Supplement for "Conditional Quantile-based Variable Screening with FDR Control in Joint Factor Models"

Han Pan, Wei Xiong and Mingyao Ai

Shandong University of Finance and Economics, University of International Business and Economics, and Peking University

Supplementary Materials

In the Supplementary Materials, we first provide some properties of MSD index under Gaussian distribution, the MSD Knockoffs procedure, more details on v-quantile knockoffs and the estimation of MSD index. Then we provide the visual representation of the convolution-type smoothed quantile loss. Next, the detailed proofs of all theoretical results are presented. We also provide additional results for numerical studies and real data analysis.

$1 Properties of MSD Index under Gaussian Distribution

The following proposition illustrates certain properties of the MSD index under the bivariate Gaussian copula distribution.

Proposition S1. *If (X, Y) follows a bivariate Gaussian copula distribution such that, after transformation via a monotone function g_1 and a linear transformation g_2, $(g_1(Y), g_2(X))$ is jointly normal with correlations $\rho = Cor(g_1(Y), g_2(X))$ and $g_1(Y), g_2(X)$ are marginally standard normal. Then for any given quantile $0 < \tau < 1$, we have the following three* *conclusions*

(i) $MSD_\tau(Y|X) = 0$ *if* $\rho = 0$ *and* $MSD_\tau(Y|X) > 0$ *otherwise.*

(ii) $MSD_\tau(Y|X)$ *can be expressed as*

$$MSD_\tau(Y|X) = \int \Big(\Phi\big(\rho x + \sqrt{1-\rho^2}\Phi^{-1}(\tau)\big) - \tau\Big)^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx,$$

where Φ *is the cumulative distribution function for the standard normal distribution. Consequently,* $MSD_\tau(Y|X)$ *is a strictly increasing function in* $|\rho|$*.*

(iii) $MSD_\tau(Y|X) = MSD_{1-\tau}(Y|X)$*.*

For a visual representation of how MSD index varies with the correlation level, Figure S1 provides an instance of (X, Y) following a bivariate normal distribution with mean zero and correlation coefficient ρ. As expected, $\mathrm{MSD}_\tau(Y|X)$ takes value zero at $\rho = 0$ and shows larger values with higher correlations. Moreover, it is noteworthy from the left panel in Figure S1 that, for a given ρ, $\mathrm{MSD}_\tau(Y|X)$ varies across the quantile range, with the difference becoming more pronounced as $|\rho|$ increases. In contrast, the graphs exhibit a flattening near 0, indicating that $\mathrm{MSD}_\tau(Y|X)$ remains considerably small when $|\rho|$ takes small values. To enhance clarity, two zoomed-in versions of the left panel are presented, focusing on regions where $|\rho| \leq 0.7$ and $|\rho| \leq 0.1$. The zoomed-in views reveal that the graphs flatten near zero, indicating that $\mathrm{MSD}\tau(Y|X)$ remains small for small values of $|\rho|$. These visualizations underscore that when τ is close to zero, MSD_τ approaches zero as $|\rho|$ becomes particularly small.

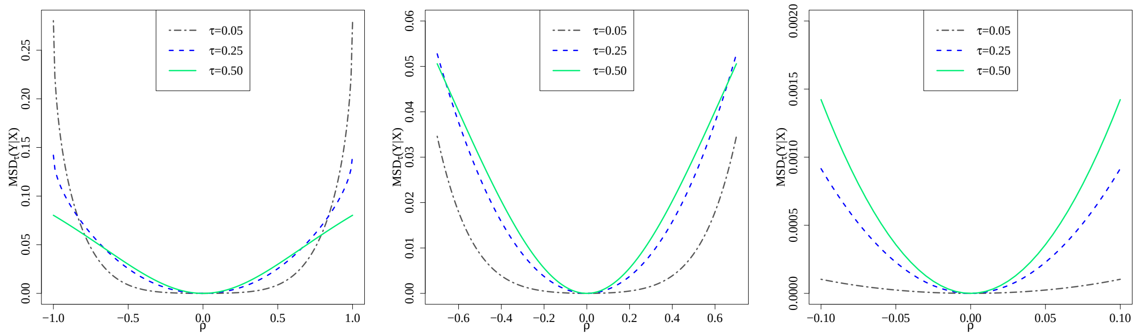

Figure S1: The MSD index as a function of the correlation coefficient ρ for a bivariate normal distribution, with $\rho \in [-1, 1]$ (left), $\rho \in [-0.7, 0.7]$ (middle) and $\rho \in [-0.1, 0.1]$ (right).

$2 Knockoffs Framework

The key idea underlying knockoffs is to generate a knockoff copy $\tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_p)$ that satisfies two conditions: (i) Swap exchangeable: the joint distribution of $(\mathbf{X}, \tilde{\mathbf{X}})$ is unchanging by swapping X_j with its knockoff counterpart $\tilde{X}_j$, for $j = 1, \ldots, p$. (ii) Nullity: $\tilde{\mathbf{X}} \perp\!\!\!\perp \boldsymbol{f} | \mathbf{X}$. Condition (i) indicates that one can not tell whether the jth column is true or a knockoff; Condition (ii) implies the knockoffs provide no further information about latent factors, this is always satisfied if $\tilde{\mathbf{X}}$ is constructed without borrowing information from $\boldsymbol{f}$.

Remark 1. Constructing knockoff features that exactly satisfy Conditions (i) and (ii) is challenging. Condition (i) also suggests that generating knockoff features requires the joint distribution of $\mathbf{X}$. If such distribution is available, Candès et al. (2018) proposed a generic algorithm sequential conditional independent pairs to construct exact knockoff features. If the distribution of $\mathbf{X}$ is unknown, Candès et al. (2018) also provided two approaches to construct the second-order knockoffs, one is equicorrelated construction and the other is semidefinite programme. Other practical construction methods can be found in Romano et al. (2020), Fan et al. (2020) and Huang and Janson (2020).

\$3 MSD Knockoffs (MSDK) Procedure

Intuitively, the MSD knockoff statistic $W_{j,\tau}$ defined in (2.9) of the main text evaluates the relative importance of the jth original variable by comparing the MSD statistic with that of its knockoff copy. A large and positive $W_{j,\tau}$ would suggest strong evidence that the original feature is relevant to the conditional quantile of $\boldsymbol{f}$ at the interested quantile. Theoretically, the MSD knockoff is verified to be a valid construction of knockoff statistic and satisfy the sign flip property in the Supplementary Materials. Thereby, $W_{j,\tau}$ can be utilized to calculate the data dependent knockoff thresholds that ensure finite sample FDR control.

The empirical counterpart of the knockoff statistic is also defined in equation (2.10) of the main text. Then we consider the relatively conservative knockoffs threshold

$$\eta_\alpha = \min\left\{T > 0 : \frac{1 + \#\{j : \widehat{W}_{j,\tau} \leq -T\}}{\#\{j : \widehat{W}_{j,\tau} \geq T\}} \leq \alpha\right\}. \tag{S3.1}$$

Once this threshold is calculated, the set of selected active features is given by

$$\hat{\mathcal{A}}_\tau(\eta_\alpha) = \{1 \leq j \leq p : \widehat{W}_{j,\tau} \geq \eta_\alpha\}. \tag{S3.2}$$

Theorem S1. *(FDR control) For any $\alpha \in [0, 1]$, the set of selected covariates $\hat{\mathcal{A}}(\eta_\alpha)$ via*

MSDK procedure satisfies

$$FDR_\tau = E\left[\frac{|\hat{\mathcal{A}}_\tau(\eta_\alpha) \cap \mathcal{A}_\tau^c|}{|\hat{\mathcal{A}}_\tau(\eta_\alpha)| \vee 1}\right] \leq \alpha.$$

$4 More Details on v-quantile Knockoffs

We introduce the v-quantile knockoffs procedure as follows. Suppose at the t-th round, given the quantile level $\tau > 0$, the feature importance statistic $\mathbf{W}_\tau^t$ has been computed. The v-quantile knockoffs algorithm starts by ordering the features according to the magnitude of $W_{j,\tau}^t$, that is,

$$|W_{r_1,\tau}^t| \geq \ldots \geq |W_{r_j,\tau}^t| \geq \ldots \geq |W_{r_p,\tau}^t| \quad \text{for some permutation } r_1, \ldots, r_p,$$

where $r_1, \ldots, r_p$ is a permutation of $1, \ldots, p$. Given a prespecified integer v, v-quantile knockoffs looks each variable one by one from r_1 to r_p according to the magnitude of each feature importance statistic until the first time there are v negative $W_{j,\tau}^t$'s. More specifically, the stopping criterion is

$$\mathcal{T}_{v,\tau}^t := \min\left\{k \in \{1, \ldots, p\} : \sum_{j=1}^{k} I\{W_{r_j,\tau}^t < 0\} \geq v\right\}.$$

The selected variable set is then defined to be

$$\hat{\mathcal{A}}_\tau^t := \left\{r_j : j < \mathcal{T}_{v,\tau}^t, W_{r_j,\tau}^t > 0\right\}.$$

Below we provide some properties of v-quantile knockoffs.

Lemma S1. *For any integer $v \geq 1$ and $\tau \in (0,1)$, the false discovery number $V(\tau) := \#\{j : j \in \hat{\mathcal{A}}_\tau \cap \mathcal{A}_\tau^c\}$ by the v-quantile knockoffs procedure is stochastically dominated by a negative binomial variable $NB(v, 1/2)$.*

Lemma S1 can be proved by using Lemma S21 and Lemma (3.1) in Janson and Su (2016). Next we present the main result, which is immediate from Lemma S1 and the negative binomial cumulative distribution function.

Lemma S2. *For any integer $k \geq 1$ and $0 < \alpha < 1$, let v be the largest integer satisfying*

$$\sum_{i=k}^{\infty} 2^{-i-v} \binom{i+v-1}{i} \leq \alpha. \tag{S4.3}$$

Then the v-quantile knockoffs procedure with parameter v controls the k-FWER at level α, that is, $P(V(\tau) \geq k) \leq \alpha$.

S5 Details for MSD Estimator

Let $h_{kj}(X_j) = \arg\min_h E[\rho_\tau(f_k - h(X_j))]$, where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the quantile loss function. It is known that $h_{kj}(X_j) = Q_\tau(f_k|X_j)$. For technical simplicity, we assume that each X_j takes values on the interval $[0,1]$. Let $\mathcal{H}_r$ be the class of functions defined in condition (C1) of Section 3. Let $\boldsymbol{\pi}(t) = (b_1(t), \ldots, b_{s_n+l+1}(t))^T$ denote a vector of normalized B-spline basis functions of order $l+1$ with s_n quasi-uniform internal knots on $[0,1]$. Then $h_{kj}(t)$ can be approximated using a linear combination of B-spline basis functions $\boldsymbol{\pi}(t)^T\boldsymbol{\beta}$, for some $\boldsymbol{\beta} \in \mathbb{R}^{s_n+l+1}$. The standard quantile regression estimator is defined as a solution to the convex optimization problem

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{s_n+l+1}} n^{-1}\sum_{i=1}^{n}\rho_\tau\big(\hat{f}_{ik}-\boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}\big).$$

However, the empirical quantile loss is not smooth and differentiable. A natural way of resolving this issue is to smooth the piecewise linear quantile loss using a kernel. Denote by $\mathcal{K}(\cdot)$ a kernel function that integrates to 1, and $h > 0$ a bandwidth, the convolution type smoothed quantile loss function is defined as

$$\mathcal{L}_{\tau,h}(u) = (\rho_\tau * \mathcal{K}_h)(u) = \int_{-\infty}^{\infty}\rho_\tau(v)\mathcal{K}_h(v-u)dv, \tag{S5.4}$$

where $*$ is a convolution operator and $\mathcal{K}_h(u) = h^{-1}K(u/h)$. Let $\hat{Q}_{\tau,h}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\mathcal{L}_{\tau,h}(\hat{f}_{ik}-\boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta})$ be the empirical convolution type smoothed quantile loss. Figure S1 indicates that $\hat{Q}_{\tau,h}(\boldsymbol{\beta})$ is twice continuously differentiable and globally convex for any $h > 0$. Let $\boldsymbol{\pi}_{ij} = \boldsymbol{\pi}(X_{ij})$, $\tilde{K}_h(u) := \int_{-\infty}^{-u/h}\mathcal{K}(t)dt$, and $\epsilon_i(\boldsymbol{\beta}) = \hat{f}_{ik}-\boldsymbol{\pi}_{ij}^T\boldsymbol{\beta}$, the gradient and hessian matrix of $\hat{Q}_{\tau,h}(\boldsymbol{\beta})$ are

$$\nabla\hat{Q}_{\tau,h}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\{\tilde{\mathcal{K}}_h(\epsilon_i(\boldsymbol{\beta}))-\tau\}\boldsymbol{\pi}_{ij}, \quad \nabla^2\hat{Q}_{\tau,h}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_h(-\epsilon_i(\boldsymbol{\beta}))\boldsymbol{\pi}_{ij}\boldsymbol{\pi}_{ij}^T. \tag{S5.5}$$

The smoothness and convexity of $\hat{Q}_{\tau,h}$ ensure the computation efficiency of gradient based algorithms for solving quantile regressions.

Let $\hat{\boldsymbol{\beta}}_{kj,h} = \arg\min_{\boldsymbol{\beta}}\sum_{i=1}^{n}\hat{Q}_{\tau,h}(\boldsymbol{\beta})$, and define $\hat{h}_{kj}(X_{ij}) = \boldsymbol{\pi}_{ij}^T\hat{\boldsymbol{\beta}}_{kj,h}$. Further define $\hat{F}_{f_k}(x) = n^{-1}\sum_{i=1}^{n}I\{\hat{f}_{ik}\le x\}$ with $I\{\cdot\}$ being indicator function, thus $n^{-1}\sum_{i=1}^{n}I\{\hat{f}_{ik}\le \hat{h}_{kj}(X_{lj})\}$ is a nonparametric estimator of $F_{\hat{f}_k}(\hat{Q}_\tau(\hat{f}_k|X_j))$. Therefore, we obtain

$$\hat{\delta}_{k,j}^\tau = \frac{1}{n}\sum_{l=1}^{n}\left\{n^{-1}\sum_{i=1}^{n}I\big(\hat{f}_{ik}\le\hat{h}_{kj}(X_{lj})\big)-\tau\right\}^2 \quad\text{and}\quad \hat{\delta}_j^\tau = \max_{1\le k\le K}\hat{\delta}_{k,j}^\tau. \tag{S5.6}$$

We expect $\hat{\delta}_j^\tau$ to be close to zero if X_j is independent of $\boldsymbol{f}$.

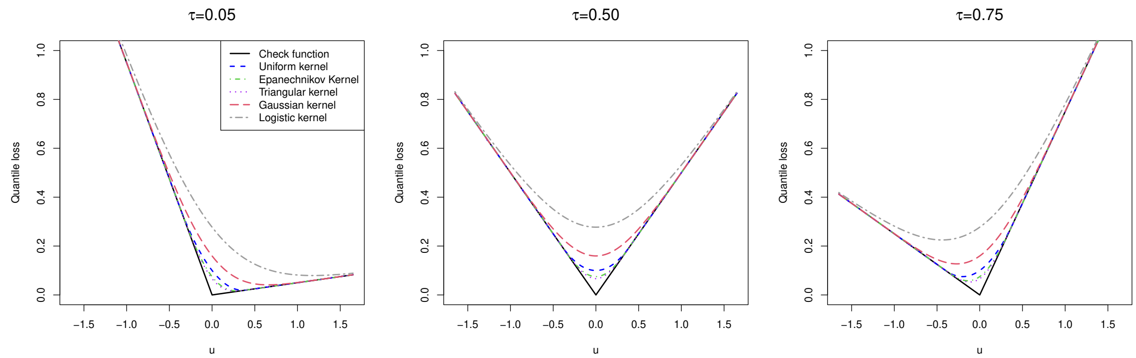

Figure S2: The convolution type quantile loss with five commonly used kernel functions under $\tau = 0.05, 0.50$ and 0.75, respectively. $h = 0.4$.

S6 Figures for Convolution-type Smoothed Quantile Loss

The key difference between the convolution type smoothed quantile loss and classical quantile loss is that the former is globally convex. Commonly used kernel functions include: (a) uniform kernel $\mathcal{K}(u) = (1/2)I(|u| \leq 1)$, (b) Gaussian kernel $\mathcal{K}(u) = \varphi(u) := (2\pi)^{-1/2}\exp(-u^2/2)$; (c) logistic kernel $\mathcal{K}(u) = e^{-u}/(1+e^{-u})^2$, (d) Epanechnikov kernel $K(u) = (3/4)(1-u^2)I(u \leq 1)$, and (e) triangular kernel $K(u) = (1-|u|)I(|u| \leq 1)$. Figure S2 provide the visual representation of the convolution-type smoothed quantile loss with bandwidth $h = 0.4$ at $\tau \in \{0.05, 0.5, 0.75\}$. It is observed that the smoothed quantile loss with the triangular kernel is more close to the quantile loss. Therefore, we use triangular kernel in our simulation studies and real data analysis.

$7 Proofs of Lemmas, Propositions and Theorems

$7.1 Proofs of Propositions 1 and S1

Proof of Proposition 1. Results (i)-(iii) are obvious. To prove result (iv), define $\tilde{X} = a + bX$, then it follows that $E_{\tilde{X}}\{F_Y(Q_\tau(Y|\tilde{X})) - \tau\}^2 = E_X\{F_Y(Q_\tau(Y|X)) - \tau\}^2$. If $g(\cdot)$ is nondecreasing, then $Q_\tau(g(Y)|\tilde{X}) = g(Q_\tau(Y|\tilde{X}))$, and $F_{g(Y)}[g(Q_\tau(Y|\tilde{X}))] = F_Y[Q_\tau(Y|\tilde{X})]$, since $P(Y \leq y|\tilde{X}) = P(g(Y) \leq g(y)|\tilde{X})$. Therefore, $E_{\tilde{X}}\{F_{g(Y)}(Q_\tau(g(Y)|\tilde{X})) - \tau\}^2 = E_X\{F_Y(Q_\tau(Y|X)) - \tau\}^2$, that is $\mathrm{MSD}_\tau(g(Y)|aX + b) = \mathrm{MSD}_\tau(Y|X)$. If $g(\cdot)$ is nonincreasing, then $Q_\tau(g(Y)|\tilde{X}) = g(Q_{1-\tau}(Y|\tilde{X}))$, and $F_{g(Y)}[g(Q_\tau(Y|\tilde{X}))] = 1 - F_Y[Q_\tau(Y|\tilde{X})]$, since $P(Y \leq y|\tilde{X}) = P(g(Y) \geq g(y)|\tilde{X})$. Therefore, $E_{\tilde{X}}\{F_{g(Y)}(Q_\tau(g(Y)|\tilde{X})) - \tau\}^2 = E_{\tilde{X}}\{1 - F_Y(Q_{1-\tau}(Y|\tilde{X})) - \tau\}^2 = E_X\{F_Y(Q_{1-\tau}(Y|X)) - (1-\tau)\}^2$, that is, $\mathrm{MSD}_\tau(g(Y)|aX + b) = \mathrm{MSD}_{1-\tau}(Y|X)$. This completes the proof. $\square$

Proof of Proposition S1. Because MSD is invariant under monotone transformations, it suffices to consider the case $g_1(t) = t$, $g_2(t) = t$, and hence X and Y are joint normal. Let $\varphi(y)$ be the probability density function of Y, which is standard normal. For the first conclusion, note that if $\rho = 0$, then X is independent of Y and $\mathrm{MSD}_\tau(Y|X) = 0$, for any τ. On the other hand, if $\rho \neq 0$, $Y|X = x \sim N(\rho x, 1 - \rho^2)$. Therefore, $F(y|x) = \Phi((y - \rho x)/\sqrt{1-\rho^2})$, $Q_\tau(Y|x) = \rho x + \Phi^{-1}(\tau)\sqrt{1-\rho^2}$, $F_Y(Q_\tau(Y|x)) = \Phi(\rho x + \Phi^{-1}(\tau)\sqrt{1-\rho^2})$. It follows that $\mathrm{MSD}_\tau(Y|X) > 0$ for any τ.

For the second conclusion, again by $Y|X \sim N(\rho X, 1 - \rho^2)$ and $F(y|x) = \Phi((y -$

$\rho x)/\sqrt{1-\rho^2})$, we have

$$\mathrm{MSD}_\tau(Y|X) = \int \Big(\Phi\big(\rho x + \sqrt{1-\rho^2}\Phi^{-1}(\tau)\big) - \tau\Big)^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx.$$

For the third conclusion,

$$\begin{aligned}
\mathrm{MSD}_{1-\tau}(Y|X) &= \int \Big(\Phi\big(\rho x + \sqrt{1-\rho^2}\Phi^{-1}(1-\tau)\big) - (1-\tau)\Big)^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx \\
&= \int \Big(\Phi\big(\rho x - \sqrt{1-\rho^2}\Phi^{-1}(\tau)\big) - 1 + \tau\Big)^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx \\
&= \int \Big(\Phi\big(-\rho x + \sqrt{1-\rho^2}\Phi^{-1}(\tau)\big) - \tau\Big)^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx \\
&= \int \Big(\Phi\big(\rho x + \sqrt{1-\rho^2}\Phi^{-1}(\tau)\big) - \tau\Big)^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx \\
&= \mathrm{MSD}_\tau(Y|X).
\end{aligned}$$

□

$7.2 Proof of Theorem 1

To derive Theorem 1, we introduce several additional conditions, which are outlined below.

(i). Letting $\mathbf{U}_i = (u_{i1}, \ldots, u_{i,d+K})^T = (f_{i1}, \ldots, f_{iK}, e_{i1}, \ldots, e_{id})^T$ and $(e_{i1}, \ldots, e_{id}) = (\varepsilon_{i1}, \ldots, \varepsilon_{id})\,\mathbf{\Psi}^{-1/2}$, $\{u_{ij}, i \in [n], j \in [d+K]\}$ are independent random variables satisfying

$$\frac{1}{n(d+K)\iota_n^4}\sum_{i=1}^{n}\sum_{j=1}^{d+K} E\left|u_{ij}^4\right| 1\left(|u_{ij}| > \iota_n\sqrt{n}\right) \to 0,$$

where $\{\iota_n\}$ is a given sequence of positive numbers satisfying $\iota_n \to 0$ and $\iota_n \log n \to +\infty$.

(ii). $\sup E\left(|u_{1j}|^{6+\gamma_0}\right)$ is bounded for all d for some $\gamma_0 > 0$.

(iii). $d/n \to \omega \in (0, \infty)$ as $n \to \infty$.

(iv). $\|[\mathrm{diag}(\mathbf{\Sigma})]^{-1}\mathbf{\Psi}\| \leq 1$ and the limiting spectral distribution (LSD) $H(t)$ of the empirical spectral distribution (ESD) $H_{d-K}(t)$ from the eigenvalues $\lambda_{K+1}(\mathbf{R}), \ldots, \lambda_d(\mathbf{R})$ of $\mathbf{R}$ exists, where

$$H_{d-K}(t) = \frac{1}{d-K} \sum_{j=K+1}^{d} 1\left(\lambda_j(\mathbf{R}) \leq t\right).$$

(v). The number K of common factors is fixed.

Condition (i) is the Lindeberg condition. The truncation parameter ι_n controls the tail behavior, preventing heavy-tailed distributions from destabilizing the sample correlation matrix. Such conditions are standard in random matrix theory to derive LSD. Condition (ii) strengthens the control over tail probabilities, which is particular important in high-dimensional settings where the number of variables d grows with n. Such conditions are common in studies involving spiked covariance models and factor analysis. Condition (iii) aligns with modern high-dimensional statistics frameworks and allows the application of Marchenko-Pastur law for sample covariance/correlation matrices. The restriction $\omega \in (0, \infty)$ avoids degenerate cases (e.g., $\omega = 0$ or $\omega = \infty$) and ensures a non-trivial LSD. Condition (iv) ensures that the scaled idiosyncratic covariance dose not dominate the factor-driven signal, preserving the identifiable of the factor structure. The existence of $H(t)$ guarantees that the ESD of the noise components converges to a deterministic limit, a cornerstone result in random matrix theory. Fixing K in Condition (v) simplifies the theoretical analysis by separating the factor structure from the high-dimensional noise. Extensions to growing K would require additional technical machinery.

Lemma S3. *For the factor model in (2.1) satisfying Conditions (i)-(v), when* $\lambda_K(\mathbf{R}) >$

$1 + \sqrt{\omega}$ *with* $\omega \in (0, \infty)$, *we have* $P\left(\hat{K}^g = K\right) \to 1$, *as* $n, d \to \infty$.

Proof of Lemma S3. **Step 1**. According to the results of Bai and Silverstein (2010), when the weights v_i are independent of $\mathbf{Z}_i$, and satisfy the moment condition, that is, $E(v_i) = 1$, $var(v_i) = \sigma_v^2 < \infty$ and $E(v_i^4) < \infty$ and when high-dimensional asymptotics are satisfied — that is, as $n, d \to \infty$, with the dimension ratio $d/n \to \omega \in (0, \infty)$— the ESD of the weighted covariance matrix

$$\hat{\mathbf{\Sigma}}_b^g = n^{-1}\mathbf{Z}^{*T}\mathbf{V}\mathbf{Z}^*,$$

converges almost surely to the same LSD as that of the unweighted covariance matrix

$$\hat{\mathbf{\Sigma}} = n^{-1}\mathbf{Z}^{*T}\mathbf{Z}^*.$$

i.e., the Marchenko-Pastur distribution or its generalized form.

In our bootstrap setting, the weights $v_i \sim Exp(1)$ fully meet the moment conditions for the weights in Theorem 3.10 of Bai and Silverstein (2010). Additionally, the conditions of independence and high-dimensional asymptotics are also satisfied in the same manner.

For the unweighted covariance matrix $\hat{\mathbf{\Sigma}}$, if the columns of $\mathbf{Z}^*$ satisfy $E(Z_{ij}^*) = 0$, $E(Z_{ij}^{*2}) = \sigma_j^2$ and $E(Z_{ij}^{*2}) < \infty$, its ESD converges to the Marchenko-Pastur law. If $\mathbf{\Sigma} = \mathbf{I}_d$, the limiting spectral density is given by

$$f_\omega(x) = \frac{\sqrt{(b-x)(x-a)}}{2\pi\omega x}\mathbf{1}_{[a,b]}(x), \quad a = (1-\sqrt{\omega})^2,\ b = (1+\sqrt{\omega})^2.$$

For the weighted covariance matrix $\hat{\mathbf{\Sigma}}_b^g$, define the Stieltjes transform

$$m(z) = \frac{1}{d}\mathrm{tr}\left(\hat{\mathbf{\Sigma}}_b^g - z\mathbf{I}_d\right)^{-1}, \quad z \in \mathbb{C}^+,$$

Using random matrix theory, the Stieltjes transform of the weighted covariance matrix satisfies the equation

$$z = \frac{1}{m(z)} + \omega \int \frac{t}{1 + tm(z)} dH(t),$$

where $H(t)$ is the limiting spectral distribution of the weights. Since the weights v_i are independent of the data and satisfy $E(v_i) = 1$, their effect is only a linear scaling of the sample covariance. Asymptotically, this scaling is absorbed into the dimension ratio ω, so the LSD remains the same as in the unweighted case. According to Theorem 3.10 in Bai and Silverstein (2010), the ESD of the weighted covariance matrix converges to the same LSD as the unweighted case. Specifically, for any continuous bounded function f,

$$\frac{1}{d} \sum_{j=1}^{d} f(\lambda_j(\hat{\mathbf{\Sigma}}_b^g)) \xrightarrow{a.s.} \int f(x) dF_\omega(x),$$

where $F_\omega(x)$ is the Marchenko-Pastur distribution. Through the above derivation, the LSD of the bootstrap covariance matrix is shown to be identical to that of the original covariance matrix. This equivalence forms the foundation for the subsequent analysis of the correlation matrix $\hat{\mathbf{R}}_b^g$.

Step 2. According to Theorem 1 in El Karoui (2008), assume that the entries of $\hat{\mathbf{\Sigma}}$ satisfy moment conditions (e.g., finite fourth moments) and the diagonal elements of $\mathbf{D} = \text{diag}(\hat{\mathbf{\Sigma}})$ converge to 1 almost surely as $d, n \to \infty$. Then, the ESD of $\hat{\mathbf{R}}$ converges almost surely to a deterministic LSD $H(t)$, which depends only on ω and the LSD of $\hat{\mathbf{\Sigma}}$. Specifically, if $\hat{\mathbf{\Sigma}}$ has an LSD $F_\omega(x)$, then $\hat{\mathbf{R}}$ has an LSD $H(t)$ satisfying

$$H(t) = F_\omega \left(t \cdot \frac{1}{d} \text{tr}(\mathbf{D}^{-1}) \right),$$

When $\frac{1}{d}\mathrm{tr}(\mathbf{D}^{-1}) \to 1$ (as $d, n \to \infty$), the LSD of $\mathbf{R}$ becomes $H(t) = F_\omega(t)$. The bootstrap correlation matrix $\hat{\mathbf{R}}_b^g$ is obtained by standardizing $\hat{\mathbf{\Sigma}}_b^g$, by Theorem 1 of El Karoui (2008), the LSD of $\hat{\mathbf{R}}_b^g$ is given by

$$H(t) = F_\omega\left(t \cdot \frac{1}{d}\mathrm{tr}(\mathbf{D}_b^{-1})\right),$$

where $\mathbf{D}_b = \mathrm{diag}(\hat{\mathbf{\Sigma}}_b^g)$. Since $\frac{1}{d}\mathrm{tr}(\mathbf{D}_b^{-1}) \to 1$ as $d, n \to \infty$, we have $H(t) = F_\omega(t)$.

Step 3. We introduce Theorem 2.1 in Bai and Ding (2012). Assume taht the population covariance matrix $\mathbf{\Sigma}$ has a spiked structure, meaning its top K eigenvalues $\lambda_1, \ldots, \lambda_K$ are significantly larger than the remaining eigenvalues (noise eigenvalues). Let $\hat{\mathbf{\Sigma}}$ be the sample covariance matrix with sample eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_d$, as $n, d \to \infty$ and $d/n \to \omega \in (0, \infty)$, the sample eigenvalues $\hat{\lambda}_j$ and the population eigenvalues λ_j satisfy the following relationship

$$\hat{\lambda}_j = \lambda_j \cdot \psi(\lambda_j) + o_p(1),$$

where $\psi(\lambda)$ is the correction function defined as

$$\psi(\lambda) = 1 + \omega \int \frac{t}{\lambda - t} dH(t),$$

and $H(t)$ is the limiting spectral distribution (LSD) of the noise eigenvalues. The corrected eigenvalues $\hat{\lambda}_j^C$ are defined via the inverse Stieltjes transform $\hat{\lambda}_j^C = -\frac{1}{\gamma_j(\hat{\lambda}_j)}$, then the corrected eigenvalues $\hat{\lambda}_j^C$ satisfy $\hat{\lambda}_j^C = \lambda_j + o_p(1)$. For the sample eigenvalues $\hat{\lambda}_j^b$ of $\hat{\mathbf{R}}_b^g$, their relationship with the population eigenvalues $\lambda_j(\mathbf{R})$ is

$$\hat{\lambda}_j^b = \lambda_j(\mathbf{R}) \cdot \psi_b(\lambda_j(\mathbf{R})) + o_p(1),$$

where $\psi_b(\lambda)$ is the correction function in the bootstrap framework, defined as

$$\psi_b(\lambda) = 1 + \omega \int \frac{t}{\lambda - t} dH_b(t).$$

Here, $H_b(t)$ is the LSD of $\hat{\mathbf{R}}_b^g$, which is identical to $H(t)$. The Stieltjes transform $m_b(z)$ of $\hat{\mathbf{R}}_b^g$ is defined as

$$m_b(z) = \frac{1}{d} \operatorname{tr} \left(\hat{\mathbf{R}}_b^g - z\mathbf{I}_p \right)^{-1}.$$

According to random matrix theory, $m_b(z)$ satisfies the equation

$$z = -\frac{1}{m_b(z)} + \omega \int \frac{t}{1 + t m_b(z)} dH_b(t).$$

For the sample eigenvalues $\hat{\lambda}_j^b$ of $\hat{\mathbf{R}}_b^g$, the corrected eigenvalues are defined as $\hat{\lambda}_j^{C,b} = -\frac{1}{\gamma_j(\hat{\lambda}_j^b)}$. Since the LSD of $\hat{\mathbf{R}}_b^g$ is identical to that of $\mathbf{R}$, the correction function $\psi_b(\lambda)$ is the same as $\psi(\lambda)$:

$$\psi_b(\lambda) = \psi(\lambda) = 1 + \omega \int \frac{t}{\lambda - t} dH(t).$$

For signal eigenvalues ($j \leq K$), the corrected eigenvalues satisfy

$$\hat{\lambda}_j^{C,b} = \lambda_j(R) + o_p(1).$$

This is because

$$\hat{\lambda}_j^{C,b} = -\frac{1}{\gamma_j(\hat{\lambda}_j^b)} \approx \lambda_j(R) \left(1 - \omega \int \frac{t}{\lambda_j(R) - t} dH(t) \right)^{-1} = \lambda_j(R) + o_p(1).$$

For noise eigenvalues ($j > K$), the corrected eigenvalues satisfy

$$\hat{\lambda}_j^{C,b} \leq 1 + \sqrt{\omega} + o_p(1).$$

This is because the LSD $H(t)$ of noise eigenvalues is supported on $[0, 1 + \sqrt{\omega}]$, and the correction process does not alter their asymptotic upper bound.

Step 4. Motivated by Theorem 6 in Fan et al. (2022), Let $\lambda_{\max}^{\text{noise}}$ be the maximum noise eigenvalue. Then by the Marchenko-Pastur law, the maximum limit of noise eigenvalues is $1+\sqrt{\omega}$. From random matrix theory, the sample maximum eigenvalue converges almost surely to this limit, thus we have

$$\lambda_{\max}^{\text{noise}} \leq 1 + \sqrt{\omega} + o_p(1).$$

Let $\lambda_{\min}^{\text{signal}} = \min_{1\leq j\leq K} \lambda_j(\mathbf{R})$. By the assumptions of the factor model, signal eigenvalues significantly exceed the upper bound of noise eigenvalues, then we have

$$\lambda_{\min}^{\text{signal}} > 1 + \sqrt{\omega} + \delta.$$

Where δ is a very small positive number. Combining the above results, we have

$$P\left(\hat{\lambda}_{K+1}^{C,b} \leq 1 + \sqrt{\omega} + \delta\right) \geq P\left(\lambda_{\max}^{\text{noise}} \leq 1 + \sqrt{\omega} + \delta\right) \to 1.$$

and

$$P\left(\hat{\lambda}_{K}^{C,b} > 1 + \sqrt{\omega} + \delta\right) \geq P\left(\lambda_{\min}^{\text{signal}} > 1 + \sqrt{\omega} + \delta\right) \to 1.$$

Therefore,

$$P\left(\hat{K}^g = K\right) \geq P\left(\hat{\lambda}_{K}^{C,b} > 1 + \sqrt{\omega} + \delta \text{ and } \hat{\lambda}_{K+1}^{C,b} \leq 1 + \sqrt{\omega} + \delta\right) \to 1.$$

Thus we complete the proof of Lemma S3. □

Proof of Theorem 1. Note that Lemma S3 holds for all $g = 1, \ldots, G$. Then Theorem 1 follows by taking the mode of $\hat{K}^1, \ldots, \hat{K}^G$. □

To establish the consistency of $\hat{K}_b$ in estimating the true number of factors when the number of factors K diverges with the dimension d, as stated in Theorem S2, an additional condition is required.

(vi). $\left\|\mathbf{B}^T[\mathrm{diag}(\mathbf{\Sigma})]^{-1}\mathbf{B}\right\| \cdot \left\|\left(\mathbf{B}^T[\mathrm{diag}(\mathbf{\Sigma})]^{-1}\mathbf{B}\right)^{-1}\right\| = O\left(d^{\varrho_2}\right)$ for $\varrho_2 < 1/3$.

Condition (vi) is equivalent to requiring that the growth rate of the condition number of the matrix $\mathbf{B}^T[\mathrm{diag}(\mathbf{\Sigma})]^{-1}\mathbf{B}$ does not exceed d^{ϱ_2}, where $\varrho_2 < 1/3$. This restriction ensures that the minimum eigenvalue $\lambda_{\min}(\mathbf{B}^T[\mathrm{diag}(\mathbf{\Sigma})]^{-1}\mathbf{B})$ does not decay too rapidly. As a result, this condition effectively rules out weak factors, i.e., those for which the columns of the factor loading matrix $\mathbf{B}$ are nearly linearly dependent or have extremely small norms.

Theorem S2. *For the factor model (2.1) satisfying Conditions (i)-(iv) and (vi), consider the case of $K = o(d^{1/3})$. When $\lambda_K(\mathbf{R}) > 1 + \sqrt{\omega} + \Upsilon$ for a very small positive constant Υ, we have $P\big(\hat{K}_b = K\big) \to 1,$ as $n, d \to \infty$.*

Proof of Theorem S2. Define

$$\mathbf{S}_b := [\mathrm{diag}(\mathbf{\Sigma})]^{-1/2}\hat{\mathbf{\Sigma}}_b^g[\mathrm{diag}(\mathbf{\Sigma})]^{-1/2},$$

then $\hat{\mathbf{R}}_b^g = [\mathrm{diag}\left(\mathbf{S}_b\right)]^{-1/2}\mathbf{S}_b\left[\mathrm{diag}\left(\mathbf{S}_b\right)\right]^{-1/2}$. Under condition (vi), for each $j \in [d]$,

$$[\mathbf{S}_b]_{jj} = \frac{1}{n}\sum_{i=1}^{n} v_i \left[\mathbf{Q}\left(\mathbf{U}_i - \overline{\mathbf{U}}\right)\left(\mathbf{U}_i - \overline{\mathbf{U}}\right)^T \mathbf{Q}^T\right]_{jj},$$

where $\mathbf{R} = \mathbf{Q}\mathbf{Q}^T$ with $\mathbf{Q} = [\mathrm{diag}(\mathbf{\Sigma})]^{-1/2}\left(\mathbf{B}, \mathbf{\Psi}^{1/2}\right)$. Since $E[v_i] = 1$ and $\mathrm{Var}(v_i) = 1$, by

Chebyshev's inequality condition on the data, we have

$$P\left(\left|[\mathbf{S}_b]_{jj} - [\mathbf{S}_n]_{jj}\right| > \epsilon \mid \mathbf{Z}\right) \leq \frac{\operatorname{Var}\left(v_i \left[\mathbf{Q}\left(\mathbf{U}_i - \overline{\mathbf{U}}\right)\left(\mathbf{U}_i - \overline{\mathbf{U}}\right)^T \mathbf{Q}^T\right]_{jj}\right)}{n\epsilon^2} = O\left(n^{-1}\right),$$

where

$$\begin{aligned}\mathbf{S}_n &= n^{-1}\sum_{i=1}^{n}[\operatorname{diag}(\mathbf{\Sigma})]^{-1/2}\left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)\left(\mathbf{Z}_i - \overline{\mathbf{Z}}\right)^T[\operatorname{diag}(\mathbf{\Sigma})]^{-1/2} \\ &= n^{-1}\sum_{i=1}^{n}\mathbf{Q}\left(\mathbf{U}_i - \overline{\mathbf{U}}\right)\left(\mathbf{U}_i - \overline{\mathbf{U}}\right)^T\mathbf{Q}^T.\end{aligned}$$

By Lemma S.8 of Fan et al. (2022), $\max_j \left|[\mathbf{S}_n]_{jj} - 1\right| = o_{\text{a.s.}}(1)$. Thus, $\max_j \left|[\mathbf{S}_b]_{jj} - 1\right| = o_p(1)$.

By the Weyl inequality, we have

$$\max_k \left|\lambda_k(\hat{\mathbf{R}}_b^g) - \lambda_k\left(\mathbf{S}_b\right)\right| \leq \left\|\hat{\mathbf{R}}_b^g - \mathbf{S}_b\right\|.$$

As $\hat{\mathbf{R}}_b^g - \mathbf{S}_b = \mathbf{D}_{\hat{\Sigma}}^{-1/2}\hat{\mathbf{\Sigma}}_b^g\mathbf{D}_{\hat{\Sigma}}^{-1/2} - \mathbf{D}^{-1/2}\hat{\mathbf{\Sigma}}_b^g\mathbf{D}^{-1/2}$, where $\mathbf{D}_{\hat{\Sigma}} = \operatorname{diag}(\hat{\mathbf{\Sigma}}_b^g)$ and $\mathbf{D} = \operatorname{diag}(\mathbf{\Sigma})$, rewrite

$$\hat{\mathbf{R}}_b^g - \mathbf{S}_b = \mathbf{D}_{\hat{\Sigma}}^{-1/2}\hat{\mathbf{\Sigma}}_b^g\left(\mathbf{D}_{\hat{\Sigma}}^{-1/2} - \mathbf{D}^{-1/2}\right) + \left(\mathbf{D}_{\hat{\Sigma}}^{-1/2} - \mathbf{D}^{-1/2}\right)\hat{\mathbf{\Sigma}}_b^g\mathbf{D}^{-1/2},$$

then we have

$$\left\|\hat{\mathbf{R}}_b^g - \mathbf{S}_b\right\| \leq \left\|\mathbf{D}_{\hat{\Sigma}}^{-1/2}\hat{\mathbf{\Sigma}}_b^g\right\| \cdot \left\|\mathbf{D}_{\hat{\Sigma}}^{-1/2} - \mathbf{D}^{-1/2}\right\| + \left\|\mathbf{D}_{\hat{\Sigma}}^{-1/2} - \mathbf{D}^{-1/2}\right\| \cdot \left\|\hat{\mathbf{\Sigma}}_b^g\mathbf{D}^{-1/2}\right\|.$$

$\mathbf{D}_{\hat{\Sigma}}^{-1/2}$ and $\mathbf{D}^{-1/2}$ are all diagonal matrices, thus

$$\begin{gathered}\left\|\mathbf{D}_{\hat{\Sigma}}^{-1/2}\hat{\mathbf{\Sigma}}_b^g\right\| \leq \left\|\mathbf{D}_{\hat{\Sigma}}^{-1/2}\right\| \cdot \left\|\hat{\mathbf{\Sigma}}_b^g\right\| = \max_j \left[\hat{\mathbf{\Sigma}}_b^g\right]_{jj}^{-1/2} \cdot \left\|\hat{\mathbf{\Sigma}}_b^g\right\|, \\ \left\|\mathbf{D}_{\hat{\Sigma}}^{-1/2} - \mathbf{D}^{-1/2}\right\| = \max_j \left|[\hat{\mathbf{\Sigma}}_b^g]_{jj}^{-1/2} - [\mathbf{\Sigma}]_{jj}^{-1/2}\right|.\end{gathered}$$

In addition, $\frac{[\hat{\mathbf{\Sigma}}_b^g]_{jj}}{[\mathbf{\Sigma}]_{jj}} = [\mathbf{S}_b]_{jj}$ and we have derived that $\max_j \left|[\mathbf{S}_b]_{jj} - 1\right| = o_p(1)$, as a result,

$$\left\|\mathbf{D}_{\hat{\Sigma}}^{-1/2} - \mathbf{D}^{-1/2}\right\| = \max_j \left|\frac{1}{\sqrt{[\hat{\mathbf{\Sigma}}_b^g]_{jj}}} - \frac{1}{\sqrt{[\mathbf{\Sigma}]_{jj}}}\right| \leq \frac{1}{2}\max_j \left|\frac{[\mathbf{\Sigma}]_{jj} - [\hat{\mathbf{\Sigma}}_b^g]_{jj}}{[\mathbf{\Sigma}]_{jj}^{3/2}}\right| = o_p(1),$$

Then

$$\left\|\hat{\mathbf{R}}_b^g - \mathbf{S}_b\right\| \leq 2 \cdot O_p(1) \cdot o_p(1) = o_p(1).$$

Thus $\lambda_k(\hat{\mathbf{R}}_b^g)$ and $\lambda_k(\mathbf{S}_b)$ share the same limit behavior.

Define the bias-corrected eigenvalues for $\hat{\mathbf{R}}_b^g$ as above: $\hat{\lambda}_j^{C,b} = -\frac{1}{\gamma_j(\hat{\lambda}_j^b)}$. Under the condition $K = o(d^{1/3})$, which ensures that the rank-K perturbation $\mathbf{Q}_1\mathbf{Q}_1^T$, with $\mathbf{Q}_1 = [\mathrm{diag}(\mathbf{\Sigma})]^{-1/2}\mathbf{B}$, does not affect the bulk spectrum asymptotics, and assuming Conditions (v)–(vii) hold, Theorem 3 of Fan et al. (2022) can be extended to the setting of $\hat{\mathbf{R}}_b^g$. Specifically, we have that the empirical Stieltjes transform converges: $|\gamma_j(z) - \underline{m}(z)| \to 0$ a.s., where $\underline{m}(z)$ solves

$$z = -\frac{1}{\underline{m}(z)} + \omega \int \frac{t}{1 + \underline{m}(z)} dH(t),$$

For $j \leq K$, the sample eigenvalues of $\hat{\mathbf{R}}_b^g$ satisfy

$$\hat{\lambda}_j^b = \lambda_j(\mathbf{R}) \cdot \psi_b(\lambda_j(\mathbf{R})) + o_p(1),$$

where $\psi_b(\lambda)$ is the correction function in the bootstrap framework, defined as

$$\psi_b(\lambda) = 1 + \omega \int \frac{t}{\lambda - t} dH_b(t).$$

Furthermore, the bias-corrected eigenvalues are consistent in the sense that: $\hat{\lambda}_j^{C,b}/\lambda_j(\mathbf{R}) \xrightarrow{p} 1$ for $j \leq K$.

For $j > K, \lambda_j(\mathbf{R}) \leq 1$ by Theorem 1 of Fan et al. (2022). Under the condition that $K = o(d^{1/3})$ and Conditions (iii), (v), the empirical spectral distribution (ESD) of $\left\{\lambda_j(\hat{\mathbf{R}}_b^g)\right\}_{j=K+1}^{d}$ converges to the same limit $H(t)$ as the original $\mathbf{R}$. By Theorem 2 of Fan et al. (2022), the Stieltjes transform $m_b(z)$ of $\hat{\mathbf{R}}_b^g$ satisfies

$$z = -\frac{1}{\underline{m}_b(z)} + \omega \int \frac{t}{1 + t\underline{m}_b(z)} dH(t) + o_p(1),$$

The right edge of the support of $H(t)$ is $(1 + \sqrt{\omega})^2$. Thus, for $j > K$,

$$\hat{\lambda}_j^b \leq (1 + \sqrt{\omega})^2 + \varsigma + o_p(1) \quad \forall \varsigma > 0,$$

The bias correction for non-spiked eigenvalues satisfies (by Lemma S.9 and Theorems 4-5 of Fan et al. (2022))

$$\hat{\lambda}_j^{C,b} \leq 1 + \sqrt{\omega} + o_p(1) \quad \text{for} \quad j > K.$$

Define $s_b = 1 + \sqrt{d/(n-1)}$, by Condition (iii), it follows that $s_b \to 1 + \sqrt{\omega}$. Moreover, we have $\lambda_K(\mathbf{R}) > 1 + \sqrt{\omega} + \Upsilon$. For $j = K$, we have established that $\hat{\lambda}_K^{C,b} \xrightarrow{p} \lambda_K(\mathbf{R}) > 1 + \sqrt{\omega} + \Upsilon$, thus

$$P\left(\hat{\lambda}_K^{C,b} > s_b\right) \to 1,$$

For $j = K + 1$, $\hat{\lambda}_{K+1}^{C,b} \leq 1 + \sqrt{\omega} + o_p(1)$. Since $s_b \to 1 + \sqrt{\omega}$,

$$P\left(\hat{\lambda}_{K+1}^{C,b} < s_b\right) \to 1,$$

The estimator is $\hat{K}^g = \max\left\{j : \hat{\lambda}_j^{C,b} > s_b\right\}$. Thus

$$P\left(\hat{K}^g = K\right) \geq P\left(\hat{\lambda}_K^{C,b} > s_b \cap \hat{\lambda}_{K+1}^{C,b} < s_b\right) \to 1.$$

Note that the above results holds for all $g = 1, \ldots, G$. Then Theorem S2 follows by taking the mode of $\hat{K}^1, \ldots, \hat{K}^G$. $\square$

§7.3 Technical Proofs for the Oracle Procedure

We next examine the oracle procedure under the assumption that the latent factors are known. Specifically, we assume that the loading matrix $\mathbf{B}$ is known and the factors $\{\boldsymbol{f}_i\}_{i=1}^n$ are observable, which serves as a heuristic device. Given a quantile $\tau > 0$, denote by $\hat{\delta}_{k,j}^{\tau,oracle} = \frac{1}{n}\sum_{l=1}^n \left\{n^{-1}\sum_{i=1}^n I\big(f_{ik} \le \hat{h}_{kj}(X_{lj})\big) - \tau\right\}^2$ and $\hat{\delta}_j^{\tau,oracle} = \max_{1\le k\le K} \hat{\delta}_{k,j}^{\tau,oracle}$ the estimators of $\mathrm{MSD}(f_k|X_j)$ and $\mathrm{MSD}(\boldsymbol{f}|X_j)$, respectively. The corresponding selected subsets are denoted as $\hat{\mathcal{A}}_{\tau,k}^{oracle}$ and $\hat{\mathcal{A}}_{\tau}^{oracle}$.

Theorem S3. *Under Conditions (C1)-(C7) in the main text, we have*

(1) For any $C > 0$ and $0 < \tau < 1$, there exist positive constants c_2' and c_3' such that

$$P\Big(\max_{1\le j\le p} |\hat{\delta}_j^{\tau,oracle} - \delta_j^\tau| \ge Cn^{-\kappa}\Big) \le Kp\Big\{\exp(-c_2' n^{1-4\kappa} + \log n) + \exp(-c_3' s_n^{-2} n^{1-2\kappa} + \log n)\Big\} \tag{S7.7}$$

for all n sufficiently large.

(2) (Sure screening property) If $\kappa < 1/4$, $s_n^2 n^{2\kappa-1} = o(1)$, take the threshold $\eta_n = c^* n^{-\kappa}$ for some constant c^*, then

$$P(\mathcal{A}_\tau \subset \hat{\mathcal{A}}_\tau^{oracle}(\eta_n)) \ge 1 - |\mathcal{A}_\tau|\{\exp(-c_2' n^{1-4\kappa} + \log n) + \exp(-c_3' s_n^{-2} n^{1-2\kappa} + \log n)\}$$

for all n sufficiently large, where $|\mathcal{A}_\tau|$ is the cardinality of $\mathcal{A}_\tau$. Therefore, $P(\mathcal{A}_\tau \subset \hat{\mathcal{A}}_\tau^{oracle}) \to 1$ as $n \to \infty$.

The results of Theorem S3 imply that we can handle the dimensionality $\log p = o(n^{1-4\kappa} + s_n^{-2} n^{1-2\kappa})$, which relies on the number of basis functions s_n and the strength of marginal signals. If we take $s_n = n^{1/(2r+1)}$(the optimal rate for spline approximation), then for $\kappa < \min(1/4, (2r-1)/(4r+2))$, we can handle the ultrahigh dimensionality, that is, p can grow at the exponential rate. To investigate the ranking consistency property of MSDS, we additionally assume the following condition.

(vii). (Condition on the minimum signal strength) $\min_{j\in\mathcal{A}_{\tau,k}} \delta_{k,j} - \max_{l\in\mathcal{A}^c_{\tau,k}} \delta_{k,l} \geq 2c_0 n^{-\kappa}$, for some $0 \leq \kappa < 1/2$ and some positive constant c_0.

Condition (vii) requires the MSD index is able to separate active and inactive covariates well in the population level. The following theorem justifies the ranking consistency property of MSDS.

Theorem S4. *(Rank consistency property) Replace Condition (C4) with Condition (vii), we have*

$$P\left(\min_{j\in\mathcal{A}_\tau} \hat{\delta}_j^{\tau,oracle} - \max_{j\in\mathcal{A}^c_\tau} \hat{\delta}_j^{\tau,oracle} > 0\right) > 1 - Kp\big\{\exp(-c_4' n^{1-4\kappa} + \log n) + \exp(-c_5' s_n^{-2} n^{1-2\kappa} + \log n)\big\},$$

where c_4' and c_5' are some positive constants. If $\log p = o(n^{1-4\kappa} + s_n^{-2} n^{1-2\kappa})$ and $\log n = o(n^{1-4\kappa} + s_n^{-2} n^{1-2\kappa})$ with $0 < \kappa < 1/4$, then we have

$$\liminf_{n\to\infty} \{\min_{j\in\mathcal{A}_\tau} \hat{\delta}_j^{\tau,oracle} - \max_{j\in\mathcal{A}^c_\tau} \hat{\delta}_j^{\tau,oracle}\} > 0, a.s.$$

We will now prove the two theorems stated above, beginning with the introduction of the following lemmas.

Lemma S4. *(Hoeffding's Inequality) Let $X_1, \ldots, X_n$ be independent random variables.*

Assume that $P(X_i \in [a_i, b_i]) = 1$ for $1 \le i \le n$, where a_i and b_i are constants. Let $\bar{X} = n^{-1}\sum_{i=1}^n X_i$. Then the following inequality holds

$$P(|\bar{X} - E(\bar{X})| \ge t) \le 2\exp\left\{-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right\}, \tag{S7.8}$$

where t is a positive constant and $E(\bar{X})$ is the expected value of $\bar{X}$.

Lemma S5. *(Bernstein's Inequality) Let $X_1, \ldots, X_n$ be independent random variables with bounded support $[-B, B]$ and zero means, then the following inequality holds*

$$P(|X_1 + \cdots + X_n| > t) \le 2\exp\left\{-\frac{t^2}{2(v + Bt/3)}\right\}, \tag{S7.9}$$

for $v \ge var(X_1 + \cdots + X_n)$.

The following notations are needed for next lemma. Let $g_{k,j} = g(f_k, X_j) = [F(Q_\tau(f_k|X_j)) - \tau]^2$, $g_{k,j}^{(i)} = g(f_{ik}, X_{ij})$, $g_k(x) = I(f_k \le x)$, $g_k^{(i)}(x) = I(f_{ik} \le x)$ for $i = 1, \ldots, n$.

Lemma S6. *For any $\epsilon \in (0, 1)$, the following inequalities hold for X_j,*

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n g_{k,j}^{(i)} - Eg_{k,j}\right| \ge \epsilon\right\} \le 2\exp\{-2n\epsilon^2\}; \tag{S7.10}$$

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n g_k^{(i)}(x) - Eg_k(x)\right| \ge \epsilon\right\} \le 2\exp\{-2n\epsilon^2\}; \tag{S7.11}$$

Proof of Lemma S6. Since $|g(X_j, f_k)| = [F(Q_\tau(f_k|X_j)) - \tau]^2 \le 1$ and $|g^k(x)| \le 1$, the inequalities follow by using Hoeffding's inequality. $\square$

For simplicity, we write $\delta_{k,j}^\tau = \delta_{k,j}$, $\hat{\delta}_{k,j}^\tau = \hat{\delta}_{k,j}$, $Q_\tau(f_k|X_j) = Q_k(X_j)$, $\mathcal{L}_{\tau,h} = \mathcal{L}_\tau$ hereinafter. Let $\tilde{K}_h(u) := \int_{-\infty}^{-u/h} \mathcal{K}(t)dt$, and $\epsilon_i(\boldsymbol{\beta}) = f_{ik} - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}$, the gradient and hessian matrix of $\hat{Q}_{\tau_m,h}(\boldsymbol{\beta})$ are

$$\nabla\hat{Q}_{\tau_m,h}(\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^{n}\{\tilde{\mathcal{K}}_h(\epsilon_i(\boldsymbol{\beta}))-\tau_m\}\boldsymbol{\pi}(X_{ij}),\ \ \nabla^2\hat{Q}_{\tau_m,h}(\boldsymbol{\beta})=\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_h(-\epsilon_i(\boldsymbol{\beta}))\boldsymbol{\pi}(X_{ij})\boldsymbol{\pi}(X_{ij})^T.$$

Denote $\nabla Q_{\tau_m,h}(\boldsymbol{\beta})=E(\nabla\hat{Q}_{\tau_m,h}(\boldsymbol{\beta}))$, and $\nabla^2 Q_{\tau_m,h}(\boldsymbol{\beta})=E(\nabla^2\hat{Q}_{\tau_m,h}(\boldsymbol{\beta}))$.

Let $B_n(\boldsymbol{\beta})=n^{-1}\sum_{i=1}^{n}\mathcal{L}_\tau(f_{ik}-\boldsymbol{\pi}(X_{ij}^T\boldsymbol{\beta}))$, $B_h(\boldsymbol{\beta})=E(\mathcal{L}_\tau(f_k-\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta})$, and $B(\boldsymbol{\beta})=E\rho_\tau(f_k-\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta})$. Then $\hat{\boldsymbol{\beta}}_{kj}=\arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{s_n+l+1}}B_n(\boldsymbol{\beta})$, $\boldsymbol{\beta}^*_{kj}$ and $\boldsymbol{\beta}^0_{kj}$ are the unique minimizers of $B_h(\boldsymbol{\beta})$ and $B(\boldsymbol{\beta})$, respectively. We can bound the difference $\|\hat{\boldsymbol{\beta}}_{kj}-\boldsymbol{\beta}^0_{kj}\|$ by $\|\hat{\boldsymbol{\beta}}_{kj}-\boldsymbol{\beta}^*_{kj}\|+\|\boldsymbol{\beta}^0_{kj}-\boldsymbol{\beta}^*_{kj}\|$. $\|\boldsymbol{\beta}^0_{kj}-\boldsymbol{\beta}^*_{kj}\|$ is dominated by $O(h^2)$ following Lemma S7. $\|\hat{\boldsymbol{\beta}}_{kj}-\boldsymbol{\beta}^*_{kj}\|$ can be bounded by their respective objective functions.

Lemma S7. *(Smoothing bias) Under Conditions (C1), (C6), and (C7), as long as n is large enough, $\|\boldsymbol{\beta}^*_{kj}-\boldsymbol{\beta}^0_{kj}\|=O(h^2)$.*

The proof of Lemma S7 can refer to online appendix in Fernandes et al. (2021).

Lemma S8. *For any $\delta>0$,*

$$P(\|\hat{\boldsymbol{\beta}}_{kj}-\boldsymbol{\beta}^*_{kj}\|\geq\delta)\leq P\left(\sup_{\|\hat{\boldsymbol{\beta}}_{kj}-\boldsymbol{\beta}^*_{kj}\|\leq\delta}|B_n(\boldsymbol{\beta})-B_h(\boldsymbol{\beta})|\geq\frac{1}{2}\inf_{\|\hat{\boldsymbol{\beta}}_{kj}-\boldsymbol{\beta}^*_{kj}\|=\delta}(B_h(\boldsymbol{\beta})-B_h(\boldsymbol{\beta}^*_{kj}))\right).\tag{S7.12}$$

Lemma S8 is a direct application of Lemma 2 of Hjort and Pollard (2011) making use of the convexity of the objective function.

The lower bound of the right-hand side of (S7.12) can be explicitly evaluated for a given $\delta>0$. This is specified in the following lemma.

Lemma S9. *Let $C>0$ be an arbitrary constant. Assume that $s_n^{-r}n^{\kappa}=o(1)$, then there* *exists a constant a_1 such that*

$$\inf_{\|\hat{\boldsymbol{\beta}}_{kj}-\boldsymbol{\beta}^*_{kj}\|=Cs_n^{1/2}} \big(B_h(\boldsymbol{\beta}) - B_h(\boldsymbol{\beta}^*_{kj})\big) \geq a_1 n^{-2\kappa},$$

for n sufficiently large.

Proof of Lemma S9. We consider $\boldsymbol{\beta} = \boldsymbol{\beta}^*_{kj} + Cs_n^{1/2}n^{-\kappa}\boldsymbol{u}$, where $\boldsymbol{u} \in \mathbb{R}^{s_n+l+1}$ satisfying $\|\boldsymbol{u}\| = 1$. We have

$$\begin{aligned} B_h(\boldsymbol{\beta}) - B_h(\boldsymbol{\beta}^*_{kj}) &= E\big\{\mathcal{L}_\tau(f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj} - Cs_n^{1/2}n^{-\kappa}\boldsymbol{u}) - \mathcal{L}_\tau(f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj})\big\} \\ &= E[-\nabla Q_{\tau,h}(\boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}^*_{kj})Cs_n^{1/2}n^{-\kappa}\boldsymbol{u}] + E[\nabla^2 Q_{\tau,h}(\boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}^*_{kj})(Cs_n^{1/2}n^{-\kappa})^2\|\boldsymbol{u}\|] \\ &= Cs_n^{1/2}n^{-\kappa}E[(\tau - \tilde{\mathcal{K}}_h(\epsilon_i(\boldsymbol{\beta}^*_{kj}))\boldsymbol{\pi}(X_j)^T\boldsymbol{u}] + E[\mathcal{K}_h(-\epsilon_i(\boldsymbol{\beta}^*_{kj}))\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T(Cs_n^{1/2}n^{-\kappa})^2] \\ &= J_1 + J_2, \end{aligned}$$

where $\epsilon_i(\boldsymbol{\beta}^*_{kj}) = f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj}$. By Hölder's inequality, we have

$$\begin{aligned} |J_1| &\leq Cs_n^{1/2}n^{-\kappa}(E(\boldsymbol{\pi}(X_j)^T\boldsymbol{u})^2)^{1/2}[E(\tau - \tilde{\mathcal{K}}_h(\epsilon_i(\boldsymbol{\beta}^*_{kj})))^2]^{1/2} \\ &\leq Cs_n^{1/2}n^{-\kappa}O(s_n^{-1/2})O(s_n^{-r}) = O(s_n^{-r}n^{-\kappa}). \end{aligned}$$

The second inequality holds by the properties of spline basis, see He et al. (2013) and Sherwood and Wang (2016).

$$J_2 = E[\mathcal{K}_h(-\epsilon_i(\boldsymbol{\beta}^*_{kj}))\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T(Cs_n^{1/2}n^{-\kappa})^2] = O(n^{-2\kappa}).$$

Note that J_2 is nonnegative and $J_1 = o(J_2)$. This completes the proof of Lemma S9. $\square$

Lemma S10. *Assume that $s_n n^{2\kappa-1} = o(1)$. For any $C > 0$, there exist positive constants a_2 and a_3 such that for n sufficiently large*

$$P(\|\hat{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\| \geq Cs_n^{1/2}n^{-\kappa}) \leq 2\exp(-a_2 n^{1-4\kappa}) + \exp(-a_3 s_n^{-2} n^{1-2\kappa}).$$

Proof of Lemma S10. Following Lemma S8 and Lemma S9, there exists some $c > 0$, for all n sufficiently large,

$$
\begin{aligned}
& P(\|\hat{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\| \geq C s_n^{1/2} n^{-\kappa}) \\
\leq \;& P\bigg(\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \leq C s_n^{1/2} n^{-\kappa}} |B_n(\boldsymbol{\beta}) - B_h(\boldsymbol{\beta})| \geq c n^{-2\kappa} \bigg) \\
\leq \;& P(|B_n(\boldsymbol{\beta}^*_{kj}) - B(\boldsymbol{\beta}^*_{kj})| \geq \frac{1}{2} c n^{-2\kappa}) \\
& + P\bigg(\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \leq C s_n^{1/2} n^{-\kappa}} |B_n(\boldsymbol{\beta}) - B_n(\boldsymbol{\beta}^*_{kj}) - B_h(\boldsymbol{\beta}) + B_h(\boldsymbol{\beta}^*_{kj})| \geq c n^{-2\kappa} \bigg) \\
= \;& L_1 + L_2.
\end{aligned}
$$

First, we evaluate L_1. Let $W_i = \boldsymbol{\mathcal{L}}_\tau(f_{ik} - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}^*_{kj})$. Then $B_n(\boldsymbol{\beta}^*_{kj}) - B_h(\beta^*_{kj}) = n^{-1} \sum_{i=1}^n (W_i - EW_i)$. Note that $|W_i| \leq C |\boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}^*_{kj}|$ for some positive constant c. By the argument of Lemma 3.1 of He et al. (2013), $\sup_t |h_j(t) - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}_{0j}| \leq c^* s_n^{-r}$. It follows that $|W_i|$'s are uniformly bounded by a constant and $\mathrm{var}(W_i) \leq \sigma^2$ for some $\sigma^2 > 0$. Applying Bernstein's inequality, we have

$$
L_1 = P(|\sum_{i=1}^n W_i| \geq c n^{1-2\kappa}/2) \leq 2 \exp\left(- \frac{c^2 n^{1-4\tau}/4}{2\sigma^2 + C c n^{-2\kappa}/3} \right) \leq 2 \exp(-a_2 n^{1-4\kappa}).
$$

Then we consider L_2. To evaluate L_2, we employ the Massart's concentration theorem (Massart, 2000). Let $V_i = \mathcal{L}_\tau(f_{ik} - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}) - \mathcal{L}_\tau(f_{ik} - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}^*_{kj})$. Then

$$
V_i = \nabla Q_{\tau,h}(\epsilon_i(\boldsymbol{\beta}^*_{kj})) \boldsymbol{\pi}(X_{ij})^T (\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj}) + \nabla^2 Q_{\tau,h}(\epsilon_i(\boldsymbol{\beta}^*_{kj}))(\boldsymbol{\pi}(X_{ij})^T (\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj}))^2.
$$

Thus, $|V_i| \leq 2 |\boldsymbol{\pi}(X_{ij})^T (\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj})| \leq c s_n n^{-\kappa}$ for some $c > 0$.

Next, let $e_1, \ldots, e_n$ be a Rademacher sequence (i.e., iid sequence taking values of ± 1

with probability 1/2) independent of $V_1, \ldots, V_n$. We have

$$\begin{aligned}
& E\Big\{ \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \le C s_n^{1/2} n^{-\kappa}} n^{-1} \Big| \sum_{i=1}^n (V_i - E(V_i)) \Big| \Big\} \le 2E\Big\{ \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \le C s_n^{1/2} n^{-\kappa}} n^{-1} \Big| \sum_{i=1}^n e_i V_i \Big| \Big\} \\
\le \;& 4E\Big\{ \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \le C s_n^{1/2} n^{-\kappa}} n^{-1} \Big| \sum_{i=1}^n e_i \boldsymbol{\pi}(X_{ij})^T (\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj}) V_i \Big| \Big\} \le 4C s_n^{1/2} n^{-\kappa} E \| n^{-1} \sum_{i=1}^n e_i \boldsymbol{\pi}(X_{ij}) \| \\
\le \;& 4C s_n^{1/2} n^{-\kappa} [E \| n^{-1} \sum_{i=1}^n e_i \boldsymbol{\pi}(X_{ij}) \|^2]^{1/2} = 4C s_n^{1/2} n^{-\kappa} \Big[n^{-2} E(\sum_{i=1}^n e_i^2 \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\pi}(X_{ij})) \Big]^{1/2} \\
\le \;& \tilde{C} s_n^{1/2} n^{-\kappa-1/2},
\end{aligned}$$

for some constant $\tilde{C}$, where the first inequality holds by the symmetrization theorem (Lemma 2.3.1, Van Der Vaart and Wellner, 1996), the second inequality holds by employing the contraction theorem (Ledoux and Talagrand, 1991), and the last inequality holds since $E(b_k^2(X_{ij})) \le c_0 s_n^{-1}$ for $1 \le k \le s_n + l + 1$, $1 \le i \le n$ and $1 \le j \le p$ (Stone, 1985).

Finally, we apply Massart's concentration theorem to calculate L_2. Let

$$A = \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \le C s_n^{1/2} n^{-\kappa}} n^{-1} \Big| \sum_{i=1}^n (V_i - E(V_i)) \Big|.$$

Then

$$L_2 = P(A \ge cn^{-2\kappa}/2) = P(Z \ge EZ + (cn^{-2\kappa}/2 - EZ)) \le \exp(-a_3 s_n^{-2} n^{1-2\kappa})$$

for some positive constant a_3 and n sufficiently large. This completes the proof. □

Next, we evaluate the term $|\hat{Q}_k(X_j) - Q_k(X_j)| = |\hat{Q}_\tau(f_k|X_j) - Q_\tau(f_k|X_j)|$.

Lemma S11. *Under Conditions (C1)-(C7), we have*

$$P(|\hat{Q}_k(x) - Q_k(x)| \ge Cn^{-\kappa}) \le 2\exp(-a_2 n^{1-4\kappa}) + \exp(-a_3 s_n^{-2} n^{1-2\kappa}).$$

Proof of Lemma S11.

$$\begin{aligned}
& |\hat{Q}_k(X_j) - Q_k(X_j)| = |\boldsymbol{\pi}(X_j)^T\hat{\boldsymbol{\beta}}_{kj} - Q_k(X_j)| \\
= \; & |\boldsymbol{\pi}(X_j)^T\hat{\boldsymbol{\beta}}_{kj} - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj}| + |\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj} - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^0_{kj}| + |\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^0_{kj} - Q_k(X_j)| \\
= \; & T_1 + T_2 + T_3.
\end{aligned}$$

By the result of Schumaker (1981), it follows that $\sup_{t\in[0,1]}|Q_k(X_j) - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^0_{kj}| = O(s_n^{-r})$. Next we evaluate T_1 and T_2.

$$T_1 = \{(\pi(X_j)^T(\hat{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}))^2\}^{1/2} = \{(\hat{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj})^T\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T(\hat{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj})\}^{1/2}.$$

Then $\sup_{x_j} T_1 \le O_p(s_n^{-1/2})\|\hat{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\|$ due to the fact that there exist two positive constants a_1^* and a_2^* such that $a_1^* s_n^{-1} \le \lambda_{\min}(E\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)) \le \lambda_{\max}(E\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)) \le a_2^* s_n^{-1}$, $\forall j$, (Zhou et al., 1998). Similarly, we have $\sup_{X_j} T_2 = O_p(s_n^{-1/2}h^2) = O_p(n^{-1/2})$ by Condition (C7). Thus,

$$\begin{aligned}
P(|\hat{Q}_k(x) - Q_k(x)| \ge Cn^{-\kappa}) \;\; & \le \;\; P(T_1 \ge Cn^{-\kappa}/3) + P(T_2 \ge Cn^{-\kappa}/3) + P(T_3 \ge Cn^{-\kappa}/3) \\
& \le \;\; P(\|\hat{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\| \ge Cs_n^{1/2}n^{-\kappa}/3) \\
& \le \;\; 2\exp(-a_2 n^{1-4\kappa}) + \exp(-a_3 s_n^{-2} n^{1-2\kappa}).
\end{aligned}$$

The second inequality holds because $P(T_2 \ge Cn^{-\kappa}/3) = 0$, and $P(T_3 \ge Cn^{-\kappa}/3) = 0$, since $0 < \kappa < 1/2$ and $s_n^{-r}n^{\kappa} = o(1)$ by Condition (C5). □

Lemma S12. *Under Conditions (C1)-(C7), there exist positive constants c_2' and c_3' such that*

$$P(|\hat{\delta}_{k,j} - \delta_{k,j}| \ge Cn^{-\kappa}) \le O(n)\exp(-c_2' n^{1-4\kappa}) + O(n)\exp(-c_3' s_n^{-2} n^{1-2\kappa}). \qquad \text{(S7.13)}$$

Proof of Lemma S12. According to the definitions of $\delta_{k,j}$ and $\hat{\delta}_{k,j}$, we have

$$\begin{aligned}
&\hat{\delta}_{k,j} - \delta_{k,j} = \frac{1}{n}\sum_{i=1}^{n}[\hat{F}_{f_k}(\hat{Q}_k(X_{ij})) - \tau]^2 - \int [F_{f_k}(Q_k(x)) - \tau]^2 dF_j(x) \\
= &\int \left[(\hat{F}_{f_k}(\hat{Q}_k(x)) - \tau)^2 - (F_{f_k}(Q_k(x)) - \tau)^2\right] d\hat{F}_j(x) + \int [F_{f_k}(Q_k(x)) - \tau]^2 d(\hat{F}_j(x) - F_j(x)) \\
= &I + II.
\end{aligned}$$

We first deal with the term I.

$$\begin{aligned}
|I| &\leq 2\int \left|\hat{F}_{f_k}(\hat{Q}_k(x)) - F_{f_k}(Q_k(x))\right| d\hat{F}_j(x) \leq 2\sup_{x\in\mathbb{R}_{X_j}}\left|\hat{F}_{f_k}(\hat{Q}_k(x)) - F_{f_k}(Q_k(x))\right| \\
&\leq 2\sup_{x\in\mathbb{R}_{X_j}}\left|\hat{F}_{f_k}(\hat{Q}_k(x)) - F_{f_k}(\hat{Q}_k(x))\right| + 2\sup_{x\in\mathbb{R}_{X_j}}\left|F_{f_k}(\hat{Q}_k(x)) - F_{f_k}(Q_k(x))\right| \\
&= 2(I_1 + I_2),
\end{aligned}$$

where the first inequality holds by $\left|\hat{F}_{f_k}(\hat{Q}_k(x)) + F_{f_k}(Q_k(x)) - 2\tau\right| \leq 2$, and the second inequality holds by $\int d\hat{F}_j(x) = 1$. Then, we first deal with the term I_1, by applying (S7.11), we can obtain that

$$P(|I_1| \geq \epsilon) = P(\sup_{x\in\mathbb{R}_{X_j}}\left|\hat{F}_{f_k}(\hat{Q}_k(x)) - F_{f_k}(\hat{Q}_k(x))\right| \geq \epsilon) \leq 2(n+1)\exp(-2n\epsilon^2).$$

Next we deal with I_2,

$$\begin{aligned}
P(|I_2| \geq \epsilon) &\leq P(\sup_{x\in\mathbb{R}_{X_j}}\left|F_{f_k}(\hat{Q}_k(x)) - F_{f_k}(Q_k(x))\right| \geq \epsilon, |\hat{Q}_k(x) - Q_k(x)| \leq Cn^{-\kappa}) \\
&\quad + P(|\hat{Q}_k(x) - Q_k(x)| \geq Cn^{-\kappa}) \\
&\leq P(\sup_{x\in\mathbb{R}_{X_j}}\left|f_k(Q_k(x))|\hat{Q}_k(x) - Q_k(x)|\right| \geq \epsilon) + P(|\hat{Q}_k(x) - Q_k(x)| \geq Cn^{-\kappa}) \\
&\leq 2(n+1)\exp(-a_2^* n\epsilon^4) + (n+1)\exp(-a_3^* s_n^{-2} n\epsilon^2) \\
&+ 2\exp(-a_2 n^{1-4\kappa}) + \exp(-a_3 s_n^{-2} n^{1-2\kappa}),
\end{aligned}$$

for some constant $a_2^* > 0$ and $a_3^* > 0$. In addition, by (S7.11), we have

$$P(II \geq \epsilon) \leq 2\exp(-2n\epsilon^2).$$

Take $\epsilon = n^{-\kappa}$, then by the above three results, we have

$$\begin{aligned}
& P(|\hat{\delta}_{k,j} - \delta_{k,j}| \geq Cn^{-\kappa}) \\
\leq\ & (4n+6)\exp(-a_2^* n^{1-4\kappa}) + (2n+3)\exp(-a_3^* s_n^{-2} n^{1-2\kappa}) + (4n+6)\exp(-2n^{1-2\kappa}) \\
\leq\ & O(n)\exp(-c_2' n^{1-4\kappa}) + O(n)\exp(-c_3' s_n^{-2} n^{1-2\kappa}),
\end{aligned}$$

for some constants $c_2' > 0$ and $c_3' > 0$. This completes the proof. □

With the above preparation, next we prove Theorem S3.

Proof of Theorem S3. (1). By lemma S12, it follows directly that

$$P(\max_{1\leq j\leq p} |\hat{\delta}_{k,j} - \delta_{k,j}| \geq Cn^{-\kappa}) \leq O(pn)\exp(-c_2' n^{1-4\kappa}) + O(pn)\exp(-c_3' s_n^{-2} n^{1-2\kappa}). \quad \text{(S7.14)}$$

$$\begin{aligned}
P(\max_{1\leq j\leq p} |\hat{\delta}_j - \delta_j| \geq Cn^{-\kappa}) &= P(\max_{1\leq j\leq p} \max_{1\leq k\leq K} |\hat{\delta}_{k,j} - \delta_{k,j}| \geq Cn^{-\kappa}) \\
&\leq O(pnK)\exp(-c_2' n^{1-4\kappa}) + O(pn)\exp(-c_3' s_n^{-2} n^{1-2\kappa}).
\end{aligned}$$

(2). The results follow by using the first result of Theorem S3. If $\mathcal{A}_{\tau,k} \not\subset \hat{\mathcal{A}}_{\tau,k}$, then there must exist some $j \in \mathcal{A}_{\tau,k}$ such that $\hat{\delta}_{kj} < c_1 n^{-\kappa}$. It follows from Condition (C4) that $|\hat{\delta}_{k,j} - \delta_{k,j}| \geq c_1 n^{-\kappa}$ for some $j \in \mathcal{A}_{\tau,k}$. Hence, $\{\max_{j\in\mathcal{A}_{\tau,k}} |\hat{\delta}_{k,j} - \delta_{k,j}| \leq c_1 n^{-\kappa}\} \subset$

$\{\mathcal{A}_{\tau,k} \subset \hat{\mathcal{A}}_{\tau,k}\}$. Consequently,

$$\begin{aligned}
P(\mathcal{A}_{\tau,k} \subset \hat{\mathcal{A}}_{\tau,k}) &\geq P\{\max_{j\in\mathcal{A}_{\tau,k}} |\hat{\delta}_{k,j} - \delta_{k,j}| \leq c_1 n^{-\kappa}\} = 1 - P\{\min_{j\in\mathcal{A}_{\tau,k}} |\hat{\delta}_{k,j} - \delta_{k,j}| \geq c_1 n^{-\kappa}\} \\
&= 1 - |\mathcal{A}_{\tau,k}| P\{|\hat{\delta}_{k,j} - \delta_{k,j}| \geq c_1 n^{-\kappa}\} \\
&\leq 1 - O\big(|\mathcal{A}_{\tau,k}|\{\exp(-c_2 n^{1-4\kappa} + \log n) + \exp(-c_3 s_n^{-2} n^{1-2\kappa} + \log n)\}\big),
\end{aligned}$$

Similarly, the result for Theorem S3(2) can be obtained. This completes the proof. □

Proof of Theorem S4.

$$\begin{aligned}
& P\Big\{\big(\min_{j\in\mathcal{A}_{\tau,k}} \hat{\delta}_{k,j} - \max_{j\in\mathcal{A}^c_{\tau,k}} \hat{\delta}_{k,j}\big) < c_1' n^{-\kappa}\Big\} \\
\leq\; & P\Big\{\big(\min_{j\in\mathcal{A}_{\tau,k}} \hat{\delta}_{k,j} - \max_{j\in\mathcal{A}^c_{\tau,k}} \hat{\delta}_{k,j}\big) - \big(\min_{j\in\mathcal{A}_{\tau,k}} \delta_{k,j} - \max_{j\in\mathcal{A}^c_{\tau,k}} \delta_{k,j}\big) < -c_1' n^{-\kappa}\Big\} \\
\leq\; & P\Big\{|\big(\min_{j\in\mathcal{A}_{\tau,k}} \hat{\delta}_{k,j} - \max_{j\in\mathcal{A}^c_{\tau,k}} \hat{\delta}_{k,j}\big) - \big(\min_{j\in\mathcal{A}_{\tau,k}} \delta_{k,j} - \max_{j\in\mathcal{A}^c_{\tau,k}} \delta_{k,j}\big)| > c_1' n^{-\kappa}\Big\} \\
\leq\; & P\Big(2\max_{1\leq j\leq p} |\hat{\delta}_{k,j} - \delta_{k,j}| > c_1' n^{-\kappa}\Big) \\
\leq\; & O(pn)\exp(-c_4 n^{1-4\kappa}) + O(pn)\exp(-c_5 s_n^{-2} n^{1-2\kappa})
\end{aligned}$$

for some constants $c_4 > 0$ and $c_5 > 0$. If we further assume $\log p = o(n^{1-4\kappa} + s_n^{-2} n^{1-2\kappa})$, thus $p < \exp\{c_4 n^{1-4\kappa}/2\}$ for large n and $p < \exp\{c_5 s_n^{-2} n^{1-2\kappa}/2\}$ for large n, similarly by assumption $\log n = o(n^{1-4\kappa} + s_n^{-2} n^{1-2\kappa})$, we have $c_4 n^{1-4\kappa}/2 > 2\log n$ and $c_5 s_n^{-2} n^{1-2\kappa}/2 > 2\log n$ for large n. Then we have for some n_0, $\sum_{n=n_0}^{\infty} np[\exp(-c_4 n^{1-4\kappa}) + \exp(-c_5 s_n^{-2} n^{1-2\kappa})] \leq \sum_{n=n_0}^{\infty} \exp(\log n - c_4 n^{1-4\kappa}/2) + \exp(\log n - c_5 s_n^{-2} n^{1-2\kappa}/2) \leq 2\sum_{n=n_0}^{\infty} n^{-2} < \infty$.

Therefore, by Borel Contelli Lemma, we obtain that

$$\lim\inf_{n\to\infty} \{\min_{j\in\mathcal{A}_{\tau,k}} \hat{\delta}_{k,j} - \max_{j\in\mathcal{A}^c_{\tau,k}} \hat{\delta}_{k,j}\} > 0, \text{a.s.}$$

Similarly, by the same technique, we can obtain another result for $\hat{\delta}_j$.

$\square$

$7.4 Technical Proofs when Loading Matrix B is Known

Next, we examine the scenario where we estimate $\boldsymbol{f}_i$ under the assumption that the loading matrix $\mathbf{B}$ is known. Assume that $\mathbf{B}$ is known, we estimate $\boldsymbol{f}_i$ by solving the optimization problem

$$\hat{\boldsymbol{f}}_i(\mathbf{B}) \in \arg\min_{\boldsymbol{f}\in\mathbb{R}^K} \sum_{j=1}^{d} \psi_{\varpi}(Z_{ij} - \hat{\mu}_j - \boldsymbol{b}_j^T \boldsymbol{f}), \tag{S7.15}$$

where $\hat{\mu}_j = \arg\min_{\theta\in\mathbb{R}} \sum_{i=1}^{n} \psi_{\varpi_j}(Z_{ij} - \theta)$, $\varpi = \varpi(n, d) > 0$ is a robustification parameter. The screening statistics based on $\hat{\boldsymbol{f}}(\mathbf{B})$ are denoted as $\hat{\delta}_{k,j}(\mathbf{B})$ and $\hat{\delta}_j(\mathbf{B})$. The subsets selected by these statistics are accordingly denoted as $\hat{\mathcal{A}}_{\tau,k}(\mathbf{B})$ and $\hat{\mathcal{A}}_{\tau}(\mathbf{B})$.

The following Lemma S14 reveals an exponential type deviation inequality for $\hat{\boldsymbol{f}}_i(\mathbf{B})$.

Lemma S13. *(Lemma C.3 in Fan et al., 2019) For every $1 \leq j \leq d$ and for any $t \geq 1$, the estimator $\hat{\mu}_j$ with $\varpi = a(n/t)^{1/2}$ and $a \geq \sigma_{jj}^{1/2}$ satisfies that as long as $n \geq 8t$,*

$$\left| \sqrt{n}(\hat{\mu}_j - \mu_j) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\varpi}(u_{ij}) \right| \leq C \frac{at}{\sqrt{n}} \tag{S7.16}$$

with probability greater than $1 - 3e^{-t}$, where $u_{ij} = Z_{ij} - \mu_j$ and $C > 0$ is an absolute constant, $\phi(u) = \min(|u|, \varpi) sign(u)$ is the derivative of $\psi_{\varpi}(u)$.

Lemma S14. *Under Condition (C8), for any $t > 0$, the estimator $\hat{\boldsymbol{f}}(\mathbf{B})$ given in (S7.15) with $\varpi = \varpi_0(d/t)^{1/2}$ for $\varpi_0 \geq \sigma_\varepsilon := \max_{1\leq j\leq d} \sigma_{\varepsilon,jj}^{1/2}$ satisfies that with probability greater* *than* $1-(2eK+1)e^{-t}$,

$$\|\hat{\boldsymbol{f}}_i(\mathbf{B})-\boldsymbol{f}_i\|_2 \leq C\varpi_0(Kt)^{1/2}d^{-1/2} \quad \text{and} \quad \max_{1\leq k\leq K}|\hat{f}_{ik}(\mathbf{B})-f_{ik}| \leq C\varpi_0 t^{1/2}d^{-1/2} \quad \text{(S7.17)}$$

as long as $d \geq \max\{\|\boldsymbol{\mu}\|_2^2/\sigma_\varepsilon^2, (\|\boldsymbol{\mu}\|_1/\sigma_\varepsilon)^2 t, K^2 t\}$, *where* $C>0$ *is constant depending on the constants in (C8).*

Proof of Lemma S14. To begin with, we introduce the following notation. Define the loss function $L_\varpi(\boldsymbol{\theta}) = d^{-1}\sum_{j=1}^d \psi_\varpi(Z_{ij}-\hat{\mu}_j-\boldsymbol{b}_j^T\boldsymbol{\theta})$ for $\boldsymbol{\theta}\in\mathbb{R}^K$, $\boldsymbol{\theta}^*=\boldsymbol{f}_i$ and $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^K} L_\varpi(\boldsymbol{\theta})$. Without loss of generality, we assume $\|\mathbf{B}\|_{\max}\leq 1$ for simplicity.

Define an intermediate estimator $\hat{\boldsymbol{\theta}}_\eta = \theta^* + \eta(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)$ such that $\|\hat{\boldsymbol{\theta}}_\eta-\boldsymbol{\theta}^*\|_2\leq r$ for some $r>0$ to be specified below (S7.24). We take $\eta=1$ if $\|\hat{\boldsymbol{\theta}}_\eta-\boldsymbol{\theta}^*\|_2\leq r$; otherwise, we choose $\eta\in(0,1)$ so that $\|\hat{\boldsymbol{\theta}}_\eta-\boldsymbol{\theta}^*\|_2 = r$. Then it follows from Lemma A.1 in Sun et al. (2020) that

$$\langle \nabla L_\varpi(\hat{\boldsymbol{\theta}}_\eta)-\nabla L_\varpi(\boldsymbol{\theta}^*), \hat{\boldsymbol{\theta}}_\eta-\boldsymbol{\theta}^*\rangle \leq \eta\langle \nabla L_\varpi(\hat{\boldsymbol{\theta}})-\nabla L_\varpi(\boldsymbol{\theta}^*), \hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\rangle, \quad \text{(S7.18)}$$

where $\nabla L_\varpi(\hat{\boldsymbol{\theta}})=0$ by the Karush-Kuhn-Tucker condition. By the mean value theorem for vector-valued functions, we have

$$\nabla L_\varpi(\hat{\boldsymbol{\theta}}_\eta)-\nabla L_\varpi(\boldsymbol{\theta}^*) = \int_0^1 \nabla^2 L_\varpi((1-t)\boldsymbol{\theta}^*+t\hat{\boldsymbol{\theta}}_\eta)dt(\hat{\boldsymbol{\theta}}_\eta-\boldsymbol{\theta}^*).$$

If, there exists some constant $a_{\min}>0$ such that

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^K:\|\boldsymbol{\theta}-\theta^*\|_2\leq r} \lambda_{\min}(\nabla L_\varpi(\boldsymbol{\theta})) \geq a_{\min}, \quad \text{(S7.19)}$$

then it follows $a_{\min}\|\hat{\boldsymbol{\theta}}_\eta-\theta^*\|_2^2 \leq -\eta\langle\nabla L_\varpi(\boldsymbol{\theta}^*), \hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\rangle \leq \|\nabla L_\varpi(\boldsymbol{\theta}^*)\|_2\|\hat{\boldsymbol{\theta}}_\eta-\boldsymbol{\theta}^*\|_2$ or equivalently,

$$a_{\min}\|\hat{\boldsymbol{\theta}}_\eta - \theta^*\|_2 \le \|\nabla L_\varpi(\boldsymbol{\theta}^*)\|_2, \tag{S7.20}$$

where $\nabla L_\varphi(\boldsymbol{\theta}^*) = -d^{-1}\sum_{j=1}^d \phi_\varpi(\mu_j + \varepsilon_{ij})\boldsymbol{b}_j$.

First we verify (S7.19). Write $S = d^{-1}\mathbf{B}\mathbf{B}$ and note that

$$\nabla^2 L_\varpi(\boldsymbol{\theta}) = \frac{1}{d}\sum_{j=1}^d \boldsymbol{b}_j\boldsymbol{b}_j^T I(|Z_{ij} - \boldsymbol{b}_j^T\boldsymbol{\theta}| \le \varpi),$$

where $Z_{ij} - \boldsymbol{b}_j^T\boldsymbol{\theta} = \boldsymbol{b}_j^T(\boldsymbol{\theta}^* - \boldsymbol{\theta}) + \mu_j + \varepsilon_{ij}$. Then for any $\boldsymbol{u} \in S^{K-1}$ and $\boldsymbol{\theta} \in \mathbb{R}^K$ satisfying $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2 \le r$,

$$\begin{aligned}
& \boldsymbol{u}^T\nabla^2 L_\varphi(\boldsymbol{\theta})\boldsymbol{u} \\
\ge\ & \boldsymbol{u}^T S\boldsymbol{u} - \frac{1}{d}(\boldsymbol{b}_j^T\boldsymbol{u})^2 I(|\varepsilon_{ij} + \mu_j| > \varpi/2) - \frac{1}{p}\sum_{j=1}^p (\boldsymbol{b}_j^T\boldsymbol{u})^2 I\{|\boldsymbol{b}_j^T(\boldsymbol{\theta}^* - \boldsymbol{\theta})| > \varpi/2\} \\
\ge\ & \boldsymbol{u}^T S\boldsymbol{u} - \max_{1\le j\le p}\|\boldsymbol{b}_j\|_2^2\left\{\frac{1}{d}\sum_{j=1}^d I(|\varepsilon_{ij} + \mu_j| > \varpi/2) + \frac{4}{\varpi^2}\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|_2^2\boldsymbol{u}^T S\boldsymbol{u}\right\}.
\end{aligned}$$

By assumption 1, $\lambda_{\min}(S) \ge c_l$ for some constant $c_l > 0$ and $\max_{1\le j\le d}\|\boldsymbol{b}_j\|_2^2 \le K$. Therefore, as long as $\varpi > 2r\sqrt{K}$ we have

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^K:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2\le r} \lambda_{\min}(\nabla^2 L_\varpi(\boldsymbol{\theta})) \ge (1 - 4\varpi^{-2}r^2K)c_l - \frac{K}{d}\sum_{j=1}^d I(|\varepsilon_{ij} + \mu_j| > \varpi/2). \tag{S7.21}$$

To bound the last term of the above inequality, by using Hoeffding's inequality, for any $t > 0$,

$$\frac{1}{d}\sum_{j=1}^d I(|\varepsilon_{ij} + \mu_j| > \varpi/2) \le \frac{1}{d}\sum_{j=1}^d P(|\varepsilon_{ij} + \mu_j| > \varpi/2) + \sqrt{\frac{t}{2d}}$$

with probability at least $1 - e^{-t}$. This, together with (S7.21) and the inequality

$$\frac{1}{d}\sum_{j=1}^d P(|\varepsilon_{ij} + \mu_j| > \varpi/2) \le \frac{4}{\varpi^2 d}\sum_{j=1}^d (\mu_j^2 + E\varepsilon_j^2) = 4\varpi^{-2}(d^{-1}\|\boldsymbol{\mu}\|_2^2 + \sigma_{\varepsilon,jj})$$

implies that, with probability greater than $1-e^{-t}$,

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^K:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2\le r}\lambda_{\min}(\nabla^2\Psi_\varpi(\boldsymbol{\theta}))\ge\frac{3}{4}c_l-K\sqrt{\frac{t}{2d}}-\frac{4K}{\varpi^2}(\frac{\|\boldsymbol{\mu}\|_2^2}{d}+\sigma_{\varepsilon,jj}) \tag{S7.22}$$

as long as $\varpi\ge 4r\sqrt{K}$.

Next we bound $\|\nabla^2\Psi_\varpi(\boldsymbol{\theta}^*)\|_2$. For each $1\le k\le K$, we write $\Phi_k=d^{-1}\sum_{j=1}^d\phi_{jk}:=d^{-1}\sum_{j=1}^d\varpi^{-1}\phi_\varpi(\mu_j+\varepsilon_{ij})b_{jk}$, such that $\|\nabla^2\Psi_\varpi(\boldsymbol{\theta}^*)\|_2\le\sqrt{K}\|\nabla^2\Psi_\varpi(\boldsymbol{\theta}^*)\|_\infty=\varpi\sqrt{K}\max_{1\le k\le K}|\Phi_k|$. Since for each $u\in\mathbb{R}$, $-\log(1-u+u^2)\le\varpi^{-1}\phi_\varpi(\varpi u)\le\log(1+u+u^2)$. After some simple algebra, we obtain that

$$\begin{aligned}e^{\phi_{jk}}&\le\{1+\varpi^{-1}(\mu_j+\varepsilon_{ij})+\varpi^{-2}(\mu_j+\varepsilon_{ij})^2\}^{b_{jk}I(b_{jk}\ge0)}\\&\quad+\{1-\varpi^{-1}(\mu_j+\varepsilon_{ij})+\varpi^{-2}(\mu_j+\varepsilon_{ij})^2\}^{-b_{jk}I(b_{jk}<0)}\\&\le 1+\varpi^{-1}(\mu_j+\varepsilon_{ij})+\varpi^{-2}(\mu_j+\varepsilon_{ij})^2.\end{aligned}$$

Taking expectation on both sides gives

$$E(e^{\phi_{jk}})\le 1+\varpi^{-1}|\mu_j|+\varpi^{-2}(\mu_j+\sigma_{\varepsilon,jj})^2.$$

Moreover, by independence and the inequality $1+t\le e^t$, we get

$$\begin{aligned}E(e^{d\Phi_k})&=\prod_{j=1}^d E(e^{\phi_{jk}})\le\exp\left\{\frac{1}{\varpi}\sum_{j=1}^d|\mu_j|+\frac{1}{\varpi^2}\sum_{j=1}^d(\mu_j^2+\sigma_{\varepsilon,jj})\right\}\\&\le\exp\left(\frac{\|\boldsymbol{\mu}\|_1}{\varpi}+\frac{\|\boldsymbol{\mu}\|_2^2}{\varphi^2}+\frac{\sigma_{\varepsilon,jj}d}{n\varpi^2}\right)\le\exp\left(\frac{\|\boldsymbol{\mu}\|_1}{\varpi}+\frac{\|\boldsymbol{\mu}\|_2^2}{\varphi^2}+\frac{\sigma_\varepsilon^2 d}{n\varpi^2}\right),\end{aligned}$$

where $\sigma_\varepsilon^2=\max_j(\sigma_{\varepsilon,jj})$. For any $t>0$, by Markov's inequality it follows that

$$P(d\Phi_k\ge 2t)\le e^{-2t}E(e^{d\Phi_k})\le\exp\left(\frac{\|\boldsymbol{\mu}\|_1}{\varpi}+\frac{\|\boldsymbol{\mu}\|_2^2}{\varphi^2}+\frac{\sigma_\varepsilon^2 d}{n\varpi^2}-2t\right)\le\exp(1-t),$$

if

$$\varpi \geq \max\left\{\|\boldsymbol{\mu}\|_1, \sigma_\varepsilon^2\sqrt{\frac{\|\boldsymbol{\mu}\|_2^2/\sigma_\varepsilon^2 + d/n}{t}}\right\}. \tag{S7.23}$$

Then it can be similarly proved that $P(-d\Phi_k \geq 2t) \leq e^{1-t}$. With all the above results, we conclude that

$$\begin{aligned} & P\left\{\|\nabla\Psi_\varpi(\boldsymbol{\theta}^*)\|_2 \geq \sqrt{K}\frac{2\varpi t}{d}\right\} \\ \leq \;& P\left\{\|\nabla\Psi_\varpi(\boldsymbol{\theta}^*)\|_\infty \geq \frac{2\varpi t}{d}\right\} \leq \sum_{k=1}^K P(|d\Phi_k| \geq 2t) \leq 2eK\exp(-t). \end{aligned} \tag{S7.24}$$

With the above preparation work, then we can prove the final conclusion. It follows (S7.22) that (S7.19) holds with $a_{\min} = c_l/4$ with probability greater than $1 - e^{-t}$, as long as $\varpi \geq 4\sqrt{K}\max\{r, c_l^{-1/2}(\|\boldsymbol{\mu}\|_2^2/d + \sigma_\varepsilon^2)^{1/2}\}$ and $d \geq 8c_l^{-2}K^2t$. Therefore, combining (S7.20) and (S7.24) with $r = \frac{\varpi}{4\sqrt{K}}$ yields that, with probability at least $1-(1+2eK)e^{-t}$, $\|\hat{\boldsymbol{\theta}}_\eta - \boldsymbol{\theta}^*\|_2 \leq 8c_l^{-1}\sqrt{K}d^{-1}\varpi t \leq r$ provided that $d > 32c_l^{-1}Kt$. By the definition of $\hat{\boldsymbol{\theta}}_\eta$, we must have $\eta = 1$ and hence $\hat{\boldsymbol{\theta}} = \hat{\theta}_\eta$, that is, $\|\hat{\boldsymbol{f}}_i(\mathbf{B}) - \boldsymbol{f}_i\|_2 \leq C\varpi_0(Kt)^{1/2}d^{-1/2}$ holds. Similarly, we have $\max_{1\leq k\leq K}|\hat{f}_{ik}(\mathbf{B}) - f_{ik}| \leq C\varpi_0 t^{1/2}d^{-1/2}$. This completes the proof of Lemma S14. □

Theorem S5. *Under Conditions (C1)-(C9), given a quantile level* $0 < \tau < 1$, *let* $\varpi = \varpi_0\{d/\log(n)\}^{1/2}$ *with* $\varpi_0 \geq \sigma_\varepsilon$, *we have*

(1) For any $C > 0$, *there exist positive constants* c_2'' *and* c_3'' *such that*

$$P\Big(\max_{1\leq j\leq p}|\hat{\delta}_j^\tau(\mathbf{B}) - \delta_j^\tau| \geq Cn^{-\kappa}\Big) \leq Kp\big\{\exp(-c_2''n^{1-4\kappa} + \log n) + \exp(-c_3''s_n^{-2}n^{1-2\kappa} + \log n)\big\} \tag{S7.25}$$

for all n *sufficiently large.*

(2) (Sure screening property) If $\kappa < 1/4$, $s_n^2 n^{2\kappa-1} = o(1)$, take the threshold $\nu_n = c^* n^{-\kappa}$ for some constant c^*, then

$$P(\mathcal{A}_\tau \subset \hat{\mathcal{A}}_\tau(\mathbf{B}, \nu_n)) \geq 1 - |\mathcal{A}_\tau| \big\{ \exp(-c_2'' n^{1-4\kappa} + \log n) + \exp(-c_3'' s_n^{-2} n^{1-2\kappa} + \log n) \big\}$$

for all n sufficiently large, where $|\mathcal{A}_\tau|$ is the cardinality of $\mathcal{A}_\tau$. Therefore, $P(\mathcal{A}_\tau \subset \hat{\mathcal{A}}_\tau(\mathbf{B})) \to 1$ as $n \to \infty$.

We will now prove Theorem S5 by beginning with the introduction of the following lemmas.

Lemma S15. *For any $\delta > 0$,*

$$P(\|\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}_{kj}^*\| \geq \delta) \leq P\Bigg(\sup_{\|\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}_{kj}^*\| \leq \delta} |B_n(\boldsymbol{\beta}) - B_h(\boldsymbol{\beta})| \geq \frac{1}{2} \inf_{\|\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}_{kj}^*\| = \delta} (B_h(\boldsymbol{\beta}) - B_h(\boldsymbol{\beta}_{kj}^*)) \Bigg). \tag{S7.26}$$

Lemma S15 is a direct application of Lemma 2 of Hjort and Pollard (2011) making use of the convexity of the objective function.

The lower bound of the right-hand side of (S7.26) can be explicitly evaluated for a given $\delta > 0$. This is specified in the following lemma.

Lemma S16. *Suppose that Conditions (C8)-(C9) hold. Let $C > 0$ be an arbitrary constant. Assume that $s_n^{-r} n^\kappa = o(1)$, $\varpi = \varpi_0 \{p/\log(n)\}^{1/2}$ then there exists a constant a_1 such that*

$$\inf_{\|\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}_{kj}^*\| = C s_n^{1/2}} \big(B_h(\boldsymbol{\beta}) - B_h(\boldsymbol{\beta}_{kj}^*)\big) \geq a_1 n^{-2\kappa},$$

for n sufficiently large.

Proof of Lemma S16. We consider $\boldsymbol{\beta} = \boldsymbol{\beta}^*_{kj} + Cs_n^{1/2}n^{-\kappa}\boldsymbol{u}$, where $\boldsymbol{u} \in \mathbb{R}^{s_n+l+1}$ satisfying $\|\boldsymbol{u}\| = 1$. We have

$$
\begin{aligned}
& B_h(\boldsymbol{\beta}) - B_h(\boldsymbol{\beta}^*_{kj}) = E\big\{\mathcal{L}_\tau(\hat{f}_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj} - Cs_n^{1/2}n^{-\kappa}\boldsymbol{u}) - \mathcal{L}_\tau(f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj})\big\} \\
= \;& E\big\{\mathcal{L}_\tau((\hat{f}_k - f_k) + f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj} - Cs_n^{1/2}n^{-\kappa}\boldsymbol{\pi}(X_j)^T\boldsymbol{u}) - \mathcal{L}_\tau(f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj})\big\} \\
= \;& E\big\{\mathcal{L}_\tau(f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj} + C\varpi_0 d^{-1/2}\log(n)^{1/2} - Cs_n^{1/2}n^{-\kappa}\boldsymbol{\pi}(X_j)^T\boldsymbol{u}) - \mathcal{L}_\tau(f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj})\big\} \\
= \;& E[-\nabla Q_{\tau,h}(\boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}^*_{kj})(Cs_n^{1/2}n^{-\kappa}\boldsymbol{u} - C\varpi_0 d^{-1/2}\log(n)^{1/2})] \\
& + E[\nabla^2 Q_{\tau,h}(\boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}^*_{kj})(Cs_n^{1/2}n^{-\kappa} - C\varpi_0 d^{-1/2}\log(n)^{1/2})^2\|\boldsymbol{u}\|] \\
= \;& (Cs_n^{1/2}n^{-\kappa} - C\varpi_0 d^{-1/2}\log(n)^{1/2})E[(\tau - \tilde{\mathcal{K}}_h(\epsilon_i(\boldsymbol{\beta}^*_{kj}))\boldsymbol{\pi}(X_j)^T\boldsymbol{u}] \\
& + E[\mathcal{K}_h(-\epsilon_i(\boldsymbol{\beta}^*_{kj}))\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T(Cs_n^{1/2}n^{-\kappa} - C\varpi_0 d^{-1/2}\log(n)^{1/2})^2] \\
= \;& J_1' + J_2',
\end{aligned}
$$

where $\epsilon_i(\boldsymbol{\beta}^*_{kj}) = f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}^*_{kj}$. By Hölder's inequality, we have

$$
\begin{aligned}
|J_1'| &\leq C|s_n^{1/2}n^{-\kappa} - \varpi_0 d^{-1/2}\log(n)^{1/2}|(E(\boldsymbol{\pi}(X_j)^T\boldsymbol{u})^2)^{1/2}[E(\tau - \tilde{\mathcal{K}}_h(\epsilon_i(\boldsymbol{\beta}^*_{kj})))^2]^{1/2} \\
&\leq Cs_n^{1/2}n^{-\kappa}O(s_n^{-1/2})O(s_n^{-r}) = O(s_n^{-r}n^{-\kappa}).
\end{aligned}
$$

The second inequality holds by the properties of spline basis and Condition (C9), see He et al. (2013) and Sherwood and Wang (2016). Similarly, by Condition (C9)

$$
J_2' = E[\mathcal{K}_h(-\epsilon_i(\boldsymbol{\beta}^*_{kj}))\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T(Cs_n^{1/2}n^{-\kappa})^2] = O(n^{-2\kappa}).
$$

Note that J_2' is nonnegative and $J_1' = o(J_2')$. This completes the proof of the lemma. □

Lemma S17. *Assume that $s_n n^{2\kappa-1} = o(1)$ and Conditions (C8)-(C9) hold. For any*

$C > 0$, there exist positive constants a_2 and a_3 such that for n sufficiently large

$$P(\|\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\| \geq C s_n^{1/2} n^{-\kappa}) \leq 2\exp(-a_2 n^{1-4\kappa}) + \exp(-a_3 s_n^{-2} n^{1-2\kappa}).$$

Proof of Lemma S17. Following Lemma S15 and Lemma S16, there exists some $c > 0$, for all n sufficiently large,

$$\begin{aligned}
& P(\|\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\| \geq C s_n^{1/2} n^{-\kappa}) \\
\leq\ & P\Big(\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj}\| \leq C s_n^{1/2} n^{-\kappa}} |B_n(\boldsymbol{\beta}) - B_h(\boldsymbol{\beta})| \geq c n^{-2\kappa} \Big) \\
\leq\ & P(|B_n(\boldsymbol{\beta}^*_{kj}) - B(\boldsymbol{\beta}^*_{kj})| \geq \frac{1}{2} c n^{-2\kappa}) \\
& + P\Big(\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj}\| \leq C s_n^{1/2} n^{-\kappa}} |B_n(\boldsymbol{\beta}) - B_n(\boldsymbol{\beta}^*_{kj}) - B_h(\boldsymbol{\beta}) + B_h(\boldsymbol{\beta}^*_{kj})| \geq c n^{-2\kappa} \Big) \\
=\ & L_1 + L_2.
\end{aligned}$$

First, we evaluate L_1. Let $W_i = \mathcal{L}_\tau(f_{ik} - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}^*_{kj})$. Then $B_n(\boldsymbol{\beta}^*_{kj}) - B_h(\boldsymbol{\beta}^*_{kj}) = n^{-1} \sum_{i=1}^n (W_i - EW_i)$. Note that $|W_i| \leq C |\boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}^*_{kj}|$ for some positive constant c. By the argument of Lemma 3.1 of He et al. (2013), $\sup_t |h_j(t) - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}_{0j}| \leq c^* s_n^{-r}$. It follows that $|W_i|$'s are uniformly bounded by a constant and $\text{var}(W_i) \leq \sigma^2$ for some $\sigma^2 > 0$. Applying Bernstein's inequality, we have

$$L_1 = P(|\sum_{i=1}^n W_i| \geq c n^{1-2\kappa}/2) \leq 2\exp\left(- \frac{c^2 n^{1-4\tau}/4}{2\sigma^2 + C c n^{-2\kappa}/3} \right) \leq 2\exp(-a_2 n^{1-4\kappa}).$$

Then we consider L_2. To evaluate L_2, we employ the Massart's concentration theorem. Let $V_i = \mathcal{L}_\tau(f_{ik} - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}) - \mathcal{L}_\tau(f_{ik} - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}^*_{kj})$. Then

$$V_i = \nabla Q_{\tau,h}(\epsilon_i(\boldsymbol{\beta}^*_{kj})) \boldsymbol{\pi}(X_{ij})^T (\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj}) + \nabla^2 Q_{\tau,h}(\epsilon_i(\boldsymbol{\beta}^*_{kj})) (\boldsymbol{\pi}(X_{ij})^T (\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj}))^2.$$

Thus, $|V_i| \le 2|\boldsymbol{\pi}(X_{ij})^T(\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj})| \le c s_n n^{-\kappa}$ for some $c > 0$.

Next, let $e_1, \ldots, e_n$ be a Rademacher sequence (i.e., iid sequence taking values of ± 1 with probability 1/2) independent of $V_1, \ldots, V_n$. We have

$$
\begin{aligned}
& E\Big\{\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \le C s_n^{1/2} n^{-\kappa}} n^{-1}\Big|\sum_{i=1}^n (V_i - E(V_i))\Big|\Big\} \le 2E\Big\{\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \le C s_n^{1/2} n^{-\kappa}} n^{-1}\Big|\sum_{i=1}^n e_i V_i\Big|\Big\} \\
\le \;& 4E\Big\{\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \le C s_n^{1/2} n^{-\kappa}} n^{-1}\Big|\sum_{i=1}^n e_i \boldsymbol{\pi}(X_{ij})^T(\boldsymbol{\beta} - \boldsymbol{\beta}^*_{kj}) V_i\Big|\Big\} \le 4C s_n^{1/2} n^{-\kappa} E\|n^{-1}\sum_{i=1}^n e_i \boldsymbol{\pi}(X_{ij})\| \\
\le \;& 4C s_n^{1/2} n^{-\kappa} [E\|n^{-1}\sum_{i=1}^n e_i \boldsymbol{\pi}(X_{ij})\|^2]^{1/2} = 4C s_n^{1/2} n^{-\kappa}\Big[n^{-2} E(\sum_{i=1}^n e_i^2 \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\pi}(X_{ij}))\Big]^{1/2} \\
\le \;& \tilde{C} s_n^{1/2} n^{-\kappa-1/2},
\end{aligned}
$$

for some constant $\tilde{C}$, where the first inequality holds by the symmetrization theorem (Lemma 2.3.1, Van Der Vaart and Wellner, 1996), the second inequality holds by employing the contraction theorem (Ledoux and Talagrand, 1991), and the last inequality holds since $E(b_k^2(X_{ij})) \le c_0 s_n^{-1}$ for $1 \le k \le s_n + l + 1$, $1 \le i \le n$ and $1 \le j \le p$ (Stone, 1985).

Finally, we apply Massart's concentration theorem to calculate L_2. Let

$$
A = \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^*_{kj}\| \le C s_n^{1/2} n^{-\kappa}} n^{-1}\Big|\sum_{i=1}^n (V_i - E(V_i))\Big|.
$$

Then

$$
L_2 = P(A \ge c n^{-2\kappa}/2) = P(Z \ge EZ + (c n^{-2\kappa}/2 - EZ)) \le \exp(-a_3 s_n^{-2} n^{1-2\kappa})
$$

for some positive constant a_3 and n sufficiently large. This completes the proof of Lemma S17. $\square$

Next, we evaluate the term $|\tilde{Q}_k(X_j) - Q_k(X_j)| = |\hat{Q}_\tau(\hat{f}_k|X_j) - Q_\tau(f_k|X_j)|$.

Lemma S18. *Under Conditions (C1)-(C7), we have*

$$P(|\tilde{Q}_k(x) - Q_k(x)| \geq Cn^{-\kappa}) \leq 2\exp(-a_2 n^{1-4\kappa}) + \exp(-a_3 s_n^{-2} n^{1-2\kappa}).$$

Proof of Lemma S18.

$$\begin{aligned} & |\tilde{Q}_k(X_j) - Q_k(X_j)| = |\boldsymbol{\pi}(X_j)^T \tilde{\boldsymbol{\beta}}_{kj} - Q_k(X_j)| \\ = \;& |\boldsymbol{\pi}(X_j)^T \tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\pi}(X_j)^T \boldsymbol{\beta}^*_{kj}| + |\boldsymbol{\pi}(X_j)^T \boldsymbol{\beta}^*_{kj} - \boldsymbol{\pi}(X_j)^T \boldsymbol{\beta}^0_{kj}| + |\boldsymbol{\pi}(X_j)^T \boldsymbol{\beta}^0_{kj} - Q_k(X_j)| \\ = \;& T_1' + T_2 + T_3. \end{aligned}$$

The terms T_2 and T_3 are the same as in Lemma S11.

$$T_1' = \{(\pi(X_j)^T(\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}))^2\}^{1/2} = \{(\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj})^T \boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T(\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj})\}^{1/2}.$$

Then $\sup_{x_j} T_1' \leq O_p(s_n^{-1/2})\|\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\|$ due to the fact that there exist two positive constants a_1^* and a_2^* such that $a_1^* s_n^{-1} \leq \lambda_{\min}(E\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)) \leq \lambda_{\max}(E\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)) \leq a_2^* s_n^{-1}$, $\forall j$, (Zhou et al., 1998). Hence

$$\begin{aligned} P(|\tilde{Q}_k(x) - Q_k(x)| \geq Cn^{-\kappa}) &\leq P(T_1' \geq Cn^{-\kappa}/3) + P(T_2 \geq Cn^{-\kappa}/3) + P(T_3 \geq Cn^{-\kappa}/3) \\ &\leq P(\|\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\| \geq C s_n^{1/2} n^{-\kappa}/3) \\ &\leq 2\exp(-a_2 n^{1-4\kappa}) + \exp(-a_3 s_n^{-2} n^{1-2\kappa}). \end{aligned}$$

The second inequality holds because $P(T_2 \geq Cn^{-\kappa}/3) = 0$, and $P(T_3 \geq Cn^{-\kappa}/3) = 0$, since $0 < \kappa < 1/2$ and $s_n^{-r} n^{\kappa} = o(1)$ by Condition (C5). □

With the above preparation work, next we prove Theorem S5.

Proof of Theorem S5. (1). By Lemma S18, and analogous to the proof of Theorem S3, it follows directly that

$$P(\max_{1\le j\le p}|\hat{\delta}_{k,j}(\mathbf{B})-\delta_{k,j}|\ge Cn^{-\kappa})\le O(pn)\exp(-c_2''n^{1-4\kappa})+O(pn)\exp(-c_3''s_n^{-2}n^{1-2\kappa}). \tag{S7.27}$$

$$\begin{aligned}P(\max_{1\le j\le p}|\hat{\delta}_{j}(\mathbf{B})-\delta_{j}|\ge Cn^{-\kappa}) &= P(\max_{1\le j\le p}\max_{1\le k\le K}|\hat{\delta}_{k,j}(\mathbf{B})-\delta_{k,j}|\ge Cn^{-\kappa})\\ &\le O(pnK)\exp(-c_2''n^{1-4\kappa})+O(pn)\exp(-c_3''s_n^{-2}n^{1-2\kappa}).\end{aligned}$$

(2). The results follow by using the first result of Theorem S5. If $\mathcal{A}_{\tau,k}\not\subset\hat{\mathcal{A}}_{\tau,k}(\mathbf{B})$, then there must exist some $j\in\mathcal{A}_{\tau,k}$ such that $\hat{\delta}_{kj}(\mathbf{B})<c_1n^{-\kappa}$. It follows from Condition (C4) that $|\hat{\delta}_{k,j}(\mathbf{B})-\delta_{k,j}|\ge c_1n^{-\kappa}$ for some $j\in\mathcal{A}_{\tau,k}$. Hence, $\{\max_{j\in\mathcal{A}_{\tau,k}}|\hat{\delta}_{k,j}(\mathbf{B})-\delta_{k,j}|\le c_1n^{-\kappa}\}\subset\{\mathcal{A}_{\tau,k}\subset\hat{\mathcal{A}}_{\tau,k}(\mathbf{B})\}$. Consequently,

$$\begin{aligned}P(\mathcal{A}_{\tau,k}\subset\hat{\mathcal{A}}_{\tau,k}(\mathbf{B})) &\ge P\{\max_{j\in\mathcal{A}_{\tau,k}}|\hat{\delta}_{k,j}(\mathbf{B})-\delta_{k,j}|\le c_1n^{-\kappa}\}=1-P\{\min_{j\in\mathcal{A}_{\tau,k}}|\hat{\delta}_{k,j}(\mathbf{B})-\delta_{k,j}|\le c_1n^{-\kappa}\}\\ &= 1-|\mathcal{A}_{\tau,k}|P\{|\hat{\delta}_{k,j}(\mathbf{B})-\delta_{k,j}|\le c_1n^{-\kappa}\}\\ &\le 1-O\big(|\mathcal{A}_{\tau,k}|\{\exp(-c_2''n^{1-4\kappa}+\log n)+\exp(-c_3''s_n^{-2}n^{1-2\kappa}+\log n)\}\big),\end{aligned}$$

Similarly, the result for Theorem S5(2) can be obtained. This completes the proof. □

$7.5 Proof of Theorem 2

We now proceed to prove Theorem 2 presented in the main text. To begin, we introduce Lemma S19 and Proposition S20.

Lemma S19. *Suppose Conditions (C8),(C10) and Condition (2.3) in the main text hold. Suppose that* $\varpi_j=t_j\sqrt{n/\log(nd)},\varpi_{jk}=t_{jk}\sqrt{n/\log(nd^2)}$ *with* $t_j\ge\sigma_{jj}^{1/2},t_{jk}\ge\operatorname{var}\left(Z_j^2\right)^{1/2}$

for $1 \leq j, k \leq d$. Then, there exist positive constants C_{B1} and C_{B2} independent of n and d such that as long as $n \geq C_{B1} \log(nd)$,

$$\max_{1 \leq j \leq d} \left\| \hat{\boldsymbol{b}}_j - \boldsymbol{b}_j \right\|_2 \leq C_{B2} \left\{ (n/\log(nd))^{-1/2} + d^{-1/2} \right\}$$

with probability greater than $1 - 4n^{-1}$, where $\hat{\boldsymbol{b}}_j$ and $\boldsymbol{b}_j$ are the jth row of $\hat{\mathbf{B}}$ and $\mathbf{B}$.

The proof of Lemma S19 can refer to online appendix in Fan et al. (2019).

Lemma S20. *Under identifiable condition (2.3), Conditions (C8) and (C10) of the main text, when $n, d \to \infty$ with $\log(nd)/n \to 0$ and $K/d \to 0$, for $\varpi \asymp \sqrt{d/\log n}$, the estimator $\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})$ given in Algorithm 2 satisfies that*

$$\left\| \hat{\boldsymbol{f}}_i(\hat{\mathbf{B}}) - \boldsymbol{f}_i \right\|_2 = o_P(1).$$

Proof of Lemma S20. Define the loss function

$$L\left(\boldsymbol{f}_i\right) = \frac{1}{d} \sum_{j=1}^{d} \psi_{\varpi} \left(Z_{ij} - \hat{\mu}_j - \widehat{\boldsymbol{b}}_j^T \boldsymbol{f}_i \right),$$

The estimator $\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})$ satisfies the following gradient condition

$$\nabla L\left(\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})\right) = -\frac{1}{d} \sum_{j=1}^{d} \ell_{\gamma} \left(Z_{ij} - \hat{\mu}_j - \widehat{\boldsymbol{b}}_j^T \hat{\boldsymbol{f}}_i(\hat{\mathbf{B}}) \right) \widehat{\boldsymbol{b}}_j = 0,$$

where $\ell_{\gamma}(u) = \min(\varpi, |u|) \cdot \operatorname{sign}(u)$ is the derivative of the Huber loss. The expression within the parentheses can be expanded as

$$Z_{ij} - \hat{\mu}_j - \widehat{\boldsymbol{b}}_j^T \hat{\boldsymbol{f}}_i(\hat{\mathbf{B}}) = \underbrace{(\mu_j - \hat{\mu}_j)}_{\Delta\mu_j} + \underbrace{\left(\boldsymbol{b}_j - \widehat{\boldsymbol{b}}_j\right)^T \boldsymbol{f}_i}_{\Delta \boldsymbol{b}_j^T \boldsymbol{f}_i} + \varepsilon_{ij} - \widehat{\boldsymbol{b}}_j^T \left(\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}}) - \boldsymbol{f}_i \right).$$

Note that the Huber estimator $\hat{\mu}_j$'s satisfy $\max_{1\le j\le d}|\hat{\mu}_j-\mu_j| = O_P(\sqrt{\log(nd)/n})$. Substitute the expanded residuals into the gradient condition, we have

$$\frac{1}{d}\sum_{j=1}^{d}\ell_{\varpi}\left(\Delta\mu_j+\Delta\boldsymbol{b}_j^T\boldsymbol{f}_i+\varepsilon_{ij}-\widehat{\boldsymbol{b}}_j^T\left(\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})-\boldsymbol{f}_i\right)\right)\widehat{\boldsymbol{b}}_j=0.$$

Assume $\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})$ is close to $\boldsymbol{f}_i$ and expand ℓ_{ϖ} around $\boldsymbol{f}_i$ to first order

$$\ell_{\varpi}\left(r_{ij}-\widehat{\boldsymbol{b}}_j^T\delta_i\right)\approx\ell_{\varpi}\left(r_{ij}\right)-\ell'_{\varpi}\left(r_{ij}\right)\widehat{\boldsymbol{b}}_j^T\delta_i.$$

where $r_{ij}=\Delta\mu_j+\Delta\boldsymbol{b}_j^T\boldsymbol{f}_i+\varepsilon_{ij},\delta_i=\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})-\boldsymbol{f}_i$, and ℓ'_{ϖ} equals 1 if $|r_{ij}|\le\varpi$ and 0 otherwise. Substituting the above result into the gradient condition, we have

$$\frac{1}{d}\sum_{j=1}^{d}\left[\ell_{\varpi}\left(r_{ij}\right)-\ell'_{\varpi}\left(r_{ij}\right)\widehat{\boldsymbol{b}}_j^T\delta_i\right]\widehat{\boldsymbol{b}}_j\approx 0,$$

By Rearranging we can yield a linear equation

$$\left(\frac{1}{d}\sum_{j=1}^{d}\ell'_{\varpi}\left(r_{ij}\right)\widehat{\boldsymbol{b}}_j\widehat{\boldsymbol{b}}_j^T\right)\delta_i\approx\frac{1}{d}\sum_{j=1}^{d}\ell_{\varpi}\left(r_{ij}\right)\widehat{\boldsymbol{b}}_j.$$

Define the matrix

$$\widehat{\mathbf{H}}_i=\frac{1}{d}\sum_{j=1}^{d}\ell'_{\varpi}\left(r_{ij}\right)\widehat{\boldsymbol{b}}_j\widehat{\boldsymbol{b}}_j^T,$$

Assume $\ell'_{\varpi}\left(r_{ij}\right)\ge c>0$ (when residuals lie mainly in the quadratic region) and $\lambda_{\min}\left(\frac{1}{d}\sum_{j=1}^{d}\boldsymbol{b}_j\boldsymbol{b}_j^T\right)\ge c_l>0$ (assumption (C8) in the main text). According to the consistency results of $\left\|\widehat{\boldsymbol{b}}_j-\boldsymbol{b}_j\right\|_2$ in lemma S19, we have,

$$\lambda_{\min}\left(\widehat{\mathbf{H}}_i\right)\ge\frac{c_l}{2}.$$

As

$$\left\|\frac{1}{d}\sum_{j=1}^{d}\ell_{\varpi}\left(r_{ij}\right)\widehat{\boldsymbol{b}}_j\right\|_2\le\frac{1}{d}\sum_{j=1}^{p}\left|\ell_{\varpi}\left(r_{ij}\right)\right|\cdot\left\|\widehat{\boldsymbol{b}}_j\right\|_2,$$

Since $|\ell_{\varpi}(r)| \leq \varpi$ and $\left\|\hat{\boldsymbol{b}}_j\right\|_2 \leq C$ (according to the result of lemma S19), applying Hoeffding's inequality gives

$$\left\|\frac{1}{d}\sum_{j=1}^{d}\ell_{\varpi}\left(r_{ij}\right)\widehat{\boldsymbol{b}}_j\right\|_2 = O_P\left(\varpi\sqrt{\frac{K\log d}{d}}\right).$$

For term $\Delta\mu_j$, we have $\max_j|\Delta\mu_j| = O_P(\sqrt{\log(nd)/n})$.

For term $\Delta\boldsymbol{b}_j^T\boldsymbol{f}_i$, we have by Cauchy-Schwarz inequality, $\max_j\left|\Delta\boldsymbol{b}_j^T\boldsymbol{f}_i\right| \leq \|\boldsymbol{f}_i\|_2 \cdot \max_j\|\Delta\boldsymbol{b}_j\|_2 = O_P\left\{\sqrt{K}\left((n/log(nd))^{-1/2} + d^{-1/2}\right)\right\}$.

Thus we have

$$\|\delta_i\|_2 = \left\|\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}}) - \boldsymbol{f}_i\right\|_2 \leq \frac{2}{c_l}\cdot O_P\left\{\varpi\sqrt{\frac{K\log d}{d}} + \sqrt{\frac{\log(nd)}{n}} + \sqrt{K}\left((n/log(nd))^{-1/2} + d^{-1/2}\right)\right\}.$$

Considering $\varpi \asymp \sqrt{d/\log n}$ and assuming $n, d \to \infty$ with $\log(nd)/n \to 0$ and $K/d \to 0$, we have

$$\left\|\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}}) - \boldsymbol{f}_i\right\|_2 = o_P(1).$$

This completes the proof of Lemma S20. □

Proof of Theorem 2. Based on Lemma S20, the proof of Theorem 2 follows a similar approach to that of Theorem S4 and Theorem S5 presented above, therefore, we omit the details here. □

The convergence property and sure screening property results presented in Theorem 2 are identical to those in Theorem S3. This similarity arises from the fact that the crucial step in proving these properties is the convergence between $\tilde{\boldsymbol{\beta}}_{kj}$ and $\boldsymbol{\beta}^*_{kj}$, where $\tilde{\boldsymbol{\beta}}_{kj} = \arg\min_{\boldsymbol{\beta}} n^{-1}\sum_{i=1}^{n}\mathcal{L}_\tau(\hat{f}_{ik} - \boldsymbol{\pi}(X_{ij}^T\boldsymbol{\beta}))$ and $\boldsymbol{\beta}^*_{kj} = \arg\min_{\boldsymbol{\beta}} E(\mathcal{L}_\tau(f_k - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta})$. Based on the consistency results between $\hat{\boldsymbol{f}}_i$ and $\boldsymbol{f}_i$ in Lemma S20, the convergence property in the Euclidean norm of $\|\tilde{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\|$ is identical to that of $\|\hat{\boldsymbol{\beta}}_{kj} - \boldsymbol{\beta}^*_{kj}\|$, where $\hat{\boldsymbol{\beta}}_{kj} = \arg\min_{\boldsymbol{\beta}} n^{-1} \sum_{i=1}^{n} \mathcal{L}_\tau(f_{ik} - \boldsymbol{\pi}(X_{ij}^T \boldsymbol{\beta}))$. Consequently, the result in Theorem 2 is consistent with the result in Theorem S3. Additional details can be found in Lemmas S15-S18 of the Supplementary Materials. Therefore, similar to the result under Theorem S3, it can be concluded that MSDS can handle the NP-dimensionality $\log p = o(n^{1-4\kappa} + s_n^{-2} n^{1-2\kappa})$ for unknown $\boldsymbol{f}$.

\$7.6 Proofs of Theorem S1 and Theorem 3

In this section, we present the theoretical proofs for Theorem S1 and Theorem 3, focusing on the control of the FDR, per family error rate, and k family-wise error rate.

Lemma S21. *For a given $\tau (0 < \tau < 1)$, let $\tilde{\mathbf{X}}$ be an exact knockoff copy of $\mathbf{X}$, and $\mathcal{A}_\tau^c = \{j_1, \ldots, j_r\}$. Then*

(i) $W_{j_k,\tau} = 0$ for all $j_k \in \mathcal{A}_\tau^c$.

(ii) Conditioning on $|\widehat{\mathbf{W}}_\tau| = (|\hat{W}_{1,\tau}|, \ldots, |\hat{W}_{p,\tau}|)^T$, $I_{j_1}, \ldots, I_{j_r}$ follow i.i.d. Bernoulli(0.5), where $I_{j_k} = 1$ if $\hat{W}_{j_k,\tau} > 0$ and 0 otherwise.

Hereafter, we denote $\mathbf{W} = \mathbf{W}_\tau$, $W_j = W_{j,\tau}$ for $j = 1, \ldots, p$.

Proof of Lemma S21. (i) For any j, let $(\mathbf{X}, \tilde{\mathbf{X}})_{(j)}$ be a vector by swapping the entries X_j and $\tilde{X}_j$ in $(\mathbf{X}, \tilde{\mathbf{X}})$. For any $S \subset \{1, \ldots, p\}$, let $(\mathbf{X}, \tilde{\mathbf{X}})_S$ be a vector by swapping the entries X_j and $\tilde{X}_j$ in $(\mathbf{X}, \tilde{\mathbf{X}})$ for all $j \in S$. Now consider $j \in \mathcal{A}_\tau^c$, denote $g_{\boldsymbol{f}|x}(\cdot|x)$ the conditional density of $\boldsymbol{f}$, then

$$g_{\boldsymbol{f}|(\mathbf{X},\tilde{\mathbf{X}})_{(j)}}(v|(u,\tilde{u})) = g_{\boldsymbol{f}|(\mathbf{X},\tilde{\mathbf{X}})}(v|(u,\tilde{u})_{(j)}) = g_{\boldsymbol{f}|x}(y|u'), \tag{S7.28}$$

where $u' = (u'_1, \dots, u'_p)$ is the first p elements in $(u,\tilde{u})_{(j)}$, i.e., $u'_k = u_k$ if $k \neq j$ and $u'_k = \tilde{u}_k$ if $k = j$. The second equality holds because $\boldsymbol{f}$ is independent of the knockoff copy $\tilde{\mathbf{X}}$. Let $\mathbf{X}_{(-j)}$ be a vector excluding the element X_j in $\mathbf{X}$. From the definition of active features, we know $\boldsymbol{f}$ is independent of X_j given $\mathbf{X}_{(-j)}$. We have

$$\begin{aligned}
g_{\boldsymbol{f}|\boldsymbol{x}}(y|u') &= g_{\boldsymbol{f}|\boldsymbol{x}}(v|u'_1, \dots, u'_p) \\
&= g_{\boldsymbol{f}|\boldsymbol{x}}(v|u_1, \dots, u_{j-1}, \tilde{u}_j, u_{j+1}, \dots, u_p) \\
&= g_{\boldsymbol{f}|\boldsymbol{x}_{(-j)}}(v|u_{(-j)}) \\
&= g_{\boldsymbol{f}|\boldsymbol{x}}(v|u) \\
&= g_{\boldsymbol{f}|(\boldsymbol{x},\tilde{\boldsymbol{x}})}(v|(u,\tilde{u}))
\end{aligned}$$

These equations together with (S7.28) implies that

$$g_{\boldsymbol{f}|(\mathbf{X},\tilde{\mathbf{X}})_{(j)}}(v|(u,\tilde{u})) = g_{\boldsymbol{f}|(\mathbf{X},\tilde{\mathbf{X}})}(v|(u,\tilde{u})).$$

This shows that

$$\boldsymbol{f}|(\mathbf{X},\tilde{\mathbf{X}})_{(j)} \stackrel{d}{=} \boldsymbol{f}|(\mathbf{X},\tilde{\mathbf{X}}).$$

By the definition of knockoffs, we know $(\mathbf{X},\tilde{\mathbf{X}})_{(j)} \stackrel{d}{=} (\mathbf{X},\tilde{\mathbf{X}})$. As a result, we have

$$(\boldsymbol{f}, (\mathbf{X},\tilde{\mathbf{X}})_{(j)}) \stackrel{d}{=} (\boldsymbol{f}, (\mathbf{X},\tilde{\mathbf{X}})),$$

which suggests that $(\boldsymbol{f}, X_j) \stackrel{d}{=} (\boldsymbol{f}, \tilde{X}_j)$. Hence, $\mathrm{MSD}_\tau(\boldsymbol{f}|X_j) = \mathrm{MSD}_\tau(\boldsymbol{f}|\tilde{X}_j)$ and $W_{j,\tau} = 0$.

(ii) Let $\widehat{\mathbf{W}} = (\hat{W}_1, \ldots, \hat{W}_p)^T$. Let $m(\cdot) : \mathbb{R}^{2p+K} \to \mathbb{R}^p$ be a function such that

$$\hat{\mathbf{W}} = m((\mathbf{X}, \tilde{\mathbf{X}}), \boldsymbol{f}).$$

We define $\epsilon_1, \ldots, \epsilon_p$ such that $\epsilon_j = 1$ for $j \in \mathcal{A}_\tau$ and $\epsilon_j \overset{iid}{\sim} \{+1, -1\}$ for $j \in \mathcal{A}_\tau^c$. By repeating the arguments in part (i), we can show that

$$(\boldsymbol{f}, (\mathbf{X}, \tilde{\mathbf{X}})) \overset{d}{=} (\boldsymbol{f}, (\mathbf{X}, \tilde{\mathbf{X}})_S) \quad \text{for any} \quad S \subset \mathcal{A}_\tau^c.$$

Now let $S = \{j : \epsilon_j = -1\}$, a subset of $\mathcal{A}^c$, then we observe that

$$(\hat{W}_1, \ldots, \hat{W}_p) = m((\mathbf{X}, \tilde{\mathbf{X}}), \boldsymbol{f}) \overset{d}{=} m((\mathbf{X}, \tilde{\mathbf{X}})_S, \boldsymbol{f}) = (\epsilon_1 \hat{W}_1, \ldots, \epsilon_p \hat{W}_p).$$

Hence $(\hat{W}_1, \ldots, \hat{W}_p) \overset{d}{=} (\epsilon_1 \hat{W}_1, \ldots, \epsilon_p \hat{W}_p)$. This completes the proof. □

Proof of Theorem S1. Without loss of generality, we assume $|\widehat{W_1}| \geq |\widehat{W_2}| \geq \cdots \geq |\widehat{W_d}| > 0$ and $|\widehat{W_{d+1}}| = \cdots = |\hat{W}_p| = 0$. Observe that

$$\begin{aligned}
\mathrm{FDR}_\tau &= E\left[\frac{|\hat{\mathcal{A}}_\tau(\eta_\alpha) \cap \mathcal{A}_\tau^c|}{|\hat{\mathcal{A}}_\tau(\eta_\alpha)| \vee 1}\right] = E\left[\frac{\#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \geq \eta_\alpha\}}{\#\{j : j \in \hat{\mathcal{A}}(\eta_\alpha)\} \vee 1}\right] \\
&= E\left[\frac{\#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \geq \eta_\alpha\}}{1 + \#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \leq -\eta_\alpha\}} \frac{1 + \#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \leq -\eta_\alpha\}}{\#\{j : j \in \hat{\mathcal{A}}(\eta_\alpha)\} \vee 1}\right] \\
&\leq E\left[\frac{1 + \#\{j : \hat{W}_j \leq -\eta_\alpha\}}{\#\{j : \hat{W}_j \geq \eta_\alpha\} \vee 1} \frac{\#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \geq \eta_\alpha\}}{1 + \#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \leq -\eta_\alpha\}}\right] \\
&\leq E\left[\alpha \frac{\#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \geq \eta_\alpha\}}{1 + \#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \leq -\eta_\alpha\}}\right]. \qquad \text{(S7.29)}
\end{aligned}$$

In order to find η_α, one can simply try different values of t starting from the smallest value, say $t = |\hat{W}_{d+1}| = 0$, then move to the second smallest value $t = |\hat{W}_d|$, then move to $t = |\hat{W}_{d-1}|$, and so on. This process terminates only if it finds a value t satisfying

(2.10). Thus, η_α can be regarded as a stopping time. More rigorously, we define, for $k = d+1, d, d-1, \dots, 1$

$$\begin{aligned} N_\tau(k) &= \frac{\#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \geq |\hat{W}_k|\}}{1 + \#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \leq -|\hat{W}_k|\}} = \frac{\#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \geq 0, j \leq k\}}{1 + \#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \leq 0, j \leq k\}} \\ &:= \frac{M_\tau^+(k)}{1 + M_\tau^-(k)}. \end{aligned}$$

Let $\mathcal{F}_k$ be the σ-algebra generated by $\{M_\tau^\pm(d+1), M_\tau^\pm(d), \dots, M_\tau^\pm(k), I_{d+1}, I_d, \dots, I_k\}$, where $I_{d+1} = 0$ and $I_j = 1$ if $j \in \mathcal{A}_\tau$ and 0 otherwise. Consequently, we have the idea whether k is in the active set $\mathcal{A}_\tau$ or not given $\mathcal{F}_k$.

Next we show that the process $N_\tau(d+1), N_\tau(d), \dots, N_\tau(1)$ is a supermartingale running backward in terms of $\mathcal{F}_{d+1} \subset \mathcal{F}_d \subset \cdots \subset \mathcal{F}_1$. If $k \in \mathcal{A}_\tau$, we have $M_\tau^+(k) = M_\tau^+(k-1)$, $M_\tau^-(k) = M_\tau^-(k-1)$ and it follows that $N_\tau(k) = N_\tau(k-1)$. If $k \in \mathcal{A}_\tau^c$, then

$$N_\tau(k-1) = \frac{M_\tau^+(k) - I(\hat{W}_k \geq 0)}{1 + M_\tau^-(k) - I(\hat{W}_k \leq 0)} = \frac{M_\tau^+(k) - I(\hat{W}_k \geq 0)}{(M_\tau^-(k) + I(\hat{W}_k \geq 0)) \vee 1}.$$

Let $d_0 = \#\{j : j \in \mathcal{A}_\tau^c\}$. From the result (ii) of Lemma S21, $\{j \in \mathcal{A}_\tau^c : I(\hat{W}_j \geq 0)\}$ are iid Bernoulli(0.5) random variables. Thus conditional on $\mathcal{F}_k$, we have

$$P(\hat{W}_k \geq 0 | \mathcal{F}_k) = P(\hat{W}_k \geq 0 | M_\tau^+(k), M_\tau^-(k)) = \frac{M_\tau^+(k)}{M_\tau^+(k) + M_\tau^-(k)}.$$

Thus in the case where $k \in \mathcal{A}_\tau^c$,

$$\begin{aligned}
E[N_\tau(k-1)|\mathcal{F}_k] &= \frac{M_\tau^+(k)}{M_\tau^+(k)+M_\tau^-(k)}\frac{M_\tau^+(k)-1}{M_\tau^-(k)+1} + \frac{M_\tau^-(k)}{M_\tau^+(k)+M_\tau^-(k)}\frac{M_\tau^+(k)}{M_\tau^-(k)\vee 1} \\
&= \begin{cases} \frac{M_\tau^+(k)}{M_\tau^-(k)+1}, & \text{if } M_\tau^-(k)>0, \\ M_\tau^+(k)-1, & \text{if } M_\tau^-(k)=0. \end{cases} \\
&= \begin{cases} N(k), & \text{if } M_\tau^-(k)>0, \\ N(k)-1, & \text{if } M_\tau^-(k)=0. \end{cases}
\end{aligned}$$

Thus, $E[N(k-1)|\mathcal{F}_k] \leq N(k)$, implying that $N(k)$, $k = d+1, d, \ldots, 1$ is a supermartingale with respect to $\{\mathcal{F}_k\}$. By the optional stopping theorem for supermartingale, we can obtain that

$$E[N(k_{\eta_\alpha})] \leq E[N(k_{d+1})] = E\left[\frac{\#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \geq 0\}}{1+\#\{j : j \in \mathcal{A}_\tau^c \text{ and } \hat{W}_j \leq 0\}}\right] = E[\frac{Z}{1+d_0-Z}],$$

where $Z = \#\{j : j \in \mathcal{A}_\tau^c, \hat{W}_j > 0\}$. Since $Z \sim Binomial(d_0, 0.5)$, we have

$$\begin{aligned}
E\left[\frac{Z}{1+d_0-Z}\right] &= \sum_{l=1}^{d_0} \binom{d_0}{l} 0.5^l 0.5^{d_0-l} \frac{l}{1+d_0-l} \\
&= \sum_{l=1}^{d_0} \binom{d_0}{l-1} 0.5^l 0.5^{d_0-l} \\
&= \sum_{l=0}^{d_0-1} \binom{d_0}{l} 0.5^{l+1} 0.5^{d_0-l-1} \\
&\leq 1.
\end{aligned}$$

Therefore, $E[N(k_{\eta_\alpha})] \leq 1$. Thus, it follows that

$$\mathrm{FDR}_\tau = E\left[\frac{|\hat{\mathcal{A}}_\tau(\eta_\alpha) \cap \mathcal{A}_\tau^c|}{|\hat{\mathcal{A}}_\tau(\eta_\alpha)| \vee 1}\right] \leq \alpha E[N(k_{\eta_\alpha})] \leq \alpha.$$

□

Proof of Theorem 3. (i) By Markov inequality, we have $E[V(\tau)] \le v/\eta_0$ directly. Then it follows that

$$E[V(\tau)] = E\Bigg[\sum_{j\in\mathcal{A}_\tau^c} I(\Pi_{j,\tau} \ge \eta_0)\Bigg] = \sum_{j\in\mathcal{A}_\tau^c} P(\Pi_{j,\tau} \ge \eta_0) \le \sum_{j\in\mathcal{A}_\tau^c} \gamma E[\Pi_{j,\tau}] = \gamma E[V_1(\tau)] \le \gamma v,$$

where $V_1(\tau)$ denotes the number of false discoveries in $\hat{S}_\tau^1(v)$ defined in Algorithm 1.

(ii) The proof of this the k-FWER control is straightforward by using the first result, since we have

$$P(V(\tau) \ge k) \le \frac{\ell E[V(\tau)]}{k} \le \frac{\ell\gamma v}{k}.$$

When the pmf of $V(\tau)$ is skewed to the left of k in the sense that

$$\sum_{u=1}^{k-1} P(V(\tau) \in [k-u, k]) \ge \sum_{u=1}^{k} P(V(\tau) \in [k, k+u]),$$

then $P(V(\tau) \ge k) \le \ell E[V(\tau)]/k$ always holds with $l = 1/2$.

In particular, by Markov's inequality, we have $l = 1$ and consequently,

$$P(V(\tau) \ge k) \le \frac{\gamma v}{k}.$$

□

$8 Additional Simulation Examples

Example S1: serial dependent factor-nonparametric model. Consider a similar joint model as in example 1, except that f_{ik}'s are generated from a stationary VAR(1) model, that is for $i \in [n]$

$$
\begin{aligned}
f_{i1} &= \boldsymbol{\gamma}_1^T f_{i-1} + (1 + X_{i,1})^2 + \xi_{i1}, \\
f_{i2} &= \boldsymbol{\gamma}_2^T f_{i-1} + X_{i3}^3 + \xi_{i2}, \\
f_{i3} &= \boldsymbol{\gamma}_3^T f_{i-1} + \exp(1 + X_{i5}) + \xi_{i3}.
\end{aligned}
$$

We set $\boldsymbol{f}_0 = 0$, and $\Gamma = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$. The (j,k)th entry of Γ is set to be 0.4 when $j = k$ and 0.3 otherwise. We consider the following cases: Case (3a): $\xi_{ik} \stackrel{iid}{\sim} N(0,1)$ for $k = 1,2,3$; Case (3b): $\xi_{i,k} = \left(\sqrt{1 + \xi_{i-1,k}^2}\right)\zeta_{ik}$, with $\boldsymbol{\xi}_0 = 0$ and $\zeta_{ik} \stackrel{iid}{\sim} N(0,1)$ for $k = 1,2,3$.

In all the settings of example S1, X_1, X_3 and X_5 are active covariates. Simulation results are provided in Tables S1 and S6 of the Supplementary Materials. All screening methods perform well when the covariates follow a multivariate normal distribution, while NIS, DCSIS and RVSIS fail to screen out inactive covariates in the presence of heavy-tailed covariates.

Example S2: factor-additive model. Consider a similar three factor model as in example 1, except that μ_j's are randomly sampled from $\{0, 0.5, 0.8\}$ for $1 \le j \le d$, $\boldsymbol{\varepsilon}_i \sim t_3(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ has the same form as in example 1, and f_{ik}'s are assumed to follow nonlinear additive models

$$
\begin{aligned}
f_{i1} &= 2(1 + X_{i1})^2 + 2X_{i2}^2 + 2\exp(1 + X_{i3}) + \xi_{i1}, \\
f_{i2} &= \exp(2X_{i2}) + (1 - X_{i4})^3 + \xi_{i2}, \\
f_{i3} &= 3X_{i3}^2 I(X_{i3} > 0) + 3/X_{i4} + 4X_{i5}^2 + \xi_{i3}.
\end{aligned}
$$

In this example, we also consider $\mathbf{X}_i$ drawn from one of the two distributions: $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$

and $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{X}})$ with $\boldsymbol{\Sigma}_{\mathbf{X}} = (0.5^{|i-j|})_{1\le i,j\le p}$. The errors $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \xi_{i3})^T$ in this model are generated from one of the following three scenarios as in example 1: (1)$\xi_{ik} \overset{iid}{\sim} N(0,1)$ for $k = 1,2,3$; (2) $\xi_{ik} \overset{iid}{\sim} t_3$ for $k = 1,2,3$; (3) heteroscedastic error, $\xi_{ik} = \exp(\sum_{j=1}^{k} X_{7+j})\zeta_{ik}$, where $\zeta_{ik} \overset{iid}{\sim} N(0, 0.7^2)$ for $k = 1,2,3$. Simulation results are shown in following table S2-S3.

For the factor-additive model, when the covariates are functionally related to the latent factors, all the other six methods completely fail. In contrast, MSDS works reasonably well in all the scenarios and outperforms the other six methods in both MMS and screening probability.

In terms of determining the number of latent factors, K, as our proposed method is a bootstrap-based eigenvalue method, we denote the proposed estimation method "BE". We compare BE with four other approaches mentioned in Section 2.5: ON (Onatski, 2009), ER (Wang, 2012), GR (Ahn and Horenstein, 2013) and ACT (Fan et al., 2022). As detailed in Tables S6-S7, methods ER, ON, and GR perform poorly in estimating the number of latent factors, as they significantly overestimate this number. Our method outperforms these competing methods across all settings.

$9 Additional Simulation Results for Screening of the Main Text

In this simulation section, we mainly provide the simulation results for examples 1-2 of the main documents in Tables S4-S5 and Tables S7-S8. These include the results for screening and estimating the number of latent factors.

$10 Simulation Results for FDR Control

This section presents the simulation results for examples 3 and 4 in Tables S9-S10 , which are used to evaluate the FDR control of the main text. For both examples 3 and 4, we consider $\tau \in \{0.50, 0.75, 0.90\}$. To evaluate the FDR control of MSDK procedure, we take $\alpha \in \{0.15, 0.2, 0.3\}$. To further reduce the variability of MSDK procedure and yield a more stable result, we employ DMSDK procedure and take $T = 50, \eta = 0.5$, and $v \in \{1, 2, 3\}$. Thus by Theorem 3, the PFER can be controlled at γv, that is, $E[V(\tau)] \leq \gamma v$. Referring to Ren et al. (2023), $\gamma \approx 1.02$ is obtained via linear programming. We also evaluate k-FWER control at nominal level 0.2, similarly by Theorem 3, it suffices to control $v \leq 0.4k/\gamma$ if we take $\ell = 1/2$, we take $k \in \{3, 4, 5\}$ for contrast analysis. In this scenario, $\gamma \approx 1.02$ can also be computed. Besides, when v is not an integer, let $\lfloor v \rfloor$ be the integer part of v and sample a random variable $U \sim Bernoulli(v - \lfloor v \rfloor)$. If $U = 1$, run the $(\lfloor v \rfloor + 1)$-knockoffs and $\lfloor v \rfloor$-knockoffs otherwise.

$11 Additional Results for Real Data Analysis

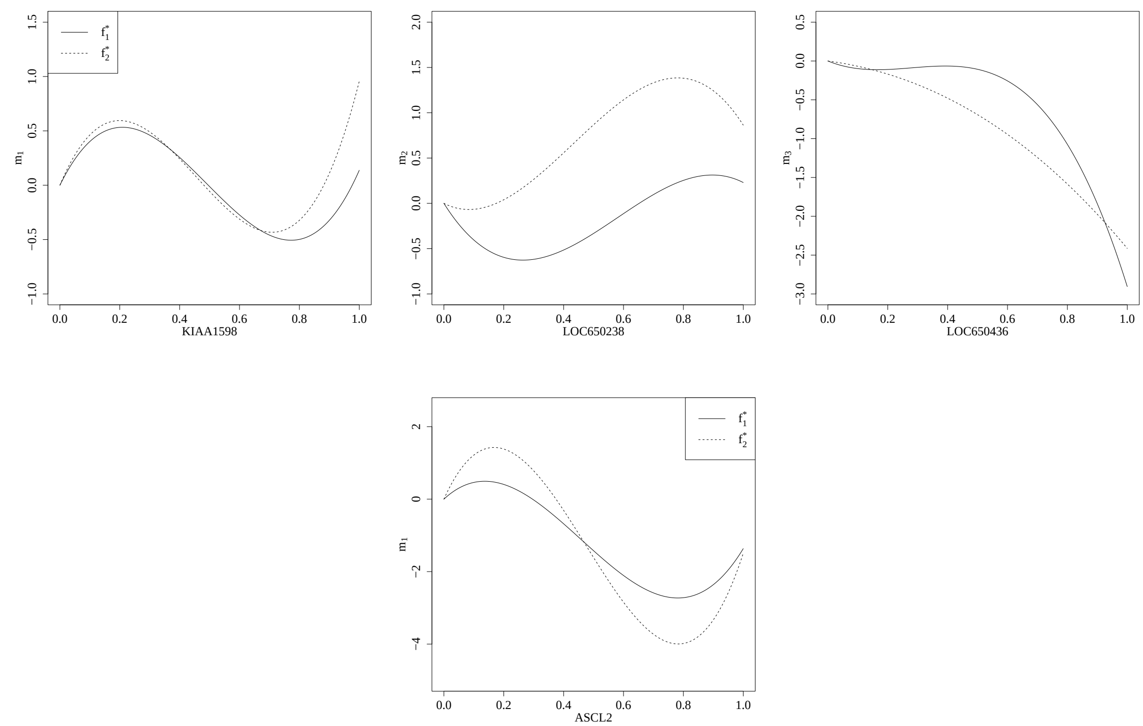

Figure S3: Trend of two latent factors associate with the selected genes at different quantiles. The first row denotes the selected genes at $\tau = 0.75$; The Second row denotes the selected gene at $\tau = 0.9$.

Bibliography

Ahn, S. C. and A. R. Horenstein (2013). Eigenvalue ratio test for the number of factors. *Econometrica 81*(3), 1203–1227.

Bai, Z. and X. Ding (2012). Estimation of spiked eigenvalues in spiked models. *Random Matrices: Theory and Applications 1*(02), 1150011.

Bai, Z. and J. W. Silverstein (2010). *Spectral analysis of large dimensional random matrices.* Springer.

Candès, E., Y. Fan, L. Janson, and J. Lv (2018). Panning for gold: model-x knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(3), 551–577.

El Karoui, N. (2008). Spectrum estimation for large dimensional covariance matrices using random matrix theory. *The Annals of Statistics 36*(6), 2757–2790.

Fan, J., J. Guo, and S. Zheng (2022). Estimating number of factors by adjusted eigenvalues thresholding. *Journal of the American Statistical Association 117*(538), 852–861.

Fan, J., Y. Ke, Q. Sun, and W.-X. Zhou (2019). Farmtest: Factor-adjusted robust multiple testing with approximate false discovery control. *Journal of the American Statistical Association 114*(528), 1880–1893.

Fan, Y., J. Lv, M. Sharifvaghefi, and Y. Uematsu (2020). Ipad: stable interpretable forecasting with knockoffs inference. *Journal of the American Statistical Association 115*(532), 1822–1834.

Fernandes, M., E. Guerre, and E. Horta (2021). Smoothing quantile regressions. *Journal of Business & Economic Statistics 39*(1), 338–357.

He, X., L. Wang, and H. G. Hong (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics 41*(1), 342–369.

Hjort, N. L. and D. Pollard (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*.

Huang, D. and L. Janson (2020). Relaxing the assumptions of knockoffs by conditioning. *The Annals of Statistics 48*(5), 3021–3042.

Janson, L. and W. Su (2016). Familywise error rate control via knockoffs. *Electronic Journal of Statistics 10*(1), 960.

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces: isoperimetry and processes*. Springer Science & Business Media.

Massart, P. (2000). About the constants in talagrand's concentration inequalities for empirical processes. *The Annals of Probability 28*(2), 863–884.

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica 77*(5), 1447–1479.

Ren, Z., Y. Wei, and E. Candès (2023). Derandomizing knockoffs. *Journal of the American Statistical Association 118*(542), 948–958.

Romano, Y., M. Sesia, and E. Candès (2020). Deep knockoffs. *Journal of the American Statistical Association 115*(532), 1861–1872.

Schumaker, L. (1981). *Spline functions: basic theory*. Wiley, New York.

Sherwood, B. and L. Wang (2016). Partially linear additive quantile regression in ultra-high dimension. *The Annals of Statistics 44*(1), 288–317.

Stone, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics 13*(2), 689–705.

Sun, Q., W.-X. Zhou, and J. Fan (2020). Adaptive huber regression. *Journal of the American Statistical Association 115*(529), 254–265.

Van Der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes*. Springer, New York.

Wang, H. (2012). Factor profiled sure independence screening. *Biometrika 99*(1), 15–28.

Zhou, S., X. Shen, and D. Wolfe (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics 26*(5), 1760–1782.

Table S1: Simulation results of example S1 with true model size 3

		MMS						MMS					
Case	Method	Median(MAD)	Mean(SD)	$\mathcal{P}_1$	$\mathcal{P}_3$	$\mathcal{P}_5$	$\mathcal{P}_{all}$	Median(MAD)	Mean(SD)	$\mathcal{P}_1$	$\mathcal{P}_3$	$\mathcal{P}_5$	$\mathcal{P}_{all}$
		$\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$						$\mathbf{X}_i \sim t_3(\mathbf{0}, \mathbf{I}_p)$					
(3a)	$\mathrm{MSDS}_{0.5}$	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	5.00(2.96)	26.56(79.74)	0.92	0.92	1.00	0.92
	$\mathrm{QaSIS}_{0.5}$	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	4.00(1.48)	5.15(2.40)	1.00	1.00	1.00	1.00
	NIS	3.00(0)	3.62(3.84)	1.00	1.00	1.00	1.00	870.00(1126.77)	1652.17(1510.62)	0.22	0.64	0.84	0.15
	SIRS	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	3.00(0)	40.64(89.82)	0.83	1.00	1.00	0.83
	DCSIS	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	20.00(25.20)	861.88(1234.65)	0.56	0.71	1.00	0.52
	RVSIS	3.00(0)	3.07(0.54)	1.00	1.00	1.00	1.00	1325.00(1482.60)	1790.84(1516.24)	0.15	0.67	0.86	0.09
	PC-Screen	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	3.00(0)	3.08(0.53)	1.00	1.00	1.00	1.00
(3b)	$\mathrm{MSDS}_{0.5}$	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	5.00(2.96)	12.61(46.69)	0.97	0.97	1.00	0.97
	$\mathrm{MSDS}_{0.75}$	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	4.00(1.48)	12.47(36.56)	0.95	0.98	1.00	0.95
	$\mathrm{MSDS}_{0.9}$	3.00(0)	3.06(0.29)	1.00	1.00	1.00	1.00	4.00(1.48)	36.35(122.61)	0.90	0.94	0.97	0.90
	$\mathrm{QaSIS}_{0.5}$	3.00(0)	3.06(0.44)	1.00	1.00	1.00	1.00	5.00(2.96)	5.74(5.38)	0.98	1.00	1.00	0.98
	$\mathrm{QaSIS}_{0.75}$	3.00(0)	3.32(0.98)	1.00	1.00	1.00	1.00	17.50(18.53)	315.33(608.81)	0.67	0.87	0.96	0.64
	$\mathrm{QaSIS}_{0.9}$	24.00(27.42)	321.14(501.12)	0.67	0.82	0.98	0.63	1295.00(1256.50)	1716.68(1308.10)	0.10	0.54	0.74	0.02
	NIS	3.00(0)	3.84(4.60)	0.98	1.00	1.00	0.98	1525.00(1964.44)	1801.56(1453.28)	0.21	0.56	0.86	0.15
	SIRS	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	4.00(1.48)	36.33(93.98)	0.82	1.00	1.00	0.82
	DCSIS	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	30.00(40.03)	662.25(949.37)	0.52	0.76	0.98	0.52
	RVSIS	3.00(0)	3.50(3.16)	1.00	1.00	1.00	1.00	1635.00(1919.96)	1925.19(1397.50)	0.17	0.62	0.86	0.13
	PC-Screen	3.00(0)	3.00(0)	1.00	1.00	1.00	1.00	3.00(0)	4.61(12.52)	1.00	1.00	0.98	0.98

Table S2: Simulation results for example S2 when $\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, p^* denotes the true model size

		MMS											
Error	Method	p^*	Median(MAD)	Mean(SD)	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_8$	$\mathcal{P}_9$	$\mathcal{P}_{10}$	$\mathcal{P}_{all}$
$N(0,1)$	$\text{MSDS}_{0.50}$	5	6.00(1.48)	10.16(22.13)	1.00	1.00	1.00	1.00	0.98	-	-	-	0.98
	$\text{QaSIS}_{0.50}$	5	22.50(24.46)	67.31(109.39)	1.00	1.00	1.00	0.98	0.60	-	-	-	0.60
	NIS	5	1418.00(490.66)	1451.57(418.55)	0.32	0.78	0.66	0.30	0.14	-	-	-	0.10
	SIRS	5	447.50(452.19)	467.36(319.36)	1.00	1.00	1.00	1.00	0.10	-	-	-	0.10
	DCSIS	5	94.50(119.34)	175.52(182.57)	0.92	0.98	0.99	0.94	0.23	-	-	-	0.23
	RVSIS	5	1596(519.50)	1552.68(580.37)	0.42	0.79	0.77	0.38	0.06	-	-	-	0.02
	PC-Screen	5	95.00(106.00)	150.90(148.81)	1.00	1.00	1.00	1.00	0.23	-	-	-	0.23
t_3	$\text{MSDS}_{0.50}$	5	5.50(0.74)	10.94(23.98)	1.00	1.00	1.00	1.00	0.98	-	-	-	0.98
	$\text{QaSIS}_{0.50}$	5	26.00(31.13)	68.27(95.89)	1.00	1.00	1.00	1.00	0.58	-	-	-	0.58
	NIS	5	2630.00(2305.44)	2609.88(1626.94)	0.25	0.61	0.57	0.28	0.14	-	-	-	0.04
	SIRS	5	2290.00(1949.61)	2364.88(1433.97)	1.00	1.00	1.00	f 1.00	0.03	-	-	-	0.03
	DCSIS	5	155.00(192.73)	218.56(201.61)	0.91	0.92	0.96	0.91	0.25	-	-	-	0.23
	RVSIS	5	3140.00(1616.03)	2960.94(1413.97)	0.36	0.65	0.67	0.34	0.08	-	-	-	0.02
	PC-Screen	5	87.00(108.97)	150.20(150.08)	1.00	1.00	1.00	1.00	0.33	-	-	-	0.33
hetero	$\text{MSDS}_{0.50}$	5	5.00(0)	5.08(0.36)	1.00	1.00	1.00	1.00	1.00				1.00
	$\text{MSDS}_{0.75}$	8	11.00(1.48)	19.35(23.50)	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.96	0.95
	$\text{MSDS}_{0.90}$	8	18.50(12.60)	46.51(77.73)	1.00	1.00	1.00	1.00	1.00	0.96	0.93	0.87	0.84
	$\text{QaSIS}_{0.50}$	5	10.00(7.41)	37.82(57.31)	1.00	1.00	1.00	1.00	0.75				0.75
	$\text{QaSIS}_{0.75}$	8	1046.00(935.37)	1181.56(1062.48)	1.00	1.00	1.00	1.00	0.61	0.32	0.32	0.17	0.03
	$\text{QaSIS}_{0.90}$	8	2830.00(1393.64)	2754.47(1177.24)	0.41	0.63	0.83	0.42	0.10	0.25	0.30	0.17	0.00
	NIS	8	3777.50(1215.73)	3438.55(1226.07)	0.50	0.75	0.73	0.46	0.11	0.15	0.21	0.17	0.00
	SIRS	8	2282.50(1174.96)	2264.64(1142.64)	1.00	1.00	1.00	1.00	0.93	0.11	0.12	0.09	0.00
	DCSIS	8	1067.50(1060.05)	1415.65(1174.22)	0.98	0.98	0.98	0.98	0.89	0.39	0.47	0.28	0.07
	RVSIS	8	3395.00(1286.15)	3135.19(1224.47)	0.72	0.80	0.81	0.56	0.23	0.17	0.17	0.15	0.00
	PC-Screen	8	3657.50(1004.46)	3552.25(1005.02)	1.00	1.00	1.00	1.00	0.23	0.05	0.02	0.05	0.00

Table S3: Simulation results for example S2 with $\mathbf{\Sigma_X} = (0.5^{|i-j|})_{1 \le i,j \le p}$, p^* denotes the true model size

		MMS											
Error	Method	p^*	Median(MAD)	Mean(SD)	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_8$	$\mathcal{P}_9$	$\mathcal{P}_{10}$	$\mathcal{P}_{all}$
$N(0,1)$	$\text{MSDS}_{0.50}$	5	5.00(0)	5.18(0.57)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	$\text{QaSIS}_{0.50}$	5	5.00(0)	16.18(45.80)	1.00	1.00	1.00	1.00	0.92	-	-	-	0.92
	NIS	5	1480.00(1355.09)	2013.33(1671.58)	0.44	0.70	0.63	0.41	0.12	-	-	-	0.11
	SIRS	5	5.00(0)	9.78(11.01)	1.00	1.00	1.00	1.00	0.97	-	-	-	0.97
	DCSIS	5	5.00(0)	33.87(80.11)	0.96	0.99	0.98	0.98	0.86	-	-	-	0.84
	RVSIS	5	1350.00(1275.24)	1705.83(1451.77)	0.61	0.78	0.74	0.51	0.18	-	-	-	0.17
	PC-Screen	5	5.00(0)	5.95(2.68)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
t_3	$\text{MSDS}_{0.50}$	5	5.00(0)	5.35(2.14)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	$\text{QaSIS}_{0.50}$	5	6.00(1.48)	9.62(11.47)	1.00	1.00	1.00	1.00	0.98	-	-	-	0.98
	NIS	5	1860.00(1651.61)	2294.00(1671.65)	0.40	0.68	0.56	0.31	0.12	-	-	-	0.10
	SIRS	5	5.00(0)	8.33(11.53)	1.00	1.00	1.00	1.00	0.98	-	-	-	0.98
	DCSIS	5	6.00(1.48)	37.63(80.95)	0.96	0.96	0.96	0.94	0.81	-	-	-	0.81
	RVSIS	5	1742.50(1567.85)	1972.88(1432.59)	0.57	0.70	0.66	0.38	0.10	-	-	-	0.09
	PC-Screen	5	5.00(0)	7.35(9.15)	1.00	1.00	1.00	1.00	0.98	-	-	-	0.98
hetero	$\text{MSDS}_{0.50}$	5	5.00(0)	6.36(4.39)	1.00	1.00	1.00	1.00	1.00				1.00
	$\text{MSDS}_{0.75}$	8	12.00(2.96)	29.02(50.42)	1.00	1.00	1.00	1.00	1.00	0.97	0.95	0.88	0.84
	$\text{MSDS}_{0.90}$	8	25.00(20.75)	47.91(58.16)	1.00	1.00	1.00	1.00	1.00	0.83	0.86	0.75	0.67
	$\text{QaSIS}_{0.50}$	5	7.00(2.96)	21.78(45.35)	1.00	1.00	1.00	1.00	0.87				0.87
	$\text{QaSIS}_{0.75}$	8	1518.00(1165.32)	1532.00(878.40)	0.95	1.00	1.00	1.00	0.54	0.22	0.25	0.15	0.05
	$\text{QaSIS}_{0.90}$	8	1842.00(1423.29)	1983.42(1023.22)	0.34	0.71	0.55	0.31	0.08	0.27	0.28	0.25	0.00
	NIS	8	3682.50(1260.21)	3445.83(1157.10)	0.53	0.77	0.71	0.38	0.05	0.14	0.14	0.21	0.00
	SIRS	8	2036.00(1165.32)	1974.22(901.87)	1.00	1.00	1.00	1.00	0.92	0.05	0.12	0.06	0.00
	DCSIS	8	518.00(521.87)	620.28(463.37)	0.97	0.97	0.98	0.97	0.82	0.41	0.45	0.23	0.11
	RVSIS	8	3745.00(1204.61)	3381.44(1261.98)	0.68	0.80	0.78	0.55	0.14	0.10	0.10	0.07	0.00
	PC-Screen	8	2462.50(1389.93)	2396.58(1145.88)	1.00	1.00	1.00	1.00	0.96	0.12	0.10	0.05	0.00

Table S4: Simulation results for example 1 with $\mathbf{\Sigma_X} = (0.5^{|i-j|})_{1 \le i,j \le p}$, p^* denotes the true model size

		MMS											
Error	Method	p^*	Median(MAD)	Mean(SD)	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_8$	$\mathcal{P}_9$	$\mathcal{P}_{10}$	$\mathcal{P}_{all}$
$N(0,1)$	$\text{MSDS}_{0.50}$	5	5.00(0)	5.00(0)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	$\text{QaSIS}_{0.50}$	5	5.00(0)	5.07(0.47)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	NIS	5	5.00(0)	5.01(0.10)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	SIRS	5	5.00(0)	5.00(0)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	DCSIS	5	5.00(0)	5.00(0)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	RVSIS	5	5.00(0)	5.00(0)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	PC-Screen	5	5.00(0)	5.00(0)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
t_3	$\text{MSDS}_{0.50}$	5	5.00(0)	5.03(0.22)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	$\text{QaSIS}_{0.50}$	5	5.00(0)	5.54(2.09)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	NIS	5	10.00(7.41)	102.90(127.06)	0.74	0.86	0.87	0.86	0.76	-	-	-	0.70
	SIRS	5	5.00(0)	5.01(0.10)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	DCSIS	5	5.00(0)	5.00(0)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	RVSIS	5	5.00(0)	42.75(137.20)	0.93	0.93	0.92	0.93	0.91	-	-	-	0.90
	PC-Screen	5	5.00(0)	5.00(0)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
hetero	$\text{MSDS}_{0.50}$	5	5.00(0)	5.73(2.12)	1.00	1.00	1.00	1.00	1.00				1.00
	$\text{MSDS}_{0.75}$	8	10.00(2.96)	22.40(36.01)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.87	0.87
	$\text{MSDS}_{0.90}$	8	29.00(25.20)	81.06(132.80)	0.93	1.00	1.00	0.98	0.85	1.00	1.00	0.78	0.61
	$\text{QaSIS}_{0.50}$	5	6.00(1.48)	7.67(6.51)	1.00	1.00	1.00	1.00	0.98				0.98
	$\text{QaSIS}_{0.75}$	8	28.50(24.46)	85.52(145.71)	0.94	1.00	1.00	0.98	0.80	1.00	1.00	0.76	0.57
	$\text{QaSIS}_{0.90}$	8	2607.50(1126.77)	2693.38(1071.41)	0.03	0.05	0.04	0.05	0.02	0.97	1.00	0.80	0.00
	NIS	8	3680.00(1071.17)	3581.50(981.38)	0.13	0.12	0.07	0.08	0.07	0.54	0.56	0.33	0.00
	SIRS	8	36.50(30.39)	67.92(79.09)	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.54	0.53
	DCSIS	8	10.00(2.96)	30.64(64.79)	0.97	0.98	0.98	0.96	0.86	1.00	1.00	0.98	0.85
	RVSIS	8	3285.00(1245.38)	3022.44(1169.79)	0.21	0.31	0.17	0.18	0.08	0.46	0.50	0.34	0.00
	PC-Screen	8	27.00(23.72)	43.05(40.86)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.63	0.63

Table S5: Simulation results for example 2 with $\mathbf{\Sigma_X} = (0.5^{|i-j|})_{1 \le i,j \le p}$, p^* denotes the true model size

		MMS ($p^* = 5$)							
ξ_{ik}	Method	Median(MAD)	Mean(SD)	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_{all}$
$N(0,1)$	$\text{MSDS}_{0.5}$	6.00(1.48)	7.09(2.86)	1.00	1.00	1.00	1.00	1.00	1.00
	$\text{QaSIS}_{0.5}$	6.00(1.48)	8.06(9.69)	0.99	1.00	1.00	1.00	1.00	0.99
	NIS	2472.50(2142.35)	2464.70(1551.20)	0.18	0.24	0.45	0.52	1.00	0.04
	SIRS	12.00(10.37)	66.64(128.47)	0.67	1.00	1.00	1.00	1.00	0.67
	DCSIS	271.00(327.65)	353.86(377.82)	0.47	0.54	0.76	0.89	1.00	0.33
	RVSIS	1875.00(1890.31)	2095.95(1450.50)	0.19	0.23	0.46	0.73	1.00	0.06
	PC-Screen	6.00(1.48)	8.46(8.67)	0.98	1.00	1.00	1.00	1.00	0.98
Cauchy	$\text{MSDS}_{0.5}$	6.00(1.48)	9.34(11.42)	1.00	1.00	0.98	1.00	1.00	0.98
	$\text{QaSIS}_{0.5}$	6.00(1.48)	10.43(12.29)	0.97	1.00	1.00	1.00	1.00	0.97
	NIS	2232.50(1719.81)	2325.80(1391.02)	0.23	0.19	0.32	0.49	0.98	0.04
	SIRS	20.00(22.23)	77.29(122.12)	0.60	1.00	0.99	1.00	1.00	0.60
	DCSIS	218.00(265.38)	355.24(369.26)	0.39	0.56	0.72	0.92	1.00	0.28
	RVSIS	1637.50(1556.73)	1990.95(1383.97)	0.22	0.21	0.37	0.66	0.99	0.06
	PC-Screen	6.00(1.48)	14.12(29.72)	0.94	1.00	1.00	1.00	1.00	0.94

Table S6: Determining the numbers of factors for example S1 by various method.

Case		BE	ACT	ER	ON	GR	BE	ACT	ER	ON	GR
				$\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$					$\mathbf{X}_i \sim t_3(\mathbf{0}, \mathbf{I}_p)$		
Case (3a)	$\text{ave}(\hat{K})$	3.00	2.97	199.00	199.00	10.36	2.53	2.42	132.12	110.63	7.05
	$\text{med}(\hat{K})$	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
	$P(\hat{K} < 3)$	0.00	0.03	0.00	0.00	0.00	0.46	0.52	0.23	0.16	0.00
	$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.03	0.00	0.00	0.66	0.55	0.03
	$P(\hat{K} = 3)$	1.00	0.97	0.00	0.00	0.97	0.54	0.48	0.11	0.29	0.97
Case (3b)	$\text{ave}(\hat{K})$	3.00	3.00	199.00	199.00	8.27	2.85	2.67	142.02	112.63	7.85
	$\text{med}(\hat{K})$	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
	$P(\hat{K} < 3)$	0.00	0.00	0.00	0.00	0.00	0.21	0.33	0.14	0.13	0.01
	$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.02	0.00	0.00	0.71	0.56	0.04
	$P(\hat{K} = 3)$	1.00	1.00	0.00	0.00	0.98	0.79	0.67	0.15	0.31	0.95

Table S7: Determining the numbers of factors for example 1 and example S2 by various methods. “hetero” denotes the heteroscedastic error case.

			Example 1					Example S2				
$\mathbf{X}_i$	ξ_{ik}		BE	ACT	ER	ON	GR	BE	ACT	ER	ON	GR
$\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$	$N(0,1)$	ave($\hat{K}$)	3.00	3.00	199.00	199.00	14.23	2.92	2.91	195.04	195.04	47.95
		med($\hat{K}$)	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
		$P(\hat{K} < 3)$	0.00	0.00	0.00	0.00	0.00	0.06	0.06	0.02	0.02	0.00
		$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.05	0.00	0.01	0.98	0.98	0.25
		$P(\hat{K} = 3)$	1.00	1.00	0.00	0.00	0.95	0.94	0.93	0.00	0.00	0.75
	t_3	ave($\hat{K}$)	3.00	3.00	199.00	199.00	43.11	2.95	2.93	199.00	199.00	42.51
		med($\hat{K}$)	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
		$P(\hat{K} < 3)$	0.00	0.00	0.00	0.00	0.00	0.03	0.04	0.00	0.00	0.01
		$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.11	0.00	0.00	1.00	1.00	0.18
		$P(\hat{K} = 3)$	1.00	1.00	0.00	0.00	0.89	0.97	0.96	0.00	0.00	0.81
	hetero	ave($\hat{K}$)	3.00	3.00	199.00	199.00	20.64	2.97	2.94	197.02	197.02	32.13
		med($\hat{K}$)	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
		$P(\hat{K} < 3)$	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.01	0.01	0.00
		$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.06	0.00	0.00	0.99	0.99	0.15
		$P(\hat{K} = 3)$	1.00	1.00	0.00	0.00	0.94	0.98	0.96	0.00	0.00	0.85
$\mathcal{N}(\mathbf{0}, \mathbf{\Sigma_X})$	$N(0,1)$	ave($\hat{K}$)	3.00	3.00	199.00	199.00	31.52	2.99	2.98	197.02	197.02	66.60
		med($\hat{K}$)	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
		$P(\hat{K} < 3)$	0.00	0.00	0.00	0.00	0.00	0.01	0.03	0.01	0.01	0.00
		$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.13	0.00	0.02	0.99	0.99	0.30
		$P(\hat{K} = 3)$	1.00	1.00	0.00	0.00	0.87	0.99	0.95	0.00	0.00	0.70
	t_3	ave($\hat{K}$)	3.00	3.00	199.00	199.00	19.67	3.01	2.99	199.00	199.00	60.37
		med($\hat{K}$)	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
		$P(\hat{K} < 3)$	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.00	0.00	0.00
		$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.07	0.04	0.03	1.00	1.00	0.28
		$P(\hat{K} = 3)$	1.00	1.00	0.00	0.00	0.93	0.94	0.92	0.00	0.00	0.72
	hetero	ave($\hat{K}$)	3.00	2.98	199.00	199.00	4.96	2.98	2.96	199.00	199.00	52.10
		med($\hat{K}$)	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
		$P(\hat{K} < 3)$	0.00	0.02	0.00	0.00	0.00	0.04	0.05	0.00	0.00	0.01
		$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.01	0.02	0.03	1.00	1.00	0.23
		$P(\hat{K} = 3)$	1.00	0.98	0.00	0.00	0.99	0.94	0.92	0.00	0.00	0.76

Table S8: Determining the numbers of factors for example 2 by various method.

ξ_{ik}		BE	ACT	ER	ON	GR	BE	ACT	ER	ON	GR
		$\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$					$\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{X}})$				
$N(0,1)$	ave$(\hat{K})$	2.22	2.15	195.04	195.04	59.18	2.22	2.18	197.02	197.02	74.59
	med$(\hat{K})$	2.00	2.00	199.00	199.00	1.00	2.00	2.00	199.00	199.00	3.00
	$P(\hat{K} < 3)$	0.55	0.60	0.02	0.02	0.66	0.64	0.66	0.01	0.01	0.43
	$P(\hat{K} > 3)$	0.01	0.00	0.98	0.98	0.27	0.00	0.00	0.99	0.99	0.33
	$P(\hat{K} = 3)$	0.44	0.40	0.00	0.00	0.07	0.36	0.34	0.00	0.00	0.24
$Cauchy(0,1)$	ave$(\hat{K})$	2.56	2.52	199.00	199.00	35.20	2.49	2.45	197.02	197.02	56.07
	med$(\hat{K})$	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
	$P(\hat{K} < 3)$	0.32	0.35	0.00	0.00	0.36	0.44	0.48	0.01	0.01	0.18
	$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.13	0.00	0.00	0.99	0.99	0.25
	$P(\hat{K} = 3)$	0.68	0.65	0.00	0.00	0.51	0.56	0.52	0.00	0.00	0.57

Table S9: Simulation results for example 3 via MSDK, PC-Knockoff and DMSDK procedures

τ	α	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_7$	$\mathcal{P}_8$	$\mathcal{P}_{all}$	$\hat{V}(\tau)(\widehat{\text{FDR}})$	Power
MSDK procedure FDR control												
0.50	0.15	0.91	0.91	0.88	0.91	0.93	0.91	0.91	0.93	0.70	1.81(0.161)	0.91
	0.20	0.88	0.98	0.95	0.95	0.95	0.95	0.98	0.90	0.71	1.76(0.180)	0.94
	0.30	0.98	0.98	0.98	0.98	0.98	0.96	1.00	1.00	0.90	4.90(0.311)	0.98
0.75	0.15	0.80	0.84	0.75	0.80	0.81	0.74	0.84	0.77	0.41	1.87(0.162)	0.79
	0.20	0.83	0.93	0.87	0.92	0.91	0.86	0.91	0.76	0.51	2.90(0.212)	0.87
	0.30	0.93	0.98	0.93	0.95	0.96	0.91	0.98	0.88	0.68	5.92(0.299)	0.94
0.90	0.15	0.22	0.31	0.24	0.28	0.27	0.27	0.27	0.23	0.00	1.25(0.296)	0.26
	0.20	0.43	0.47	0.46	0.44	0.48	0.40	0.55	0.32	0.05	2.90(0.272)	0.44
	0.30	0.63	0.76	0.65	0.78	0.78	0.64	0.69	0.54	0.14	9.67(0.361)	0.68
PC-Knockoff procedure FDR control												
	α	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_7$	$\mathcal{P}_8$	$\mathcal{P}_{all}$	$\hat{V}(\widehat{\text{FDR}})$	Power
	0.15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.34(0.114)	1.00
	0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.20(0.159)	1.00
	0.30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	4.16(0.284)	1.00
DMSDK procedure PFER control												
τ	v	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_7$	$\mathcal{P}_8$	$\mathcal{P}_{all}$	$\hat{V}(\tau)(\widehat{\text{FDR}})$	Power
0.50	1	0.96	1.00	0.96	1.00	1.00	0.96	0.99	0.95	0.90	0.40(0.048)	0.97
	2	1.00	1.00	1.00	1.00	1.00	0.95	1.00	0.96	0.92	1.33(0.142)	0.98
	3	0.96	1.00	0.97	1.00	1.00	1.00	1.00	0.98	0.94	1.48(0.156)	0.99
0.75	1	0.90	0.95	0.87	0.94	0.93	0.86	0.92	0.83	0.45	0.74(0.085)	0.90
	2	0.90	0.98	0.97	0.98	0.97	0.94	0.95	0.95	0.71	1.29(0.139)	0.96
	3	0.93	0.98	0.93	1.00	1.00	0.89	0.98	0.96	0.73	2.45(0.234)	0.96
0.90	1	0.47	0.77	0.50	0.68	0.72	0.63	0.76	0.51	0.07	0.66(0.076)	0.63
	2	0.63	0.80	0.65	0.73	0.78	0.54	0.82	0.63	0.11	1.39(0.148)	0.69
	3	0.70	0.83	0.66	0.80	0.82	0.60	0.82	0.66	0.24	2.47(0.236)	0.73
DMSDK procedure k-FWER control at 0.20												
τ	$k(v)$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_7$	$\mathcal{P}_8$	$\mathcal{P}_{all}$	$\hat{V}(\tau)(\widehat{\text{FWER}})$	Power
0.50	3(1.17)	0.97	1.00	0.97	0.98	0.99	0.99	1.00	0.96	0.91	0.78(0.000)	0.98
	4(1.56)	0.96	1.00	0.97	1.00	1.00	1.00	1.00	0.99	0.92	0.81(0.000)	0.99
	5(1.96)	0.97	1.00	1.00	0.99	0.99	0.99	1.00	0.99	0.96	1.35(0.019)	0.99
0.75	3(1.17)	0.92	0.97	0.91	0.91	0.98	0.90	0.94	0.78	0.50	0.63(0.011)	0.90
	4(1.56)	0.86	0.97	0.92	0.97	0.96	0.92	0.96	0.86	0.60	0.77(0.011)	0.93
	5(1.96)	0.87	0.98	0.93	1.00	0.98	0.94	0.97	0.90	0.67	1.33(0.020)	0.95
0.90	3(1.17)	0.49	0.66	0.53	0.75	0.65	0.69	0.83	0.54	0.06	0.81(0.056)	0.64
	4(1.56)	0.53	0.67	0.52	0.77	0.70	0.72	0.87	0.59	0.06	1.27(0.067)	0.67
	5(1.96)	0.62	0.71	0.60	0.79	0.74	0.74	0.89	0.59	0.09	1.84(0.078)	0.71

Table S10: Determining the numbers of factors for examples 3-4 by various methods

	Example 3					Example 4				
	BE	ACT	ER	ON	GR	BE	ACT	ER	ON	GR
ave($\hat{K}$)	3.00	3.00	199.00	199.00	14.58	3.00	2.98	199.00	199.00	46.65
med($\hat{K}$)	3.00	3.00	199.00	199.00	3.00	3.00	3.00	199.00	199.00	3.00
$P(\hat{K} < 3)$	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
$P(\hat{K} > 3)$	0.00	0.00	1.00	1.00	0.05	0.00	0.00	1.00	1.00	0.20
$P(\hat{K} = 3)$	1.00	1.00	0.00	0.00	0.95	1.00	0.98	0.00	0.00	0.80

Table S11: Selection results for human well-being data by various screening approaches and DMSDK with $V(\tau) = 1$

$MSDS_{0.25}$	$MSDS_{0.50}$	$MSDS_{0.75}$	$MSDS_{0.90}$	$QaSIS_{0.50}$	NIS	SIRS	DCSIS	RVSIS	PC-Screen	$DMSDK_{0.25}$	$DMSDK_{0.50}$	$DMSDK_{0.75}$	$DMSDK_{0.90}$
LOC650238	ANKRD33	NCK2	BLOC1S1	CLDND1	LOC650436	XRCC6BP1	LOC650238	LOC650436	LOC650238	LOC650238	LOC650238	KIAA1598	ASCL2
XRCC6BP1	LOC650238	BLOC1S1	ASCL2	LOC650238	AES	LOC650238	ANKRD33	LOC650238	ZC3H4	XRCC6BP1	MED28	LOC650238	
DCAKD	CTSD	LOC650238	GRN	SET	CA4	ZC3H4	LOC650436	ANKRD33	ANKRD33			LOC650436	
RPS8	BLOC1S1	PRDM8	FAM129C	ASCL2	GFOD1	LOC650436	CTSD	CTSD	KIAA0174				
CD69	VCAN	ASCL2	PARVB	PLA2G4C	ITGB1BP1	COQ10A	TFG	LILRA6	CTSD				
SSR1	UCP2	LACTB	LOC100127993	ZC3H4	LIN54	TUFM	ZC3H4	C14ORF2	XRCC6BP1				
KLRG1	LOC388339	FKBP15	ENDOD1	TGOLN2	LOC650238	TFG	LOC388588	LIPA	CD69				
LOC653071	LPAR2	ALOX12	GIMAP7	SLC46A2	C14ORF2	DCAKD	COQ10A	OSBP2	COQ10A				
WAC	MED28	LOC650436	LOC650436	PPP1R2	CTSD	NCKAP1L	KIAA0174	FKBP15	WAC				
C6ORF192	LACTB	FAM129C	DDEF2	NDEL1	ANKRD33	C14ORF2	WAC	FECH	LPAR2				
GFM1	RPS26L	GRN	LACTB	FLVCR2	CD36	LIPA	ADAP2	WDR40A	TFG				
CDC42SE1	LIN54	JAM3	BLK	PSMF1	GTSCR1	PEPD	FECH	TFDP1	AKR7A2				
ARHGDIB	WAC	KIAA1598	MPEG1	LILRA6	PRDM8	UCP2	ASCL2	IFIT1L	GTSCR1				
RPS25	RPS8	COX7A2L	PDLIM1	LASS2	NDEL1	CAPN1	XRCC6BP1	CAPN1	ARHGDIB				
NCR3	ZC3H4	HSPA1B	LOC646753	C16ORF80	LOC641825	TUBA1B	GTSCR1	PEPD	SLU7				
COQ10A	F2R	MPEG1	ATG16L1	TNFRSF25	TFG	FKBP15	C1ORF131	ADAP2	NCKAP1L				
LOC652755	C16ORF80	PARVB	LOC388588	GYG1	LILRA6	ANKRD33	CAPN1	TFG	IL1R2				

Table S12: Coefficients of Linear quantile regression with selected genes by DMSDK method as inputs at each quantile

		LOC650238	LOC650436	XRCC6BP1	MED28	KIAA1598	ASCL2
$\tau = 0.25$	f_1^*	0.59		-0.50			
	f_2^*	0.47		-0.18			
$\tau = 0.50$	f_1^*	0.43			0.07		
	f_2^*	0.42			0.34		
$\tau = 0.75$	f_1^*	0.32	-0.19			-0.18	
	f_2^*	0.42	-0.47			-0.21	
$\tau = 0.90$	f_1^*						-0.24
	f_2^*						-0.19