respectively.

Lemma A3. *Suppose that Condition 5 holds. If $n \gtrsim d^{4\alpha+2}\log(pd)$, then there exist some positive constants $\tilde{c}_1$ and $\tilde{c}_2$ such that, with probability greater than $1-\tilde{c}_1(pd)^{-\tilde{c}_2}$, the estimates $\{\hat{\sigma}_{jklm}\}$ satisfy*

$$\max_{\substack{j,k\in[2p]\\ l,m\in[d]}} \frac{|\hat{\sigma}_{jklm} - \sigma_{jklm}|}{(l \vee m)^{\alpha+1}\omega_{jl}^{1/2}\omega_{km}^{1/2}} \lesssim \sqrt{\frac{\log(pd)}{n}}. \tag{S.4}$$

Suppose that Conditions 3 and 5 hold . If $n \gtrsim d^{3\alpha+2}\log(pd)$, then there exist some positive constants $\tilde{c}_3$ and $\tilde{c}_4$ such that, with probability greater than $1-\tilde{c}_3(pd)^{-\tilde{c}_4}$, the estimates $\{\hat{\sigma}_{jl}^{X,Y}\}$ satisfy

$$\max_{\substack{j\in[2p]\\ l\in[d]}} \frac{|\hat{\sigma}_{jl}^{X,Y} - \sigma_{jl}^{X,Y}|}{l^{\alpha+1}\omega_{jl}^{1/2}} \lesssim \sqrt{\frac{\log(pd)}{n}}. \tag{S.5}$$

Proof. (S.4) is a direct deduction from Theorem 4 in Guo and Qiao (2023)

$$\max_{\substack{j,k\in[2p]\\ l,m\in[d]}} \frac{|\hat{\sigma}_{jklm} - \sigma_{jklm}|}{(l \vee m)^{\alpha+1}\omega_{jl}^{1/2}\omega_{km}^{1/2}} \lesssim \mathcal{M}_1^X\sqrt{\frac{\log(pd)}{n}},$$

where, under no serial dependence, the functional stability measure of $\{\mathbf{X}_i(\cdot)\}_{i\in n}$ is $\mathcal{M}_1^X = 1$.

(S.5) can be derived from Proposition 1 in Fang et al. (2022) as

$$\max_{\substack{j\in[2p]\\ l\in[d]}} \frac{|\hat{\sigma}_{jl}^{X,Y}-\sigma_{jl}^{X,Y}|}{l^{\alpha+1}\omega_{jl}^{1/2}} \lesssim \mathcal{M}_{X,Y}\sqrt{\frac{\log(pd)}{n}},$$

where the measure of dependence between $\{\mathbf{X}_i(\cdot)\}_{i\in[n]}$ and $\{Y_i\}_{i\in[n]}$ is $\mathcal{M}_{X,Y}=\mathcal{M}_1^X+\mathcal{M}_1^Y+\mathcal{M}_{1,1}^{X,Y}$. Under no serial dependence, $\mathcal{M}_1^X=\mathcal{M}_1^Y=1$. It then suffices to establish the boundedness of $\mathcal{M}_{1,1}^{X,Y}$ to verify the validity of (S.5). Define that $\mathbf{\Sigma}_{XY}(\cdot)=\mathrm{Cov}(\mathbf{X}(\cdot),Y)$ and $\Sigma_{YY}=\mathrm{Var}(Y)$. By (3), the cross-spectral stability measure $\mathcal{M}_{1,1}^{X,Y}$ satisfies

$$\begin{aligned}\mathcal{M}_{1,1}^{X,Y} &= \underset{\mathbf{\Phi}\in\boldsymbol{\mathcal{H}}_0,\|\mathbf{\Phi}\|_0\leqslant 1,v\in\mathbb{R}_0}{esssup} \frac{\big|\langle\mathbf{\Phi},\mathbf{\Sigma}_{XY}v\rangle\big|}{\sqrt{\langle\mathbf{\Phi},\mathbf{\Sigma}_{XX}(\mathbf{\Phi})\rangle}\sqrt{\Sigma_{YY}v^2}} \\ &= \underset{\mathbf{\Phi}\in\boldsymbol{\mathcal{H}}_0,\|\mathbf{\Phi}\|_0\leqslant 1,v\in\mathbb{R}_0}{esssup} \frac{\big|\mathrm{Cov}\big(\sum_{j=1}^p\sum_{l=1}^\infty\langle\phi_{jl},\Phi_j\rangle\xi_{jl},Yv\big)\big|}{\sqrt{\mathrm{Var}\big(\sum_{j=1}^p\sum_{l=1}^\infty\langle\phi_{jl},\Phi_j\rangle\xi_{jl}\big)}\sqrt{\mathrm{Var}\big(Yv\big)}}\leqslant 1,\end{aligned} \tag{S.6}$$

where $\mathbf{\Phi}=(\Phi_1,\ldots,\Phi_p)^{\mathrm{T}}$, $\boldsymbol{\mathcal{H}}_0=\big\{\mathbf{\Phi}\in\boldsymbol{\mathcal{H}}:\langle\mathbf{\Phi},\mathbf{\Sigma}_{XX}(\mathbf{\Phi})\rangle\in(0,\infty)\big\}$, $\|\mathbf{\Phi}\|_0=\sum_{j=1}^p I(\|\Phi_j\|_{\mathcal{S}}\neq 0)$, and $\mathbb{R}_0=\big\{v\in\mathbb{R}:\Sigma_{YY}v^2\in(0,\infty)\big\}$. We complete the proof of this lemma. □

Lemma A4. *Suppose that Conditions 4–5 hold. Denote* $\boldsymbol{Z}=\big(\Xi,\widetilde{\Xi}\big)\in\mathbb{R}^{n\times 2pd}$ *and* $\widehat{\boldsymbol{Z}}=\big(\widehat{\Xi},\breve{\Xi}\big)\in\mathbb{R}^{n\times 2pd}$. *If* $n\gtrsim d^{4\alpha+2}\log(pd)$, *then there exist some positive constants* $c_z,\tilde{c}_5,\tilde{c}_6$ *such that*

$$n^{-1}\boldsymbol{\theta}^{\mathrm{T}}\big\{\widehat{\mathbf{D}}^{-1}\big(\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\big)\widehat{\mathbf{D}}^{-1}\big\}\boldsymbol{\theta}\geqslant\underline{\mu}\big\|\boldsymbol{\theta}\big\|^2-c_zd^{\alpha+1}\big\{\log(pd)/n\big\}^{1/2}\big\|\boldsymbol{\theta}\big\|_1^2,\quad\forall\boldsymbol{\theta}\in\mathbb{R}^{2pd},$$

with probability greater than $1-\tilde{c}_5(pd)^{-\tilde{c}_6}$.

Proof. Let $\widehat{\mathbf{\Gamma}}=n^{-1}\widehat{\mathbf{D}}^{-1}(\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}})\widehat{\mathbf{D}}^{-1}$ and $\mathbf{\Gamma}=n^{-1}\mathbf{D}^{-1}\mathbb{E}[\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}]\mathbf{D}^{-1}$. It is evident that $\boldsymbol{\theta}^{\mathrm{T}}\widehat{\mathbf{\Gamma}}\boldsymbol{\theta}=\boldsymbol{\theta}^{\mathrm{T}}\mathbf{\Gamma}\boldsymbol{\theta}+\boldsymbol{\theta}^{\mathrm{T}}(\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma})\boldsymbol{\theta}$. Consequently, we have

$$\boldsymbol{\theta}^{\mathrm{T}}\widehat{\mathbf{\Gamma}}\boldsymbol{\theta}\geqslant\boldsymbol{\theta}^{\mathrm{T}}\mathbf{\Gamma}\boldsymbol{\theta}-\|\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}\|_{\max}\|\boldsymbol{\theta}\|_1^2. \tag{S.7}$$

It follows from Condition 4 and (S.7) that $\boldsymbol{\theta}^{\mathrm{T}}\widehat{\mathbf{\Gamma}}\boldsymbol{\theta}\geqslant\underline{\mu}\|\boldsymbol{\theta}\|^2-\|\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}\|_{\max}\|\boldsymbol{\theta}\|_1^2$. By Lemma 5 of Guo and Qiao, (2023), we obtain that, if $n\gtrsim d^{4\alpha+2}\log(pd)$, then with probability greater than $1-\tilde{c}_5(pd)^{-\tilde{c}_6}$,

$$\|\widehat{\mathbf{\Gamma}}-\mathbf{\Gamma}\|_{\max}\leqslant c_zd^{\alpha+1}\big\{\log(pd)/n\big\}^{1/2}. \tag{S.8}$$

Combining the above results, we complete the proof of this lemma. □

Lemma A5. *Suppose that Condition 5 holds. If $n \gtrsim d^{4\alpha+2}\log(pd)$, then there exist some positive constants $\tilde{c}_5, \tilde{c}_6$ such that*

$$\max_{\substack{j\in[2p]\\ l\in[d]}} \left| \frac{\hat{\omega}_{jl}^{-1/2} - \omega_{jl}^{-1/2}}{\omega_{jl}^{-1/2}} \right| \lesssim \sqrt{\frac{\log(pd)}{n}},$$

with probability greater than $1 - \tilde{c}_5(pd)^{-\tilde{c}_6}$.

Proof. By Proposition 3 in Guo and Qiao (2023) and $\mathcal{M}_1^Y = 1$ under no serial dependence, we complete the proof of this lemma. □

Lemma A6. *Suppose that Conditions 3, 5–6 hold. If $n \gtrsim d^{4\alpha+2}\log(pd)$, then there exist some positive constants $c_e, \tilde{c}_5, \tilde{c}_6$ such that*

$$n^{-1}\|\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\big(\mathbf{Y} - \widehat{\boldsymbol{Z}}\mathbf{b}\big)\|_{\max}^{(d\times 1)} \leqslant c_e s\big(d^{\alpha+2}\big\{\log(pd)/n\big\}^{1/2} + d^{1-\tau}\big)$$

with probability greater than $1 - \tilde{c}_5(pd)^{-\tilde{c}_6}$.

Proof. Note that

$$\begin{aligned} & n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}(\mathbf{Y} - \widehat{\boldsymbol{Z}}\mathbf{b}) \\ = & n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\mathbf{Y} - \mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\mathbf{Y}] + \mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\mathbf{Y}] - n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{b} \\ = & n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\mathbf{Y} - \mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\mathbf{Y}] + \mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\mathbf{b}] - n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{b} + \mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\epsilon}], \end{aligned} \tag{S.9}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^{\mathrm{T}}$ is the truncation error.

First, we show the deviation bounds of $n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\mathbf{Y} - \mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\mathbf{Y}]$, which can be decomposed as $\widehat{\mathbf{D}}^{-1}\big(n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\mathbf{Y} - \mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\mathbf{Y}]\big) + (\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1})\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\mathbf{Y}]$. By Lemmas A3 and A5, we have

$$\|n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\mathbf{Y} - \mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\mathbf{Y}]\|_{\max}^{(d\times 1)} \lesssim d^{\alpha+3/2}\big\{\log(pd)/n\big\}^{1/2}. \tag{S.10}$$

Second, we write $\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\mathbf{b}] - n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{b} = (\boldsymbol{\Gamma} - \widehat{\boldsymbol{\Gamma}})\mathbf{D}\mathbf{b} + \widehat{\boldsymbol{\Gamma}}(\mathbf{D} - \widehat{\mathbf{D}})\mathbf{b}$. By $\|\mathbf{D}\mathbf{b}\|_1^{(d\times 1)} = O(s)$ and (S.8), we obtain that

$$\|(\boldsymbol{\Gamma} - \widehat{\boldsymbol{\Gamma}})\mathbf{D}\mathbf{b}\|_{\max}^{(d\times 1)} \lesssim s d^{\alpha+2}\big\{\log(pd)/n\big\}^{1/2}. \tag{S.11}$$

By Lemma A5 and $\|\mathbf{D}\mathbf{b}\|_1^{(d\times 1)} = O(s)$, we have

$$\|\widehat{\boldsymbol{\Gamma}}(\mathbf{D}-\widehat{\mathbf{D}})\mathbf{b}\|_{\max}^{(d\times 1)} = \|\widehat{\boldsymbol{\Gamma}}(\mathbf{D}-\widehat{\mathbf{D}})\mathbf{D}^{-1}\mathbf{D}\mathbf{b}\|_{\max}^{(d\times 1)} \lesssim sd\big\{\log(pd)/n\big\}^{1/2}. \tag{S.12}$$

Note (S.10), (S.11) and (S.12) all hold with probability greater than $1-\tilde{c}_5(pd)^{-\tilde{c}_6}$. By Condition 6(i) and Lemma 23 in Fang et al. (2022), we have that $\big\|\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\epsilon}]\big\|_{\max} \leqslant O(sd^{1/2-\tau})$, which implies that $\big\|\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\epsilon}]\big\|_{\max}^{(d\times 1)} \lesssim sd^{1-\tau}$. Combining this with (S.9), (S.10), (S.11) and (S.12), we complete the proof of this lemma. □

Lemma A7. *Suppose that Conditions 3–6 hold, $d^{\alpha}s\lambda_n \to 0$ as $n,p,d\to\infty$, and the regularization parameter $\lambda_n \geqslant 2c_e s\|\widehat{\mathbf{D}}\|_{\max}\big(d^{\alpha+2}\{\log(pd)/n\}^{1/2} + d^{1-\tau}\big)$. Then there exist some positive constants $\tilde{c}_5,\tilde{c}_6$ such that, with probability greater than $1-\tilde{c}_5(pd)^{-\tilde{c}_6}$,*

$$\sum_{j=1}^{2p}\|\mathbf{b}_j - \widehat{\mathbf{b}}_j\| \lesssim d^{\alpha}s\lambda_n.$$

Proof. Given that $\widehat{\mathbf{b}}_j$ is the solution to the minimization problem in (10), we have

$$\begin{aligned}&-n^{-1}\mathbf{Y}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\widehat{\mathbf{b}} + \frac{1}{2}\widehat{\mathbf{b}}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\widehat{\mathbf{b}} + \lambda_n\|\widehat{\mathbf{b}}\|_1^{(d\times 1)}\\ \leqslant& -n^{-1}\mathbf{Y}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{b} + \frac{1}{2}\mathbf{b}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{b} + \lambda_n\|\mathbf{b}\|_1^{(d\times 1)}.\end{aligned}$$

Let $\boldsymbol{\Delta} = \widehat{\mathbf{b}} - \mathbf{b}$ and $\bar{S}^c$ represents the complement of S within the set $[2p]$. Consequently, we have

$$\begin{aligned}&\frac{1}{2}\boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\boldsymbol{\Delta}\\ \leqslant& -\boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{b} + \boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\mathbf{Y} + \lambda_n(\|\mathbf{b}\|_1^{(d\times 1)} - \|\widehat{\mathbf{b}}\|_1^{(d\times 1)})\\ \leqslant& \boldsymbol{\Delta}^{\mathrm{T}}(n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\mathbf{Y} - n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{b}) + \lambda_n(\|\boldsymbol{\Delta}_S\|_1^{(d\times 1)} - \|\boldsymbol{\Delta}_{\bar{S}^c}\|_1^{(d\times 1)}).\end{aligned} \tag{S.13}$$

By Lemma A6 and the choice of λ_n, we obtain that, with probability greater than $1-\tilde{c}_5(pd)^{-\tilde{c}_6}$,

$$\begin{aligned}\big|\boldsymbol{\Delta}^{\mathrm{T}}(n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\mathbf{Y} - n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{b})\big| &= \big|\boldsymbol{\Delta}^{\mathrm{T}}\widehat{\mathbf{D}}\{n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}(\mathbf{Y}-\widehat{\boldsymbol{Z}}\mathbf{b})\}\big|\\ &\leqslant \|n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}(\mathbf{Y}-\widehat{\boldsymbol{Z}}\mathbf{b})\|_{\max}^{(d\times 1)}\|\widehat{\mathbf{D}}\|_{\max}\|\boldsymbol{\Delta}\|_1^{(d\times 1)}\\ &\leqslant \frac{\lambda_n}{2}(\|\boldsymbol{\Delta}_S\|_1^{(d\times 1)} + \|\boldsymbol{\Delta}_{\bar{S}^c}\|_1^{(d\times 1)}),\end{aligned} \tag{S.14}$$

Combining (S.13) and (S.14), we have

$$\frac{3\lambda_n}{2}\|\mathbf{\Delta}_S\|_1^{(d\times 1)} - \frac{\lambda_n}{2}\|\mathbf{\Delta}_{\bar{S}^c}\|_1^{(d\times 1)} \geqslant \frac{1}{2}\mathbf{\Delta}^{\mathrm{T}} n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{\Delta} \geqslant 0,$$

which indicates that $3\|\mathbf{\Delta}_S\|_1^{(d\times 1)} \geqslant \|\mathbf{\Delta}_{\bar{S}^c}\|_1^{(d\times 1)}$. By Condition 4, $d^\alpha s\lambda_n \to 0$ and Lemma A4, we can let $\underline{\mu} \geqslant 32ds[c_z d^{\alpha+1}\{\log(pd)/n\}^{1/2}] = o(1)$, which ensures that $\mathbf{\Delta}^{\mathrm{T}} n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}} \geqslant \underline{\mu}\|\mathbf{\Delta}\widehat{\mathbf{D}}\|^2/2$. Combining this with Lemma A4 and Condition 5, we have

$$\begin{aligned}\mathbf{\Delta}^{\mathrm{T}} n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{\Delta} &\geqslant \underline{\mu}\|\mathbf{\Delta}\widehat{\mathbf{D}}\|^2 - 16c_z s d^{\alpha+2}\{\log(pd)/n\}^{1/2}\|\mathbf{\Delta}\widehat{\mathbf{D}}\|^2 \\ &\geqslant \underline{\mu}\|\mathbf{\Delta}\widehat{\mathbf{D}}\|^2/2 \geqslant \underline{\mu}c_0\alpha^{-1}d^{-\alpha}\|\mathbf{\Delta}\|^2/2.\end{aligned}$$

Note the facts that $\|\mathbf{\Delta}\|_1^{(d\times 1)} = \|\mathbf{\Delta}_S\|_1^{(d\times 1)} + \|\mathbf{\Delta}_{\bar{S}^c}\|_1^{(d\times 1)} \leqslant 4\|\mathbf{\Delta}_S\|_1^{(d\times 1)} \leqslant 4s^{1/2}\|\mathbf{\Delta}\|$ and $3\lambda_n\|\mathbf{\Delta}_S\|_1^{(d\times 1)} \geqslant \mathbf{\Delta}^{\mathrm{T}} n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{\Delta}$. Hence, $6s^{1/2}\lambda_n\|\mathbf{\Delta}\| \geqslant 3\lambda_n\|\mathbf{\Delta}_S\|_1^{(d\times 1)}/2 \geqslant \underline{\mu}c_0\alpha^{-1}d^{-\alpha}\|\mathbf{\Delta}\|^2/4$. Then we have

$$\|\mathbf{\Delta}\| \leqslant 24\alpha d^\alpha s^{1/2}\lambda_n/(\underline{\mu}c_0) \quad\text{and}\quad \|\mathbf{\Delta}\|_1^{(d\times 1)} \leqslant 96\alpha d^\alpha s\lambda_n/(\underline{\mu}c_0),$$

with probability greater than $1-\tilde{c}_5(pd)^{-\tilde{c}_6}$. The proof of this lemma is completed. □

Lemma A8. *Suppose that Condition 6(iii) holds. Then there exists some constant $c_1' \in \big(2(qs)^{-1}, 1\big)$ such that, with probability greater than $1-\tilde{c}_5(pd)^{-\tilde{c}_6}$, $|\widehat{S}_\delta| \geqslant c_1' s$ for $\widehat{S}_\delta$ defined in (4).*

Proof. By Lemma A7, we have that, with probability greater than $1-\tilde{c}_5(pd)^{-\tilde{c}_6}$,

$$\max_{j\in[p]}\|\mathbf{b}_j - \widehat{\mathbf{b}}_j\| \leqslant 24\alpha d^\alpha s^{1/2}\lambda_n/(\underline{\mu}c_0) \text{ and } \max_{j\in[p]}\|\widehat{\mathbf{b}}_{j+p}\| \leqslant 24\alpha d^\alpha s^{1/2}\lambda_n/(\underline{\mu}c_0).$$

Hence, for any $j \in [p]$, we have that

$$W_j = \|\widehat{\mathbf{b}}_j\| - \|\widehat{\mathbf{b}}_{j+p}\| \geqslant -\|\widehat{\mathbf{b}}_{j+p}\| \geqslant -24\alpha d^\alpha s^{1/2}\lambda_n/(\underline{\mu}c_0). \tag{S.15}$$

This implies $T_\delta \leqslant 24\alpha d^\alpha s^{1/2}\lambda_n/(\underline{\mu}c_0)$. Otherwise, if $T_\delta > 24\alpha d^\alpha s^{1/2}\lambda_n/(\underline{\mu}c_0)$, by (S.15), we have $\{j \in [p] : W_j < -T_\delta\}$ is a null set. Under Condition 6(iii), if $j \in S_2 = \big\{j \in [p] : \|\mathbf{b}_j\| \gg 24\alpha d^\alpha s^{1/2}\lambda_n/(\underline{\mu}c_0)\big\}$, we have $W_j = \|\widehat{\mathbf{b}}_j\| - \|\widehat{\mathbf{b}}_{j+p}\| \geqslant \|\mathbf{b}_j\| - \|\widehat{\mathbf{b}}_j - \mathbf{b}_j\| - \|\widehat{\mathbf{b}}_{j+p}\| \gg 24\alpha d^\alpha s^{1/2}\lambda_n/(\underline{\mu}c_0)$. Therefore, $S_2 \subseteq \widehat{S}_\delta = \big\{j \in [p] : W_j \geqslant T_\delta\big\}$. Combing this with Condition 6(iii), we complete the proof of this lemma. □

The proof strategy for Theorem 3 closely resembles that for Theorem 5 based on the above technical lemmas. Hence we will only provide detailed proof of Theorem 5 in Section A.4 and omit the detailed proof of Theorem 3 here.

A.3 Proof of Theorem 4

To prove Theorem 4, we firstly present some technical lemmas with their proofs.

Corollary 2. *For any subset* $G \subseteq S^c$, $\big(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \breve{\boldsymbol{\xi}}_i^{\mathrm{T}}\big) \,\Big|\, \widehat{\boldsymbol{\eta}}_i \overset{D}{=} \big(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \breve{\boldsymbol{\xi}}_i^{\mathrm{T}}\big)_{swap(G)} \,\Big|\, \widehat{\boldsymbol{\eta}}_i$.

Proof. In a similar way to the proof of Lemma A1, we obtain that $\big\{\mathbf{X}_i(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}_i(\cdot)^{\mathrm{T}}, Y_i(\cdot)\big\} \overset{D}{=} \big(\{\mathbf{X}_i(\cdot)^{\mathrm{T}}, \widetilde{\mathbf{X}}_i(\cdot)^{\mathrm{T}}\}_{\mathrm{swap}(G)}, Y_i(\cdot)\big)$. This implies that, for any $\mathbf{t} \in (\boldsymbol{\mathcal{H}}^2, \mathcal{H}_Y)$,

$$\mathbb{E}\Big[\exp\Big\{\iota\Big\langle \mathbf{t}, (\mathbf{X}_i^{\mathrm{T}}, \widetilde{\mathbf{X}}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}\Big\rangle\Big\}\Big] = \mathbb{E}\Big[\exp\Big\{\iota\Big\langle \mathbf{t}, \big\{(\mathbf{X}_i^{\mathrm{T}}, \widetilde{\mathbf{X}}_i^{\mathrm{T}})_{\mathrm{swap}(G)}, Y_i\big\}^{\mathrm{T}}\Big\rangle\Big\}\Big].$$

Given $\mathbf{t} = (\mathbf{c}_1^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_1, \dots, \mathbf{c}_p^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_p, \widetilde{\mathbf{c}}_1^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_1, \dots, \widetilde{\mathbf{c}}_p^{\mathrm{T}}\widehat{\boldsymbol{\phi}}_p, \mathbf{c}_Y^{\mathrm{T}}\widehat{\boldsymbol{\psi}})^{\mathrm{T}}$, where $\mathbf{c}_j = (c_{j1}, \dots, c_{jd_j}, 0, 0, \dots)^{\mathrm{T}}$, $\widetilde{\mathbf{c}}_j = (\tilde{c}_{j1}, \dots, \tilde{c}_{jd_j}, 0, 0, \dots)^{\mathrm{T}}$, $\mathbf{c}_Y = (c_{Y,1}, \dots, c_{Y,\tilde{d}}, 0, 0, \dots)^{\mathrm{T}}$, $\widehat{\boldsymbol{\phi}}_j = (\hat{\phi}_{j1}, \hat{\phi}_{j2}, \dots)^{\mathrm{T}}$ for $j \in [p]$ and $\widehat{\boldsymbol{\psi}} = (\hat{\psi}_1, \hat{\psi}_2, \dots)^{\mathrm{T}}$, it follows that

$$\mathbb{E}\Big[\exp\big\{\iota\big\langle \mathbf{c}, (\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \breve{\boldsymbol{\xi}}_i^{\mathrm{T}}, \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}})^{\mathrm{T}}\big\rangle\big\}\Big|\mathbf{t}\Big] = \mathbb{E}\Big[\exp\big\{\iota\big\langle \mathbf{c}, \{(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \breve{\boldsymbol{\xi}}_i^{\mathrm{T}})_{\mathrm{swap}(G)}, \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}\}^{\mathrm{T}}\big\rangle\big\}\Big|\mathbf{t}\Big], \tag{S.16}$$

for any $\mathbf{c} = (\bar{\mathbf{c}}_1^{\mathrm{T}}, \dots, \bar{\mathbf{c}}_p^{\mathrm{T}}, \bar{\widetilde{\mathbf{c}}}_1^{\mathrm{T}}, \dots, \bar{\widetilde{\mathbf{c}}}_p^{\mathrm{T}}, \bar{\mathbf{c}}_Y^{\mathrm{T}})^{\mathrm{T}}$, where $\bar{\mathbf{c}}_j = (c_{j1}, \dots, c_{jd_j})^{\mathrm{T}}$, $\bar{\widetilde{\mathbf{c}}}_j = (\tilde{c}_{j1}, \dots, \tilde{c}_{jd_j})^{\mathrm{T}}$, $\bar{\mathbf{c}}_Y = (c_{Y,1}, \dots, c_{Y,\tilde{d}})^{\mathrm{T}}$. By (S.16) and the total expectation formula, the joint characteristic function of $(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \breve{\boldsymbol{\xi}}_i^{\mathrm{T}}, \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}})$ is equal to that of $\big\{(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \breve{\boldsymbol{\xi}}_i^{\mathrm{T}})_{\mathrm{swap}(G)}, \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}\big\}$, which implies that $(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \breve{\boldsymbol{\xi}}_i^{\mathrm{T}}, \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}) \overset{D}{=} \big\{(\widehat{\boldsymbol{\xi}}_i^{\mathrm{T}}, \breve{\boldsymbol{\xi}}_i^{\mathrm{T}})_{\mathrm{swap}(G)}, \widehat{\boldsymbol{\eta}}_i^{\mathrm{T}}\big\}$, and thus completes our proof of this lemma. □

By Corollary 2, we next prove Lemma 1 in the context of FFLR.

Proof of Lemma 1 in FFLR. Denote $\widehat{\boldsymbol{\Upsilon}} = (\widehat{\boldsymbol{\eta}}_1, \dots, \widehat{\boldsymbol{\eta}}_n)^{\mathrm{T}} \in \mathbb{R}^{n \times \tilde{d}}$. First, since $W_j = \|\widehat{\mathbf{B}}_j\|_{\mathrm{F}} - \|\widehat{\mathbf{B}}_{p+j}\|_{\mathrm{F}}$ for each $j \in [p]$, the flip-sign property of $W_j = W_j(\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}, \widehat{\boldsymbol{\Upsilon}})$ holds. Second, denote $\boldsymbol{W} = (W_1, \dots, W_p)^{\mathrm{T}}$ and take any subset $G \subseteq S^c$ of null. By swapping variables in G, we define

$$\boldsymbol{W}_{\mathrm{swap}(G)} \triangleq \Big(W_1\big(\{\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}\}_{\mathrm{swap}(G)}, \widehat{\boldsymbol{\Upsilon}}\big), \dots, W_p\big(\{\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}\}_{\mathrm{swap}(G)}, \widehat{\boldsymbol{\Upsilon}}\big)\Big)^{\mathrm{T}}.$$

According to Corollary 2, we have $(\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}}, \widehat{\boldsymbol{\Upsilon}}) \overset{D}{=} \big((\widehat{\boldsymbol{\Xi}}, \breve{\boldsymbol{\Xi}})_{\mathrm{swap}(G)}, \widehat{\boldsymbol{\Upsilon}}\big)$, implying $\boldsymbol{W} \overset{D}{=} \boldsymbol{W}_{\mathrm{swap}(G)}$.

Lastly, let $S^c_- = \{j \in S^c : \delta_j = -1\}$, where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^{\mathrm{T}}$ is a sequence of independent random variables. Each δ_j follows a Rademacher distribution if $j \in S^c$, and $\delta_j = 1$ otherwise. In the first step, we establish $\boldsymbol{W}_{\mathrm{swap}(S^c_-)} = (\delta_1 W_1, \ldots, \delta_p W_p)^{\mathrm{T}}$. Following the second step, we obtain $\boldsymbol{W}_{\mathrm{swap}(S^c_-)} \overset{D}{=} \boldsymbol{W}$. Combing the above results, we have $(\delta_1 W_1, \ldots, \delta_p W_p)^{\mathrm{T}} \overset{D}{=} \boldsymbol{W}$, which completes the proof of this lemma. □

Lemma A9. *Suppose that Condition 2 holds. Then $\|\beta_j\|_{\mathcal{S}} = 0$ if and only if $j \in S^c$, where S^c is defined in (1).*

Proof. On the one hand, assume that $\|\beta_j\|_{\mathcal{S}} = 0$. For any $\big(t_Y(\cdot), t_j(\cdot)\big) \in (\mathcal{H}_Y, \mathcal{H}_j)$, the joint characteristic function of $\big(Y(\cdot), X_j(\cdot)\big)$ conditional on $X_{-j}(\cdot)$ can be factorized as

$$\begin{aligned}
& \mathbb{E}\big[\exp\big\{\iota\big\langle (t_Y, t_j)^{\mathrm{T}}, (Y, X_j)^{\mathrm{T}}\big\rangle\big\} \mid X_{-j}\big] \\
= \; & \mathbb{E}\big[\exp\big\{\iota\big\langle (t_Y, t_j)^{\mathrm{T}}, \big\{\textstyle\sum_{k\neq j}\int_{\mathcal{U}} \beta_k(\cdot, u) X_k(u) du + \varepsilon, X_j\big\}^{\mathrm{T}}\big\rangle\big\} \mid X_{-j}\big] \\
= \; & \mathbb{E}\big[\exp\big\{\iota\langle t_Y, \textstyle\sum_{k\neq j}\int_{\mathcal{U}} \beta_k(\cdot, u) X_k(u) du + \varepsilon\rangle\big\} \mid X_{-j}\big]\mathbb{E}\big[\exp\big\{\iota\langle t_j, X_j\rangle\big\} \mid X_{-j}\big] \\
= \; & \mathbb{E}\big[\exp\big\{\iota\langle t_Y, Y\rangle\big\} \mid X_{-j}\big]\mathbb{E}\big[\exp\big\{\iota\langle t_j, X_j\rangle\big\} \mid X_{-j}\big],
\end{aligned}$$

where the second equality comes from the fact that $\sum_{k\neq j}\int_{\mathcal{U}} \beta_k(\cdot, u) X_k(u) du + \varepsilon$ and X_j are independent conditional on X_{-j}. Hence it implies that $j \in S^c$.

On the other hand, assume that Y and X_j are conditionally independent, i.e., $j \in S^c$. Then the joint characteristic function conditional on $X_{-j}(\cdot)$ can be factorized as

$$\mathbb{E}\big[\exp\big\{\iota\big\langle (t_Y, t_j)^{\mathrm{T}}, (Y, X_j)^{\mathrm{T}}\big\rangle\big\} \mid X_{-j}\big] = \mathbb{E}\big[\exp\big\{\iota\langle t_Y, Y\rangle\big\} \mid X_{-j}\big]\mathbb{E}\big[\exp\big\{\iota\langle t_j, X_j\rangle\big\} \mid X_{-j}\big].$$

In FFLR, it is worth noting that the left-hand side involves an interaction term, i.e., $\mathbb{E}\Big[\exp\big\{\iota\big\langle t_Y, \int_{\mathcal{U}} \beta_j(\cdot, u) X_j(u) du\big\rangle\big\} \,\Big|\, X_{-j}\Big]$, which needs to be a constant. Condition 2 implies that the interaction term is a constant only when $\|\beta_j\|_{\mathcal{S}} = 0$. Combing the above results, we complete the proof of this lemma. □

Proof of Theorem 4. Provided that Lemma 1 applies to FFLR, this confirms that the signs of null statistics are distributed as i.i.d. coin flips. By the Theorem 3.4 in Candès et al.

(2018), we have

$$\mathbb{E}\left[\frac{|\widehat{S}_1 \cap S^c|}{|\widehat{S}_1|}\right] \leqslant q \quad \text{and} \quad \mathbb{E}\left[\frac{|\widehat{S}_0 \cap S^c|}{|\widehat{S}_0| + 1/q}\right] \leqslant q,$$

where S^c is defined as (1). By Lemma A9, we establish the equivalence between S^c in (1) and the set $\{j : \|\beta_j\|_{\mathcal{S}} = 0\}$, whose complement is defined in (16). This equivalence demonstrates that FDR in FFLR is effectively controlled, which completes our proof. □

A.4 Proof of Theorem 5

Before presenting technical lemmas used in the proof of Theorem 5, we begin with introducing some notation. For $j \in [p]$, denote $\sigma_{jlm}^{X,Y} = \mathbb{E}[\xi_{ijl}\eta_{im}]$ and its estimator $\hat{\sigma}_{jlm}^{X,Y} = n^{-1}\sum_{i=1}^n \hat{\xi}_{ijl}\hat{\eta}_{im}$ for $l \in [d]$, $m \in [\tilde{d}]$. For $j \in [2p]\backslash[p]$, denote $\sigma_{jlm}^{X,Y} = \mathbb{E}[\tilde{\xi}_{i(j-p)l}\eta_{im}]$ and its estimator $\hat{\sigma}_{jlm}^{X,Y} = n^{-1}\sum_{i=1}^n \check{\xi}_{i(j-p)l}\hat{\eta}_{im}$. Let $\boldsymbol{\Upsilon} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n)^{\mathrm{T}} \in \mathbb{R}^{n \times \tilde{d}}$ with the estimator $\widehat{\boldsymbol{\Upsilon}} = (\widehat{\boldsymbol{\eta}}_1, \ldots, \widehat{\boldsymbol{\eta}}_n)^{\mathrm{T}}$.

Lemma A10. *Suppose that Conditions 5, 7, 8 hold. If $n \gtrsim (d^{4\alpha+2} \vee \tilde{d}^{4\tilde{\alpha}+2})\log(pd\tilde{d})$, then there exist some positive constants $\tilde{c}_7, \tilde{c}_8$ such that, with probability greater than $1-\tilde{c}_7(pd\tilde{d})^{-\tilde{c}_8}$, the estimates $\{\hat{\sigma}_{jlm}^{X,Y}\}$ satisfy*

$$\max_{\substack{j\in[2p] \\ l\in[d], m\in[\tilde{d}]}} \frac{|\hat{\sigma}_{jlm}^{X,Y} - \sigma_{jlm}^{X,Y}|}{(l^{\alpha+1} \vee m^{\tilde{\alpha}+1})\omega_{jl}^{1/2}\tilde{\omega}_m^{1/2}} \lesssim \sqrt{\frac{\log(pd\tilde{d})}{n}}.$$

Proof. The proof of this lemma is similar to that of Lemma A3, we thus omit the proof. □

Lemma A11. *Suppose that Conditions 5, 7–9 hold. If $n \gtrsim (d^{4\alpha+2} \vee \tilde{d}^{4\tilde{\alpha}+2})\log(pd\tilde{d})$, then there exist some positive constants $c_e, \tilde{c}_9, \tilde{c}_{10}$ such that*

$$n^{-1}\|\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}(\widehat{\boldsymbol{\Upsilon}} - \widehat{\boldsymbol{Z}}\mathbf{B})\|_{\max}^{(d\times\tilde{d})} \leqslant c_e s d^{1/2}\big(\{d^{\alpha+3/2} \vee \tilde{d}^{\tilde{\alpha}+3/2}\}\{\log(pd\tilde{d})/n\}^{1/2} + d^{1/2-\tau}\big)$$

with probability greater than $1 - \tilde{c}_9(pd\tilde{d})^{-\tilde{c}_{10}}$.

Proof. Note that

$$
\begin{aligned}
& n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}(\widehat{\boldsymbol{\Upsilon}}-\widehat{\boldsymbol{Z}}\mathbf{B}) \\
&= n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{\Upsilon}}-\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Upsilon}]+\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Upsilon}]-n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{B} \\
&= n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{\Upsilon}}-\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Upsilon}]+\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\mathbf{B}]-n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{B}+\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\epsilon}],
\end{aligned} \tag{S.17}
$$

where $\boldsymbol{\epsilon}=(\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_n)^{\mathrm{T}}\in\mathbb{R}^{n\times\tilde{d}}$ is the truncation error.

First, we show the deviation bounds of $n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{\Upsilon}}-\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Upsilon}]$, which can be decomposed as $\widehat{\mathbf{D}}^{-1}\big(n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{\Upsilon}}-\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Upsilon}]\big)+(\widehat{\mathbf{D}}^{-1}-\mathbf{D}^{-1})\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Upsilon}]$. Then, by Condition 8, Lemmas A5 and A10, we have

$$
\|n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{\Upsilon}}-\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\Upsilon}]\|_{\max}^{(d\times\tilde{d})}\lesssim d^{1/2}(d^{\alpha+1}\vee\tilde{d}^{\tilde{\alpha}+1})\big\{\log(pd\tilde{d})/n\big\}^{1/2}. \tag{S.18}
$$

Second, we write $\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}\mathbf{B}]-n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{B}=(\boldsymbol{\Gamma}-\widehat{\boldsymbol{\Gamma}})\mathbf{D}\mathbf{B}+\widehat{\boldsymbol{\Gamma}}(\mathbf{D}-\widehat{\mathbf{D}})\mathbf{B}$. By $\|\mathbf{D}\mathbf{B}\|_1^{(d\times\tilde{d})}=O(s)$ and (S.8), we have

$$
\|(\boldsymbol{\Gamma}-\widehat{\boldsymbol{\Gamma}})\mathbf{D}\mathbf{B}\|_{\max}^{(d\times\tilde{d})}\lesssim sd^{\alpha+2}\big\{\log(pd)/n\big\}^{1/2}. \tag{S.19}
$$

By Lemma A5 and $\|\mathbf{D}\mathbf{B}\|_1^{(d\times\tilde{d})}=O(s)$, we have

$$
\|\widehat{\boldsymbol{\Gamma}}(\mathbf{D}-\widehat{\mathbf{D}})\mathbf{B}\|_{\max}^{(d\times\tilde{d})}\lesssim sd\big\{\log(pd)/n\big\}^{1/2}. \tag{S.20}
$$

Note (S.18), (S.19) and (S.20) all hold with probability greater than $1-\tilde{c}_9(pd)^{-\tilde{c}_{10}}$. By Condition 9(i) and Proposition 4 in Guo and Qiao (2023), we have $\big\|\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\epsilon}]\big\|_{\max}\leqslant O(sd^{1/2-\tau})$, which implies that $\|\mathbf{D}^{-1}\mathbb{E}[n^{-1}\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\epsilon}]\|_{\max}^{(d\times\tilde{d})}\lesssim sd^{1-\tau}$. Combing this with (S.17), (S.18), (S.19) and (S.20), we complete the proof of this lemma. □

Lemma A12. *Suppose that Conditions 4–5, 7–9 hold, $sd^{\alpha}\lambda_n\to 0$ as $n,p,d\to\infty$, and the regularization parameter $\lambda_n\geqslant 2c_e s\|\widehat{\mathbf{D}}\|_{\max}d^{1/2}\big[(d^{\alpha+3/2}\vee\tilde{d}^{\tilde{\alpha}+3/2})\{\log(pd\tilde{d})/n\}^{1/2}+d^{1/2-\tau}\big]$. Then there exist some positive constants $\tilde{c}_9,\tilde{c}_{10}$ such that, with probability greater than $1-\tilde{c}_9(pd\tilde{d})^{-\tilde{c}_{10}}$,*

$$
\sum_{j=1}^{2p}\|\mathbf{B}_j-\widehat{\mathbf{B}}_j\|_F\lesssim sd^{\alpha}\lambda_n.
$$

Proof. Since $\widehat{\mathbf{B}}_j$ is the minimizer of (17), we have

$$\begin{aligned}&-\operatorname{tr}(n^{-1}\widehat{\boldsymbol{\Upsilon}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\widehat{\mathbf{B}})+\frac{1}{2}\operatorname{tr}(\widehat{\mathbf{B}}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\widehat{\mathbf{B}})+\lambda_n\|\widehat{\mathbf{B}}\|_1^{(d\times\tilde{d})}\\ \leqslant&-\operatorname{tr}(n^{-1}\widehat{\boldsymbol{\Upsilon}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{B})+\frac{1}{2}\operatorname{tr}(\mathbf{B}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{B})+\lambda_n\|\mathbf{B}\|_1^{(d\times\tilde{d})}.\end{aligned}$$

Let $\boldsymbol{\Delta}=\widehat{\mathbf{B}}-\mathbf{B}$ and $\bar{S}^c$ represents the complement of S in the set $[2p]$. Consequently, we obtain that

$$\begin{aligned}&\frac{1}{2}\operatorname{tr}(\boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\boldsymbol{\Delta})\\ \leqslant&-\operatorname{tr}(\boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{B})+\operatorname{tr}(\boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{\Upsilon}})+\lambda_n(\|\mathbf{B}\|_1^{(d\times\tilde{d})}-\|\widehat{\mathbf{B}}\|_1^{(d\times\tilde{d})})\\ \leqslant&\operatorname{tr}\big\{\boldsymbol{\Delta}^{\mathrm{T}}(n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{\Upsilon}}-n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{B})\big\}+\lambda_n\big(\|\boldsymbol{\Delta}_S\|_1^{(d\times\tilde{d})}-\|\boldsymbol{\Delta}_{\bar{S}^c}\|_1^{(d\times\tilde{d})}\big).\end{aligned}\tag{S.21}$$

By Lemma A11 and the choice of λ_n, we obtain that, with probability greater than $1-\tilde{c}_9(pd)^{-\tilde{c}_{10}}$,

$$\begin{aligned}|\operatorname{tr}\{\boldsymbol{\Delta}^{\mathrm{T}}(n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{\Upsilon}}-n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\mathbf{B})\}|&\leqslant\|n^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}(\widehat{\boldsymbol{\Upsilon}}-\widehat{\boldsymbol{Z}}\mathbf{B})\|_{\max}^{(d\times\tilde{d})}\|\widehat{\mathbf{D}}\|_{\max}\|\boldsymbol{\Delta}\|_1^{(d\times\tilde{d})}\\ &\leqslant\frac{\lambda_n}{2}(\|\boldsymbol{\Delta}_S\|_1^{(d\times\tilde{d})}+\|\boldsymbol{\Delta}_{\bar{S}^c}\|_1^{(d\times\tilde{d})}).\end{aligned}\tag{S.22}$$

Combining (S.21) and (S.22), we have

$$\frac{3\lambda_n}{2}\|\boldsymbol{\Delta}_S\|_1^{(d\times\tilde{d})}-\frac{\lambda_n}{2}\|\boldsymbol{\Delta}_{\bar{S}^c}\|_1^{(d\times\tilde{d})}\geqslant\frac{1}{2}\operatorname{tr}(\boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\boldsymbol{\Delta})\geqslant 0,$$

which indicates that $3\|\boldsymbol{\Delta}_S\|_1^{(d\times\tilde{d})}\geqslant\|\boldsymbol{\Delta}_{\bar{S}^c}\|_1^{(d\times\tilde{d})}$. By Condition 4, $sd^{\alpha}\lambda_n\to 0$ and Lemma A4, we can let $\underline{\mu}\geqslant 32d\tilde{d}s[c_z d^{\alpha+1}\{\log(pd)/n\}^{1/2}]=o(1)$, which ensures that $\operatorname{tr}(\boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\boldsymbol{\Delta})\geqslant\underline{\mu}\|\boldsymbol{\Delta}\widehat{\mathbf{D}}\|_{\mathrm{F}}^2/2$. Combining this with Lemma A4 and Condition 5, we have

$$\begin{aligned}\operatorname{tr}(\boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\boldsymbol{\Delta})&\geqslant\underline{\mu}\|\boldsymbol{\Delta}\widehat{\mathbf{D}}\|_{\mathrm{F}}^2-16c_z\tilde{d}d^{\alpha+2}\big\{\log(pd)/n\big\}^{1/2}\|\boldsymbol{\Delta}\widehat{\mathbf{D}}\|_{\mathrm{F}}^2\\ &\geqslant\underline{\mu}\|\boldsymbol{\Delta}\widehat{\mathbf{D}}\|_{\mathrm{F}}^2/2\geqslant\underline{\mu}c_0\alpha^{-1}d^{-\alpha}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2/2.\end{aligned}$$

Note the facts that $\|\boldsymbol{\Delta}\|_1^{(d\times\tilde{d})}=\|\boldsymbol{\Delta}_S\|_1^{(d\times\tilde{d})}+\|\boldsymbol{\Delta}_{\bar{S}^c}\|_1^{(d\times\tilde{d})}\leqslant 4\|\boldsymbol{\Delta}_S\|_1^{(d\times\tilde{d})}\leqslant 4s^{1/2}\|\boldsymbol{\Delta}\|_{\mathrm{F}}$ and $3\lambda_n\|\boldsymbol{\Delta}_S\|_1^{(d\times\tilde{d})}\geqslant\operatorname{tr}(\boldsymbol{\Delta}^{\mathrm{T}}n^{-1}\widehat{\boldsymbol{Z}}^{\mathrm{T}}\widehat{\boldsymbol{Z}}\boldsymbol{\Delta})$. Hence, $6s^{1/2}\|\boldsymbol{\Delta}\|_{\mathrm{F}}\geqslant 3\lambda_n\|\boldsymbol{\Delta}_S\|_1^{(d\times\tilde{d})}/2\geqslant\underline{\mu}c_0\alpha^{-1}d^{-\alpha}\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2/4$, which implies that

$$\|\boldsymbol{\Delta}\|_{\mathrm{F}}\leqslant 24\alpha s^{1/2}d^{\alpha}\lambda_n/(\underline{\mu}c_0)\text{ and }\|\boldsymbol{\Delta}\|_1^{(d\times\tilde{d})}=\sum_{j=1}^{2p}\|\mathbf{B}_j-\widehat{\mathbf{B}}_j\|_{\mathrm{F}}\leqslant 96\alpha sd^{\alpha}\lambda_n/(\underline{\mu}c_0),$$

with probability greater than $1-\tilde{c}_9(pd)^{-\tilde{c}_{10}}$. The proof is completed. □

Lemma A13. *Suppose that Condition 9(iii) holds. Then there exists some constant* $c_2' \in (2(qs)^{-1}, 1)$ *such that* $|\widehat{S}_\delta| \geqslant c_2's$ *for* $\widehat{S}_\delta$ *defined in (4), with probability greater than* $1 - \tilde{c}_9(pd\tilde{d})^{-\tilde{c}_{10}}$.

Proof. By Lemma A12, we have that, with probability greater than $1 - \tilde{c}_9(pd\tilde{d})^{-\tilde{c}_{10}}$,

$$\max_{j\in[p]} \|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} \leqslant 24\alpha s^{1/2} d^\alpha \lambda_n/(\underline{\mu} c_0) \text{ and } \max_{j\in[p]} \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \leqslant 24\alpha s^{1/2} d^\alpha \lambda_n/(\underline{\mu} c_0).$$

Hence, for each $j \in [p]$, we have that

$$W_j = \|\widehat{\mathbf{B}}_j\|_{\mathrm{F}} - \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \geqslant -\|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \geqslant -24\alpha s^{1/2} d^\alpha \lambda_n/(\underline{\mu} c_0). \tag{S.23}$$

which implies $T_\delta \leqslant 24\alpha s^{1/2} d^\alpha \lambda_n/(\underline{\mu} c_0)$. Otherwise $\{j \in [p] : W_j < -T_\delta\}$ constitutes a null set. By Condition 9(iii), we have that if $j \in S_2 = \{j \in [p] : \|\mathbf{B}_j\|_{\mathrm{F}} \gg 24\alpha s^{1/2} d^\alpha \lambda_n/(\underline{\mu} c_0)\}$, then $W_j = \|\widehat{\mathbf{B}}_j\|_{\mathrm{F}} - \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \geqslant \|\mathbf{B}_j\|_{\mathrm{F}} - \|\widehat{\mathbf{B}}_j - \mathbf{B}_j\|_{\mathrm{F}} - \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \gg 24\alpha s^{1/2} d^\alpha \lambda_n/(\underline{\mu} c_0)$. This implies that $S_2 \subseteq \widehat{S}_\delta = \{j \in [p] : W_j \geqslant T_\delta\}$. Combing this with Condition 9(iii), we complete the proof of this lemma. □

We are now ready to prove Theorem 5.

Proof of Theorem 5. Consider the ordered statistics $|W_{(1)}| \geqslant |W_{(2)}| \geqslant \cdots \geqslant |W_{(p)}|$. Let j^* denote the index such that the threshold $T_\delta = |W_{(j^*)}|$. By definition of T_δ, $-T_\delta < W_{(j^*+1)} \leqslant 0$. We will establish Theorem 5 by investigating two scenarios: $-T_\delta < W_{(j^*+1)} < 0$ and $W_{(j^*+1)} = 0$.

Scenario 1. For $-T_\delta < W_{(j^*+1)} < 0$, given the definition of T_δ, we have

$$\frac{|\{j \in [p] : W_j \leqslant -T_\delta\}| + 2}{|\{j \in [p] : W_j \geqslant T_\delta\}|} > q. \tag{S.24}$$

This result holds because otherwise $|W_{(j^*+1)}|$ would serve as the new lower threshold for knockoffs. According to (S.24) and Lemma A13, we have

$$\big|\{j \in [p] : W_j \leqslant -T_\delta\}\big| > q\big|\{j \in [p] : W_j \geqslant T_\delta\}\big| - 2 \geqslant qc_2's - 2,$$

for $c_2' \in \big(2(qs)^{-1}, 1\big)$. By Lemma A12 with $\sum_{j=1}^{2p} \|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} \leqslant 96\alpha s d^\alpha \lambda_n/(\underline{\mu} c_0)$, and considering $\|\mathbf{B}_j\|_{\mathrm{F}} = 0$ for $j = p+1, \ldots, 2p$, we obtain

$$
\begin{aligned}
96\alpha s d^\alpha \lambda_n/(\underline{\mu} c_0) = \sum_{j=1}^{2p} \|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} &\geqslant \sum_{j \in \{j \in [p] : W_j \leqslant -T_\delta\}} \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \\
&\geqslant T_\delta \cdot \big|\{j \in [p] : W_j \leqslant -T_\delta\}\big| \geqslant T_\delta \cdot (q c_2' s - 2),
\end{aligned} \tag{S.25}
$$

where, the second inequality holds because for $j \in \{j \in [p] : W_j \leqslant -T_\delta\}$, we have $\|\widehat{\mathbf{B}}_j\|_{\mathrm{F}} - \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \leqslant -T_\delta$, implying $\|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \geqslant T_\delta$. By (S.25), we have $T_\delta \leqslant 96\alpha s d^\alpha \lambda_n \big\{\underline{\mu} c_0 (q c_2' s - 2)\big\}^{-1}$. Similarly, by Lemma A12, $\|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \geqslant \|\widehat{\mathbf{B}}_j\|_{\mathrm{F}} - T_\delta$ for $j \in \widehat{S}_\delta^c$, the triangle inequality, and $\min_{j \in S} \|\mathbf{B}_j\|_{\mathrm{F}} \geqslant \kappa_n d^\alpha \lambda_n/\underline{\mu}$, we have

$$
\begin{aligned}
96\alpha s d^\alpha \lambda_n/(\underline{\mu} c_0) &= \sum_{j=1}^{2p} \|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} = \sum_{j=1}^{p} \big(\|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} + \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}}\big) \\
&\geqslant \sum_{j \in \widehat{S}_\delta^c \cap S} \big(\|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} + \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}}\big) \\
&\geqslant \sum_{j \in \widehat{S}_\delta^c \cap S} \big(\|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} + \|\widehat{\mathbf{B}}_j\|_{\mathrm{F}} - T_\delta\big) \\
&\geqslant \sum_{j \in \widehat{S}_\delta^c \cap S} \big(\|\mathbf{B}_j\|_{\mathrm{F}} - T_\delta\big) \geqslant \big(\kappa_n d^\alpha \lambda_n/\underline{\mu} - T_\delta\big) \cdot |\widehat{S}_\delta^c \cap S|.
\end{aligned} \tag{S.26}
$$

For large enough κ_n such that $T_\delta \leqslant 96\alpha s d^\alpha \lambda_n \big\{\underline{\mu} c_0 (q c_2' s - 2)\big\}^{-1} \leqslant \kappa_n d^\alpha \lambda_n/(2\underline{\mu})$ and (S.26), it holds, with probability greater than $1 - \tilde{c}_9 (p d \tilde{d})^{-\tilde{c}_{10}}$, that

$$
\frac{|\widehat{S}_\delta \cap S|}{|S| \vee 1} = 1 - \frac{|\widehat{S}_\delta^c \cap S|}{|S| \vee 1} \geqslant 1 - \frac{192\alpha}{c_0} \kappa_n^{-1}.
$$

Scenario 2. For $W_{(j^*+1)} = 0$, we have $\widehat{S}_\delta = \{j \in [p] : W_j > 0\}$ and $\{j \in [p] : W_j \leqslant -T_\delta\} = \{j \in [p] : W_j < 0\}$.

If $|\{j \in [p] : W_j < 0\}| > c_n s$ with $c_n = 192\alpha/(c_0 \kappa_n)$, it follows from

$$
96\alpha s d^\alpha \lambda_n/(\underline{\mu} c_0) = \sum_{j=1}^{2p} \|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} \geqslant \sum_{j \in \{j \in [p] : W_j \leqslant -T_\delta\}} \|\widehat{\mathbf{B}}_{j+p}\|_{\mathrm{F}} \geqslant T_\delta \cdot |\{j \in [p] : W_j \leqslant -T_\delta\}|,
$$

that

$$
T_\delta \leqslant \frac{\sum_{j=1}^{2p} \|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}}}{|\{j \in [p] : W_j \leqslant -T_\delta\}|} = \frac{\sum_{j=1}^{2p} \|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}}}{|\{j \in [p] : W_j < 0\}|} < \frac{\kappa_n d^\alpha \lambda_n}{2\underline{\mu}}.
$$

Consequently, the argument simplifies to Scenario 1, and the subsequent analysis follows.

If $|\{j \in [p] : W_j < 0\}| \leqslant c_n s$, by $\widehat{S}_\delta = \mathrm{supp}(W_j) \backslash \{j \in [p] : W_j < 0\}$ with $\mathrm{supp}(W_j) = \{j \in [p] : W_j \neq 0\}$, we have

$$\left|\widehat{S}_\delta \cap S\right| = \left|\mathrm{supp}(W_j) \cap S\right| - \left|\{j \in [p] : W_j < 0\} \cap S\right| \geqslant \left|\mathrm{supp}(W_j) \cap S\right| - c_n s. \tag{S.27}$$

Define $\widehat{S}^c_{\mathrm{GL}} = \big\{j \in [p] : \|\widehat{\mathbf{B}}_j\|_{\mathrm{F}} = 0\big\}$. As stated in Fan et al. (2020a), we have to assume that there are no ties in the magnitude of the nonzero components of the group lasso solution. Then we can conclude that $\big\{j \in [p] : W_j = 0\big\} \subseteq \widehat{S}^c_{\mathrm{GL}}$, which shows that $[p] \backslash \widehat{S}^c_{\mathrm{GL}} \subseteq \mathrm{supp}(W_j)$. By Lemma A12, we have

$$\begin{aligned} 96\alpha s d^\alpha \lambda_n/(\underline{\mu} c_0) &= \sum_{j=1}^{2p} \|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} \geqslant \sum_{j \in \widehat{S}^c_{\mathrm{GL}} \cap S} \|\mathbf{B}_j - \widehat{\mathbf{B}}_j\|_{\mathrm{F}} \\ &= \sum_{j \in \widehat{S}^c_{\mathrm{GL}} \cap S} \|\mathbf{B}_j\|_{\mathrm{F}} \geqslant |\widehat{S}^c_{\mathrm{GL}} \cap S| \min_{j \in S} \|\mathbf{B}_j\|_{\mathrm{F}}. \end{aligned} \tag{S.28}$$

Note that $\min_{j \in S} \|\mathbf{B}_j\|_{\mathrm{F}} \geqslant \kappa_n d^\alpha \lambda_n / \underline{\mu}$. Then we can get $|\widehat{S}^c_{\mathrm{GL}} \cap S| \leqslant 96\alpha s/(c_0 \kappa_n)$ from (S.28). Therefore, we have

$$\left|([p] \backslash \widehat{S}^c_{\mathrm{GL}}) \cap S\right| \geqslant s\big\{1 - 96\alpha/(c_0\kappa_n)\big\}.$$

Combining (S.27) and $[p] \backslash \widehat{S}^c_{\mathrm{GL}} \subseteq \mathrm{supp}(W_j)$, we have $|\widehat{S}_\delta \cap S| \geqslant \left|\big([p] \backslash \widehat{S}^c_{\mathrm{GL}}\big) \cap S\right| - c_n s = s\big\{1 - 96\alpha/(c_0\kappa_n) - 192\alpha/(c_0\kappa_n)\big\}$, which shows that

$$\frac{|\widehat{S}_\delta \cap S|}{|S| \vee 1} \geqslant 1 - \frac{192\alpha}{c_0} \kappa_n^{-1}$$

holds with probability greater than $1 - \tilde{c}_9 (pd\tilde{d})^{-\tilde{c}_{10}}$.

Combing the above results under two scenarios, we have the power

$$\mathrm{Power}(\widehat{S}) = E\left[\frac{|\widehat{S}_\delta \cap S|}{|S| \vee 1}\right] \to 1,$$

which completes the proof of Theorem 5. □

A.5 Proof of Lemma 2

We will show that the set E defined in (18) is equivalent to the set $\big\{(j,k) \in [p]^2 : k \in S_j\big\}$ defined in Lemma 2.

Proof. Let β_{jk} and C_{jk} represent operators induced from the coefficient function in (19) and the conditional covariance function, respectively. Note that we use β_{jk} and C_{jk} to denote both the operators and the kernel functions for notational economy. To demonstrate Lemma 2, we will prove that $\|C_{jk}\|_{\mathcal{S}} = 0 \Longleftrightarrow \|\beta_{jk}\|_{\mathcal{S}} = 0$. Note the fact that $\|C_{jk}\|_{\mathcal{S}} = 0$ if and only if $\big\langle f, C_{jk}(g)\big\rangle = 0$ for any $f \in \mathcal{H}_j$ and $g \in \mathcal{H}_k$. By Condition 10 and Lemma S5 in Solea and Li (2022), we have $\big\langle f, C_{jk}(g)\big\rangle = 0 \Longleftrightarrow \mathrm{Cov}\big(\langle f, X_j\rangle, \langle g, X_k\rangle \mid X_{-\{j,k\}}\big) = 0 \Longleftrightarrow \mathrm{Cov}\big(\langle f, \sum_{l\neq j} \beta_{jl}(X_l) + \varepsilon_j\rangle, \langle g, X_k\rangle \mid X_{-\{j,k\}}\big) = 0$. Since ε_j and X_k are independent and for any $l \in [p]\backslash\{j,k\}$ and $\mathrm{Cov}\big(\big\langle f, \beta_{jl}(X_l)\big\rangle, \langle g, X_k\rangle | X_{-\{j,k\}}\big) = 0$, we obtain that

$$
\begin{aligned}
\big\langle f, C_{jk}(g)\big\rangle = 0 &\Longleftrightarrow \mathrm{Cov}\big(\big\langle f, \beta_{jk}(X_k)\big\rangle, \langle g, X_k\rangle \mid X_{-\{j,k\}}\big) = 0 \\
&\Longleftrightarrow \|\beta_{jk}\|_{\mathcal{S}} = 0, \text{ provided that } \beta_{jk} \text{ is a linear operator,}
\end{aligned}
$$

which implies that $\|C_{jk}\|_{\mathcal{S}} = 0$ if and only if $\|\beta_{jk}\|_{\mathcal{S}} = 0$ and thus completes our proof. □

A.6 Proof of Theorem 6

Note that we have already established the validity of Lemma 1 within the FFLR framework, which can be directly extended to each row of $\mathbf{W}$ within the FGGM framework. We are now ready to prove Theorem 6.

Proof. Provided the validity of Lemma 1 in FFLR, it can be similarly demonstrated that the knockoff statistics $\mathbf{W}$ satisfy the sign-flip property on the neighborhood set NE_j for each $j \in [p]$ at the rowwise level, where $NE_j = \big\{k \in [p]\backslash\{j\} : \|C_{jk}\|_{\mathcal{S}} \neq 0\big\}$. By Theorem 3.1 in Li and Maathuis (2021), we have

$$
\mathbb{E}\left[\frac{|\widehat{E}_{\mathrm{A},1} \cap E^c|}{|\widehat{E}_{\mathrm{A},1}| \vee 1}\right] \leqslant q, \quad \text{and} \quad \mathbb{E}\left[\frac{|\widehat{E}_{\mathrm{O},1} \cap E^c|}{|\widehat{E}_{\mathrm{O},1}| \vee 1}\right] \leqslant q,
$$

where E is defined in (18), which implies that $\mathrm{FDR}_{\mathrm{A}} \leqslant q$, and $\mathrm{FDR}_{\mathrm{O}} \leqslant q$ in GGM. Similarly, drawing from Theorem 3.2 in Li and Maathuis (2021), we establish that the modified FDR can be controlled, i.e., $\mathrm{mFDR}_{\mathrm{A}} \leqslant q$, and $\mathrm{mFDR}_{\mathrm{O}} \leqslant q$. □

A.7 Proof of Theorem 7

To prove Theorem 7, we firstly present several technical lemmas with their proofs. In the following lemmas, let $\boldsymbol{Z}_{-j} = (\boldsymbol{\Xi}_{-j}, \widetilde{\boldsymbol{\Xi}}_{-j}) \in \mathbb{R}^{n\times 2(p-1)d}$, $\boldsymbol{\Xi}_{-j} = (\boldsymbol{\xi}_{1(-j)}, \ldots, \boldsymbol{\xi}_{n(-j)})^{\mathrm{T}} \in \mathbb{R}^{n\times(p-1)d}$, $\widetilde{\boldsymbol{\Xi}}_{-j} = (\widetilde{\boldsymbol{\xi}}_{1(-j)}, \ldots, \widetilde{\boldsymbol{\xi}}_{n(-j)})^{\mathrm{T}} \in \mathbb{R}^{n\times(p-1)d}$, $\boldsymbol{\xi}_{i(-j)} = (\boldsymbol{\xi}_{i1}^{\mathrm{T}}, \ldots, \boldsymbol{\xi}_{i(j-1)}^{\mathrm{T}}, \boldsymbol{\xi}_{i(j+1)}^{\mathrm{T}}, \ldots, \boldsymbol{\xi}_{ip}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{(p-1)d}$, $\widetilde{\boldsymbol{\xi}}_{i(-j)} = (\widetilde{\boldsymbol{\xi}}_{i1}^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\xi}}_{i(j-1)}^{\mathrm{T}}, \widetilde{\boldsymbol{\xi}}_{i(j+1)}^{\mathrm{T}}, \ldots, \widetilde{\boldsymbol{\xi}}_{ip}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{(p-1)d}$, $\boldsymbol{\Xi}_j = (\boldsymbol{\xi}_{1j}, \ldots, \boldsymbol{\xi}_{nj})^{\mathrm{T}}$, $\mathbf{D}_{-j} = \mathrm{diag}(\mathbf{D}_1, \ldots, \mathbf{D}_{j-1}, \mathbf{D}_{j+1}, \ldots, \mathbf{D}_p, \mathbf{D}_1, \ldots, \mathbf{D}_{j-1}, \mathbf{D}_{j+1}, \ldots, \mathbf{D}_p) \in \mathbb{R}^{2(p-1)d\times 2(p-1)d}$, and $\mathbf{B}_{j(-j)} = (\mathbf{B}_{j1}^{\mathrm{T}}, \ldots, \mathbf{B}_{j(j-1)}^{\mathrm{T}}, \mathbf{B}_{j(j+1)}^{\mathrm{T}}, \ldots, \mathbf{B}_{jp}^{\mathrm{T}}, \mathbf{B}_{j(p+1)}^{\mathrm{T}}, \ldots, \mathbf{B}_{j(p+j-1)}^{\mathrm{T}}, \mathbf{B}_{j(p+j+1)}^{\mathrm{T}}, \ldots, \mathbf{B}_{j(2p)}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2(p-1)d\times d}$ with estimates $\widehat{\boldsymbol{Z}}_{-j} = (\widehat{\boldsymbol{\Xi}}_{-j}, \breve{\boldsymbol{\Xi}}_{-j})$, $\widehat{\boldsymbol{\Xi}}_{-j} = (\widehat{\boldsymbol{\xi}}_{1(-j)}, \ldots, \widehat{\boldsymbol{\xi}}_{n(-j)})^{\mathrm{T}}$, $\breve{\boldsymbol{\Xi}}_{-j} = (\breve{\boldsymbol{\xi}}_{1(-j)}, \ldots, \breve{\boldsymbol{\xi}}_{n(-j)})^{\mathrm{T}}$, $\widehat{\boldsymbol{\xi}}_{i(-j)} = (\widehat{\boldsymbol{\xi}}_{i1}^{\mathrm{T}}, \ldots, \widehat{\boldsymbol{\xi}}_{i(j-1)}^{\mathrm{T}}, \widehat{\boldsymbol{\xi}}_{i(j+1)}^{\mathrm{T}}, \ldots, \widehat{\boldsymbol{\xi}}_{ip}^{\mathrm{T}})^{\mathrm{T}}$, $\breve{\boldsymbol{\xi}}_{i(-j)} = (\breve{\boldsymbol{\xi}}_{i1}^{\mathrm{T}}, \ldots, \breve{\boldsymbol{\xi}}_{i(j-1)}^{\mathrm{T}}, \breve{\boldsymbol{\xi}}_{i(j+1)}^{\mathrm{T}}, \ldots, \breve{\boldsymbol{\xi}}_{ip}^{\mathrm{T}})^{\mathrm{T}}$, $\widehat{\boldsymbol{\Xi}}_j = (\widehat{\boldsymbol{\xi}}_{1j}, \ldots, \widehat{\boldsymbol{\xi}}_{nj})^{\mathrm{T}}$, $\widehat{\mathbf{D}}_{-j} = \mathrm{diag}(\widehat{\mathbf{D}}_1, \ldots, \widehat{\mathbf{D}}_{j-1}, \widehat{\mathbf{D}}_{j+1}, \ldots, \widehat{\mathbf{D}}_p, \widehat{\mathbf{D}}_1, \ldots, \widehat{\mathbf{D}}_{j-1}, \widehat{\mathbf{D}}_{j+1}, \ldots, \widehat{\mathbf{D}}_p)$, and $\widehat{\mathbf{B}}_{j(-j)} = (\widehat{\mathbf{B}}_{j1}^{\mathrm{T}}, \ldots, \widehat{\mathbf{B}}_{j(j-1)}^{\mathrm{T}}, \widehat{\mathbf{B}}_{j(j+1)}^{\mathrm{T}}, \ldots, \widehat{\mathbf{B}}_{jp}^{\mathrm{T}}, \widehat{\mathbf{B}}_{j(p+1)}^{\mathrm{T}}, \ldots, \widehat{\mathbf{B}}_{j(p+j-1)}^{\mathrm{T}}, \widehat{\mathbf{B}}_{j(p+j+1)}^{\mathrm{T}}, \ldots, \widehat{\mathbf{B}}_{j(2p)}^{\mathrm{T}})^{\mathrm{T}}$ for $i \in [n]$ and $j \in [p]$.

Lemma A14. *Suppose that Conditions 4–5 hold. If $n \gtrsim d^{4\alpha+2}\log(pd)$, for each $j \in [p]$, there exist some positive constants $c_z, \tilde{c}_{11}, \tilde{c}_{12}$ such that, with probability greater than $1-\tilde{c}_{11}(pd)^{-\tilde{c}_{12}}$,*

$$\boldsymbol{\theta}^{\mathrm{T}}\big\{n^{-1}\widehat{\mathbf{D}}_{-j}^{-1}(\widehat{\boldsymbol{Z}}_{-j}^{\mathrm{T}}\widehat{\boldsymbol{Z}}_{-j})\widehat{\mathbf{D}}_{-j}^{-1}\big\}\boldsymbol{\theta} \geqslant \underline{\mu}\|\boldsymbol{\theta}\|^2 - c_z d^{\alpha+1}\big\{\log(pd)/n\big\}^{1/2}\|\boldsymbol{\theta}\|_1^2, \quad \forall \boldsymbol{\theta} \in \mathbb{R}^{2(p-1)d}.$$

Proof. The proof of this lemma for FGGM is similar to that of Lemma A4 for FFLR and hence is omitted here. It is noteworthy that, in an analogy to the infimum $\underline{\mu}$ defined in Condition 4 for FFLR, we can define the corresponding infimum for FGGM, which is used in our proof of this lemma and is no less than $\underline{\mu}$ for FFLR. Hence, the result with the presence of $\underline{\mu}$ in this lemma remains valid. □

Lemma A15. *Suppose that Conditions 5 and 11 hold. If $n \gtrsim d^{4\alpha+2}\log(pd)$, for each $j \in [p]$, there exist some positive constants $\tilde{c}_e, \tilde{c}_{11}, \tilde{c}_{12}$ such that*

$$\|n^{-1}\widehat{\mathbf{D}}_{-j}^{-1}\widehat{\boldsymbol{Z}}_{-j}^{\mathrm{T}}(\widehat{\boldsymbol{\Xi}}_j - \widehat{\boldsymbol{Z}}_{-j}\mathbf{B}_{j(-j)})\|_{\max}^{(d\times d)} \leqslant \tilde{c}_e s(d^{\alpha+2}\{\log(pd)/n\}^{1/2} + d^{1-\tau}),$$

with probability greater than $1-\tilde{c}_{11}(pd)^{-\tilde{c}_{12}}$.

Proof. The proof of this lemma is similar to that of Lemma A11, thus being omitted here. □

Lemma A16. *Suppose that Conditions 4–5 and 11 hold, $s_j d^\alpha \lambda_{nj} \to 0$ as $n, p, d \to \infty$, and any regularization parameter $\lambda_{nj} \geqslant 2c_e s_j \|\widehat{\mathbf{D}}_{-j}\|_{\max}(d^{\alpha+2}\{\log(pd)/n\} + d^{1-\tau})$ for each $j \in [p]$. Then for each $j \in [p]$, there exist some positive constants $\tilde{c}_{11}, \tilde{c}_{12}$ such that*

$$\sum_{k \in [2p] \backslash \{j, p+j\}} \|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_F \lesssim s_j d^\alpha \lambda_{nj},$$

with probability greater than $1 - \tilde{c}_{11}(pd)^{-\tilde{c}_{12}}$.

Proof. The proof of this lemma is similar to that of Lemma A12, thus being omitted here. Specifically, we obtain that $\sum_{k \in [2p] \backslash \{j,p+j\}} \|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} \leqslant 96\alpha s_j d^\alpha \lambda_n / (\underline{\mu} c_0)$. □

Lemma A17. *Suppose that Condition 11(iii) holds, then there exists $c' \in \big((1+a)c_a p(qs)^{-1}, 1\big)$ such that $|\widehat{E}_{O,\delta}| \geqslant c'|E|$, with probability greater than $1 - \tilde{c}_{11}(pd)^{-\tilde{c}_{13}}$.*

Proof. First, applying Condition 11(iii) along with Lemmas A13 and A16, we obtain that $|\widehat{S}_{\delta,j}| \geqslant c_j s_j$ holds with probability greater than $1 - \tilde{c}_{11}(pd)^{-\tilde{c}_{12}}$ for each $j \in [p]$, where $c_j \in \big((1+a)c_a p(qs)^{-1}, 1\big)$. By Bonferroni inequality, $|\widehat{S}_{\delta,j}| \geqslant c_j s_j$ holds simultaneously across $j \in [p]$ with probability greater than $1 - \tilde{c}_{11}(pd)^{-\tilde{c}_{13}}$, which can be achieved for sufficiently large n. Then under the OR rule, we have

$$|\widehat{E}_{\mathrm{O},\delta}| \geqslant \sum_{j=1}^{p} |\widehat{S}_{\delta,j}| \geqslant \sum_{j=1}^{p} c_j s_j \geqslant c' \sum_{j=1}^{p} s_j = c'|E|,$$

where $c' = \inf_{j \in [p]} c_j \in \big((1+a)c_a p(qs)^{-1}, 1\big)$. The proof is completed. □

We are now ready to prove Theorem 7.

Proof of Theorem 7. First, it is essential to note that for each j, the selected threshold $T_{\delta,j}$ represents the minimum positive number that satisfies the constraints in the optimization problem (22). Let $|W_{j(1)}| \geqslant |W_{j(2)}| \geqslant \cdots \geqslant |W_{j(p)}|$ represent the ordered statistics. The index at which the threshold $T_{\delta,j} = |W_{j(k_j^*)}|$ is reached is denoted by k_j^*. Similar to Theorem 5, we will prove this theorem in two scenarios: $-T_{\delta,j} < W_{j(k_j^*+1)} < 0$ and $W_{j(k_j^*+1)} = 0$, respectively.

Scenario 1. For $-T_{\delta,j} < W_{j(k_j^*+1)} < 0$, by the definition of $T_{\delta,j}$, we obtain that

$$\frac{\big|\big\{k \in [p] \backslash \{j\} : W_{jk} \leqslant -T_{\delta,j}\big\}\big| + 1 + a}{|\widehat{E}_{\mathrm{O},\delta}|} > \frac{q}{c_a p}. \tag{S.29}$$

The result in (S.29) holds because otherwise $|W_{j(k_j^*+1)}|$ would represent the new lower threshold for knockoffs. By (S.29), we have $\big|\{k \in [p]\backslash\{j\} : W_{jk} \leqslant -T_{\delta,j}\}\big| > q|\widehat{E}_{\mathrm{O},\delta}|/(c_a p) - 1 - a$. Combining this with the result in Lemma A17, we have $q|\widehat{E}_{\mathrm{O},\delta}|/(c_a p) - 1 - a \geqslant qc'|E|/(c_a p) - 1 - a$. Then, we obtain that

$$\big|\{k \in [p]\backslash\{j\} : W_{jk} \leqslant -T_{\delta,j}\}\big| > qc'|E|/(c_a p) - 1 - a, \tag{S.30}$$

where $c' \in \big((1+a)c_a p(qs)^{-1}, 1\big)$. It follows from Lemma A16 that $\sum_{k\in[2p]\backslash\{j,p+j\}} \|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} \leqslant 96 s_j d^{\alpha}\lambda_{nj}/(c_0\underline{\mu})$. For each $j \in [p]$ and $k \in \{p+1,\ldots,2p\}\backslash\{j+p\}$, we have $\|\mathbf{B}_{jk}\|_{\mathrm{F}} = 0$. Then we obtain that

$$\begin{aligned} 96 s_j d^{\alpha}\lambda_{nj}/(\underline{\mu}c_0) &= \sum_{k\in[2p]\backslash\{j,p+j\}} \|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} \\ &\geqslant \sum_{k\in\{k\in[p]\backslash\{j\}:W_{jk}\leqslant -T_{\delta,j}\}} \|\widehat{\mathbf{B}}_{j(k+p)}\|_{\mathrm{F}} \\ &\geqslant T_{\delta,j} \cdot |\{k \in [p]\backslash\{j\} : W_{jk} \leqslant -T_{\delta,j}\}, \end{aligned} \tag{S.31}$$

where the last inequality holds since when $W_{jk} \leqslant -T_{\delta,j}$, it implies that $\|\widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} - \|\widehat{\mathbf{B}}_{j(k+p)}\|_{\mathrm{F}} \leqslant -T_{\delta,j}$ and then follows that $\|\widehat{\mathbf{B}}_{j(k+p)}\|_{\mathrm{F}} \geqslant T_{\delta,j}$. Combing the results in (S.30) and (S.31), we obtain that

$$T_{\delta,j} \leqslant \frac{96 s_j d^{\alpha}\lambda_{nj}}{(qc's/c_a p - 1 - a)\underline{\mu}c_0}. \tag{S.32}$$

Similarly, by Lemma A16, the triangle inequality, and Condition 11(ii), we have

$$\begin{aligned} 96\alpha s_j d^{\alpha}\lambda_{nj}/(\underline{\mu}c_0) &= \sum_{k\in[2p]\backslash\{j,p+j\}} \|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} \\ &= \sum_{k\in[p]\backslash\{j\}} \big\{\|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} + \|\widehat{\mathbf{B}}_{j(k+p)}\|_{\mathrm{F}}\big\} \\ &\geqslant \sum_{k\in\widehat{S}^c_{\delta,j}\cap S_j} \big\{\|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} + \|\widehat{\mathbf{B}}_{j(k+p)}\|_{\mathrm{F}}\big\} \\ &\geqslant \sum_{k\in\widehat{S}^c_{\delta,j}\cap S_j} \big\{\|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} + \|\widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} - T_{\delta,j}\big\} \\ &\geqslant \sum_{k\in\widehat{S}^c_{\delta,j}\cap S_j} \|\mathbf{B}_{jk}\|_{\mathrm{F}} - T_{\delta,j} \geqslant \{\kappa_n d^{\alpha}\lambda_{nj}/\underline{\mu} - T_{\delta,j}\} \cdot |\widehat{S}^c_{\delta,j} \cap S_j|, \end{aligned} \tag{S.33}$$

where the second inequality holds since $\|\widehat{\mathbf{B}}_{j(k+p)}\|_{\mathrm{F}} \geqslant \|\widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} - T_{\delta,j}$ for $j \in \widehat{S}_{\delta,j}^c$. For sufficiently large κ_n, by (S.32), we obtain that $T_{\delta,j} \leqslant 96 s_j d^\alpha \lambda_{nj} \big\{(qc's/c_a p - 1 - a)\underline{\mu} c_0\big\}^{-1} \leqslant \kappa_n d^\alpha \lambda_{nj}/(2\underline{\mu})$. Combining this with (S.33) yields that $|\widehat{S}_{\delta,j}^c \cap S_j| \leqslant 192\alpha s_j/c_0$, which implies that

$$\frac{|\widehat{S}_{\delta,j} \cap S_j|}{|E| \vee 1} = \frac{|S_j|}{|E| \vee 1} - \frac{|\widehat{S}_{\delta,j}^c \cap S_j|}{|E| \vee 1} \geqslant \frac{s_j}{s \vee 1} - \frac{192\alpha s_j}{c_0 s}\kappa_n^{-1} \tag{S.34}$$

holds with probability greater than $1 - \tilde{c}_{11}(pd)^{-\tilde{c}_{12}}$.

Scenario 2. For $W_{j(k_j^*+1)} = 0$, we have $\widehat{S}_{\delta,j} = \big\{k \in [p]\backslash\{j\} : W_{jk} > 0\big\}$ and $\big\{k \in [p]\backslash\{j\} : W_{jk} \leqslant -T_{\delta,j}\big\} = \big\{k \in [p]\backslash\{j\} : W_{jk} < 0\big\}$.

If $\big|\big\{k \in [p]\backslash\{j\} : W_{jk} < 0\big\}\big| > c_{jn}s_j$ for each $j \in [p]$, where $c_{jn} = 192\alpha/(c_0\kappa_n)$, it follows from

$$\begin{aligned} 96\alpha s_j d^\alpha \lambda_{nj}/(\underline{\mu} c_0) &\geqslant \sum_{k \in [2p]\backslash\{j, p+j\}} \|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} \\ &\geqslant \sum_{\{k \in [p]\backslash\{j\} : W_{jk} \leqslant -T_{\delta,j}\}} \|\widehat{\mathbf{B}}_{jk+p}\|_{\mathrm{F}} \\ &\geqslant T_{\delta,j} \cdot \big|\big\{k \in [p]\backslash\{j\} : W_{jk} \leqslant -T_{\delta,j}\big\}\big|, \end{aligned}$$

that $T_{\delta,j} \leqslant (\kappa_n d^\alpha \lambda_{nj})/(2\underline{\mu})$. Consequently, the argument simplifies to Scenario 1, and the subsequent analysis follows.

If $|\{k \in [p]\backslash\{j\} : W_{jk} < 0\}| \leqslant c_{jn}s_j$, by $\widehat{S}_{\delta,j} = \mathrm{supp}(W_{jk})\backslash\{k \in [p]\backslash\{j\} : W_{jk} < 0\}$, where $\mathrm{supp}(W_{jk}) = \{k \in [p]\backslash\{j\} : W_{jk} \neq 0\}$, we have

$$\begin{aligned} |\widehat{S}_{\delta,j} \cap S_j| &= |\mathrm{supp}(W_{jk}) \cap S_j| - \big|\big\{k \in [p]\backslash\{j\} : W_{jk} < 0\big\} \cap S_j\big| \\ &\geqslant |\mathrm{supp}(W_{jk}) \cap S_j| - c_{jn}s_j \\ &\geqslant \big|\big\{[p]\backslash\widehat{S}_{\mathrm{GL},j}^c\big\} \cap S_j\big| - c_{jn}s_j, \end{aligned} \tag{S.35}$$

where $\widehat{S}_{\mathrm{GL},j}^c = \big\{k \in [p]\backslash\{j\} : \|\widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} = 0\big\}$. The last inequality in (S.35) holds since $\big\{k \in [p]\backslash\{j\} : W_{jk} = 0\big\} \subseteq \widehat{S}_{\mathrm{GL},j}^c$, i.e., $[p]\backslash\widehat{S}_{\mathrm{GL},j}^c \subseteq \mathrm{supp}(W_{jk})$. By Lemma A16, we have

$$\begin{aligned} 96\alpha s_j d^\alpha \lambda_{nj}/(\underline{\mu} c_0) &= \sum_{k \in [2p]\backslash\{j, p+j\}} \|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} \geqslant \sum_{j \in \widehat{S}_{\mathrm{GL},j}^c \cap S_j} \|\mathbf{B}_{jk} - \widehat{\mathbf{B}}_{jk}\|_{\mathrm{F}} \\ &= \sum_{j \in \widehat{S}_{\mathrm{GL},j}^c \cap S_j} \|\mathbf{B}_{jk}\|_{\mathrm{F}} \geqslant |\widehat{S}_{\mathrm{GL},j}^c \cap S_j| \min_{k \in S_j} \|\mathbf{B}_{jk}\|_{\mathrm{F}}. \end{aligned} \tag{S.36}$$

By (S.36) and Condition 11(ii), we can get $|\widehat{S}^c_{\mathrm{GL},j} \cap S_j| \leqslant 96\alpha s_j/(c_0\kappa_n)$, which indicates that $\left|\{[p]\backslash\widehat{S}^c_{\mathrm{GL},j}\} \cap S_j\right| \geqslant s_j\{1 - 96\alpha/(c_0\kappa_n)\}$. Combining this with the result in (S.35), we have $|\widehat{S}_j \cap S_j| = s_j\{1 - 96\alpha/(c_0\kappa_n) - 192\alpha/(c_0\kappa_n)\}$, which means that, with probability greater than $1 - \tilde{c}_{11}(pd)^{-\tilde{c}_{12}}$,

$$\frac{|\widehat{S}_{\delta,j} \cap S_j|}{|E| \vee 1} \geqslant \frac{s_j}{s}\{1 - 192\alpha/(c_0\kappa_n)\}. \tag{S.37}$$

Finally, combing results in (S.34) and (S.37) with Lemma A17, we have that

$$\sum_{j=1}^{p} \frac{|\widehat{S}_{\delta,j} \cap S_j|}{|E| \vee 1} \geqslant \sum_{j=1}^{p} \frac{s_j}{s}\{1 - 192\alpha/(c_0\kappa_n)\}$$

holds with probability greater than $1 - \tilde{c}_{11}(pd)^{-\tilde{c}_{13}}$. Then it follows that

$$E\left[\sum_{j=1}^{p} \frac{|\widehat{S}_{\delta,j} \cap S_j|}{|E| \vee 1}\right] \geqslant \left[\sum_{j=1}^{p} \frac{s_j}{s}\{1 - 192\alpha/(c_0\kappa_n)\}\right]\{1 - \tilde{c}_{11}(pd)^{-\tilde{c}_{13}}\} \to 1. \tag{S.38}$$

By Lemma 2 and the OR rule, we have $|\widehat{E}_{\mathrm{O},\delta} \cap E| \geqslant \sum_{j=1}^{p} |\widehat{S}_{\delta,j} \cap S_j|$, which implies that

$$E\left[\frac{|\widehat{E}_{\mathrm{O},\delta} \cap E|}{|E| \vee 1}\right] \geqslant E\left[\sum_{j=1}^{p} \frac{|\widehat{S}_{\delta,j} \cap S_j|}{|E| \vee 1}\right]. \tag{S.39}$$

Combining (S.38) and (S.39), we complete the proof of Theorem 7. □

A.8 Proof of Lemma 3

Proof of Lemma 3. With $\Sigma_{X_jX_k} = \Sigma^{1/2}_{X_jX_j} C_{X_jX_k} \Sigma^{1/2}_{X_kX_k}$, $\Sigma^{1/2}_{X_jX_j} = \sum_{l=1}^{\infty} \omega^{1/2}_{jl} \phi_{jl} \otimes \phi_{jl}$, and $C_{X_jX_k} = \sum_{l=1}^{\infty}\sum_{m=1}^{\infty} \mathrm{Corr}(\xi_{jl}, \xi_{km})(\phi_{jl} \otimes \phi_{km})$, we can prove Lemma 3 akin to Theorem 2 in Solea and Li (2022). Hence, the proof is omitted. □

B Additional derivations

B.1 Simplified objective functions

In Section 4.1, we give the corresponding equivalent forms of the objective function in (25), under E1, E2 and E3. In this section, we will provide detailed derivations for these simplified forms. Consider any $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}} \in \boldsymbol{\mathcal{H}}$, where each $x_j(\cdot) = \sum_{l=1}^{\infty} c_{jl}\phi_{jl}(\cdot) \in \mathcal{H}_j$.

E1: Consider $R_{X_jX_j} = \sum_{l=1}^{\infty} r(\phi_{jl} \otimes \phi_{jl})$ for $r \in [0,1]$. By Lemma 3 and the definition of operator norm, we have

$$
\begin{aligned}
\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{L}} &= \sup_{\|x_j\| \leqslant 1} \left\|(C_{X_jX_j} - R_{X_jX_j})(x_j)\right\| = \sup_{\|x_j\| \leqslant 1} \left\|\sum_{l=1}^{\infty}(1-r)(\phi_{jl} \otimes \phi_{jl})(x_j)\right\| \\
&= \sup_{\|x_j\| \leqslant 1} \left\|\sum_{l=1}^{\infty}(1-r)\langle \phi_{jl}, x_j\rangle \phi_{jl}\right\| = \sup_{\|x_j\| \leqslant 1} \left\{\sum_{l=1}^{\infty}(1-r)^2 \langle \phi_{jl}, x_j\rangle^2\right\}^{1/2} \\
&= |1-r| \sup_{\|x_j\| \leqslant 1} \left\{\sum_{l=1}^{\infty}\langle \phi_{jl}, x_j\rangle^2\right\}^{1/2} = |1-r| \sup_{\|x_j\| \leqslant 1} \|x_j\| \\
&= |1-r| = 1-r,
\end{aligned}
$$

which means the objective function can be simplified to $\min_r(1-r)$.

E2: Consider $R_{X_jX_j} = \sum_{l=1}^{\infty} r_j(\phi_{jl} \otimes \phi_{jl})$ for $r_j \in [0,1]$. By the similar arguments as above, we can obtain that

$$
\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{L}} = 1 - r_j,
$$

which means the the objective functions can be simplified to $\min_{(r_1,\ldots,r_p)} \sum_j (1-r_j)$.

E3: Consider $R_{X_jX_j} = \sum_{l=1}^{\infty} r_{jl}(\phi_{jl} \otimes \phi_{jl})$ for $r_{jl} \in [0,1]$. Likewise, we obtain that

$$
\begin{aligned}
\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{L}} &= \sup_{\|x_j\| \leqslant 1} \left\|(C_{X_jX_j} - R_{X_jX_j})(x_j)\right\| = \sup_{\|x_j\| \leqslant 1} \left\|\sum_{l=1}^{\infty}(1-r_{jl})(\phi_{jl} \otimes \phi_{jl})(x_j)\right\| \\
&= \sup_{\|x_j\| \leqslant 1} \left\|\sum_{l=1}^{\infty}(1-r_{jl})\langle \phi_{jl}, x_j\rangle \phi_{jl}\right\| = \sup_{\|x_j\| \leqslant 1} \left\{\sum_{l=1}^{\infty}(1-r_{jl})^2 \langle \phi_{jl}, x\rangle^2\right\}^{1/2} \\
&= \sup_{\|x_j\| \leqslant 1} \left\{\sum_{l=1}^{\infty}(1-r_{jl})^2 c_{jl}^2\right\}^{1/2} \\
&\leqslant \sup_l |1-r_{jl}| \sup_{\|x_j\| \leqslant 1} \left(\sum_{l=1}^{\infty} c_{jl}^2\right)^{1/2} = \sup_l |1-r_{jl}|,
\end{aligned}
$$

which indicates that $\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{L}} \leqslant \sup_l |1-r_{jl}|$. Next we will prove that $\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{L}} = \sup_l |1-r_{jl}|$. Let's consider two scenarios. First, when $\sup_l |1-r_{jl}| = \max_l |1-r_{jl}|$, we set $c_{jl*} = 1$ for l^* such that $|1-r_{jl*}| = \max_l |1-r_{jl}|$. Consequently, we obtain $\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{L}} = \max_l |1-r_{jl}| = \sup_l |1-r_{jl}|$. Second, if $\sup_l |1-r_{jl}| \notin \{|1-r_{jl}| : l \geqslant 1\}$, let $\sup_l |1-r_{jl}| = |1-r^*|$. In this case, there exists a subsequence l_m for $m \geqslant 1$ such that $r_{jl_m} \to r^*$ as $m \to \infty$. Under this scenario, we set $c_{jl_m} \to 1$ as $m \to \infty$. Consequently, we conclude that $\|C_{X_jX_j} - R_{X_jX_j}\|_{\mathcal{L}} = \sup_l |1 - r_{jl}|$. Combining the results under two scenarios, we complete the proof.

For (25) in E3, the objective function is simplified as: $\min_{(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)} \sum_j \sup_l |1 - r_{jl}|$. This corresponds to a non-smooth positive semidefinite programming problem, which poses a computationally challenging task. For the sake of computational simplicity, we instead consider solving the following smooth positive semidefinite programming problem:

$$\begin{aligned} &\min_{(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)} \sum_{j=1}^{p}\sum_{l=1}^{k_n}(1 - r_{jl}) \\ &\text{subject to } r_{jl} \in [0,1],\ 2\widehat{\mathbf{C}}_{XX} - \widehat{\mathbf{R}}_{XX} \succeq 0. \end{aligned} \tag{S.40}$$

Remark 7. *We aim to show that the solution set of the programming problem with objective function* $\min_{(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)} \sum_j \sup_l |1 - r_{jl}|$ *is equivalent to that of the programming problem with objective function* $\min_{(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)} \sum_{j=1}^{p}\sum_{l=1}^{k_n}(1-r_{jl})$. *Let* $\bar{\mathbf{r}}^* = (\bar{\mathbf{r}}_1^*, \ldots, \bar{\mathbf{r}}_p^*)$ *be the optimal solution to* $\min_{(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)} \sum_j \sup_l (1 - r_{jl})$ *with the same constraints in the optimization problem (S.40), and* $\bar{\mathbf{r}}^\star = (\bar{\mathbf{r}}_1^\star, \ldots, \bar{\mathbf{r}}_p^\star)$ *be the optimal solution to (S.40). Define the set* $F = \{\bar{\mathbf{r}} = (\bar{\mathbf{r}}_1, \ldots, \bar{\mathbf{r}}_p) : r_{jl} \in [0,1]$ *and* $2\widehat{\mathbf{C}}_{XX} - \widehat{\mathbf{R}}_{XX} \succeq 0\}$ *as the feasible domain. On the one hand, when solving* $\min_{(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)} \sum_j \sup_l (1 - r_{jl})$, *each* r_{jl} *tends to a higher value within* F, *which implies that* $r_{jl}^* \geqslant r_{jl}^\star$ *for each* $j \in [p]$ *and* $l \in [k_n]$. *On the other hand, since* $\bar{\mathbf{r}}^\star = (\bar{\mathbf{r}}_1^\star, \ldots, \bar{\mathbf{r}}_p^\star)$ *is the optimal solution to* $\min_{(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)} \sum_j \sum_l (1 - r_{jl})$, *we have* $\sum_j \sum_l (1 - r_{jl}^\star) \leqslant \sum_j \sum_l (1 - r_{jl}^*)$. *This implies that* $\sum_j \sum_l r_{jl}^* \leqslant \sum_j \sum_l r_{jl}^\star$. *Both sides hold if and only if* $r_{jl}^* = r_{jl}^\star$ *for each* $j \in [p]$ *and* $l \in [k_n]$.

B.2 Coordinate mapping

As demonstrated in Section 4.2, $\mathcal{H}_j$ is spanned by a finite set of functions $\mathcal{B}_j = \{b_{j1}, \ldots, b_{jk_n}\}$. Each X_{ij} can be expressed as a linear combination: $X_{ij} = c_{ij1}b_{j1} + \cdots + c_{ijk_n}b_{jk_n}$, and its coordinate can be represented as $[X_{ij}]_{\mathcal{B}_j}$. Likewise, the coordinate of any operator $K : \mathcal{H}_j \to \mathcal{H}_k$ is denoted as ${}_{\mathcal{B}_k}[K]_{\mathcal{B}_j}$. This mapping, $K \to {}_{\mathcal{B}_k}[K]_{\mathcal{B}_j}$, is referred to as the coordinate mapping. There are five main properties of the coordinate mapping (Li and Solea, 2018) that are crucial for subsequent analysis. Here, K_1 and K_2 represent operators mapping from $\mathcal{H}_j$

to $\mathcal{H}_k$, a and b are real numbers, and $\boldsymbol{\mathcal{B}}$ denotes the Cartesian product of $\mathcal{B}_1, \ldots, \mathcal{B}_p$.

P1. linearity: ${}_{\mathcal{B}_k}[aK_1 + bK_2]_{\mathcal{B}_j} = a\big({}_{\mathcal{B}_k}[K_1]_{\mathcal{B}_j}\big) + b\big({}_{\mathcal{B}_k}[K_2]_{\mathcal{B}_j}\big)$;

P2. tensor product: ${}_{\mathcal{B}_k}[X_{ik} \otimes X_{ij}]_{\mathcal{B}_j} = [X_{ik}]_{\mathcal{B}_k}[X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}}\mathbf{G}_j$;

P3. operator calculation: $[K(X_{ij})]_{\mathcal{B}_k} = \big({}_{\mathcal{B}_k}[K]_{\mathcal{B}_j}\big)[X_{ij}]_{\mathcal{B}_j}$;

P4. inner product: $\langle X_{ij}, \widetilde{X}_{ij}\rangle = [X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}}\mathbf{G}_j[\widetilde{X}_{ij}]_{\mathcal{B}_j}$;

P5. operator matrix: ${}_{\boldsymbol{\mathcal{B}}}[\mathbf{C}_{XX}]_{\boldsymbol{\mathcal{B}}} = \big({}_{\mathcal{B}_j}[C_{X_jX_k}]_{\mathcal{B}_k}\big)_{j,k\in[p]}$.

First, we will present the Karhunen-Loève expansion by coordinate mapping. By P1 and P2, we can deduce that ${}_{\mathcal{B}_j}[\widehat{\Sigma}_{X_jX_j}]_{\mathcal{B}_j} = (n^{-1}\sum_{i=1}^n [X_{ij}]_{\mathcal{B}_j}[X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}})\mathbf{G}_j$. By P3, $(\hat{\omega}_{jl}, \hat{\phi}_{jl}(\cdot))_{l\geqslant 1}$ is the eigenvalue/eigenfunction pair of $\widehat{\Sigma}_{X_jX_j}$ if and only if ${}_{\mathcal{B}_j}([\widehat{\Sigma}_{X_jX_j}]_{\mathcal{B}_j})[\hat{\phi}_{jl}]_{\mathcal{B}_j} = \hat{\omega}_{jl}[\hat{\phi}_{jl}]_{\mathcal{B}_j}$, which indicates that

$$n^{-1}\mathbf{G}_j^{1/2}\sum_{i=1}^n \big([X_{ij}]_{\mathcal{B}_j}[X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}}\big)\mathbf{G}_j^{1/2}\big(\mathbf{G}_j^{1/2}[\hat{\phi}_{jl}]_{\mathcal{B}_j}\big) = \hat{\omega}_{jl}\big(\mathbf{G}_j^{1/2}[\hat{\phi}_{jl}]_{\mathcal{B}_j}\big),$$

i.e., $(\hat{\omega}_{jl}, \mathbf{G}_j^{1/2}[\hat{\phi}_{jl}]_{\mathcal{B}_j})_{l\geqslant 1}$ is the eigenvalue/eigenvector pair of the matrix

$$n^{-1}\mathbf{G}_j^{1/2}\sum_{i=1}^n \big([X_{ij}]_{\mathcal{B}_j}[X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}}\big)\mathbf{G}_j^{1/2}.$$

Hence, we can obtain the coordinate of $\hat{\phi}_{jl}$ as $[\hat{\phi}_{jl}]_{\mathcal{B}_j} = \mathbf{G}_j^{\dagger 1/2}\mathbf{v}_{jl}$, where $\mathbf{v}_{jl}$ is the eigenvector of $n^{-1}\mathbf{G}_j^{1/2}\sum_{i=1}^n \big([X_{ij}]_{\mathcal{B}_j}[X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}}\big)\mathbf{G}_j^{1/2}$. Finally, we can obtain the empirical Karhunen-Loève expansion of (3) as $X_{ij} - \hat{\mu}_j = \sum_{l=1}^{k_n}\hat{\xi}_{ijl}\hat{\phi}_{jl}$ by P4, where

$$\hat{\xi}_{ijl} = \langle X_{ij} - \hat{\mu}_j, \hat{\phi}_{jl}\rangle = [X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}}\mathbf{G}_j[\hat{\phi}_{jl}]_{\mathcal{B}_j} = [X_{ij}]_{\mathcal{B}_j}^{\mathrm{T}}\mathbf{G}_j^{1/2}\mathbf{v}_{jl}. \tag{S.41}$$

Second, we will derive that $2\widehat{\mathbf{C}}_{XX} - \widehat{\mathbf{R}}_{XX} \geq 0$ reduces to $2\widehat{\boldsymbol{\Omega}}_C - \widehat{\boldsymbol{\Omega}}_R \geq 0$. As shown in Section 4.2, $\widehat{C}_{X_jX_k} = (1-\gamma_n)\sum_{l=1}^{k_n}\sum_{m=1}^{k_n}\widehat{\Theta}_{jklm}^{\mathrm{S}}(\hat{\phi}_{jl}\otimes\hat{\phi}_{km}) + \gamma_n I(j=k)\widehat{I}_{X_jX_j}$, with $\widehat{\Theta}_{jklm}^{\mathrm{S}} = n^{-1}\sum_{i=1}^n(\widehat{\xi}_{ijl} - n^{-1}\sum_{i=1}\widehat{\xi}_{ijl})(\widehat{\xi}_{ikm} - n^{-1}\sum_{i=1}\widehat{\xi}_{ikm})/(\hat{\omega}_{jl}^{1/2}\hat{\omega}_{km}^{1/2})$ and $\hat{I}_{X_jX_j} = \sum_{l=1}^{k_n}\hat{\phi}_{jl}\otimes\hat{\phi}_{jl}$. By

P1 and P2, the coordinate of $\widehat{C}_{X_jX_k}$ can be represented as

$$\begin{aligned}
{}_{\mathcal{B}_j}[\widehat{C}_{X_jX_k}]_{\mathcal{B}_k} &= {}_{\mathcal{B}_j}\Big[\sum_{l=1}^{k_n}\sum_{m=1}^{k_n}\{(1-\gamma_n)\widehat{\Theta}^{\mathrm{S}}_{jklm}+\gamma_n I(l=m)I(j=k)\}(\hat{\phi}_{jl}\otimes\hat{\phi}_{km})\Big]_{\mathcal{B}_k} \\
&= \sum_{l=1}^{k_n}\sum_{m=1}^{k_n}\big\{(1-\gamma_n)\widehat{\Theta}^{\mathrm{S}}_{jklm}+\gamma_n I(l=m)I(j=k)\big\}\big([\hat{\phi}_{jl}]_{\mathcal{B}_j}[\hat{\phi}_{km}]^{\mathrm{T}}_{\mathcal{B}_k}\big)\mathbf{G}_k \\
&= \widehat{\mathbf{\Phi}}_j\widehat{\mathbf{\Theta}}_{C_{jk}}\widehat{\mathbf{\Phi}}_k^{\mathrm{T}}\mathbf{G}_k,
\end{aligned} \tag{S.42}$$

where $\widehat{\mathbf{\Theta}}_{C_{jk}} = (1-\gamma_n)(\widehat{\Theta}^{\mathrm{S}}_{jklm})_{k_n\times k_n}+\gamma_n I(j=k)\mathbf{I}_{k_n}\in\mathbb{R}^{k_n\times k_n}$ and $\widehat{\mathbf{\Phi}}_j\in\mathbb{R}^{k_n\times k_n}$ with its lth column $[\hat{\phi}_{jl}]_{\mathcal{B}_j}$ for $l\in[k_n]$. Similarly, for $j\in[p]$, we can get the the coordinate of $\widehat{R}_{X_jX_j}$ as

$$ {}_{\mathcal{B}_j}[\widehat{R}_{X_jX_j}]_{\mathcal{B}_j} = \widehat{\mathbf{\Phi}}_j\widehat{\mathbf{\Theta}}_{R_{jj}}\widehat{\mathbf{\Phi}}_j^{\mathrm{T}}\mathbf{G}_j, \tag{S.43}$$

where $\widehat{\mathbf{\Theta}}_{R_{jj}} = r\mathbf{I}_{k_n}\in\mathbb{R}^{k_n\times k_n}$ under E1, $\widehat{\mathbf{\Theta}}_{R_{jj}} = r_j\mathbf{I}_{k_n}\in\mathbb{R}^{k_n\times k_n}$ under E2, and $\widehat{\mathbf{\Theta}}_{R_{jj}} = \mathrm{diag}(r_{j1},\ldots,r_{jk_n})\in\mathbb{R}^{k_n\times k_n}$ under E3. By (S.42), (S.43), P3–P5, we have

$$\begin{aligned}
&\big\langle(2\widehat{\mathbf{C}}_{XX}-\widehat{\mathbf{R}}_{XX})(\mathbf{x}),\mathbf{x}\big\rangle \geqslant 0 \\
\Longleftrightarrow\;& \big[(2\widehat{\mathbf{C}}_{XX}-\widehat{\mathbf{R}}_{XX})(\mathbf{x})\big]^{\mathrm{T}}_{\mathcal{B}}(\oplus_{j\in[p]}\mathbf{G}_j)[\mathbf{x}]_{\mathcal{B}} \geqslant 0 \\
\Longleftrightarrow\;& [\mathbf{x}]^{\mathrm{T}}_{\mathcal{B}}\;{}_{\mathcal{B}}[2\widehat{\mathbf{C}}_{XX}-\widehat{\mathbf{R}}_{XX}]^{\mathrm{T}}_{\mathcal{B}}\;(\oplus_{j\in[p]}\mathbf{G}_j)\;[\mathbf{x}]_{\mathcal{B}} \geqslant 0 \\
\Longleftrightarrow\;& [\mathbf{x}]^{\mathrm{T}}_{\mathcal{B}}\;(\oplus_{j\in[p]}\mathbf{G}_j)^{\mathrm{T}}\;(\oplus_{j\in[p]}\widehat{\mathbf{\Phi}}_j)(2\widehat{\mathbf{\Theta}}_C-\widehat{\mathbf{\Theta}}_R)(\oplus_{j\in[p]}\widehat{\mathbf{\Phi}}_j)^{\mathrm{T}}\;(\oplus_{j\in[p]}\mathbf{G}_j)\;[\mathbf{x}]_{\mathcal{B}} \geqslant 0,
\end{aligned} \tag{S.44}$$

where $[\mathbf{x}]_{\mathcal{B}} = \big([x_1]^{\mathrm{T}}_{\mathcal{B}_1},\ldots,[x_p]^{\mathrm{T}}_{\mathcal{B}_p}\big)^{\mathrm{T}}$, $\widehat{\mathbf{\Theta}}_C = (\widehat{\mathbf{\Theta}}_{C_{jk}})_{j,k\in[p]}$, $\widehat{\mathbf{\Theta}}_R = \mathrm{diag}(r\mathbf{I}_{k_n},\ldots,r\mathbf{I}_{k_n})$ under E1, $\widehat{\mathbf{\Theta}}_R = \mathrm{diag}(r_1\mathbf{I}_{k_n},\ldots,r_p\mathbf{I}_{k_n})\in\mathbb{R}^{pk_n\times pk_n}$ under E2, $\widehat{\mathbf{\Theta}}_R = \mathrm{diag}(r_{11},\ldots,r_{1k_n},\ldots,r_{p1},\ldots,r_{pk_n})\in\mathbb{R}^{pk_n\times pk_n}$ under E3, $\oplus_{j\in[p]}\widehat{\mathbf{\Phi}}_j = \mathrm{diag}(\widehat{\mathbf{\Phi}}_1,\ldots,\widehat{\mathbf{\Phi}}_p)$, and $\oplus_{j\in[p]}\mathbf{G}_j = \mathrm{diag}(\mathbf{G}_1,\ldots,\mathbf{G}_p)$. (S.44) means that $2\widehat{\mathbf{C}}_{XX}-\widehat{\mathbf{R}}_{XX}\succeq 0$ if and only if

$$(\oplus_{j\in[p]}\mathbf{G}_j)^{\mathrm{T}}\;(\oplus_{j\in[p]}\widehat{\mathbf{\Phi}}_j)(2\widehat{\mathbf{\Theta}}_C-\widehat{\mathbf{\Theta}}_R)(\oplus_{j\in[p]}\widehat{\mathbf{\Phi}}_j)^{\mathrm{T}}\;(\oplus_{j\in[p]}\mathbf{G}_j)\succeq 0. \tag{S.45}$$

Note that $\widehat{\mathbf{\Omega}}_C = \widehat{\mathbf{A}}\widehat{\mathbf{\Theta}}_C\widehat{\mathbf{A}}^{\mathrm{T}}$ and $\widehat{\mathbf{\Omega}}_R = \widehat{\mathbf{A}}\widehat{\mathbf{\Theta}}_R\widehat{\mathbf{A}}^{\mathrm{T}}$, where $\widehat{\mathbf{A}} = \mathrm{diag}(\mathbf{G}_1\widehat{\mathbf{\Phi}}_1,\ldots,\mathbf{G}_p\widehat{\mathbf{\Phi}}_p)\in\mathbb{R}^{pk_n\times pk_n}$. (S.45) implies that $2\widehat{\mathbf{C}}_{XX}-\widehat{\mathbf{R}}_{XX}\succeq 0$ if and only if $2\widehat{\mathbf{\Omega}}_C-\widehat{\mathbf{\Omega}}_R\succeq 0$.

B.3 Algorithms

The construction of functional Model-X knockoffs utilizes the Karhunen-Loève expansion. As outlined in Section 4.3, Algorithm 2 comprises three main steps. In Step 1, three expressions of $\widehat{\mathbf{\Theta}}_R$ are obtained by determining the parameters of r, $(r_1,\ldots,r_p)$, and $(\bar{\mathbf{r}}_1,\ldots,\bar{\mathbf{r}}_p)$

which involve solving the optimization problems in (28), (29), and (30), respectively. The second step involves constructing the FPC scores of $\widetilde{\mathbf{X}}_i$ using the estimated FPC scores of $\mathbf{X}_i$ and the procedure outlined in Algorithm 1. The expression of $C_{X_jX_k}$ in Lemma 3 involves the correlations between the FPC scores, which are equivalent to the covariances between the normalized FPC scores. Therefore, our focus is on the distribution of the estimated normalized FPC scores. Considering $\big(\mathbf{X}_i^{\mathrm{T}}, \widetilde{\mathbf{X}}_i^{\mathrm{T}}\big)$ as a MGP for each $i \in [n]$ and using (S.41), (24), and (26), we derive that the estimated normalized FPC scores satisfy

$$
\begin{aligned}
&\operatorname{diag}\big(\widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{A}}^{\mathrm{T}}, \widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{A}}^{\mathrm{T}}\big)\big([X_{i1}]_{\mathcal{B}_1}^{\mathrm{T}}, \ldots, [X_{ip}]_{\mathcal{B}_p}^{\mathrm{T}}, [\widetilde{X}_{i1}]_{\mathcal{B}_1}^{\mathrm{T}}, \ldots, [\widetilde{X}_{ip}]_{\mathcal{B}_p}^{\mathrm{T}}\big)^{\mathrm{T}} \\
=&\operatorname{diag}\big(\widehat{\mathbf{W}}^{-1/2}, \widehat{\mathbf{W}}^{-1/2}\big)\big(\hat{\xi}_{i1l}, \ldots, \hat{\xi}_{i1k_n}, \ldots, \hat{\xi}_{ipl}, \ldots, \hat{\xi}_{ipk_n}, \check{\xi}_{i1l}, \ldots, \check{\xi}_{i1k_n}, \ldots, \check{\xi}_{ipl}, \ldots, \check{\xi}_{ipk_n}\big) \\
\sim&\mathcal{N}(\mathbf{0}_{2pk_n}, \widehat{\boldsymbol{\Theta}}),
\end{aligned}
$$

where $\widehat{\mathbf{A}} = \operatorname{diag}(\mathbf{G}_1\widehat{\boldsymbol{\Phi}}_1, \ldots, \mathbf{G}_p\widehat{\boldsymbol{\Phi}}_p) \in \mathbb{R}^{pk_n \times pk_n}$, $\widehat{\mathbf{W}} = \operatorname{diag}(\hat{\omega}_{11}, \ldots, \hat{\omega}_{1k_n}, \ldots, \hat{\omega}_{p1}, \ldots, \hat{\omega}_{pk_n})$ is a normalization matrix, and

$$
\widehat{\boldsymbol{\Theta}} = \begin{pmatrix} \widehat{\boldsymbol{\Theta}}_C & \widehat{\boldsymbol{\Theta}}_C - \widehat{\boldsymbol{\Theta}}_R \\ \widehat{\boldsymbol{\Theta}}_C - \widehat{\boldsymbol{\Theta}}_R & \widehat{\boldsymbol{\Theta}}_C \end{pmatrix}. \tag{S.46}
$$

Note $\widehat{\boldsymbol{\Theta}}_C \in \mathbb{R}^{pk_n \times pk_n}$ and $\widehat{\boldsymbol{\Theta}}_R \in \mathbb{R}^{pk_n \times pk_n}$ in (S.46) are obtained from (S.44) and (S.45). Then for $i \in [n]$, by conditional distribution under multivariate Gaussianity, we can obtain that

$$
\widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{A}}^{\mathrm{T}}\big([\widetilde{X}_{i1}]_{\mathcal{B}_1}^{\mathrm{T}}, \ldots, [\widetilde{X}_{ip}]_{\mathcal{B}_p}^{\mathrm{T}}\big)^{\mathrm{T}}\Big|\widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{A}}^{\mathrm{T}}([X_{i1}]_{\mathcal{B}_1}^{\mathrm{T}}, \ldots, [X_{ip}]_{\mathcal{B}_p}^{\mathrm{T}})^{\mathrm{T}} \sim \mathcal{N}(\widehat{\boldsymbol{\mu}}_{\tilde{X}|X}, \widehat{\boldsymbol{\Theta}}_{\tilde{X}|X}),
$$

where $\widehat{\boldsymbol{\mu}}_{\tilde{X}|X} = (\mathbf{I}_{pk_n} - \widehat{\boldsymbol{\Theta}}_R\widehat{\boldsymbol{\Theta}}_C^{-1})\widehat{\mathbf{W}}^{-1/2}\widehat{\mathbf{A}}^{\mathrm{T}}([X_{i1}]_{\mathcal{B}_1}^{\mathrm{T}}, \ldots, [X_{ip}]_{\mathcal{B}_p}^{\mathrm{T}})^{\mathrm{T}}$ and $\widehat{\boldsymbol{\Theta}}_{\tilde{X}|X} = 2\widehat{\boldsymbol{\Theta}}_R - \widehat{\boldsymbol{\Theta}}_R\widehat{\boldsymbol{\Theta}}_C^{-1}\widehat{\boldsymbol{\Theta}}_R$. This conditional distribution forms the foundation for sampling the estimated FPC scores of $\widetilde{\mathbf{X}}_i$. Finally, we construct the functional knockoffs $\check{\mathbf{X}}_i = (\check{X}_{i1}, \ldots, \check{X}_{ip})^{\mathrm{T}}$ from these estimated FPC scores using the Karhunen-Loève expansion.

C Additional empirical results

C.1 Additional simulation results

Table 4 and 5 present the empirical power and FDR for partially observed functional data under the model settings of SFLR and FGGM in Section 5, respectively. Considering the similar conclusions drawn from SFLR and FFLR for fully observed functional data, we only apply comparison methods to SFLR for partially observed functional data due to computational efficiency.

p	n	KF1		KF2		KF3		TF		GL	
		FDR	Power	FDR	Power	FDR	Power	FDR	Power	FDR	Power
50	100	0.18	0.94	0.18	0.96	0.19	0.96	0.21	0.79	0.28	1.00
	200	0.16	0.99	0.15	0.98	0.16	0.98	0.18	0.85	0.26	1.00
100	100	0.17	0.92	0.15	0.92	0.15	0.92	0.16	0.62	0.47	1.00
	200	0.20	1.00	0.20	1.00	0.18	1.00	0.12	0.93	0.31	1.00
150	100	0.09	0.99	0.09	0.99	0. 10	0.99	0.14	0.70	0.68	1.00
	200	0.13	1.00	0.12	1.00	0.12	1.00	0.08	0.96	0.37	1.00

Table 4: The empirical power and FDR in SFLR for partially observed functional data.

C.2 Specific ROIs

Table 6 presents 34 regions of interest (ROIs) and the associated labelling index for the emotion related fMRI dataset in Section 6.1.

p	n	KF1		TF		GL	
		FDR	Power	FDR	Power	FDR	Power
50	100	0.18	0.74	0.15	0.50	0.23	0.75
	200	0.20	0.94	0.20	0.79	0.22	0.94
100	100	0.17	0.63	0.12	0.48	0.20	0.68
	200	0.19	0.84	0.18	0.73	0.22	0.86

Table 5: The empirical power and FDR in FGGM for partially observed functional data.

Index	Region	Index	Region
1	bankssts	2	caudal anterior cingulate
3	caudal middle frontal	4	cuneus
5	entorhinal	6	fusiform
7	inferior parietal	8	inferior temporal
9	isthmus cingulate	10	lateral occipital
11	lateral orbitofrontal	12	lingual
13	medial orbito frontal	14	middle temporal
15	parahippocampal	16	paracentral
17	parsopercularis	18	parsorbitalis
19	parstriangularis	20	pericalcarine
21	postcentral	22	posterior cingulate
23	precentral	24	precuneus
25	rostral anterior cingulate	26	rostral middle frontal
27	superior frontal	28	superior parietal
29	superior temporal	30	supramarginal
31	frontal pole	32	temporal pole
33	transverse temporal	34	insula

Table 6: The labeling index of ROIs in Section 6.1.

References

Candès, E., Fan, Y., Janson, L. and Lv, J. (2018). Panning for gold: 'model-X' knockoffs forhigh dimensional controlled variable selection, *Journal of the Royal Statistical Society: Series B* **80**: 551–577.

Fan, Y., Demirkaya, E., Li, G. and Lv, J. (2020a). Rank: Large-scale inference with graphical nonlinear knockoffs, *Journal of the American Statistical Association* **115**: 362–379.

Fang, Q., Guo, S. and Qiao, X. (2022). Finite sample theory for high-dimensional functional/scalar time series with applications, *Electronic Journal of Statistics* **16**: 527–591.

Guo, S. and Qiao, X. (2023). On consistency and sparsity for high-dimensional functional time series with application to autoregressions, *Bernoulli* **29**: 451–472.

Li, B. and Solea, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI, *Journal of the American Statistical Association* **113**: 1637–1655.

Li, J. and Maathuis, M. H. (2021). GGM knockoff filter: False discovery rate control for Gaussian graphical models, *Journal of the Royal Statistical Society: Series B* **83**: 534–558.

Solea, E. and Li, B. (2022). Copula Gaussian graphical models for functional data, *Journal of the American Statistical Association* **117**: 781–793.