Statistica Sinica Preprint No: SS-2025-0167	
Title	Conditional Quantile-based Variable Screening with FDR Control in Joint Factor Models
Manuscript ID	SS-2025-0167
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0167
Complete List of Authors	Han Pan, Wei Xiong and Mingyao Ai
Corresponding Authors	Mingyao Ai
E-mails	myai@math.pku.edu.cn
Notice: Accepted author version.	

Statistica Sinica

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

Abstract: Joint factor models are commonly adopted to relate unobservable factors with covariates. Traditional approaches to joint models often assume linear relationships between latent factors and covariates, require prior knowledge of the number of latent factors, and typically fail to address heavy-tailedness or high-dimensional covariates. To overcome these challenges, we propose a general factor-covariate model and introduce a new variable selection procedure to broaden the scope of application and to alleviate the curse of dimensionality. The procedure is unfolded in three steps: robust estimation of factors via Huber regression, feature screening using an index of mean squared deviation (MSD) of conditional quantile and false discovery rate (FDR) control based on derandomized quantile knockoffs. To facilitate implementation, we employ smoothing quantile regression and apply a modified bootstrap-based eigenvalue method to determine the number of factors. Theoretical justifications on the sure screening property as well as the control of FDR, per family error rate and k family-wise error rate are provided. The superiority of our proposed procedure over existing methods is demonstrated by numerical studies on simulated and real datasets.

Key words: Derandomized knockoffs, False discovery rate, High-dimensional screening, Joint models, Smoothing quantile regression.

1. Introduction

Latent variables are prevalent across numerous scientific disciplines, including financial engineering, sociology, psychology, and biomedical research, among others. These variables cannot be directly measured by observed variables but are instead characterized by multiple observable surrogates. For instance, in medical studies, traits such as depression or overall adverse effects, are examples of latent variables. To understand the relationships between the covariates (e.g., age, gender) and these medical traits, latent variable analysis is crucial for uncovering the hidden patterns.

A conventional method for analyzing latent variables involves regressing the observable surrogates that characterize corresponding traits on the covariates of interest. However, this straightforward approach has limitations, as the observable surrogates may be imperfect representations of latent variables. To facilitate broader applications, existing literature has explored two-stage joint modeling approaches. In the first stage, latent variables are characterized by multiple observed variables through a factor model. In the second stage, a latent response-on-scalar regression is utilized to examine the potential covariates for latent responses. Roy and Lin (2000) combined factor analysis with a linear mixed model to evaluate the effectiveness of methadone treatment in reducing illicit drug use. Ouyang et al. (2018) integrated factor model with a semiparametric failure time model to analyze multivariate censored data. However, these approaches face challenges when handling high-dimensional covariates. In a more recent study, Yu et al. (2023) introduced a joint model combining a factor model with linear regression to investigate the relationship between

psychological well-being and ultrahigh dimensional human genome. Their method assumes linearity between covariates and latent factors, which may not hold or be challenging to validate in practice. To address these challenges, we propose a more general factor-covariate model that integrates a factor model with a nonparametric model to capture nonlinear effects of high-dimensional covariates on latent factors. This joint model is motivated by an empirical study of Fredrickson et al. (2013), which aimed to explore the relationship between measured genes and two types of well-being. These well-beings are unobserved latent factors that must be derived from data collected via a questionnaire survey. Given that genes may have nonlinear effects on the well-being, a joint analysis that integrates factor model for extracting latent responses and a nonparametric model for examining the relationship between genes and these latent factors is crucial for this investigation.

Several critical issues need to be addressed within the joint factor-covariate model. The first concern is the estimation of latent factors. While methods based on least squares regression or EM algorithms have been proposed (Bai, 2003; Desai and Storey, 2012), these methods assume joint normality of factors and noise, a condition rarely met in practice. The second issue arises from the potential ultrahigh dimensionality of covariates, leading to the curse of dimensionality for nonparametric model. To address this, additional assumptions on the regression function or covariates are required, with the sparsity assumption being a common strategy, which posits that only a small number of covariates are relevant to the latent responses. Our objective here is to propose novel methods to identify the covariates that significantly contribute to the multivariate latent responses. The third issue involves

controlling the number of false discoveries, mathematically formulated as controlling the false discovery rate (FDR), which is crucial for reliable feature selection. Existing screening methods (He et al., 2013) tend to sacrifice FDR control for the sure screening property by choosing a conservative threshold, resulting in an inflated model size. Additionally, most current joint modeling techniques assume prior knowledge of the number of latent factors. Hence, it is essential to develop robust methods for estimation, screening, and selection that handle multiple latent factors and heterogeneous effects simultaneously.

In response to these considerations, this paper introduces a quantile-based mean squared deviation (MSD) index and a MSD-Select procedure tailored for factor-covariate model. We gradually unveil the whole procedure in three steps. First, we apply a Huber regression to estimate latent factors robustly. Next, building on MSD index, we develop a novel quantile-adaptive screening procedure for multivariate latent factors, allowing the active covariates to vary across quantiles. Following the screening step, a quantile-adaptive derandomized knockoffs procedure is introduced to further control FDR while maintaining high power at the targeted quantile levels. Existing FDR control methods generally fall into two categories: p-value based and knockoffs based methods. The classical p-value based approach (Benjamini and Hochberg, 1995) requires exact p-values for FDR control, in contrast, recent knockoffs algorithms use synthetic knockoff features to control FDR (Barber and Candès, 2015; Candès et al., 2018; Liu et al., 2022). Relatively little work has focused on FDR control in a quantile-adaptive manner. To bridge this gap, we extend the knockoffs to the quantile framework, ensuring a parsimonious model with guaranteed FDR control.

The primary contributions of this paper are as follows. First, we propose a general factor-covariate model that uncovers the hidden relationship between observed variables and covariates. We further introduce a bootstrap-based eigenvalue criterion to determine the number of latent factors, which is demonstrated to be theoretically consistent and empirically effective. Second, we develop a new MSD index to quantify the association between two random vectors from a quantile perspective. Within quantile framework, Li et al. (2015) proposed a quantile correlation to measure the linear quantile relationship between a univariate response Y and covariates X. Shao and Zhang (2014) extended this concept to multivariate $\mathbf{X}$ by introducing a martingale difference divergence, assuming that X has finite second-order moments. Liu et al. (2022) further introduced a projection quantile correlation, which does not require the moment condition. The MSD index improves upon these measures by quantifying the quantile dependence between a multivariate $\mathbf{Y}$ and a multivariate $\mathbf{X}$, without the need for moment conditions. The MSD index has several appealing properties: it equals zero if and only if the quantile independence holds and is robust to heavy-tailed data and outliers since it is invariant under monotone variable transformations. Additionally, it has a low computational cost of $O(n^2)$, compared with the $O(n^3)$ of the projection quantile correlation. The index is estimated using smoothing quantile regression techniques (Fernandes et al., 2021), effectively addressing challenges such as the nondifferentiability of the quantile loss function and the curse of dimensionality. We also derive useful exponential bounds to establish the sure screening property of the MSD index. Third, we formulate a quantile-adaptive procedure to control FDR, leveraging the concept

of derandomized knockoffs (Ren et al., 2023). We prove that the proposed procedure can simultaneously control the FDR, the per family error rate (PFER), and the k family-wise error rate (k-FWER). Simulation studies demonstrate that this method effectively controls the FDR more tightly while maintaining high power in practical scenarios.

The rest of the paper is organized as follows. In Section 2, we introduce the factor-covariate joint model and develop a quantile-adaptive MSD-Select procedure. A bootstrap-based eigenvalue method for determining the number of latent factors is also proposed. Section 3 establishes the theoretical guarantees, including sure screening properties and control of FDR, PFER, and k-FWER. The superiority of our new procedure over existing methods is demonstrated through numerical studies on both simulated and real datasets in Sections 4 and 5. Section 6 concludes with some discussions. All technical proofs and additional numerical studies are provided in the Supplementary Materials.

2. MSD-Select Procedure

2.1 General Factor-Covariate Model

To establish the functional associations between observed variables and potential covariates, we develop the factor-covariate model, which comprises two major components: a factor model and a nonparametric model, i.e.,

$$\begin{cases} \mathbf{Z} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{f} + \boldsymbol{\varepsilon}, \\ \boldsymbol{f} = \boldsymbol{m}(X_1, \ldots, X_p) + \boldsymbol{\xi}, \end{cases} \tag{2.1}$$

where $\mathbf{Z} = (Z_1, \ldots, Z_d)^T$ is a $d \times 1$ vector of observable variables with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{jk})_{1 \leq j,k \leq d}$. The dependence structure of $\mathbf{Z}$ is captured by the $K \times 1$ latent factors $\boldsymbol{f} = (f_1, \ldots, f_K)^T$ with zero mean. $\mathbf{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)^T$ is a $d \times K$ loading matrix and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)^T$ is a $d \times 1$ vector of idiosyncratic random errors, independent of $\boldsymbol{f}$. Additionally, $\mathbf{X} = (X_1, \ldots, X_p)^T$ is a $p \times 1$ vector of covariates, $\boldsymbol{m}(\cdot) = (m_1(\cdot), \ldots, m_K(\cdot))^T$ with $m_k(\cdot) : \mathbb{R}^p \to \mathbb{R}$ is the regression function corresponding to the kth latent factor and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)^T$ is a $K \times 1$ vector of random errors.

Let $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ be n independent and identically distributed (iid) observations, then $\mathbf{Z}_i = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{f}_i + \boldsymbol{\varepsilon}_i$, for $i = 1, \ldots, n$. Since only $\mathbf{Z}_i$ is observable, $\mathbf{B}$ and $\boldsymbol{f}_i$ cannot be estimated separately as they are both unobserved. For any $K \times K$ nonsingular matrix $\mathbf{D}$, one can choose $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{D}$ and $\tilde{\boldsymbol{f}}_i = \mathbf{D}^{-1}\boldsymbol{f}_i$ such that $\tilde{\mathbf{B}}\tilde{\boldsymbol{f}}_i = \mathbf{B}\boldsymbol{f}_i$. To make both $\mathbf{B}$ and $\boldsymbol{f}_i$ identifiable, we follow Fan et al. (2019) and impose the following conditions

$$\boldsymbol{\Sigma}_f = \mathbf{I}_K \quad \text{and} \quad \mathbf{B}^T\mathbf{B} \text{ is diagonal}, \tag{2.2}$$

where $\boldsymbol{\Sigma}_f$ is the covariance matrix of $\boldsymbol{f}$ and $\mathbf{I}_K$ is a $K \times K$ identity matrix. Other identifiability conditions can be found in Bai and Li (2012), Fan et al. (2013). Let $\text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_\varepsilon = (\sigma_{\varepsilon,ij})_{1 \leq i,j \leq d}$, under conditions (2.2), the covariance matrix of $\mathbf{Z}$ is $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + \boldsymbol{\Sigma}_\varepsilon$.

Remark 1. *In model (2.1), the number of latent factors K is unknown and must be learned from data. Methods for determining the number of latent factors are generally based on the eigenvalues of covariance or correlation matrix of the observable variables. Following Bai and Ng (2002), Fan et al. (2022), we define $K = rank(\mathbf{B})$. Further details on the selection of K will be provided in Section 2.5.*

In model (2.1), both d and p are allowed to diverge with the sample size n, and $\boldsymbol{\xi}$ may exhibit skewness or heavy tails. When $p \gg n$, directly fitting the factor-covariate model would be problematic due to the curse of dimensionality, leading to the simultaneous challenges to computational expediency, statistical accuracy, and algorithmic stability. These concerns, coupled with the heterogeneity nature of ultrahigh dimensional data, motivate the development of a robust screening procedure for the factor-covariate model.

2.2 Screening with a Conditional Quantile-based Index

The goal of this section is to identify a sparse set of ultrahigh dimensional covariates $\mathbf{X}$ that are relevant for modeling the conditional quantile of multiple latent factors $\boldsymbol{f}$. We advocate a quantile-adaptive screening procedure that allows the sparse set to vary across quantiles. Note that the multivariate conditional quantile function is not a trivial extension of its univariate counterpart, as the notion of multivariate quantile function is not uniquely defined. To address, given a quantile $\tau \in (0,1)$, we define the set of active variables as

$$\mathcal{A}_\tau = \bigcup_{k=1}^{K} \big\{1 \le j \le p : \ Q_\tau(f_k|\mathbf{X}) \text{ functionally depends on } X_j\big\}, \tag{2.3}$$

where $\mathcal{A}_{\tau,k} := \big\{1 \le j \le p : Q_\tau(f_k|\mathbf{X}) \text{ functionally depends on } X_j\big\}$ is the active set for the kth factor f_k, and $Q_\tau(f_k|\mathbf{X}) = \inf\{y \in \mathbb{R} : P(f_k \le y|\mathbf{X}) \ge \tau\}$ is the τth conditional quantile of f_k given $\mathbf{X}$. We use $\mathcal{A}_\tau^c$ and $\mathcal{A}_{\tau,k}^c$ to denote the index sets of inactive covariates for $\boldsymbol{f}$ and f_k, respectively. Denote by $|\mathcal{A}|$ the cardinality of $\mathcal{A}$, then $|\mathcal{A}_\tau| \le \sum_{k=1}^K |\mathcal{A}_{\tau,k}|$. Throughout this paper, $|\mathcal{A}_{\tau,k}|$ is assumed to be smaller than the sample size.

Remark 2. *A permutation matrix* $\mathbf{P}$ *and its inverse can be substituted in the factor model*

to yield $\mathbf{Z} = \boldsymbol{\mu} + \mathbf{B}\mathbf{P}^{-1}\mathbf{P}\boldsymbol{f} + \boldsymbol{\varepsilon}$. *The elements of* $\mathbf{P}\boldsymbol{f}$ *correspond to original factors, but in another order. This complicates the task of accurately identifying* $\mathcal{A}_{\tau,k}$, *as* $\boldsymbol{f}$ *is identified up to an invertible matrix. Hence, our focus is mainly on recovering* $\mathcal{A}_\tau$ *rather than* $\mathcal{A}_{\tau,k}$.

Denote by $F_{f_k}(\cdot)$ and $Q_\tau(f_k)$ the distribution function and unconditional quantile of f_k. f_k and X_j are independent if and only if $Q_\tau(f_k|X_j) = Q_\tau(f_k)$ holds for all $\tau \in (0,1)$. Similarly, given $\tau \in (0,1)$, if $F_{f_k}\{Q_\tau(f_k|x_j)\} = \tau$ holds for all $x_j \in \mathbb{R}_{X_j}$, i.e., $Q_\tau(f_k|X_j) = Q_\tau(f_k)$, f_k is τ-quantile independent of X_j, and $X_j \in \mathcal{A}^c_{\tau,k}$. This motivates us to develop a mean squared deviation (MSD) index, defined by

$$\mathrm{MSD}_\tau(f_k|X_j) = E_{X_j}\big\{F_{f_k}\big[Q_\tau(f_k|X_j)\big] - \tau\big\}^2 \tag{2.4}$$

to measure the τ-quantile dependence between f_k and X_j. Proposition 1 states that the MSD index possesses several appealing properties in quantifying quantile independence.

Proposition 1. *Let* $X \in \mathbb{R}_X$ *and* $Y \in \mathbb{R}_Y$ *be two continuous random variables, then*

(i) $MSD_\tau(Y|X) = E_X\{F_Y(Q_\tau(Y|X)) - F_Y(Q_\tau(Y))\}^2 = \int\{F_Y(Q_\tau(Y|x)) - \tau\}^2 dF(x)$.

(ii) It holds that $MSD_\tau(Y|X) = 0$ *for all* $\tau \in (0,1)$ *if and only if* X *and* Y *are independent.*

(iii) For a given $\tau \in (0,1)$, $0 \le MSD_\tau(Y|X) \le max\{\tau^2, (1-\tau)^2\}$, *and* $MSD_\tau(Y|X) = 0$ *if and only if* $Q_\tau(Y|X) = Q_\tau(Y)$ *almost surely.*

(iv) $MSD_\tau(Y|X)$ *is invariant under monotone variable transformation, that is, for* $a, b \in \mathbb{R}$ $(b \neq 0)$ *and any strictly monotone transformation* $g(\cdot)$, $MSD_\tau(Y|X) = MSD_\tau\{g(Y)|a+bX\}$ *if* $g(\cdot)$ *is nondecreasing, and* $MSD_\tau(Y|X) = MSD_{1-\tau}\{g(Y)|a+bX\}$ *if* $g(\cdot)$ *is nonincreasing.*

Proof of Proposition 1 is provided in the Supplementary Materials. Result (i) shows

that the MSD index, resembling a mean squared error, is fully nonparametric and model-free. This together with result (ii) motivates us to utilize this index to screen out inactive covariates at the interested quantiles. In general, $\mathrm{MSD}_\tau(Y|X) \neq \mathrm{MSD}_\tau(X|Y)$. But when X and Y are jointly normal, $\mathrm{MSD}_\tau(X|Y) = \mathrm{MSD}_\tau(Y|X)$. Result (iv) indicates that the MSD index is invariant under strictly increasing monotone variable transformations, whereas such property is not shared by most of the screening indices (e.g., Pearson correlation and distance correlation). This invariance property indicates that the MSD index is robust against model misspecification. Proposition S1 of the Supplementary Materials further elucidates specific properties of the MSD index under bivariate Gaussian copula distribution.

Intuitively, based on (2.3), we can employ

$$\mathrm{MSD}_\tau(\boldsymbol{f}|X_j) := \max_{1\le k\le K} \mathrm{MSD}_\tau(f_k|X_j) \tag{2.5}$$

to measure the importance of X_j for $\boldsymbol{f}$. To perform feature screening, we use the sample counterpart, $\widehat{\mathrm{MSD}}_\tau(\boldsymbol{f}|X_j)$, defined in Section 2.4. Hence, the set of active features is obtained as

$$\hat{\mathcal{A}}_\tau = \big\{1 \le j \le p : \widehat{\mathrm{MSD}}_\tau(\boldsymbol{f}|X_j) \ge \eta\big\}, \tag{2.6}$$

where $\eta \ge 0$ is a prespecified threshold. We refer to this procedure as MSD-based screening (MSDS). With a proper choice of η, we show that MSDS enjoys sure screening property.

2.3 FDR Control via Derandomized Quantile Knockoffs

Next, we introduce how to enhance the performance of MSDS to achieve FDR control while maintaining a high power using knockoffs procedure. In practice, all active covariates can

be included with high probability by employing a conservative threshold. However, there is no guarantee of the accuracy of $\hat{\mathcal{A}}_\tau$ with no falsely selected covariates. This section focuses on identifying $\hat{\mathcal{A}}_\tau$ with a theoretically guaranteed error rate. To be more specific, we introduce a quantile knockoffs approach to select active variables while controlling FDR_τ at a prespecified target level $\alpha \in (0, 1)$, where

$$\mathrm{FDR}_\tau := E(\mathrm{FDP}_\tau) \quad \text{and} \quad \mathrm{FDP}_\tau := \frac{|\hat{\mathcal{A}}_\tau \cap \mathcal{A}_\tau^c|}{|\hat{\mathcal{A}}_\tau| \vee 1}.$$

The FDP_τ represents the false discovery proportion at a given quantile τ. Denote by $V(\tau) = |\hat{\mathcal{A}}_\tau \cap \mathcal{A}_\tau^c|$ the number of false discoveries with respect to τ, then the per family error rate (PFER) is the expected number of $V(\tau)$, that is, $\mathrm{PFER}_\tau = E[V(\tau)]$. We refer to Barber and Candès (2015) and Candès et al. (2018) for knockoffs framework. More details of the knockoffs can also be found in the Supplementary Materials. Given a $\tau \in (0, 1)$, we propose to measure the dependence between X_j and $\boldsymbol{f}$ by the following knockoff statistic

$$W_{j,\tau} := \mathrm{MSD}_\tau(\boldsymbol{f}|X_j) - \mathrm{MSD}_\tau(\boldsymbol{f}|\tilde{X}_j), \quad j = 1, \ldots, p, \tag{2.7}$$

where $\tilde{X}_j$ is a knockoff copy of X_j, $\mathrm{MSD}_\tau(\boldsymbol{f}|X_j)$ and $\mathrm{MSD}_\tau(\boldsymbol{f}|\tilde{X}_j)$ are defined in 2.5. The empirical counterpart of the knockoff statistic (2.7) is

$$\widehat{W}_{j,\tau} := \widehat{\mathrm{MSD}}_\tau(\hat{\boldsymbol{f}}|X_j) - \widehat{\mathrm{MSD}}_\tau(\hat{\boldsymbol{f}}|\tilde{X}_j), \quad j = 1, \ldots, p. \tag{2.8}$$

In the Supplementary Materials, we introduce the MSD Knockoffs (MSDK) procedure, which serves as the foundation for the DMSDK procedure outlined below, and demonstrate in Theorem S1 that the MSDK procedure effectively controls the FDR.

Derandomized MSD knockoffs (DMSDK) procedure. The MSDK procedure based on the model-X knockoffs is a randomized procedure. To yield more stable results and enhance statistical power, we are inspired by the derandomized idea (Ren et al., 2023) and introduce a derandomized quantile knockoffs that aggregate the results across multiple quantile knockoffs realizations. By using this method, the FDR along with the PFER and k-FWER are verified to be controlled, as demonstrated in Theorem 3 of Section 3. Details of the DMSDK procedure are summarized in Algorithm 1.

Algorithm 1: Derandomized MSD Knockoffs (DMSDK) procedure

Input : Matrix of covariates $\mathbf{X} \in \mathbb{R}^{n\times p}$; latent factors $\boldsymbol{f} \in \mathbb{R}^{n\times K}$; number of realizations T; selection threshold η_0; integer $v \geq 1$; a quantile level $\tau \in (0,1)$.

for $t = 1, \ldots T$ **do**

  i. Generate a knockoff copy $\tilde{\mathbf{X}}^{(t)}$.

  ii. Run the MSDK base procedure with $\tilde{\mathbf{X}}^{(t)}$ and obtain $\widehat{\mathbf{W}}_\tau^{(t)} = (\widehat{W}_{1,\tau}^{(t)}, \ldots, \widehat{W}_{p,\tau}^{(t)})^T$.

  iii. Rank the features by magnitudes of $\widehat{\mathbf{W}}_\tau^{(t)}$, that is, $|\widehat{W}_{r_1,\tau}^{(t)}| \geq \cdots \geq |\widehat{W}_{r_p,\tau}^{(t)}|$ for some permutation $r_1, \ldots, r_p$.

  iv. Construct a stopping criterion: $\mathcal{T}_\tau^{(t)}(v) = \min\left\{1 \leq k \leq p : \sum_{j=1}^{k} I\{\widehat{W}_{r_j,\tau}^{(t)} < 0\} \geq v\right\}$, and obtain the selection set $\hat{S}_\tau^{(t)}(v) = \left\{r_j : j < \mathcal{T}_\tau^{(t)}(v), \widehat{W}_{rj,\tau} > 0\right\}$.

end

Calculate the selection probability $\Pi_{j,\tau}(v) = \frac{1}{T}\sum_{t=1}^{T} I\{j \in \hat{S}_\tau^{(t)}(v)\}$.

Output: selection set $\hat{\mathcal{A}}_\tau(v, \eta_0) = \{1 \leq j \leq p : \Pi_{j,\tau}(v) \geq \eta_0\}$.

In Algorithm 1, η_0 controls how many times a variable needs to be selected to appear in the final selection set. The larger η_0 is, the fewer variables enter the selection set. Following Ren et al. (2023), we set $\eta_0 = 0.5$ throughout this paper. The parameter v is chosen to control the PFER, and the PFER control holds regardless of the choice of T.

2.4 A Generic Robust Estimation and Selection Procedure

We now propose a MSD based selection procedure named MSD-Select under the factor-covariate model which includes the following steps: (i) Robust estimation of latent factors;

(ii) Feature screening via MSD index; (iii) Quantile-adaptive FDR control (including PFER and k-FWER control) via DMSDK method. Details are summarized in Algorithm 2.

Algorithm 2: MSD-Select Procedure

Input : Observed data $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{id})^T \in \mathbb{R}^d$; covariates $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^T \in \mathbb{R}^p$ for $i = 1, \ldots, n$; integers $K \geq 1$,$T \geq 1$, $v \geq 1$, $\kappa_n < n/2$; bandwidth $h > 0$; a prespecified level $\alpha \in (0, 1)$; selection threshold $\eta_0 > 0$; a quantile level τ.

Procedure:

Step 1 (Estimation). Obtain an estimator $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, for example, the sample mean and covariance matrix or their robust versions. Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \hat{\lambda}_K$ be the top K eigenvalues of $\hat{\boldsymbol{\Sigma}}$, and $\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \ldots, \hat{\boldsymbol{v}}_K$ be their corresponding eigenvectors. Define $\hat{\mathbf{B}} = (\tilde{\lambda}_1^{1/2}\hat{\boldsymbol{v}}_1, \ldots, \tilde{\lambda}_K^{1/2}\hat{\boldsymbol{v}}_K) \in \mathbb{R}^{d\times K}$, where $\tilde{\lambda}_k = \max(\hat{\lambda}_k, 0)$. Let $\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_d \in \mathbb{R}^K$ be the d rows of $\hat{\mathbf{B}}$, and $\varpi > 0$ be the robustification parameter depends on both n and d, obtain

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

Step 2 (Screening). For $j = 1, \ldots, p$, construct the sample MSD statistic by using convolution type kernel smooth quantile regression approach, $\hat{\delta}_j^\tau := \widehat{\mathrm{MSD}}_\tau(\hat{\boldsymbol{f}}|X_j)$. Select the top κ_n covariates, i.e., $\hat{\mathcal{A}}_\tau = \{j : \hat{\delta}_j^\tau \text{ is among the } \kappa_n\text{th largest}\}$.

Step 3 (Selection). Construct second-order knockoff features $\tilde{\mathbf{X}}_{\hat{\mathcal{A}}_\tau}$ for $\mathbf{X}_{\hat{\mathcal{A}}_\tau}$. For all $j \in \hat{A}_\tau$, compute $\widehat{W}_{j,\tau} = \widehat{\mathrm{MSD}}_\tau(\hat{\boldsymbol{f}}|X_{\hat{\mathcal{A}}_\tau,j}) - \widehat{\mathrm{MSD}}_\tau(\hat{\boldsymbol{f}}|\tilde{X}_{\hat{\mathcal{A}}_\tau,j})$. Run DMSDK procedure proposed in Algorithm 1. Calculate the stopping criterion, $\mathcal{T}_\tau^{(t)}(v) = \min\left\{1 \leq k \leq \kappa_n : \sum_{j=1}^k I\{\widehat{W}_{r_j,\tau}^{(t)} < 0\} \geq v\right\}$, and obtain the final selection set $\hat{\mathcal{A}}_\tau(v, \eta_0) = \{j \in \hat{\mathcal{A}}_\tau : \Pi_{j,\tau}(v) \geq \eta_0\}$ based on the selection probability $\Pi_{j,\tau}(v)$ given in Algorithm 1.

Output : $\hat{\boldsymbol{f}} = (\hat{f}_1, \ldots, \hat{f}_K)^T$, $\hat{\mathbf{B}} \in \mathbb{R}^{d\times K}$ and final selection set $\hat{\mathcal{A}}_\tau(v, \eta_0)$.

Remark 3. *In Algorithm 2, a fixed number of factors, K, must be specified in advance. Since the true value of K is typically unknown, Section 2.5 will propose a novel and consistent approach for determining an appropriate K, which can then be used as input for Algorithm 2. Other reliable methods for determining K are also recommended.*

In step 1, we borrow the idea of Fan et al. (2019) to robustly estimate $\boldsymbol{\mu} = (\mu_j)_{1\leq j\leq d}$ and $\boldsymbol{\Sigma} = (\sigma_{jk})_{1\leq j,k\leq d}$ of $\mathbf{Z}$ by using a Huber loss, with $\mu_j = E(X_j)$ and $\sigma_{jk} = E(X_jX_k) - \mu_j\mu_k$. The reason why we take Huber loss largely lies in the good property of Huber estimator

for heavy-tailed data under mild moment conditions. It is shown that a sub-Gaussian type deviation bound is allowed by employing Huber estimator with a properly diverging parameter (Fan et al., 2019). The Huber loss $\psi_\varpi(\cdot)$ (Huber, 1964) is defined as

$$\psi_\varpi(u) = \begin{cases} u^2/2, & \text{if } |u| \le \varpi, \\ \varpi|u| - \varpi^2/2, & \text{if } |u| > \varpi, \end{cases} \tag{2.9}$$

where $\varpi > 0$ is a robustification parameter to tradeoff between bias and robustness. Then

$$\hat{\mu}_j = \arg\min_{\theta\in\mathbb{R}} \sum_{i=1}^{n} \psi_{\varpi_j}(Z_{ij} - \theta), \tag{2.10}$$

$$\hat{\sigma}_{jk} = \hat{\theta}_{jk} - \hat{\mu}_j\hat{\mu}_k \quad \text{with} \quad \hat{\theta}_{jk} = \arg\min_{\theta\in\mathbb{R}} \sum_{i=1}^{n} \psi_{\varpi_{jk}}(Z_{ij}Z_{ik} - \theta), \tag{2.11}$$

where $\varpi_j > 0$ and $\varpi_{jk} > 0$ are robustification parameters. Equations (2.10) and (2.11) yield the robust mean estimator $\hat{\boldsymbol{\mu}}$ and covariance estimator $\hat{\boldsymbol{\Sigma}}$, respectively.

In step 2, denote by $\delta_{k,j}^\tau = \mathrm{MSD}_\tau(f_k|X_j)$ and $\hat{\delta}_{k,j}^\tau = \widehat{\mathrm{MSD}}_\tau(\hat{f}_k|X_j)$ the population and sample MSD indices between the jth covariate and the kth latent factor, respectively; denote by $\delta_j^\tau = \mathrm{MSD}_\tau(\boldsymbol{f}|X_j)$ and $\hat{\delta}_j^\tau = \widehat{\mathrm{MSD}}_\tau(\hat{\boldsymbol{f}}|X_j)$ the population and sample MSD indices between the jth covariate and the multiple latent factors, respectively. By definition (2.5), $\delta_j^\tau = \max_{1\le k\le K} \delta_{k,j}^\tau$ and $\hat{\delta}_j^\tau = \max_{1\le k\le K} \hat{\delta}_{k,j}^\tau$. To obtain the empirical $\hat{\delta}_{j,k}^\tau$, we use B-spline approximation and employ the convolution type kernel smoothing quantile regression to further reduce the variability of standard quantile regression estimators (Fernandes et al., 2021). More details can be found in Section \$5 of the Supplementary Materials. In practice, we often rank all the covariates according to $\hat{\delta}_j^\tau$ and keep the top $\lfloor n/\log(n)\rfloor$ covariates, i.e., $\kappa_n = \lfloor n/\log(n)\rfloor$, where $\lfloor a\rfloor$ denotes the integer part of a.

2.5 Determining the Number of Latent Factors

This section introduces a method for determining K, a critical step for both Algorithm 2 and for estimating the loading matrix and factors in model (2.1).

The number K can be generally estimated by eigenvalue based criteria (Onatski, 2010; Ahn and Horenstein, 2013; Fan et al., 2022), information criteria (Bai and Ng, 2002; Bai et al., 2018), cross validation (Owen and Wang, 2016) and parallel analysis (Dobriban and Owen, 2019). Let $\hat{\lambda}_j$ denote the jth largest eigenvalue of the sample covariance matrix. Onatski (2009) proposed to estimate the number of factors by $\hat{K}_{\mathrm{ON}} = \arg\max_{r_{\min}<i\le r_{\max}}(\hat{\lambda}_i-\hat{\lambda}_{i+1})/(\hat{\lambda}_{i+1}-\hat{\lambda}_{i+2})$, where $r_{\min}$ and $r_{\max}$ are the predetermined lower and upper bounds of K. Wang (2012) estimated the number of factors by the ratios of two adjacent eigenvalues, $\hat{K}_{\mathrm{ER}} = \arg\max_{1\le i\le r_{\max}} \hat{\lambda}_i/\hat{\lambda}_{i+1}$. Ahn and Horenstein (2013) considered using $\hat{K}_{\mathrm{GR}} = \arg\max_{1\le i\le r_{\max}} \log(V_{i-1}/V_i)/\log(V_i/V_{i+1})$, where $V_i = \sum_{j=i+1}^{d} \hat{\lambda}_j$. A drawback of the above covariance-based methods is that they do not take into account the scales of the observed variables, thus can be inconsistent. Another option is to follow the idea of Fan et al. (2022), who used sample correlation matrix. Let $\mathbf{R}$ be the $d \times d$ dimensional correlation matrix of $\mathbf{\Sigma}$, that is, $\mathbf{R} = [\mathrm{diag}(\mathbf{\Sigma})]^{-1/2}\mathbf{\Sigma}[\mathrm{diag}(\mathbf{\Sigma})]^{-1/2}$, where $\mathrm{diag}(\mathbf{\Sigma})$ is the diagonal matrix obtained by replacing the off-diagonal components of $\mathbf{\Sigma}$ with zeros. Then the estimator of K is

$$\hat{K}_{\mathrm{ACT}} = \max\left\{j : \hat{\lambda}_j^C(\mathbf{R}) > 1 + \sqrt{d/(n-1)}\right\}, \tag{2.12}$$

where $\hat{\lambda}_j^C(\mathbf{R}) = -1/\gamma_j(\hat{\lambda}_j(\mathbf{R}))$ is a bias corrected estimator of $\lambda_j(\mathbf{R})$ with $\gamma_j(z) = -(1-$

$(d-j)/(n-1))z^{-1}+(n-1)^{-1}\big[\sum_{l=j+1}^{d}(\hat{\lambda}_l(\mathbf{R})-z)^{-1}+((3\hat{\lambda}_j(\mathbf{R})+\hat{\lambda}_{j+1}(\mathbf{R}))/4-z)^{-1}\big]$. The idea behind equation (2.12) is that $K \geq \max\{1 \leq j \leq d : \lambda_j(\mathbf{R}) > 1\}$, i.e., the number of eigenvalues greater than 1 should not exceed the number of factors. However, this approach requires the signal of factors to be strong. For those observable variables with weak factors, the estimated $\hat{\lambda}_j^C$ could be quite smaller than 1 and be unstable, making the estimated $\hat{K}$ be much smaller than K. To detect weak factors and yield more stable results, we propose in Algorithm 3 a modified bootstrap-based method for choosing K.

Algorithm 3: A bootstrap-based eigenvalue method to determine K

Input : Observable matrix $\mathbf{Z} \in \mathbb{R}^{n\times d}$; a diagonal weight matrix $\mathbf{V} \in \mathbb{R}^{n\times n}$; number of bootstrap samples G.

for $g = 1, \ldots G$ **do**

  i. Let v_i's be diagonal elements of $\mathbf{V}$, $\bar{\mathbf{Z}} = n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i$ and $\mathbf{Z}^*$ be the centered sample matrix. Obtain the sample covariance matrix $\hat{\mathbf{\Sigma}}_b^g = n^{-1}\sum_{i=1}^{n} v_i(\mathbf{Z}_i-\bar{\mathbf{Z}})(\mathbf{Z}_i-\bar{\mathbf{Z}})^T = n^{-1}\mathbf{Z}^{*T}\mathbf{V}\mathbf{Z}^*$ and correlation matrix $\hat{\mathbf{R}}_b^g$, respectively.

  ii. Calculate the bias corrected eigenvalues $\hat{\lambda}_j^C(\hat{\mathbf{R}}_b^g)$ for $j = 1, \ldots, d$.

  iii. Obtain $\hat{K}^g = \max\big\{j : \hat{\lambda}_j^C(\hat{\mathbf{R}}_b^g) > 1 + \sqrt{d/(n-1)}\big\}$.

end

Output: $\hat{K}_b = \text{Mode}\{\hat{K}^g : g = 1, \ldots, G\}$, namely, the mode of $\{\hat{K}^g\}_{g=1}^G$.

In Algorithm 3, the weight matrix $\mathbf{V}$ can be chosen by drawing each diagonal element independently from exponential distribution exp(1), this ensures $\hat{\mathbf{\Sigma}}_b$ positive semidefinite. Other distribution families can also be considered for constructing $\mathbf{V}$, while we do not specify here. The number of factors K is approximated as the mode of the resulting set. Alternatively, one can choose the average or the median of the set $\{\hat{K}^g : g = 1, \ldots, G\}$.

Under additional assumptions that $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\varepsilon}) = \mathbf{\Psi} > \mathbf{0}_{p\times p}$, where $\mathbf{\Psi}$ is diagonal or more generally satisfies $\max_{i\leq d}\sum_{j\leq d}|\sigma_{\varepsilon,ij}|^q = o(d)$ for some $q \in [0,1]$, the proposed method provides a reliable estimator of the number of factors. Specifically, Theorem 1 establishes that the estimator $\hat{K}_b$ obtained from Algorithm 3 consistently estimates K

under moderate regularity conditions when K is fixed. Theorem S2 of the Supplementary Materials further extends this result to the case where K diverges as $d \to \infty$.

Theorem 1. *For the factor model (2.1) satisfying Conditions (i)-(v) in §7.2 of the Supplementary Materials, when* $\lambda_K(\mathbf{R}) > 1 + \sqrt{\omega}$ *with* $\omega \in (0, \infty)$, *we have* $P\big(\hat{K}_b = K\big) \to 1$, *as* $n, d \to \infty$.

3. Theoretical Results

To fully understand the proposed procedures, we successively establish the theoretical results through several steps, starting with an oracle procedure that assumes the loading matrix $\mathbf{B}$ is known and the factors $\{\boldsymbol{f}_i\}_{i=1}^n$ are observable, which serves as a heuristic device. Second, we keep $\mathbf{B}$ known but treat $\boldsymbol{f}$ as latent, and study the estimator $\hat{\boldsymbol{f}}(\mathbf{B})$. We then analyze the difference between the true $\mathbf{B}$ and its estimate $\hat{\mathbf{B}}$, which leads to the theoretical properties of $\hat{\boldsymbol{f}}(\hat{\mathbf{B}})$ when both $\boldsymbol{f}$ and $\mathbf{B}$ are unknown. For brevity, this section presents only the results based on $\hat{\boldsymbol{f}}(\hat{\mathbf{B}})$, while additional results are provided in the Supplementary Materials. The screening statistics constructed from $\hat{\boldsymbol{f}}(\hat{\mathbf{B}})$ are denoted by $\hat{\delta}_{k,j}$ and $\hat{\delta}_j$, with the corresponding selected subsets $\hat{\mathcal{A}}_{\tau,k}$ and $\hat{\mathcal{A}}_\tau$. Further denote $[K] = \{1, \ldots, K\}$. To facilitate technical derivations, we first impose the following regularity conditions.

(C1) (Condition on the conditional quantile) For $k \in [K]$, the conditional quantile function $Q_\tau(f_k|X_j)$ belongs to $\mathcal{H}_r$, where $\mathcal{H}_r$ is the class of functions defined on $[0, 1]$ whose mth derivative satisfies a Lipschitz condition of order v: $|h_{kj}^{(m)}(s) - h_{kj}^{(m)}(t)| \le C|s - t|^v$, for some positive constant C, $s, t \in [0, 1]$, where m is a nonnegative integer and $v \in (0, 1]$

satisfies $r = m + v > 0.5$.

(C2) (Condition on latent factors) The conditional density $g_{f_k|X_j}(t)$ is bounded away from 0 and ∞ on $[Q_\tau(f_k|X_j) - \xi_k, Q_\tau(f_k|X_j) + \xi_k]$, for some $\xi_k > 0$, uniformly in X_j. The marginal density $g_{f_k}(t)$ is bounded away from 0 and ∞, for $k = 1, \ldots, K$.

(C3) (Condition on marginal covariates) The marginal density function g_j of X_j, for $j \in [p]$, are uniformly bounded away from 0 and ∞.

(C4) (Condition on signal size) $\min_{j \in \mathcal{A}_{\tau,k}} \delta_{k,j} \geq 2c_1 n^{-\kappa}$ for some $0 \leq \kappa < 1/2$ and some positive constant c_1.

(C5) (Condition on the basis function) The number of basis functions s_n satisfies $s_n^{-r} n^{\kappa} = o(1)$ and $s_n n^{2\kappa - 1} = o(1)$ as $n \to \infty$.

(C6) (Condition on the kernel function) The kernel function $\mathcal{K}(\cdot)$ is integrable, twice differentiable with bounded first and second derivatives, satisfying $\int \mathcal{K}(u)du = 1$ and $0 < \int_0^\infty \mathcal{K}(u)\{1 - \mathcal{K}(u)\}du < \infty$.

(C7) (Condition on the bandwidth) The bandwidth $0 < h < 1$, and $h = O\big((s_n/n)^{1/4}\big)$.

(C8) (Condition on the idiosyncratic error $\boldsymbol{\varepsilon}$ and loading matrix $\mathbf{B}$) The idiosyncratic errors $\varepsilon_1, \ldots, \varepsilon_d$ are mutually independent, and there exist constants $C_\varepsilon, c_\varepsilon > 0$ such that $c_\varepsilon \leq \min_{1 \leq j \leq d} \sigma_{\varepsilon,jj}^{1/2} \leq \max_{1 \leq j \leq d} \left(E\varepsilon_j^4\right)^{1/4} \leq C_\varepsilon$. There exist constants $c_l, c_u > 0$ such that $\lambda_{\min}(d^{-1}\mathbf{B}^T\mathbf{B}) \geq c_l$ and $\|\mathbf{B}\|_{\max} \leq c_u$.

(C9) (Condition on dimensionality) (n, d) satisfies that $n^{4\kappa - 1} \log n = o(d)$ as $n, d \to \infty$ for some $0 \leq \kappa < 1/2$.

(C10) (Pervasiveness) There exist positive constants c_{B1}, c_{B2} and c_{B3} such that $c_{B1} d \leq$

$\bar{\lambda}_k - \bar{\lambda}_{k+1} \leq c_{B2}d$ for $k = 1, \ldots, K$ with $\bar{\lambda}_{K+1} := 0$, and $\|\mathbf{\Sigma}_\varepsilon\| \leq c_{B3} < \bar{\lambda}_K$, where $\bar{\lambda}_k$ is the top kth eigenvalues of $\mathbf{B}\mathbf{B}^{\mathrm{T}}$.

Condition (C1) assumes that the conditional quantile function $Q_\tau(f_k|X_j)$ belongs to a class of smooth functions. This condition is standard for nonparametric spline approximation. Condition (C2) is a standard condition on random errors in the theory for quantile regression. It relaxes the usual sub-Gaussian assumptions that are needed in literature on high dimensional inference. Condition (C3) is similar to Condition (B) of Fan et al. (2011) and Condition (C4) of He et al. (2013). Note that Condition (C3) is not restrictive when X_j is supported on a bounded interval, say [0,1]. When X_j has an unbounded support (e.g., normal), we can view X_j as coming from a truncated distribution. Specifically, if X_j contains outliers or follows a heavy-tailed distribution, we can improve performance by removing the outliers or transforming X_j to a uniform distribution on $[0, 1]$. Condition (C4) assumes that the active covariates at quantile level τ have strong enough marginal signals, a smaller κ indicates a stronger marginal signal. This condition is crucial as it ensures marginal utilities carry information about the covariates in the active set. Condition (C5) describes how fast the number of basis functions is allowed to grow with the sample size. Conditions (C6) and (C7) guarantee that the smoothing bias can be ignored when sample size is large enough. Condition (C8) is a standard assumption in factor model. Condition (C9) enables the factor-covariate model to be high-dimensional with respect to both d and p. Condition (C10) is required for high-dimensional spiked covariance model with the first several eigenvalues well separated and significantly larger than the rest.

Theorem 2. *Under Conditions (C1)-(C10), for a given* $\tau \in (0,1)$, *if* $n^{-1}\log(nd) \to 0$, $K/d \to 0$, *and* $\varpi \asymp \sqrt{d/\log n}$ *as* $n, d \to \infty$, *then*

(i) for any $C > 0$, *there exist positive constants* c_2 *and* c_3 *such that for* n *sufficiently large*

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

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

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

where $|\mathcal{A}_\tau|$ *is the cardinality of* $\mathcal{A}_\tau$. *Further assuming that* $\log n = o(n^{1-4\kappa} + s_n^{-2} n^{1-2\kappa})$, *we have* $P(\mathcal{A}_\tau \subset \hat{\mathcal{A}}_\tau) \to 1$ *as* $n \to \infty$.

(iii) (Rank consistency property) Replace Condition (C4) with $\min_{j\in\mathcal{A}_{\tau,k}} \delta_{k,j} - \max_{l\in\mathcal{A}_{\tau,k}^c} \delta_{k,l} \ge 2c_1 n^{-\kappa}$, *for some* $0 \le \kappa < 1/2$ *and some positive constant* c_1, *we have*

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

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* $0 < \kappa < 1/4$ *and* $s_n^2 n^{2\kappa-1} = o(1)$, *then*

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

The sure screening and rank consistency properties based on the oracle procedure, where the true factors $\boldsymbol{f}$ and loading matrix $\mathbf{B}$ are assumed known, are straightforward to derive and are provided in Theorems S3-S4 of the Supplementary Materials. Building upon these

oracle results, together with the consistency of $\hat{\boldsymbol{f}}(\mathbf{B})$ for $\boldsymbol{f}$, the consistency of $\hat{\mathbf{B}}$ for $\mathbf{B}$, and the consistency of $\hat{\boldsymbol{f}}(\hat{\mathbf{B}})$ for $\boldsymbol{f}$ (as shown in Lemma S14 and Lemmas S19-S20), we derive Theorem 2, which establishes the sure screening and rank consistency properties when both $\mathbf{B}$ and $\boldsymbol{f}$ are unknown. Specifically, Theorem 2 (i) suggests that we can handle the dimensionality $\log(np) = o(n^{1-4\kappa} + s_n^{-2}n^{1-2\kappa})$. This dimensionality depends on the number of basis functions s_n and the strength of the marginal signals $n^{-\kappa}$. If we take $s_n = n^{1/(2r+1)}$ (the optimal rate for spline approximation), then for $\kappa < \min(1/4, (r-1/2)/(2r+1))$, we can handle ultrahigh dimensionality, that is, p can grow at the exponential rate. In Theorem 2 (ii), when we take $\kappa < 1/4$ and $s_n^2 n^{2\kappa-1} = o(1)$, the condition $\log n = o(n^{1-4\kappa} + s_n^{-2}n^{1-2\kappa})$ typically holds, which guarantees that $P(\mathcal{A}_\tau \subset \hat{\mathcal{A}}_\tau) \to 1$ as $n \to \infty$, that is, all active covariates can be selected with high probability. The rank consistency result in Theorem 2 (iii) strengthens the sure screening property in (ii) by imposing a stronger assumption on the signal gap between active and inactive covariates, i.e, $\min_{j\in\mathcal{A}_{\tau,k}} \delta_{k,j} - \max_{l\in\mathcal{A}_{\tau,k}^c} \delta_{k,l} \geq 2c_1 n^{-\kappa}$. Provided that $\log p = o(n^{1-4\kappa} + s_n^{-2}n^{1-2\kappa})$ and $\log n = o(n^{1-4\kappa} + s_n^{-2}n^{1-2\kappa})$, the active covariates are always ranked ahead of inactive ones with high probability.

Remark 4. *Here we make some remarks on the estimation of* $\hat{\boldsymbol{f}}$ *and* $\hat{\mathbf{B}}$*, along with their convergence rates. (i) Existing approaches for estimating the factor matrix* $\mathbf{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n)^T$ *and the loading matrix* $\mathbf{B}$ *typically assume that the errors in (2.1) are i.i.d. sub-Gaussian. Under this assumption,* $\mathbf{F}$ *and* $\mathbf{B}$ *can be obtained via constrained least squares (Bai, 2003; Fan et al., 2013). The resulting estimators* $\widetilde{\mathbf{F}}$ *and* $\widetilde{\mathbf{B}}$ *are given by* $\widetilde{\mathbf{B}} = n^{-1}\widetilde{\mathbf{F}}^T\mathbf{X}$*, where the columns of* $\widetilde{\mathbf{F}}/\sqrt{n}$ *are the eigenvectors corresponding to the largest* K *eigenval-*

ues of $\mathbf{X}\mathbf{X}^T$. *Fan et al. (2013) showed that* $\widetilde{\mathbf{F}}$ *consistently estimates* $\mathbf{F}$ *up to a rotation. Under conditions such as* $\sqrt{d}\log d = o(n)$, *the optimal rate for* $\widetilde{\boldsymbol{f}}_i$ *is* $1/\sqrt{d}$ *(Bai, 2003; Fan et al., 2013; Li et al., 2018). By comparison, our estimator* $\hat{\boldsymbol{f}}_i(\mathbf{B})$ *achieves the rate* $\sqrt{\log n/d}$ *with a properly chosen robustification parameter* ϖ. *Although slightly slower, this rate reflects the tradeoff for robustness, as our method is designed to handle heavy-tailed errors by relaxing the sub-Gaussian assumption and requiring only finite fourth moments* $(E(\varepsilon^4) < \infty)$. *(ii) Our robust procedure begins with constructing a covariance estimator* $\hat{\mathbf{\Sigma}}$, *which yields robust loadings* $\hat{\boldsymbol{b}}_j$*'s. Based on* $\hat{\mathbf{B}} = (\hat{\boldsymbol{b}}_1, \ldots, \hat{\boldsymbol{b}}_d)^T$, $\hat{\boldsymbol{f}}_i$ *is estimated robustly via Huber regression, denoted as* $\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})$. *Under the condition* $\log d = o(n)$, *traditional estimators* $\widetilde{\boldsymbol{b}}_j$*'s converge at the rate of* $O_p(\sqrt{\log d/n})$, *whereas our estimator achieves the rate of* $O_p(\sqrt{\log(nd)/n} + 1/\sqrt{d})$, *as shown in Lemma S19 of the Supplementary Materials. This difference stems from the employment of the adaptive Huber covariance estimator (Fan et al., 2019), which requires weaker assumptions than the sample covariance and is broadly applicable to heavy-tailed settings. (iii) The estimation error of* $\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})$ *can be decomposed as* $\|\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}}) - \boldsymbol{f}_i\|_2 \le \|\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}}) - \hat{\boldsymbol{f}}_i(\mathbf{B})\|_2 + \|\hat{\boldsymbol{f}}_i(\mathbf{B}) - \boldsymbol{f}_i\|_2$. *This implies that the convergence rate of* $\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})$ *depends jointly on the estimation accuracy of* $\hat{\boldsymbol{f}}_i(\mathbf{B})$ *and* $\hat{\mathbf{B}}$. *Improvements in either component lead to faster convergence of* $\hat{\boldsymbol{f}}_i(\hat{\mathbf{B}})$.

Theorem 3 further establishes that the proposed DMSDK procedure controls both the PFER and the k-FWER.

Theorem 3. *(PFER and k-FWER control) Consider the DMSDK procedure (Algorithm 1) with a base procedure satisfying* $PFER_\tau \le v$, *under the condition* $P(\Pi_{j,\tau}(v) \ge \eta_0) \le$

$\gamma E(\Pi_{j,\tau}(v))$ for every $j \in \mathcal{A}_\tau^c$, we have

(i) $E[V(\tau)] \leq \gamma v$.

(ii) Further assuming that $P(V(\tau) \geq k) \leq \ell E[V(\tau)]/k$ *for each* $k \geq 1$, *we have*

$$P(V(\tau) > k) \leq \ell\gamma v/k.$$

The base procedure we adopt here is the v-quantile knockoffs, as specified in the Supplementary Materials. From Theorem 3 (i), the upper bound for PFER_τ is relatively conservative. Take $\eta_0 = 0.5$, we have $E[V(\tau)] \leq 2v$, implying $\gamma = 2$. This bound can be improved by introducing additional assumptions. Specifically, if the knockoff variables are conditionally iid, the number of selections $T\Pi_{j,\tau}$ follows a binomial distribution conditional on $\mathbf{X}$ and $\boldsymbol{f}$ (by the law of larger numbers), that is, $T\Pi_{j,\tau}|\mathbf{X}, \boldsymbol{f} \sim Bin(T, P(j \in \hat{S}_\tau^{(1)}|\mathbf{X}, \boldsymbol{f}))$, where $\hat{S}_\tau^{(1)}$ is defined in Algorithm 1. Consequently, the PFER_τ is calculated directly via

$$E[V(\tau)] = E\Big[\sum_{j\in\mathcal{A}_\tau^c} P(j \in \hat{\mathcal{A}}_\tau(v, \eta_0)|\mathbf{X}, \boldsymbol{f})\Big] = E\Big[\sum_{j\in\mathcal{A}_\tau^c} P(T\Pi_{j,\tau} \geq T\eta_0|\mathbf{X}, \boldsymbol{f})\Big].$$

Take $T = 3$ and let $p_j = P(j \in \hat{S}_\tau^{(1)}|\mathbf{X}, \boldsymbol{f})$, then $E[V(\tau)] \leq 1.125v$. We refer to Ren et al. (2023) for more details on choosing parameters v, k, ℓ, and γ to help control the k-FWER.

4. Simulation Studies

In this section, we conduct simulation studies to investigate the performance of the proposed procedures, including MSDS for feature screening, and MSDK and DMSDK for FDR control. We first evaluate the sure screening property of MSDS and compare it with six other popular screening procedures in literature: quantile-adaptive sure independence screening

(QaSIS, He et al., 2013), nonparametric independence screening (NIS, Fan et al., 2011), sure independent ranking and screening (SIRS, Zhu et al., 2011), distance correlation based screening (DCSIS, Li et al., 2012), RV correlation based screening (RVSIS, Yu et al., 2023) and projection correlation based screening (PC-Screen, Liu et al., 2022). DCSIS, RVSIS and PC-Screen can handle multiple responses, while the others are intended for univariate response. To adapt QaSIS, NIS, and SIRS for multiple responses, we adopt the same construction logic as in MSDS, but replace MSD with other screening statistics in (2.5). For MSDS, the robustification parameters involved in the Huber loss are selected by five fold cross validation as in Fan et al. (2019). The kernel function is set to be the triangular kernel function and we take $h = \max\{0.05, \{(S + \log n)/n\}^{2/5}\}$ as suggested by He et al. (2023), where S is the number of basis. In computing MSDS, QaSIS, and NIS, we set $S = 3$. To mimic real dataset, we consider $(d, K, p, n) = (50, 3, 5000, 200)$. We also evaluate the performance of our proposed method for determining the number of latent factors (K) and compare it with several competitors mentioned in Section 2.5. Due to space limitations, results for this part are provided in the Supplementary Materials.

We first simulate two examples to evaluate screening performance and adopt the following criteria: (1) minimum model size (MMS) to include all active covariates: we report the mean of MMS with its standard error (SD), and the median of MMS with its median absolute deviation (MAD) over 200 replications; (2) $\mathcal{P}_j$: the proportion including a single X_j for a given model size $\lfloor n/\log n \rfloor$; (3) $\mathcal{P}_{all}$: the proportion including all active covariates for a given model size $\lfloor n/\log n \rfloor$. In examples 1-2, $\mathbf{X}_i$ is drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ and $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\mathbf{X})$,

where $\mathbf{\Sigma_X} = (0.5^{|i-j|})_{1\le i,j\le p}$. In example 3, $\mathbf{X}_i$ is drawn from $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ and $t_3(\mathbf{0}, \mathbf{I}_p)$. For heteroscedastic error cases, we take $\tau \in \{0.50, 0.75, 0.90\}$, otherwise, $\tau = 0.50$. Other examples, including factor-additive model and serial dependent factor-nonparametric model, are provided in the Supplementary Materials.

Example 1: factor-linear model. Consider a three factor model $\mathbf{Z}_i = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{f}_i + \boldsymbol{\varepsilon}_i, i = 1, \ldots, n$, where $\mathbf{B} = (b_{jl})_{1\le j\le p, 1\le l\le 3}$ has iid entries b_{jl}'s generated from the uniform distribution $U(-2, 2)$, and $\boldsymbol{\varepsilon}_i$'s are drawn from multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{\varepsilon}})$ with $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ a sparse matrix whose diagonal entries being 3 and off-diagonal entries independently drawn from $0.3 \times Bernoulli(0.05)$. We set the mean $\boldsymbol{\mu} = 0.5 \times \mathbf{1}_d$ and $\boldsymbol{f}_i$'s are assumed to come from a linear model, i.e., $\boldsymbol{f}_i = \mathbf{A}\mathbf{X}_i + \boldsymbol{\xi}_i$ for $i = 1, \ldots, n$, where the error term $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \xi_{i3})^T$ is generated from the following distributions to model symmetric, heavy-tailed and heteroscedastic error cases, respectively: $\xi_{ik} \overset{iid}{\sim} N(0, 1)$, $\xi_{ik} \overset{iid}{\sim} t_3$, and $\xi_{ik} = \exp(\sum_{j=1}^{k} X_{7+j})\zeta_{ik}$, where $\zeta_{ik} \overset{iid}{\sim} N(0, 0.7^2)$ for $k = 1, 2, 3$. $\mathbf{A}$ is defined as

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & \cdots & 0 \end{pmatrix}_{3\times p}.$$

Example 2: factor-nonparametric model. We also consider a three factor model, similar to example 1, but with all weak factors. The b_{jl}'s are iid from $N(0, 0.04)$, and ε_{ij}'s are iid from $N(0, 150)$. Besides, μ_j's are randomly sampled from $\{0, 0.5, 0.8\}$ for $1 \le j \le d$ and

f_{ik}'s are assumed to follow nonlinear nonparametric models,

$$f_{i1} = \exp(2 - X_{i2} - X_{i4}) + \xi_{i1}, \quad f_{i2} = (2X_{i1} + 3X_{i3})^2 \log(|X_{i1}|) + \xi_{i2},$$

$$f_{i3} = 3(-X_{i1}X_{i3} + X_{i5})^3 \exp(-2X_{i5}) + \xi_{i3},$$

where the error term $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \xi_{i3})^T$ is considered to come from two different distributions: $\xi_{ik} \overset{iid}{\sim} N(0,1)$ and $\xi_{ik} \overset{iid}{\sim} Cauchy$ for $k = 1, 2, 3$.

Note that in example 1, the true model size is 5 for homogeneous errors. However, for heteroscedastic errors, X_8, X_9, X_{10} become active when $\tau \in \{0.75, 0.9\}$, leading to a true model size of 8 for these two quantiles. Simulation results for these two examples are summarized in Tables 1-2, Tables S4-S5 and S7-S8 of the Supplementary Materials. Note that we focus on a relative high-dimensional factor model with $d = 500$ to test our procedure for estimating the number of latent factors. When $\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, we observe that in the factor-linear benchmark model with normal errors and strong factors, all competing methods perform well. However, in the presence of heavy-tailed errors, NIS and RVSIS struggle to accurately identify all active covariates. Other methods, such as QaSIS, SIRS, DCSIS and PC-Screen experience significant performance deterioration and require much larger model sizes to recover the active set when heteroscedastic errors are present. For the factor-nonparametric model with weak factors, all methods, except for our proposed MSDS and QaSIS, fail to correctly identify covariates X_1 and X_3. MSDS, however, shows the highest probability of including all active covariates. Notably, when ξ_{ik} follows Cauchy distribution, MSDS significantly outperforms QaSIS. When $\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{X}})$, similar results are observed in Tables S4-S5 of the Supplementary Materials, demonstrating that MSDS

exhibits the best screening performance among all the competitors. This highlights MSDS as a powerful tool for addressing a wide range of factor-covariate models, effectively handling both strong and weak factors in high-dimensional settings.

Table 1: Simulation results for example 1 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	5.00(0)	6.42(3.74)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	$\text{QaSIS}_{0.50}$	5	6.00(1.48)	13.51(34.10)	0.96	1.00	1.00	1.00	0.97	-	-	-	0.94
	NIS	5	5.00(0)	5.04(0.20)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	SIRS	5	5.00(0)	5.03(0.23)	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)	5.03(0.18)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
t_3	$\text{MSDS}_{0.50}$	5	5.00(0)	7.40(6.27)	0.99	1.00	1.00	1.00	0.99	-	-	-	0.98
	$\text{QaSIS}_{0.50}$	5	9.00(5.93)	25.46(43.30)	0.94	1.00	1.00	1.00	0.89	-	-	-	0.83
	NIS	5	24.00(28.16)	104.60(171.36)	0.74	0.85	0.85	0.86	0.68	-	-	-	0.58
	SIRS	5	5.00(0)	6.92(10.05)	1.00	1.00	1.00	1.00	0.96	-	-	-	0.96
	DCSIS	5	5.00(0)	6.69(15.08)	0.98	0.98	1.00	1.00	0.98	-	-	-	0.98
	RVSIS	5	5.00(0)	51.02(126.54)	0.85	0.90	0.93	0.85	0.87	-	-	-	0.78
	PC-Screen	5	5.00(0)	5.33(2.20)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
hetero	$\text{MSDS}_{0.50}$	5	5.00(0)	5.71(1.63)	1.00	1.00	1.00	1.00	1.00	-	-	-	1.00
	$\text{MSDS}_{0.75}$	8	10.00(1.48)	124.94(292.98)	0.93	0.92	0.92	0.92	0.92	0.88	0.93	0.91	0.83
	$\text{MSDS}_{0.90}$	8	11.00(4.44)	111.66(269.43)	0.92	0.95	0.99	0.96	0.92	0.99	0.99	0.97	0.85
	$\text{QaSIS}_{0.50}$	5	59.00(72.64)	105.85(134.09)	1.00	1.00	1.00	0.96	0.45	-	-	-	0.45
	$\text{QaSIS}_{0.75}$	8	1015.00(1104.53)	1203.17(1022.68)	0.98	1.00	1.00	0.91	0.36	0.85	0.92	0.60	0.17
	$\text{QaSIS}_{0.90}$	8	2025.00(1186.08)	2067.41(1105.33)	0.31	0.51	0.38	0.13	0.06	0.98	0.98	0.72	0.00
	NIS	8	3395.00(1593.79)	3135.05(1300.74)	0.31	0.44	0.35	0.18	0.08	0.49	0.51	0.48	0.00
	SIRS	8	37.00(28.17)	300.82(276.77)	1.00	1.00	1.00	1.00	1.00	0.89	0.89	0.56	0.51
	DCSIS	8	9.00(1.85)	147.82(397.33)	0.98	1.00	1.00	0.98	0.84	1.00	1.00	0.96	0.81
	RVSIS	8	2585.00(1149.01)	2522.76(1132.08)	0.60	0.69	0.67	0.45	0.17	0.43	0.45	0.36	0.00
	PC-Screen	8	1295.00(893.26)	1476.50(975.89)	1.00	1.00	1.00	1.00	0.93	0.83	0.50	0.10	0.03

Next, we simulate two additional examples to evaluate the FDR control performance of the proposed MSDK and DMSDK procedures. For comparison, we also include the PC-Knockoff procedure introduced in Liu et al. (2022). Here we set $\mathbf{\Sigma}_{\mathbf{X}} = (0.25^{|i-j|})_{1\le i,j\le p}$.

Example 3: FDR control for factor-linear model. Consider the factor model as defined in example 1. Let $\boldsymbol{f}_i = \boldsymbol{\beta}^T\mathbf{X}_i + \boldsymbol{\xi}_i$ for $i = 1, \ldots, n$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$ with $\boldsymbol{\beta}_1 =$

Table 2: Simulation results for example 2 when $\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, p^* denotes the true model size

		MMS ($p^* = 5$)							
ξ_{ik}	Method	Median(MAD)	Mean(SD)	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_{all}$
$N(0,1)$	$\text{MSDS}_{0.5}$	10.00(5.93)	39.12(92.08)	0.92	0.99	0.89	0.99	1.00	0.82
	$\text{QaSIS}_{0.5}$	11.00(8.89)	39.32(91.22)	0.93	1.00	0.85	1.00	1.00	0.78
	NIS	3980.00(900.67)	3765.70(1015.43)	0.14	0.04	0.06	0.01	0.99	0.00
	SIRS	2788.00(1221.66)	2600.80(1009.58)	0.02	0.99	0.05	1.00	1.00	0.00
	DCSIS	1736.00(1108.98)	1701.04(892.07)	0.15	0.25	0.08	0.25	1.00	0.00
	RVSIS	3885.00(1030.40)	3611.00(996.92)	0.10	0.04	0.05	0.01	0.99	0.00
	PC-Screen	395.00(318.75)	660.40(682.53)	0.68	1.00	0.32	1.00	1.00	0.24
Cauchy	$\text{MSDS}_{0.5}$	9.50(6.67)	34.75(78.05)	0.92	1.00	0.88	1.00	1.00	0.82
	$\text{QaSIS}_{0.5}$	16.50(17.04)	60.88(103.03)	0.92	0.97	0.74	1.00	1.00	0.64
	NIS	3992.50(963.69)	3755.05(989.09)	0.10	0.05	0.09	0.05	0.96	0.00
	SIRS	2336.00(1313.58)	2339.64(1060.46)	0.04	1.00	0.08	1.00	1.00	0.00
	DCSIS	1676.00(1031.89)	1794.84(906.18)	0.12	0.25	0.15	0.23	0.99	0.02
	RVSIS	3912.50(848.78)	3612.00(1040.81)	0.08	0.09	0.11	0.05	0.96	0.00
	PC-Screen	515.00(530.02)	702.90(672.75)	0.66	1.00	0.30	1.00	1.00	0.22

$(\mathbf{1}_3^T, \mathbf{0}_{p-3}^T)^T$, $\boldsymbol{\beta}_2 = (\mathbf{0}_3^T, \mathbf{1}_2^T, \mathbf{0}_{p-5}^T)^T$, $\boldsymbol{\beta}_3 = (\mathbf{0}_5^T, \mathbf{1}_3^T, \mathbf{0}_{p-8}^T)^T$, and $\mathbf{X}_i$ is drawn from a mixture distribution $0.9\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{X}) + 0.1t_2(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{X})$. $\xi_{ik} \overset{iid}{\sim} t_3$ for $k = 1, 2, 3$. Therefore, X_1-X_8 are active with $|\mathcal{A}_\tau| = 8$.

Example 4: FDR control for factor-mixed model with weak factors. We consider a similar three factor model as in example 1, except that the entries of factor loading matrix b_{jl}'s are iid from $N(0, 0.04)$ for $j \neq l$. We draw ε_{ij}'s iid from $N(0, v_j^2)$ and $v_j^2 \sim U(0, 5.5)$ for $j = 1, \ldots, d$. In this scenario, the signals of latent factors are very weak. The latent factors are assumed to follow a mixed model with both additive and single index components

$$f_{i1} = 2 + 3X_{i1} + 3X_{i2} + \xi_{i1}, \quad f_{i2} = (2 + 3X_{i3})^2 + 2(X_{i4} - 1)^3 + \xi_{i2},$$

$$f_{i3} = \exp(X_{i5} + X_{i6}) + \xi_{i3},$$

where $\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{X})$ and $\xi_{ik} \overset{iid}{\sim} N(0, 1)$ for $k = 1, 2, 3$. Thus, X_1-X_6 are active, $|\mathcal{A}_\tau| = 6$.

The settings and implementation details for examples 3-4 are provided in Section $10

of the Supplementary Materials. We summarize the results for examples 3-4 in Table 3 and Tables S9-S10 of the Supplementary Materials, in which α denotes the prespecified FDR level, v is the prespecified PFER level, and k-FWER is defined as k-FWER $= P(V(\tau) \geq k)$. Here, $\hat{V}(\tau)$ and $\hat{V}$ are the average number of false discoveries, $\widehat{\text{FDR}}$ is the empirical FDR, that is, the average empirical FDP, $\widehat{\text{FWER}}$ is the empirical k-FWER, i.e., the average empirical $\hat{P}(V(\tau) \geq k)$, and Power refers to the average empirical power.

It is observed that our MSDK and DMSDK procedures effectively control the FDR at the prespecified level α, as well as the PFER and k-FWER. The PC-Knockoff procedure performs well in example 3. However, when applied to nonlinear models with weak factors, its empirical power and selection probability decline significantly. Additionally, the base MSDK procedure has an average execution time of approximately 45 seconds, while the PC-Knockoff takes about 50 minutes on Windows machines with 2.4 GHz CPUs and 16 GB of memory running R software. Consequently, the PC-Knockoff procedure is less suitable for scenarios with large n and p, particularly when n is large. In addition, the results for MSDK procedure at higher quantiles, such as $\tau = 0.90$, seem to be out of control, likely due to data sparsity at higher tails. In contrast, the derandomized version consistently achieves much higher selection probability and power while effectively maintaining the sure screening property across all settings. This demonstrates that the DMSDK procedure is particularly effective for high-dimensional factor-covariate models with both strong and weak factors, not only identifying variables across quantiles with FDR control but also handling large datasets efficiently with reasonable computational time.

Table 3: Simulation results for example 4 via MSDK, PC-Knockoff and DMSDK procedures

MSDK procedure FDR control										
τ	α	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_{all}$	$\hat{V}(\tau)(\widehat{\text{FDR}})$	Power
0.50	0.15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.02(0.145)	1.00
	0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.48(0.198)	1.00
	0.30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.41(0.287)	1.00
0.75	0.15	0.93	0.93	0.93	0.93	0.93	0.92	0.92	0.94(0.135)	0.93
	0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.56(0.195)	1.00
	0.30	1.00	1.00	1.00	1.00	1.00	0.98	0.98	2.53(0.297)	0.99
0.90	0.15	0.84	0.84	0.83	0.84	0.80	0.82	0.76	1.42(0.191)	0.83
	0.20	0.96	0.95	0.94	0.97	0.89	0.92	0.78	1.77(0.220)	0.94
	0.30	0.97	0.96	0.96	0.97	0.83	0.84	0.73	3.26(0.331)	0.92
PC-Knockoff procedure FDR control										
	α	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_{all}$	$\hat{V}(\widehat{\text{FDR}})$	Power
	0.15	0.48	0.48	0.47	0.48	0.47	0.47	0.45	1.22(0.124)	0.47
	0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.54(0.173)	1.00
	0.30	1.00	1.00	1.00	1.00	1.00	1.00	1.00	3.26(0.277)	1.00
DMSDK procedure PFER control										
τ	v	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_{all}$	$\hat{V}(\tau)(\widehat{\text{FDR}})$	Power
0.50	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.32(0.051)	1.00
	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.10(0.154)	1.00
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.82(0.232)	1.00
0.75	1	1.00	1.00	1.00	1.00	0.98	1.00	0.98	0.30(0.047)	0.99
	2	1.00	1.00	1.00	1.00	0.98	1.00	0.98	1.14(0.159)	0.99
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.01(0.250)	1.00
0.90	1	1.00	0.98	0.98	1.00	0.99	0.91	0.80	0.46(0.071)	0.95
	2	0.99	0.99	1.00	1.00	0.90	0.98	0.89	1.34(0.182)	0.97
	3	0.99	0.99	1.00	1.00	0.92	0.98	0.90	2.28(0.275)	0.98
DMSDK procedure k-FWER control at 0.20										
τ	$k(v)$	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$	$\mathcal{P}_6$	$\mathcal{P}_{all}$	$\hat{V}(\tau)(\widehat{\text{FWER}})$	Power
0.50	3(1.17)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.64(0.000)	1.00
	4(1.56)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.71(0.000)	1.00
	5(1.96)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.31(0.000)	1.00
0.75	3(1.17)	1.00	1.00	1.00	1.00	1.00	0.98	0.98	0.54(0.000)	0.99
	4(1.56)	1.00	1.00	1.00	1.00	0.98	0.98	0.97	0.66(0.000)	0.99
	5(1.96)	1.00	1.00	1.00	1.00	0.98	0.98	0.97	1.11(0.000)	0.99
0.90	3(1.17)	1.00	0.97	0.98	1.00	0.86	0.93	0.77	0.51(0.030)	0.95
	4(1.56)	1.00	0.97	0.98	1.00	0.89	0.93	0.79	0.76(0.020)	0.96
	5(1.96)	1.00	0.97	0.98	1.00	0.91	0.94	0.82	1.17(0.010)	0.97

5. Real Data Analysis

We apply the proposed procedures to a human well-being dataset studied by Fredrickson et al. (2013) and Yu et al. (2023), which is available at `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45330`. The dataset includes questionnaire responses from 84 healthy adults and gene expression measurements for 34,591 genes. All participants were

between 35 and 64 years old, could read and write in English, and had no chronic diseases or disabilities. To assess both hedonic and eudaimonic well-being, these participants were asked to respond to the Mental Health Continuum Short Form (MHC-SF) that comprises 14 questions reflecting hedonic and eudemonic well-being. These questions can be found in the Supplementary Materials of Yu et al. (2023). Respondents rated the frequency of each feeling they experienced over the past few weeks on a scale from 0 to 5 (0:never, 1:once or twice, 2:approximately once per week, 3:two or three times per week, 4:almost every day and 5:every day). Out of 84 participants, only 76 samples were valid due to missing responses. The main aim of this application is to explore the biological implications of hedonic and eudaimonic well-being through human genome. Understanding whether these two well-beings engage similar biological processes is considered an important yet challenging task.

We apply the proposed factor-covariate model to investigate the intrinsic relationship between the assayed genes and the two types of well-being. Since these well-being types are summarized by the 14 items, we set $K = 2$ and $d = 14$ in the factor model. First, we estimate two latent factors, denoted by f_1^* and f_2^*, respectively. Next, we investigate the functional relationships between the two estimated latent factors and the human genes ($p = 34,591$) using a nonparametric model. To facilitate this analysis, we begin by selecting the top 5,000 genes with the largest variance in expression values and standardize their expression measurements to have zero mean and unit variance. Active genes associated with the two latent factors are identified using our proposed MSDS method, followed by the DMSDK procedure to control both the FDR and PFER. We focus on

$\tau \in \{0.25, 0.50, 0.75, 0.90\}$. Additionally, we apply other screening methods discussed in the simulation studies for comparison. The top 17 genes identified by the various screening methods and the selected genes based on DMSDK procedure with $v = 1$ are listed in Table S11 of the Supplementary Materials. It is clear that the sets of active genes selected by different methods have multiple overlaps, suggesting several key findings. First, among all the screening methods, the genes LOC650238 and LOC650436 are selected most frequently. Second, genes selected by quantile-based procedures differ substantially from those selected by other methods. This indicates that certain genes may show strong associations with latent factors only at specific quantiles of the conditional distribution, such as the upper or lower tails, which other methods might overlook. Third, a more detailed conclusion can be conducted when we compare the genes selected at different quantiles. When $\tau \in \{0.25, 0.50\}$, only three out of the 17 genes overlap, and only three genes (LOC650238, BLOC1S1, LACTB) are selected at three or more quantiles. This highlights the heterogeneity in the data. By further conducting the DMSDK procedure, approximately two genes are selected as relevant at each quantile.

These selected genes are further served as inputs to fit both linear and additive models. This approach helps mitigate the curse of dimensionality while also facilitating the exploration of functional associations of each identified gene. The results, presented in Table S12, Figure 1, and Figure S3 of the Supplementary Materials, show that the two well-beings under linear model engage similar biological processes, as the selected genes are either positively or negatively correlated with both well-being. For additive model, Figure 1

and Figure S3 further suggests that the two well-beings exhibit similar biological structures against the selected genes. These findings support the conclusions of Fredrickson et al. (2013), which indicate that the hedonic and eudaimonic well-being share similar affective correlates and are strongly positively correlated. In contrast, Yu et al. (2023) examined the same dataset but focused solely on the mean of the conditional distribution, which fails to capture genes that are functionally associated with well-being across different quantiles.

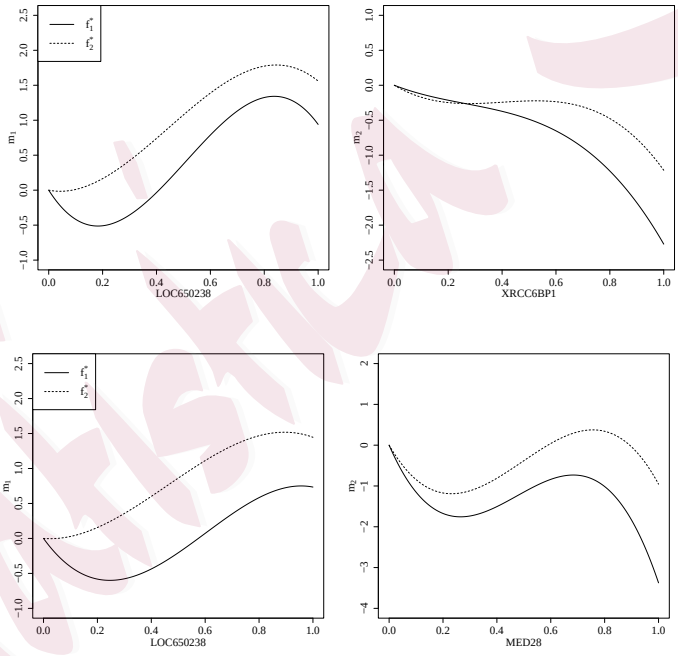

Figure 1: Estimated additive functions of the identified genes from our method at different quantiles for the two latent factors. The first row shows the results for $\tau = 0.25$, while the second row shows the results for $\tau = 0.5$.

To further investigate the predictive performance of the DMSDK procedure, and compare with the MSDS-SCAD method. The dataset is randomly partitioned into a training

set of 60 samples and a testing set of 16 samples. A five fold cross validation is applied to the training data to select the tuning parameters. The average number of selected genes (Size) over 100 replications is reported, with the corresponding standard errors in parentheses. We then assess the performance on the test set for each partition. The prediction error (PE) is defined as $\sum_{i=1}^{16}\sum_{k=1}^{2}\mathcal{L}_{\tau,h}(f_{ik}^{*}-\hat{f}_{ik})$, where $\mathcal{L}_{\tau,h}(\cdot)$ is the convolution type smoothed quantile loss function. The numbers in parentheses represent the corresponding standard errors across 100 partitions. The results, shown in Table 4, clearly indicate that the analysis based on MSDS followed by DMSDK (MSD-select) achieves strong predictive power, with a smaller model size and lower prediction error at each quantile level.

Table 4: Prediction performance for human well-being data

	Linear Model				Additive Model			
	DMSDK		MSDS-SCAD		DMSDK		MSDS-SCAD	
τ	Size	PE	Size	PE	Size	PE	Size	PE
0.25	2.40(0.98)	0.62(0.14)	5.66(2.06)	0.66(0.14)	2.88(1.80)	0.79(0.22)	6.88(1.62)	1.08(0.50)
0.50	3.18(2.08)	0.67(0.07)	6.34(1.95)	0.71(0.10)	3.22(1.37)	1.97(3.11)	4.56(1.29)	1.45(2.51)
0.75	3.52(1.64)	0.50(0.08)	5.06(1.82)	0.57(0.10)	1.66(0.51)	1.10(1.41)	4.46(1.32)	1.39(1.18)
0.90	4.70(2.27)	0.29(0.11)	2.47(1.27)	0.29(0.12)	1.44(0.54)	1.60(3.68)	6.14(2.48)	2.26(2.19)

6. Discussions and Extensions

In this paper, we develop a MSD-Select procedure, which integrates estimation, screening, and selection with FDR control for joint modeling involving latent factors. This procedure leverages joint information across latent factors, utilizing the quantile-adaptive MSD index for screening and the quantile-adaptive DMSDK procedure for selection. The procedure exhibits several notable advantages: First, it allows both the dimensions of observed

variables and covariates to diverge with sample size n, thus offering broader applicability. Second, to ensure robustness against heavy-tailed errors or covariates, it employs a model-free MSD index that provide valuable insights into heterogeneity in the relationship between active covariates and latent factors. By employing smoothing quantile regression techniques in place of traditional quantile regression methods and B-spline approximation in estimation, the nondifferentiability of the quantile loss function and curse of dimensionality of the nonparametric function can be further circumvented. Additionally, we establish the sure screening properties under mild conditions. Third, robustness in FDR control is achieved by extending the classical knockoffs procedure into the quantile regression framework. Numerical studies demonstrate that the MSDS screening method surpasses existing methods, and the proposed DMSDK procedure achieves tighter FDR control, along with PFER control and k-FWER control, while maintaining higher power.

In the factor-covariate model, latent factors need to be linked to the covariates through a nonparametric model. However, this model presumes that covariates are continuously distributed and does not accommodate categorical variables. To overcome this limitation, we propose employing a semiparametric partial linear model as an alternative to the purely nonparametric model. Nonetheless, the task of identifying which covariates should be modeled linearly versus nonlinearly remains complex, especially in high dimensional contexts. This challenge is slated for exploration in our future research endeavors.

One limitation of the MSD-Select method is that it does not directly account for the quantile association between a multivariate latent factor and a covariate. This arises from

the fact that the conditional quantile of a multivariate vector is not uniquely defined and is challenging to handle. Future research could explore directly using the multivariate conditional quantile, potentially based on measure-transportation-based concepts, to define a screening index for multivariate latent factors. An additional avenue for extending the factor-covariate model involves the incorporation of sparsity constraints within the factor loading matrix. This modification has the potential to enhance both interpretability and predictive performance in high-dimensional settings. Furthermore, integrating temporal dependencies into the factor-covariate modelrelevant for time series data or longitudinal studies, can be achieved by extending the factor model to either a state-space model or a dynamic factor model. Additionally, Ren et al. (2024) introduced a derandomized knockoff procedure by aggregating e-values from multiple knockoff realizations. Future research could explore extensions that integrate this new derandomized knockoff methodology with our proposed MSD index. Finally, the development of robust methods for controlling FDR, such as using data splitting or Gaussian mirror approaches within the quantile regression framework, offers another promising direction for future investigation.

Supplementary Materials

The properties of the MSD index under Gaussian distribution, details on the MSD knockoffs procedure, v-quantile knockoffs and estimation of the MSD index, figures for convolution-type smoothed quantile loss, as well as all technical proofs, additional results from numerical studies are provided in the online Supplementary Materials.

Acknowledgments

The authors would like to thank the editor, associate editor and referees for their insightful comments and suggestions that have significantly improved the paper. The first two authors contribute equally to this work, and Mingyao Ai is the corresponding author. Xiong's work was supported by the Funds for Central Universities in UIBE CXTD14-05. Ai's work was supported by NSFC grants 12131001 and W2412023, and LMEQF.

References

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

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and K. Li (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics 40*(1), 436–465.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, Z., K. P. Choi, and Y. Fujikoshi (2018). Consistency of aic and bic in estimating the number of significant components in high-dimensional principal component analysis. *The Annals of Statistics 46*(3), 1050–1076.

Barber, R. F. and E. J. Candès (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics 43*(5), 2055–2085.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: a practical and powerful approach to

multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 57*(1), 289–300.

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.

Desai, K. H. and J. D. Storey (2012). Cross-dimensional inference of dependent high-dimensional data. *Journal of the American Statistical Association 107*(497), 135–151.

Dobriban, E. and A. B. Owen (2019). Deterministic parallel analysis: an improved method for selecting factors and principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 81*(1), 163–183.

Fan, J., Y. Feng, and R. Song (2011). Nonparametric independence screening in sparse ultra-high-dimensional additive models. *Journal of the American Statistical Association 106*(494), 544–557.

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, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

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

Fredrickson, B. L., K. M. Grewen, K. A. Coffey, S. B. Algoe, A. M. Firestine, J. M. Arevalo, J. Ma, and S. W. Cole (2013). A functional genomic perspective on human well-being. *Proceedings of the National Academy of*

Sciences 110(33), 13684–13689.

He, X., X. Pan, K. M. Tan, and W.-X. Zhou (2023). Smoothed quantile regression with large-scale inference. *Journal of Econometrics 232*(2), 367–388.

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.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics 35*(1), 73–101.

Li, G., Y. Li, and C.-L. Tsai (2015). Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association 110*(509), 246–261.

Li, Q., G. Cheng, J. Fan, and Y. Wang (2018). Embracing the blessing of dimensionality in factor models. *Journal of the American Statistical Association 113*(521), 380–389.

Li, R., W. Zhong, and L. Zhu (2012). Feature screening via distance correlation learning. *Journal of the American Statistical Association 107*(499), 1129–1139.

Liu, J., Y. Si, Y. Niu, and R. Zhang (2022). Projection quantile correlation and its use in high-dimensional grouped variable screening. *Computational Statistics & Data Analysis 167*, 107369.

Liu, W., Y. Ke, J. Liu, and R. Li (2022). Model-free feature screening and fdr control with knockoff features. *Journal of the American Statistical Association 117*(537), 428–443.

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

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of*

Economics and Statistics 92(4), 1004–1016.

Ouyang, M., X. Wang, C. Wang, and X. Song (2018). Bayesian semiparametric failure time models for multivariate censored data with latent variables. *Statistics in Medicine 37*(28), 4279–4297.

Owen, A. B. and J. Wang (2016). Bi-cross-validation for factor analysis. *Statistical Science 31*(1), 119–139.

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

Roy, J. and X. Lin (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics 56*(4), 1047–1054.

Shao, X. and J. Zhang (2014). Martingale difference correlation and its use in high-dimensional variable screening. *Journal of the American Statistical Association 109*(507), 1302–1318.

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

Yu, C., W. Guo, X. Song, and H. Cui (2023). Feature screening with latent responses. *Biometrics 79*(2), 878–890.

Zhu, L., L. Li, R. Li, and L. Zhu (2011). Model-free feature screening for ultrahigh-dimensional data. *Journal of the American Statistical Association 106*(496), 1464–1475.

Han Pan, School of Statistics and Mathematics, Shandong University of Finance and Economics, Jinan, China

E-mail: scott_pan@163.com

Wei Xiong, School of Statistics, University of International Business and Economics, Beijing, China

E-mail: xiongwei@uibe.edu.cn

Mingyao Ai, School of Mathematical Sciences and Center for Statistical Science, Peking University, Beijing, China.

E-mail: myai@pku.edu.cn