Statistica Sinica **31** (2021), 421-447
doi:https://doi.org/10.5705/ss.202018.0219

COPULA-BASED PARTIAL CORRELATION SCREENING: A JOINT AND ROBUST APPROACH

Xiaochao Xia[1,2] and Jialiang Li[2,3,4]

[1]*Huazhong Agricultural University,* [2]*National University of Singapore,*
[3]*Duke-NUS Graduate Medical School and* [4]*Singapore Eye Research Institute*

Abstract: Screening for ultrahigh-dimensional features becomes difficult in the presence of outlying observations, heterogeneous or heavy-tailed distributions, multicollinearity, and confounding effects. Standard correlation-based marginal screening methods may offer a weak solution to these problems. We contribute a novel robust joint screener that safeguards against outliers and distribution misspecification of both the response variable and the covariates, and accounts for external variables at the screening step. Specifically, we introduce a copula-based partial correlation (CPC) screener. We show that the empirical process of the estimated CPC converges weakly to a Gaussian process. Furthermore, we establish the sure screening property for the CPC screener under very mild technical conditions, which need not require a moment condition, and are weaker than existing alternatives in the literature. Moreover, from a theoretical perspective, our approach allows for a diverging number of conditional variables. Extensive simulation studies and two data applications demonstrate the effectiveness of the proposed screening method.

Key words and phrases: Copula partial correlation, outlier, sure independent screening.

1. Introduction

Ultrahigh-dimensional data have followed as a natural consequence of the increasing availability of big data in many business and scientific research fields, including medicine, genetics, finance, and economics. Such massive data usually share two features: (i) the number of predictors or features can be very high, diverging to infinity with the sample size; and (ii) the data distribution is very likely to be heteroscedastic and heavy-tailed, for both the response and the covariates. These two features are observed in the two real-data sets investigated in Section 5. A variable screener helps identify important predictors from among numerous candidates. However, note that such large scale data still require using

Corresponding author: Xiaochao Xia, College of Science, Huazhong Agricultural University, Yifu Building A303, No.1, Shizishan Street, Hongshan District, Wuhan, Hubei Province, China, 430070. E-mail: xia_xiao_chao@126.com.

a comprehensive model to accurately predict a future outcome. Thus, the popular pure marginal screening approaches may not be adequate for model-building purposes. Therefore, we contribute a new screening method that addresses the above issues and complements the existing methodology.

Variable screening serves as a fast and efficient computing device. Numerous feature-screening methods have been proposed, including sure independence screening (SIS) by Fan and Lv (2008), who first established the sure screening property under a Gaussian linear model, sure independent ranking screening (SIRS, Zhu et al. (2011)), Kendall's τ-based screening (Kendall-SIS, Li et al. (2012)), distance correlation-based screening (DC-SIS, Li, Zhong and Zhu (2012)), quantile-adaptive screening (QaSIS, He, Wang and Hong (2013)), empirical likelihood screening (Chang, Tang and Wu (2013, 2016)), censored rank independence screening for lifetime data (CRIS, Song et al. (2014)), screening based on a quantile correlation (QC-SIS, Li, Li and Tsai (2015)), conditional quantile screening (CQ-SIS, Wu and Yin (2015)), survival impaction index screening (SII, Li et al. (2016)), and nonparametric independence screening (NIS, Fan, Feng and Song. (2011); Cheng et al. (2014); Xia, Yang and Li (2016)), among many others. These screening tools potentially suffer from two drawbacks. First, most evaluate a marginal association between the response and the predictors, without adjusting the external variables. Therefore, jointly important predictors may be screened out incorrectly if their marginal signal is not as strong as the spurious predictors in the ranked list. On the other hand, marginally important variables may be jointly ineffective; hence, including them in a multivariate model may lead to a less convincing prediction (e.g. Xia et al. (2016)). To take into account joint effects, marginal feature screening is usually followed by an iterative calculation, such as the iterative SIS (ISIS) in Fan and Lv (2008), which is computationally expensive, and does not guarantee theoretical success. Second, the distributions of the response and the predictors may be rather different from the light-tailed symmetric normal distribution, and very often, outliers affect the computed screening indices. Some of the aforementioned procedures address the robustness of the response. However, to the best of our knowledge, none address the robustness of the covariates, which is a harder problem, with a higher dimension.

We aim to solve the aforementioned two problems using a new screener. Specifically, to address the first, we develop a joint feature screening method in which we incorporate additional information. Several conditional feature screening methods have been proposed. For instance, Liu, Li and Wu (2014) considered

a sure independence screening procedure that uses a conditional Pearson correlation coefficient and a kernel smoothing. Their method can handle ultrahigh-dimensional varying-coefficient feature variables, which are investigated in Fan, Ma and Dai (2014) and Cheng et al. (2014) as well. In addition, Xia, Li and Fu (2019) considered a robust screening method based on a conditional quantile correlation, as a generalized conception of Li, Li and Tsai (2015). However, these authors consider only a single conditional variable. As an extension to multivariate conditional variables, Chu, Li and Reimherr (2016) studied several confounding variables. Barut, Fan and Verhasselt (2016) extended the approach of Fan and Song (2010)) to allow for a portion of the predictors as conditional variables. Our work provides a more general framework, in which all ultrahigh-dimensional predictors and low-dimensional confounders can be considered jointly during the screening process. To address the second problem, we incorporate a robust copula-based correlation and a partial correlation in our screening methods. The nonparametric copula is a well-known distribution-free summary measure that leads naturally to a screener that is robust against outliers and distribution misspecification. To the best of our knowledge, very few works apply this classical dependence concept in a high-dimensional setting. Xia, Li and Fu (2019) proposed a robust conditional feature screening approach; however, their method performs robustly against the response, but not against the covariates. Another relevant recent work is that of Ma, Li and Tsai (2017).

This study contributes to the literature as follows. First, we propose a doubly robust copula-based correlation (CC). Copula functions are very popular bivariate functions used to model the nonlinear dependence between paired variates; see Nelsen (2007) for an introduction to copula functions. A CC characterizes the empirical dependence between two random variables evaluated at a level pair, and is invariant under a monotone transformation for both variables. We study the asymptotic process properties of the CC. A marginal variable screening approach using the CC (CC-SIS) achieves the desired sure screening consistency (Fan and Lv (2008)). Second, by extending the copula-based correlation to a copula-based partial correlation (CPC), we construct a more general framework for joint screening. The importance of each predictor is evaluated in the presence of conditional variables, providing a fast conditional feature-screening method. The CPC is also robust, owing to its construction from a nonparametric estimation, and thus may be more reliable than the similar approach of Ma, Li and Tsai (2017), and have a broader range of application. We provide both theoretical and numerical support for the proposed screening method. Our data analysis

indicates that the final multivariate regression models built after our screening approach indeed predict outcomes with improved accuracy.

The rest of the paper is organized as follows. Sections 2 and 3 present the methodologies and large-sample properties for the CC and CC-SIS and for the CPC and CPC-SIS, respectively. Section 4 discusses further implementation details and cases for the CPC-SIS. Simulations and two real-data applications are presented in Section 5. Section 6 discusses choosing the parameters for the method. Section 7 concludes the paper. All technical proofs and additional simulations are relegated to the online Supplementary Material.

2. Copula-Based Correlation and Variable Screening

Consider two continuous random variables, X and Y. Let F_X be the cumulative distribution function (CDF) of X, which is assumed to be right continuous; $F_X^{-1}(\tau) = \inf\{x : F_X(x) \geq \tau\}$ is the τ-quantile of F_X, $F_{Y,X}$ is the joint CDF of Y and X, and $F_{Y|X}$ is the conditional distribution function of Y, given X, with density $f_{Y|X}$. We use $F_{n,X}$, $F_{n,X}^{-1}$, and $F_{n,Y,X}$ to denote empirical versions of F_X, F_X^{-1} and $F_{Y,X}$, respectively, based on a sample of size n. Let $D[a,b]$ be the Banach space of all càdlàg functions $z : [a,b] \mapsto \mathbb{R}$ on an interval $[a,b] \subset \bar{\mathbb{R}}$ equipped with the uniform norm, and let $\ell^\infty([a,b]^2)$ denote the collection of all bounded functions $z : [a,b]^2 \mapsto \mathbb{R}$. We use $\to^d$ to denote convergence in distribution.

2.1. Copula-based correlation

We propose the following CC

$$\varrho_{Y,X}(\tau,\iota) = \frac{F_{Y,X}(F_Y^{-1}(\tau), F_X^{-1}(\iota)) - \tau\iota}{\sqrt{\tau(1-\tau)\iota(1-\iota)}}, \quad 0 \leq \tau, \iota \leq 1, \tag{2.1}$$

where the first term in the numerator is a copula function $C(u,v) = F_{U,V}(u,v)$, with $U = F_Y(Y)$ and $V = F_X(X)$, evaluated at $(u,v) = (\tau,\iota)$ (see Corollary 2.3.7 of (Nelsen, 2007, p.22)). From simple algebra, we have $\varrho_{Y,X}(\tau,\iota) = \mathrm{E}[\psi_\tau(Y - F_Y^{-1}(\tau))\psi_\iota(X - F_X^{-1}(\iota))]/\sqrt{\tau(1-\tau)\iota(1-\iota)} = \mathrm{cov}(\psi_\tau(Y - F_Y^{-1}(\tau)), \psi_\iota(X - F_X^{-1}(\iota)))/\sqrt{\tau(1-\tau)\iota(1-\iota)}$, where $\psi_\tau(u) = \tau - I(u \leq 0)$ and $I(\cdot)$ is the indicator function. Because $\mathrm{var}(\psi_\iota(X - F_X^{-1}(\iota))) = \iota(1-\iota)$ and $\mathrm{var}(\psi_\tau(Y - F_Y^{-1}(\tau))) = \tau(1-\tau)$, $\varrho_{Y,X}(\tau,\iota)$ given in (2.1) is indeed a legitimate correlation coefficient that lies between -1 and 1. Like other known correlation measures, the CC is equal to zero if X and Y are independent.

The CC can measure the nonlinear dependence between X and Y, and thus

incorporates various kinds of bivariate joint distributions of X and Y. In addition, because the indicator function is unaffected by outliers and extreme values, the CC is robust for certain heavy-tailed distributions for both Y and X. Note that a monotone transformation of X and Y does not alter the value of the CC.

Given a sample of independent and identically distributed (i.i.d.) observations $\{(X_i, Y_i), i = 1, \ldots, n\}$, we can construct an empirical estimate of $\varrho_{Y,X}(\tau, \iota)$ as

$$\begin{aligned}\widehat{\varrho}_{Y,X}(\tau,\iota) &= \frac{F_{n,Y,X}(F_{n,Y}^{-1}(\tau), F_{n,X}^{-1}(\iota)) - \tau\iota}{\sqrt{\tau(1-\tau)\iota(1-\iota)}} \\ &= \frac{n^{-1}\sum_{i=1}^{n}\psi_\tau(Y_i - F_{n,Y}^{-1}(\tau))\psi_\iota(X_i - F_{n,X}^{-1}(\iota))}{\sqrt{\tau(1-\tau)\iota(1-\iota)}}. \end{aligned} \tag{2.2}$$

Let $\sigma_{Y,X}(\tau,\iota) = F_{Y,X}(F_Y^{-1}(\tau), F_X^{-1}(\iota))$, $\sigma_{X|Y}(\tau,\iota) = F_{X|Y=F_Y^{-1}(\tau)}(F_X^{-1}(\iota))$, and $\sigma_{Y|X}(\tau,\iota) = F_{Y|X=F_X^{-1}(\iota)}(F_Y^{-1}(\tau))$. In the following, we fix the level at (τ,ι), and write $\sigma_{Y,X}$, $\sigma_{X|Y}$, and $\sigma_{Y|X}$, respectively, for simplicity. Furthermore, define

$$\begin{aligned}\xi(Y,X;\tau,\iota) = \frac{1}{\sqrt{\tau(1-\tau)\iota(1-\iota)}}\Big[&I(Y \le F_Y^{-1}(\tau), X \le F_X^{-1}(\iota)) \\ &-\sigma_{X|Y}(\tau,\iota)I(Y \le F_Y^{-1}(\tau)) - \sigma_{Y|X}(\tau,\iota)I(X \le F_X^{-1}(\iota))\Big].\end{aligned}$$

The weak convergence result for $\widehat{\varrho}_{Y,X}(\tau,\iota)$ is established in the next theorem.

Theorem 1. *Let $0 < a < b < 1$, and suppose that the marginal distributions F_X and F_Y are continuously differentiable on the intervals $[F_X^{-1}(a) - \varepsilon, F_X^{-1}(b) + \varepsilon]$ and $[F_Y^{-1}(a) - \varepsilon, F_Y^{-1}(b) + \varepsilon]$ with positive derivatives f_X and f_Y, respectively, for some $\varepsilon > 0$. Furthermore, assume that the conditional density functions $f_{Y|X}$ and $f_{X|Y}$ are continuous on the product of these intervals. Then,*

$$\sqrt{n}\{\widehat{\varrho}_{Y,X}(\tau,\iota) - \varrho_{Y,X}(\tau,\iota)\} \overset{w}{\rightsquigarrow} \mathbb{G}_{Y,X}(\tau,\iota)$$

in $\ell^\infty([a,b]^2)$, where $\overset{w}{\rightsquigarrow}$ denotes "converge weakly," and $\mathbb{G}_{Y,X}(\tau,\iota)$ is Gaussian process with mean zero and covariance function $\Omega_1(\tau_1,\iota_1;\tau_2,\iota_2) \equiv \mathrm{E}\{[\xi(Y,X;\tau_1, \iota_1) - \mathrm{E}\xi(Y,X;\tau_1,\iota_1)] \times [\xi(Y,X;\tau_2,\iota_2) - \mathrm{E}\xi(Y,X;\tau_2,\iota_2)]\}$.

We can write the covariance function explicitly as $\Omega_1(\tau_1,\iota_1;\tau_2,\iota_2) = \{F_{Y,X}(F_Y^{-1}(\tau_1 \wedge \tau_2), F_X^{-1}(\iota_1 \wedge \iota_2)) - F_{Y,X}(F_Y^{-1}(\tau_1), F_X^{-1}(\iota_1))F_{Y,X}(F_Y^{-1}(\tau_2), F_X^{-1}(\iota_2)) - \sigma_{X|Y}(\tau_2,\iota_2)[F_{Y,X}(F_Y^{-1}(\tau_1\wedge\tau_2), F_X^{-1}(\iota_1)) - F_{Y,X}(F_Y^{-1}(\tau_1), F_X^{-1}(\iota_1))\tau_2] - \sigma_{Y|X}(\tau_2,\iota_2)$ $[F_{Y,X}(F_Y^{-1}(\tau_1), F_X^{-1}(\iota_1\wedge\iota_2)) - F_{Y,X}(F_Y^{-1}(\tau_1), F_X^{-1}(\iota_1))\iota_2] - \sigma_{X|Y}(\tau_1,\iota_1)[F_{Y,X}(F_Y^{-1}$

$(\tau_1), F_X^{-1}(\iota_2)) - F_{Y,X}(F_Y^{-1}(\tau_2), F_X^{-1}(\iota_2))\tau_1] + \sigma_{X|Y}(\tau_1,\iota_1)\sigma_{X|Y}(\tau_2,\iota_2)(\tau_1\wedge\tau_2 - \tau_1\tau_2) + \sigma_{X|Y}(\tau_1,\iota_1)\sigma_{Y|X}(\tau_2,\iota_2)[F_{Y,X}(F_Y^{-1}(\tau_1), F_X^{-1}(\iota_2)) - \tau_1\iota_2] - \sigma_{Y|X}(\tau_1,\iota_1)\times[F_{Y,X}(F_Y^{-1}(\tau_2), F_X^{-1}(\iota_1\wedge\iota_2)) - F_{Y,X}(F_Y^{-1}(\tau_2), F_X^{-1}(\iota_2))\iota_1] + \sigma_{Y|X}(\tau_1,\iota_1)\sigma_{X|Y}(\tau_2,\iota_2)\times[F_{Y,X}(F_Y^{-1}(\tau_2), F_X^{-1}(\iota_1)) - \tau_2\iota_1] + \sigma_{Y|X}(\tau_1,\iota_1)\sigma_{Y|X}(\tau_2,\iota_2)(\iota_1\wedge\iota_2 - \iota_1\iota_2)\}/[\tau_1(1-\tau_1)\iota_1(1-\iota_1)\tau_2(1-\tau_2)\iota_2(1-\iota_2)]^{1/2}$. In particular, at fixed (τ,ι), if $\varrho_{Y,X}(\tau,\iota) = 0$, then $\sqrt{n}\widehat{\varrho}_{Y,X}(\tau,\iota) \to^d N(0,\Omega_1)$, where $\Omega_1 \equiv \Omega_1(\tau,\iota;\tau,\iota) = \{\sigma_{Y,X} - \sigma_{Y,X}^2 + (\tau - \tau^2)\sigma_{X|Y}^2 + (\iota-\iota^2)\sigma_{Y|X}^2 - 2(1-\tau)\sigma_{Y,X}\sigma_{X|Y} - 2(1-\iota)\sigma_{Y,X}\sigma_{Y|X} + 2[\sigma_{Y,X} - \tau\iota]\sigma_{X|Y}\sigma_{Y|X}\}/[\tau(1-\tau)\iota(1-\iota)]$. If Y and X are independent, then $\Omega_1 = 1$, producing the same null distribution used in classical correlation and autocorrelation studies. In contrast to the work of Li, Li and Tsai (2015), our result is free of the moment conditions on X; Li, Li and Tsai (2015) require the existence of a fourth-order moment on X to achieve the convergence in law. The justifications for this theorem rely on empirical process techniques (Billingsley (1999); van der Vaart and Wellner (1996); Kosorok (2008)).

In order to make a statistical inference, for example, constructing a confidence interval for $\varrho_{Y,X}(\tau,\iota)$ and testing a hypothesis such as $H_0 : \varrho_{Y,X}(\tau,\iota) = 0$, we need to estimate the covariance function $\Omega_1(\tau_1,\iota_1;\tau_2,\iota_2)$. To this end, denote $m_1(y) = \mathrm{E}\{I(X \le F_X^{-1}(\iota))|Y = y\}$ and $m_2(x) = \mathrm{E}\{I(Y \le F_Y^{-1}(\tau))|X = x\}$. We can use a nonparametric approach, for example, the Nadaraya–Watson (NW) method (Nadaraya (1964) and Watson (1964)), to obtain estimates $\widehat{m}_1(y)$ and $\widehat{m}_2(x)$ for $m_1(y)$ and $m_2(x)$, respectively, where the unknown $F_X^{-1}(\iota)$ and $F_Y^{-1}(\tau)$ are replaced by $F_{n,X}^{-1}(\iota)$ and $F_{n,Y}^{-1}(\tau)$, respectively. Therefore, we obtain the estimates $\widehat{\sigma}_{X|Y}(\tau,\iota) = \widehat{m}_1(F_{n,Y}^{-1}(\tau))$ and $\widehat{\sigma}_{Y|X}(\tau,\iota) = \widehat{m}_2(F_{n,X}^{-1}(\iota))$. Next, we estimate $\Omega_1(\tau_1,\iota_1;\tau_2,\iota_2)$. Denote $\widehat{\xi}_n(Y_i,X_i;\tau,\iota) = \big[I(Y_i \le F_{n,Y}^{-1}(\tau), X_i \le F_{n,X}^{-1}(\iota)) - \widehat{\sigma}_{X|Y}(\tau,\iota)I(Y_i \le F_{n,Y}^{-1}(\tau)) - \widehat{\sigma}_{Y|X}(\tau,\iota)I(X_i \le F_{n,X}^{-1}(\tau))\big]/\sqrt{\tau(1-\tau)\iota(1-\iota)}$ and $\overline{\xi}_n(Y,X;\tau,\iota) = n^{-1}\sum_{i=1}^n \widehat{\xi}_n(Y_i,X_i;\tau,\iota)$. Then, we obtain the consistent estimate $\Omega_1(\tau_1,\iota_1;\tau_2,\iota_2)$ as $\widehat{\Omega}_1(\tau_1,\iota_1;\tau_2,\iota_2) = n^{-1}\sum_{i=1}^n[\widehat{\xi}_n(Y_i,X_i;\tau_1,\iota_1) - \overline{\xi}_n(Y,X;\tau_1,\iota_1)]\times[\widehat{\xi}_n(Y_i,X_i;\tau_2,\iota_2) - \overline{\xi}_n(Y,X;\tau_2,\iota_2)]$.

In practice, we usually encounter a situation where Y is univariate, but X is multivariate. As as extension to Theorem 1, and to compare the strength of the dependence of the two random variables X_1 and X_2 on Y, we examine the difference $\varrho_{Y,X_1}(\tau,\iota) - \varrho_{Y,X_2}(\tau,\iota)$. In particular, we can test a hypothesis using this difference. Given a sample $\{(Y_i,X_{i1},X_{i2}), i = 1,\ldots,n\}$, similarly to (2.2), we can define $\widehat{\varrho}_{Y,X_1}(\tau,\iota)$ and $\widehat{\varrho}_{Y,X_2}(\tau,\iota)$. The following theorem applies.

Theorem 2. *Let* $0 < a < b < 1$, *and suppose that the marginal distributions* F_{X_k} *and* F_Y *are continuously differentiable on the intervals* $[F_{X_k}^{-1}(a) - \varepsilon, F_{X_k}^{-1}(b) + \varepsilon]$

and $[F_Y^{-1}(a)-\varepsilon, F_Y^{-1}(b)+\varepsilon]$ *with positive derivatives* f_{X_k} *and* f_Y*, respectively, for some* $\varepsilon>0$ *and* $k=1,2$*. Furthermore, assume that the conditional density functions* $f_{Y|X_k}$ *and* $f_{X_k|Y}$*, for* $k=1,2$*, are continuous on the product of these intervals. Then, we have*

$$\sqrt{n}\{[\widehat{\varrho}_{Y,X_1}(\tau,\iota)-\widehat{\varrho}_{Y,X_2}(\tau,\iota)]-[\varrho_{Y,X_1}(\tau,\iota)-\varrho_{Y,X_2}(\tau,\iota)]\} \overset{w}{\rightsquigarrow} \mathbb{G}_{Y,X_1,X_2}(\tau,\iota)$$

in $\ell^\infty([a,b]^2)$*, where* $\mathbb{G}_{Y,X_1,X_2}(\tau,\iota)$ *is a Gaussian process with mean zero and covariance function* $\Xi_1(\tau_1,\iota_1;\tau_2,\iota_2)\equiv \mathrm{E}\{[\eta(Y,X_1,X_2;\tau_1,\iota_1)-\mathrm{E}\eta(Y,X_1,X_2;\tau_1,\iota_1)]\times [\eta(Y,X_1,X_2;\tau_2,\iota_2)-\mathrm{E}\eta(Y,X_1,X_2;\tau_2,\iota_2)]\}$*,* $\eta(Y,X_1,X_2;\tau,\iota)=\xi(Y,X_1;\tau,\iota)-\xi(Y,X_2;\tau,\iota)$ *and* $\xi(Y,X;\tau,\iota)$ *is given as in Theorem* 1*.*

It follows from Theorem 2 that for a fixed pair (τ,ι), if $\varrho_{Y,X_1}(\tau,\iota)=\varrho_{Y,X_2}(\tau,\iota)$, then $\sqrt{n}\{[\widehat{\varrho}_{Y,X_1}(\tau,\iota)-\widehat{\varrho}_{Y,X_2}(\tau,\iota)]\overset{d}{\to}N(0,\Xi_1)$. Here, $\Xi_1\equiv\Xi_1(\tau,\iota;\tau,\iota)=\Omega_1^{(1)}+\Omega_1^{(2)}-2A_{12}$, where $\Omega_1^{(k)}$ is the same as Ω_1, except that X in Ω_1 is replaced by X_k, for $k=1,2$. In addition, $A_{12}\equiv A_{12}(\tau,\iota)=\big\{[\sigma_{Y,X_1,X_2}(\tau,\iota)-\sigma_{Y,X_1}\sigma_{Y,X_2}]-(1-\tau)\sigma_{X_2|Y}\sigma_{Y,X_1}-\sigma_{Y|X_2}[\sigma_{Y,X_1,X_2}(\tau,\iota)-\iota\sigma_{Y,X_1}]-(1-\tau)\sigma_{X_1|Y}\sigma_{Y,X_2}+\tau(1-\tau)\sigma_{X_1|Y}\sigma_{X_2|Y}+\sigma_{X_1|Y}\sigma_{Y|X_2}(\sigma_{Y,X_2}-\tau\iota)-\sigma_{Y|X_1}[\sigma_{Y,X_1,X_2}(\tau,\iota)-\iota\sigma_{Y,X_2}]+\sigma_{Y|X_1}\sigma_{X_2|Y}(\sigma_{Y,X_1}-\tau\iota)+\sigma_{Y|X_1}\sigma_{Y|X_2}[\sigma_{X_1,X_2}(\iota,\iota)-\iota^2]\big\}/\sqrt{\tau(1-\tau)\iota(1-\iota)}$, where $\sigma_{Y,X_1,X_2}(\tau,\iota)=F_{Y,X_1,X_2}(F_Y^{-1}(\tau),F_{X_1}^{-1}(\iota),F_{X_2}^{-1}(\iota))$ and $\sigma_{X_1,X_2}(\iota,\iota)=F_{X_1,X_2}(F_{X_1}^{-1}(\iota),F_{X_2}^{-1}(\iota))$. If Y,X_1, and X_2 are mutually independent, then $\Xi_1=2$. Next, we estimate the covariance function $\Xi_1(\tau_1,\iota_1;\tau_2,\iota_2)$. Let $\widehat{\eta}_n(Y_i,X_{i1},X_{i2};\tau,\iota)=\widehat{\xi}_n(Y_i,X_{i1};\tau,\iota)-\widehat{\xi}_n(Y_i,X_{i2};\tau,\iota)$, where $\widehat{\xi}_n(Y_i,X_i;\tau,\iota)$ is given as before, and $\overline{\eta}_n(Y,X_1,X_2;\tau,\iota)=n^{-1}\sum_{i=1}^n\widehat{\eta}_n(Y_i,X_{i1},X_{i2};\tau,\iota)$. Then, $\Xi_1(\tau_1,\iota_1;\tau_2,\iota_2)$ can be estimated as $\widehat{\Xi}_1(\tau_1,\iota_1;\tau_2,\iota_2)=n^{-1}\sum_{i=1}^n[\widehat{\eta}_n(Y_i,X_{i1},X_{i2};\tau_1,\iota_1)-\overline{\eta}_n(Y,X_1,X_2;\tau_1,\iota_1)]\times[\widehat{\eta}_n(Y_i,X_{i1},X_{i2};\tau_2,\iota_2)-\overline{\eta}_n(Y,X_1,X_2;\tau_2,\iota_2)]$.

2.2. CC-based variable screening

Suppose we have a sample $\{(Y_i,\mathbf{X}_i),i=1,\dots,n\}$ consisting of n independent copies of $(Y,\mathbf{X})$, where Y is the response variable and $\mathbf{X}=(X_1,\dots,X_p)^T$ is a vector of p predictors. When the number of predictors, p, is of an exponential order of the sample size n, that is, the so-called ultrahigh dimension, and most of the p predictors are irrelevant, we can use the CC as a screener to identify the sparse set of informative predictors. We write p_n instead of p to emphasize the dependence on the sample size. An empirical estimate for the CC between Y and X_j is given by

$$\widehat{\varrho}_{Y,X_j}(\tau,\iota)=\frac{n^{-1}\sum_{i=1}^{n}\psi_\tau(Y_i-F_{n,Y}^{-1}(\tau))\psi_\iota(X_{ij}-F_{n,X_j}^{-1}(\iota))}{\sqrt{\tau(1-\tau)\iota(1-\iota)}}. \tag{2.3}$$

Then, we may select an empirical active set as

$$\widehat{\mathcal{M}}_a=\big\{j:|\widehat{\varrho}_{Y,X_j}(\tau,\iota)|\geq\nu_n,1\leq j\leq p_n\big\}, \tag{2.4}$$

where ν_n is a user-specified threshold parameter that controls the size of the final screened model. Using the CC for variable screening can lead to the sure independence screening (SIS) property; here, we refer to this procedure as CC-SIS.

Denote the true active set by $\mathcal{M}_a^*=\{j:|\varrho_{Y,X_j}(\tau,\iota)|>0,j=1,\ldots,p_n\}$. Write $F_{Y|\mathbf{X}}^{-1}(\tau)=\inf\{y:P(Y\leq y|\mathbf{X})\geq\tau\}$, $u_j=|\varrho_{Y,X_j}(\tau,\iota)|$, and $\widehat{u}_j=|\widehat{\varrho}_{Y,X_j}(\tau,\iota)|$. To establish the screening consistency, we need the following conditions.

(C1) In a neighborhood of $F_Y^{-1}(\tau)$, the density $f_Y(y)$ of Y is uniformly bounded away from zero and infinity, and has a bounded derivative. For every $1\leq j\leq p_n$, in a neighborhood of $F_{X_j}^{-1}(\iota)$, the density $f_{X_j}(x)$ of X_j is uniformly bounded away from zero and infinity, and has a bounded derivative.

(C2) $\min_{j\in\mathcal{M}_a^*}u_j\geq C_0n^{-\kappa}$, for some $\kappa>0$ and $C_0>0$.

Theorem 3. *(Screening Property for CC-SIS) Suppose that condition* (C1) *holds. Then:*
(i) For any constant $C>0$, there exists some positive constant $\tilde{c}_1$, such that for sufficiently large n,

$$P\bigg(\max_{1\leq j\leq p_n}\big|\widehat{u}_j-u_j\big|\geq Cn^{-\kappa}\bigg)\leq 6p_n\exp(-\tilde{c}_1n^{1-2\kappa}).$$

(ii) In addition, if condition (C2) *is satisfied, by choosing $\nu_n=C_1n^{-\kappa}$ with $C_1\leq C_0/2$, we have*

$$P\big(\mathcal{M}_a^*\subset\widehat{\mathcal{M}}_a\big)\geq 1-6s_n\exp(-\tilde{c}_1n^{1-2\kappa})$$

for sufficiently large n, where $s_n=|\mathcal{M}_a^*|$ is the cardinality of set $\mathcal{M}_a^*$.

This result implies that the CC-SIS can select all truly active predictors with very high probability. The dimensionality can be as high as $p_n=o(\exp(n^{1-2\kappa}))$, which is similar to those of other model-free feature screening methods (e.g., Li et al. (2012) and Wu and Yin (2015)). Moreover, its nonparametric nature means

our result requires fewer condition on both the predictors and the response. In reality, no moment assumption is imposed on the predictors or the response.

In practice, the threshold parameter ν_n plays an important role in producing a satisfactory model. A small value of ν_n will result in a large number of predictors after screening, which, in turn, leads to many incorrect positives. Here, we employ a data-driven procedure, in which we control the false discovery rates (FDR) to determine the threshold for the CC-SIS. From Theorem 1, for covariate j such that $\varrho_{Y,X_j}(\tau,\iota)=0$, it follows that, asymptotically, $\sqrt{n}[\widehat{\Omega}_1(\tau,\iota;\tau,\iota)]^{-1/2}\widehat{\varrho}_{Y,X_j}(\tau,\iota) \sim N(0,1)$. We can use high-criticism t-tests to select the variables $\widehat{\mathcal{M}}_{a,\delta}=\{j:\sqrt{n}[\widehat{\Omega}_1(\tau,\iota;\tau,\iota)]^{-1/2}|\widehat{\varrho}_{Y,X_j}(\tau,\iota)|\geq\delta\}$, for a small $\delta>0$. This controls the FDR $\mathrm{E}\{|\widehat{\mathcal{M}}_{a,\delta}\cap(\mathcal{M}_a^*)^c|/|(\mathcal{M}_a^*)^c|\}$, defined by Zhao and Li (2012). The following proposition justifies this FDR procedure.

Proposition 1. *(FDR Property) Assume conditions* (C1)–(C2) *and the condition of Theorem 1 hold. Then, if we choose* $\delta=\Phi^{-1}(1-\bar{d}_n/(2p_n))$, *where* $\Phi(\cdot)$ *is the CDF of the standard normal variable and* $\bar{d}_n$ *is the number of false positives that can be tolerated, then for some constant* $c_a>0$, *we have*

$$\mathrm{E}\left\{\frac{|\widehat{\mathcal{M}}_{a,\delta}\cap(\mathcal{M}_a^*)^c|}{|(\mathcal{M}_a^*)^c|}\right\}\leq\frac{\bar{d}_n}{p_n}+\frac{c_a}{\sqrt{n}}.$$

3. Copula-Based Partial Correlation and Variable Screening

3.1. Copula-based partial correlation, CPC

To facilitate a joint screening procedure (Ma, Li and Tsai (2017)), we define a CPC for Y and X, conditional on a q-dimensional random vector $\mathbf{Z}$, as

$$\varrho_{Y,X|\mathbf{Z}}(\tau,\iota)=\frac{\mathrm{E}\{\psi_\tau(Y-\mathbf{Z}^T\boldsymbol{\alpha}^0)\psi_\iota(X-\mathbf{Z}^T\boldsymbol{\theta}^0)\}}{\sqrt{\tau(1-\tau)\iota(1-\iota)}},\tag{3.1}$$

where $\boldsymbol{\alpha}^0=\mathrm{argmin}_{\boldsymbol{\alpha}}\mathrm{E}\{\rho_\tau(Y-\mathbf{Z}^T\boldsymbol{\alpha})\}$ and $\boldsymbol{\theta}^0=\mathrm{argmin}_{\boldsymbol{\theta}}\mathrm{E}\{\rho_\iota(X-\mathbf{Z}^T\boldsymbol{\theta})\}$, where $\rho_w(u)=u[w-I(u\leq 0)]$, for $w=\tau$ or ι. Note that this implies that $\mathbf{Z}^T\boldsymbol{\alpha}^0=F^{-1}_{Y|\mathbf{Z}}(\tau)$ and $\mathbf{Z}^T\boldsymbol{\theta}^0=F^{-1}_{X|\mathbf{Z}}(\iota)$. The parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ can be interpreted as the marginal increment on the conditional quantiles of Y and X, given $\mathbf{Z}$, respectively, when increasing by a unit of $\mathbf{Z}$. The CPC is actually the CC between Y and X_j, after removing the confounding effects of $\mathbf{Z}$. Linear partial correlation is widely used in regression diagnostics, and describes the association of the response and the predictor, conditional on specific values of the other predictors. The unconditional $\varrho_{Y,X}(\tau,\iota)$-value may be spurious, owing to lurking variables,

and does not necessarily imply the same $\varrho_{Y,X|\mathbf{Z}}(\tau,\iota)$-value conditional on $\mathbf{Z}$. Our copula-based version is relatively more robust for a real-data analysis. The CC is a special case of the CPC when $\mathbf{Z}$ is a constant.

With sample observations $\{(Y_i, X_i, \mathbf{Z}_i), i = 1, \ldots, n\}$, we obtain the following estimate of $\varrho_{Y,X|\mathbf{Z}}(\tau,\iota)$. Let $\widehat{\boldsymbol{\alpha}} = \operatorname{argmin}_{\boldsymbol{\alpha}}(1/n)\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{Z}_i^T\boldsymbol{\alpha})$ and $\widehat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}}(1/n)\sum_{i=1}^n \rho_\iota(X_i - \mathbf{Z}_i^T\boldsymbol{\theta})$. Both can be obtained from a quantile regression straightforwardly. An empirical estimator for $\varrho_{Y,X|\mathbf{Z}}(\tau,\iota)$ is

$$\widehat{\varrho}_{Y,X|\mathbf{Z}}(\tau,\iota) = \frac{n^{-1}\sum_{i=1}^n \psi_\tau(Y_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}})\psi_\iota(X_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}})}{\sqrt{\tau(1-\tau)\iota(1-\iota)}}. \tag{3.2}$$

To study the asymptotic property of $\widehat{\varrho}_{Y,X|\mathbf{Z}}(\tau,\iota)$, we denote

$$\begin{aligned}
&\Delta_{11} = \mathrm{E}\{f_{Y|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0)\mathbf{Z}\mathbf{Z}^T\},\ \Delta_{12} = \mathrm{E}\big\{F_{X|\mathbf{Z},Y=\mathbf{Z}^T\boldsymbol{\alpha}^0}(\mathbf{Z}^T\boldsymbol{\theta}^0)f_{Y|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0)\mathbf{Z}\big\},\\
&\Delta_{21} = \mathrm{E}\big\{F_{Y|\mathbf{Z},X=\mathbf{Z}^T\boldsymbol{\theta}^0}(\mathbf{Z}^T\boldsymbol{\alpha}^0)f_{X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}\big\},\ \Delta_{22} = \mathrm{E}\{f_{X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}\mathbf{Z}^T\},\\
&\Sigma_{11} = \mathrm{E}\{F_{Y,X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0, \mathbf{Z}^T\boldsymbol{\theta}^0)\}[1 - \mathrm{E}\{F_{Y,X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0, \mathbf{Z}^T\boldsymbol{\theta}^0)\}],\\
&\Sigma_{22} = \mathrm{E}\{\psi_\tau^2(Y - \mathbf{Z}^T\boldsymbol{\alpha}^0)\mathbf{Z}\mathbf{Z}^T\},\ \Sigma_{33} = \mathrm{E}\{\psi_\iota^2(X - \mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}\mathbf{Z}^T\},\\
&\Sigma_{12} = \mathrm{E}\{F_{Y,X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0, \mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}\},\ \Sigma_{23} = \mathrm{E}\{\psi_\tau(Y - \mathbf{Z}^T\boldsymbol{\alpha}^0)\psi_\iota(X - \mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}\mathbf{Z}^T\},
\end{aligned}$$

where $\boldsymbol{\alpha}^0$ and $\boldsymbol{\theta}^0$ are defined in (3.1). We have the following asymptotic result.

Theorem 4. *Let $0 < a < b < 1$. Suppose Δ_{11} and Δ_{22} are uniformly positive-definite matrices in τ and ι, and there exists a constant $\pi > 0$, such that $f_{Y|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0 + \cdot)$, $f_{Y|\mathbf{Z},X}(\mathbf{Z}^T\boldsymbol{\alpha}^0 + \cdot)$, $f_{X|\mathbf{Z},Y}(\mathbf{Z}^T\boldsymbol{\theta}^0 + \cdot)$, and $f_{X|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\theta}^0 + \cdot)$ are uniformly integrable on $[-\pi,\pi]$, and uniformly bounded away from zero and infinity in τ and ι. Then,*

$$\sqrt{n}\{\widehat{\varrho}_{Y,X|\mathbf{Z}}(\tau,\iota) - \varrho_{Y,X|\mathbf{Z}}(\tau,\iota)\} \overset{w}{\rightsquigarrow} \mathbb{G}_{Y,X|\mathbf{Z}}(\tau,\iota)$$

in $\ell^\infty([a,b]^2)$, where $\mathbb{G}_{Y,X|\mathbf{Z}}(\tau,\iota)$ is a Gaussian process with mean zero and covariance function $\Omega_2(\tau_1,\iota_1;\tau_2,\iota_2) \equiv \mathrm{E}\{[\zeta(Y,X,\mathbf{Z};\tau_1,\iota_1) - \mathrm{E}\zeta(Y,X,\mathbf{Z};\tau_1,\iota_1)] \times [\zeta(Y,X,\mathbf{Z};\tau_2,\iota_2) - \mathrm{E}\zeta(Y,X,\mathbf{Z};\tau_2,\iota_2)]\}$, and $\zeta(Y,X,\mathbf{Z};\tau,\iota) = [I(Y \le \mathbf{Z}^T\boldsymbol{\alpha}^0, X \le \mathbf{Z}^T\boldsymbol{\theta}^0) - \Delta_{12}^T\Delta_{11}^{-1}I(Y \le \mathbf{Z}^T\boldsymbol{\alpha}^0)\mathbf{Z} - \Delta_{21}^T\Delta_{22}^{-1}I(X \le \mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}]/\sqrt{\tau(1-\tau)\iota(1-\iota)}$.

If $\mathbf{Z} \equiv 1$ (i.e., no conditional variable is available), the asymptotic distribution in Theorem 4 reduces to that in Theorem 1. The above result implies that for a fixed pair (τ,ι), if $\varrho_{Y,X|\mathbf{Z}}(\tau,\iota) = 0$, then $\sqrt{n}\widehat{\varrho}_{Y,X|\mathbf{Z}}(\tau,\iota) \to^d N(0,\Omega_2)$, where $\Omega_2 \equiv \Omega_2(\tau,\iota;\tau,\iota) = \mathrm{E}\{[\zeta(Y,X,\mathbf{Z};\tau,\iota) - \mathrm{E}\zeta(Y,X,\mathbf{Z};\tau,\iota)]^2\} = (1/(\tau(1-\tau)\iota(1-\iota)))[\Sigma_{11} + \Delta_{12}^T\Delta_{11}^{-1}\Sigma_{22}\Delta_{11}^{-1}\Delta_{12} + \Delta_{21}^T\Delta_{22}^{-1}\Sigma_{33}\Delta_{22}^{-1}\Delta_{21} - 2(1-\tau)\Delta_{12}^T\Delta_{11}^{-1}\Sigma_{12} - 2(1-$

$\iota)\Delta_{21}^T\Delta_{22}^{-1}\Sigma_{12} + 2\Delta_{12}^T\Delta_{11}^{-1}\Sigma_{23}\Delta_{22}^{-1}\Delta_{21}]$.

This theorem can be used for a statistical inference if we can find a consistent estimate for Ω_2. To this end, let $e_1^* = Y - \mathbf{Z}^T\boldsymbol{\alpha}^0$ and $e_2^* = X - \mathbf{Z}^T\boldsymbol{\theta}^0$, and assume that the random vectors $(e_1^*, \mathbf{Z}, X)$ and $(e_2^*, \mathbf{Z}, Y)$ have joint densities $f_{e_1^*,\mathbf{Z},X}$ and $f_{e_2^*,\mathbf{Z},Y}$, respectively. Denote by $f_{e_1^*}$, $f_{e_2^*}$, $f_{e_1^*|\mathbf{Z}}$, $f_{e_1^*|\mathbf{Z},X}$, $f_{e_2^*|\mathbf{Z}}$, and $f_{e_2^*|\mathbf{Z},Y}$, the marginal densities of e_1^* and e_2^*, the conditional densities of e_1^* given $\mathbf{Z}$ and $(\mathbf{Z}, X)$, and the conditional densities of e_2^* given $\mathbf{Z}$ and $(\mathbf{Z}, Y)$, respectively. Then, we can verify that $\Delta_{11} = \mathrm{E}\{f_{e_1^*|\mathbf{Z}}(0)\mathbf{Z}\mathbf{Z}^T\} = f_{e_1^*}(0)\mathrm{E}\{\mathbf{Z}\mathbf{Z}^T|e_1^* = 0\}$ and, similarly, $\Delta_{12} = f_{e_1^*}(0)\mathrm{E}\{I(X \le \mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}|e_1^* = 0\}$, $\Delta_{21} = f_{e_2^*}(0)\mathrm{E}\{I(Y \le \mathbf{Z}^T\boldsymbol{\alpha}^0)\mathbf{Z}|e_2^* = 0\}$, and $\Delta_{22} = f_{e_2^*}(0)\mathrm{E}\{\mathbf{Z}\mathbf{Z}^T|e_2^* = 0\}$. To estimate these quantities, we first calculate the quantile regression estimates $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\theta}}$, and then obtain the corresponding quantile residuals $\hat{e}_{1i}^* = Y_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}}$ and $\hat{e}_{2i}^* = X_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}$, for $i = 1, \ldots, n$. Next, we estimate Δ_{12}. The estimates for Δ_{11}, Δ_{21}, and Δ_{22} can be obtained similarly. We use the nonparametric NW method used to estimate $\sigma_{X|Y}(\tau, \iota)$ and $\sigma_{Y|X}(\tau, \iota)$ in Section 2.1, to obtain estimates for each component of $\mathbf{m}(s) = \mathrm{E}\{I(X \le \mathbf{Z}^T\boldsymbol{\theta}^0)\mathbf{Z}|e_1^* = s\}$ from the data $\{(\hat{e}_{1i}^*, I(X_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}})\mathbf{Z}_i), i = 1, \ldots, n\}$; denote these estimates by $\widehat{\mathbf{m}}(s)$. Then, we obtain $\widehat{\Delta}_{12} = \widehat{f}_{e_1^*}(0)\widehat{\mathbf{m}}(0)$, where $\widehat{f}_{e_1^*}(0)$ is the nonparametric kernel density estimate for $f_{e_1^*}(0)$ in Δ_{12}, based on $\{\hat{e}_{1i}^*, i = 1, \ldots, n\}$. It can be shown that $\widehat{\Delta}_{12}$, is consistent under some regularity conditions. For the other unknown terms in Ω_2, we have $\widehat{\Sigma}_{11} = n^{-1}\sum_{i=1}^n I(Y_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}}, X_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}) - [n^{-1}\sum_{i=1}^n I(Y_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}}, X_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}})]^2$, $\widehat{\Sigma}_{22} = n^{-1}\sum_{i=1}^n \psi_\tau^2(Y_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}})\mathbf{Z}_i\mathbf{Z}_i^T$, $\widehat{\Sigma}_{33} = n^{-1}\sum_{i=1}^n \psi_\iota^2(X_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}})\mathbf{Z}_i\mathbf{Z}_i^T$, $\widehat{\Sigma}_{12} = n^{-1}\sum_{i=1}^n I(Y_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}}, X_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}})\mathbf{Z}_i$, and $\widehat{\Sigma}_{23} = n^{-1}\sum_{i=1}^n \psi_\tau(Y_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}})\psi_\iota(X_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}})\mathbf{Z}_i\mathbf{Z}_i^T$. We use the plug-in approach to obtain a consistent estimate of Ω_2, which we denote by $\widehat{\Omega}_2$.

The next theorem tests whether $\varrho_{\tau,\iota}(Y, X_1|\mathbf{Z}) = \varrho_{\tau,\iota}(Y, X_2|\mathbf{Z})$ for two random variables X_1 and X_2. Write $\boldsymbol{\theta}_k^0 = \mathrm{argmin}_{\boldsymbol{\theta}}\mathrm{E}\{\rho_\iota(X_k - \mathbf{Z}^T\boldsymbol{\theta})\}$, for $k = 1, 2$, and let $\Delta_{12}^{(k)}$ be Δ_{12}, where X and $\boldsymbol{\theta}^0$ are replaced by X_k and $\boldsymbol{\theta}_k^0$, respectively, for $k = 1, 2$. In the same manner, define $\Delta_{21}^{(k)}$, $\Delta_{22}^{(k)}$, $\Sigma_{11}^{(k)}$, $\Sigma_{33}^{(k)}$, $\Sigma_{12}^{(k)}$, $\Sigma_{23}^{(k)}$ and, accordingly, $\Omega_2^{(k)}$, for $k = 1, 2$. In addition, write $\Delta_{31} = \mathrm{E}\{F_{Y,X_1,X_2|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0, \mathbf{Z}^T\boldsymbol{\theta}_1^0, \mathbf{Z}^T\boldsymbol{\theta}_2^0)\} - \mathrm{E}\{F_{Y,X_1|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0, \mathbf{Z}^T\boldsymbol{\theta}_1^0)\}\mathrm{E}\{F_{Y,X_2|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0, \mathbf{Z}^T\boldsymbol{\theta}_2^0)\}$, $\Delta_{32} = \mathrm{E}\{F_{Y,X_1,X_2|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0, \mathbf{Z}^T\boldsymbol{\theta}_1^0, \mathbf{Z}^T\boldsymbol{\theta}_2^0)\mathbf{Z}\}$, and $\Delta_{33} = \mathrm{E}\{\psi_\iota(X_1 - \mathbf{Z}^T\boldsymbol{\theta}_1^0)\psi_\iota(X_2 - \mathbf{Z}^T\boldsymbol{\theta}_2^0)\mathbf{Z}\mathbf{Z}^T\}$.

Theorem 5. *Let $0 < a < b < 1$. Suppose that matrices Δ_{11} and $\Delta_{22}^{(k)}$, for $k = 1, 2$, are uniformly positive definite in τ and ι, and there exists a constant $\pi > 0$, such that $f_{Y|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\alpha}^0 + \cdot)$, $f_{Y|\mathbf{Z},X_k}(\mathbf{Z}^T\boldsymbol{\alpha}^0 + \cdot)$, $f_{X_k|\mathbf{Z},Y}(\mathbf{Z}^T\boldsymbol{\theta}^0 + \cdot)$, and $f_{X_k|\mathbf{Z}}(\mathbf{Z}^T\boldsymbol{\theta}^0 + \cdot)$ are uniformly integrable on $[-\pi, \pi]$, for $k = 1, 2$, and uniformly bounded away from zero and infinity in τ and ι. Then,*

$$\sqrt{n}\{[\widehat{\varrho}_{Y,X_1|\mathbf{Z}}(\tau,\iota)-\widehat{\varrho}_{Y,X_2|\mathbf{Z}}(\tau,\iota)]-[\varrho_{Y,X_1|\mathbf{Z}}(\tau,\iota)-\varrho_{Y,X_2|\mathbf{Z}}(\tau,\iota)]\} \overset{w}{\rightsquigarrow} \mathbb{G}_{Y,X_1,X_2|\mathbf{Z}}(\tau,\iota)$$

in $\ell^\infty([a,b]^2)$*, where* $\mathbb{G}_{Y,X_1,X_2|\mathbf{Z}}(\tau,\iota)$ *is a Gaussian process with mean zero and covariance function* $\Xi_2(\tau_1,\iota_1;\tau_2,\iota_2) \equiv \mathrm{E}\{[\beta(Y,X_1,X_2,\mathbf{Z};\tau_1,\iota_1) - \mathrm{E}\beta(Y,X_1,X_2,\mathbf{Z};\tau_1,\iota_1)]\times[\beta(Y,X_1,X_2,\mathbf{Z};\tau_2,\iota_2)-\mathrm{E}\beta(Y,X_1,X_2,\mathbf{Z};\tau_2,\iota_2)]\}$*, and* $\beta(Y,X_1,X_2,\mathbf{Z};\tau,\iota) = \zeta(Y,X_1,\mathbf{Z};\tau,\iota)-\zeta(Y,X_2,\mathbf{Z};\tau,\iota)$*, where* $\zeta(Y,X_1,\mathbf{Z};\tau,\iota)$ *is given as in Theorem* 4.

For fixed (τ,ι), if $\varrho_{\tau,\iota}(Y,X_1|\mathbf{Z}) = \varrho_{\tau,\iota}(Y,X_2|\mathbf{Z})$, then $\sqrt{n}[\widehat{\varrho}_{Y,X_1|\mathbf{Z}}(\tau,\iota) - \widehat{\varrho}_{Y,X_2|\mathbf{Z}}(\tau,\iota)] \xrightarrow{d} N(0,\Xi_2)$, where $\Xi_2 \equiv \Xi_2(\tau,\iota;\tau,\iota) = \Omega_2^{(1)} + \Omega_2^{(2)} - 2B_{12}$ and $B_{12} \equiv B_{12}(\tau,\iota) = 1/(\tau(1-\tau)\iota(1-\iota))[\Delta_{31} - (1-\tau)(\Delta_{12}^{(2)})^T\Delta_{11}^{-1}\Sigma_{12}^{(1)} - (1-\tau)(\Delta_{12}^{(1)})^T\Delta_{11}^{-1}\Sigma_{12}^{(2)} + (\Delta_{12}^{(1)})^T\Delta_{11}^{-1}\Sigma_{22}\Delta_{11}^{-1}\Delta_{12}^{(2)} + (\Delta_{12}^{(1)})^T\Delta_{11}^{-1}\times\Sigma_{23}^{(2)}(\Delta_{22}^{(2)})^{-1}\Delta_{21}^{(2)} + (\Delta_{21}^{(1)})^T(\Delta_{22}^{(1)})^{-1}\Delta_{32} + (\Delta_{21}^{(1)})^T(\Delta_{22}^{(1)})^{-1}\Sigma_{23}^{(1)}\Delta_{11}^{-1}\Delta_{12}^{(2)} + (\Delta_{21}^{(2)})^T(\Delta_{22}^{(2)})^{-1}(\iota\Sigma_{12}^{(1)} - \Delta_{32}) + (\Delta_{21}^{(1)})^T(\Delta_{22}^{(1)})^{-1}\Delta_{33}(\Delta_{22}^{(2)})^{-1}\Delta_{21}^{(2)}]$. Given a sample of observations $\{(Y_i, X_{i1}, X_{i2}, \mathbf{Z}_i), i = 1,\ldots,n\}$, the asymptotic variance Ξ_2 can be estimated as $\widehat{\Xi}_2 = \widehat{\Omega}_2^{(1)} + \widehat{\Omega}_2^{(2)} - 2\widehat{B}_{12}$, where $\widehat{\Omega}_2^{(1)}$ and $\widehat{\Omega}_2^{(2)}$ are defined as $\widehat{\Omega}_2$ given above. To obtain the estimate $\widehat{B}_{12}$, we need only estimate Δ_{31}, Δ_{32}, and Δ_{33}, because the other unknown quantities in B_{12} can be estimated using the previous methods. Specifically, we use the following estimates: $\widehat{\Delta}_{31} = n^{-1}\sum_{i=1}^n I(Y_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}}, X_{i1} \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_1, X_{i2} \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_2) - \big[n^{-1}\sum_{i=1}^n I(Y_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}}, X_{i1} \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_1)\big]\big[n^{-1}\sum_{i=1}^n I(Y_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}}, X_{i2} \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_2)\big]$, $\widehat{\Delta}_{32} = n^{-1}\sum_{i=1}^n I(Y_i \le \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}}, X_{i1} \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_1, X_{i2} \le \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_2)\mathbf{Z}_i$, and $\widehat{\Delta}_{33} = n^{-1}\sum_{i=1}^n \psi_\iota(X_{i1} - \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_1)\psi_\iota(X_{i2} - \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_2)\mathbf{Z}_i\mathbf{Z}_i^T$, where $\widehat{\boldsymbol{\theta}}_k = \operatorname{argmin}_{\boldsymbol{\theta}} n^{-1}\sum_{i=1}^n \rho_\iota(X_{ik} - \mathbf{Z}_i^T\boldsymbol{\theta})$, for $k = 1, 2$.

3.2. CPC-based variable screening

We now propose a joint robust screening using the CPC. There are two practical scenarios that favor joint screening over marginal screening. First, we may acquire low-dimensional variables $\mathbf{W} \in \mathbb{R}^r$, in addition to ultrahigh-dimensional covariates $\mathbf{X}$. For example, when studying the relationship between a disease phenotype Y and genetic variables $\mathbf{X}$, we may also have patient demographic information or environmental variables, which we include in $\mathbf{W}$. Consequently, we have the data set $\{(Y_i, \mathbf{X}_i, \mathbf{W}_i), i = 1,\ldots,n\}$. Second, even if there is no external variable $\mathbf{W}$, it may still be necessary to consider joint screening by removing the effects from correlated components in $\mathbf{X}$. For instance, some covariates, $\mathbf{X}_{\mathcal{S}_j}$, may be closely correlated to X_j and, thus, influence the observed correlation between Y and X_j indirectly, where $\mathcal{S}_j$ is a subset of $\{1,\ldots,p_n\} \setminus \{j\}$. Ma, Li and Tsai (2017) considered a set $\mathcal{S}_j$ that they refer to as a conditional set

with relatively small size ($< n$). To account for both scenarios, we consider the conditional variables $\mathbf{Z} = (\mathbf{W}^T, \mathbf{X}_{\mathcal{S}_j}^T)^T$, which we allow to vary with j. However, for simplicity of presentation, we use $\mathbf{Z}$ instead of $\mathbf{Z}_j$, and denote by q_n the dimension of $\mathbf{Z}$. In principle, we only need $q_n = \max_{1\le j\le p_n}(r + |\mathcal{S}_j|)$ for sure screening. In practice, we can select a proper $\mathcal{S}_j$ as follows: Treat X_j as the response and $\mathbf{X}_{-j} = \{X_k, k \neq j, 1 \le k \le p_n\}$ as the predictors. Then, apply any sensible marginal screening method, such as the CC-SIS, to select the top ℓ most important predictors, and set these as the conditional variables.

For ultrahigh-dimensional covariates $\mathbf{X} = (X_1, \ldots, X_{p_n})^T$, we can define the CPC between Y and the jth covariate X_j, given $\mathbf{Z}$, in the same way as in (3.1); that is,

$$\varrho_{Y,X_j|\mathbf{Z}}(\tau,\iota) = \frac{\mathrm{E}\{\psi_\tau(Y - \mathbf{Z}^T\boldsymbol{\alpha}^0)\psi_\iota(X_j - \mathbf{Z}^T\boldsymbol{\theta}_j^0)\}}{\sqrt{\tau(1-\tau)\iota(1-\iota)}}, \tag{3.3}$$

where $\boldsymbol{\alpha}^0 = \operatorname{argmin}_{\boldsymbol{\alpha}}\mathrm{E}\{\rho_\tau(Y - \mathbf{Z}^T\boldsymbol{\alpha})\}$ and $\boldsymbol{\theta}_j^0 = \operatorname{argmin}_{\boldsymbol{\theta}_j}\mathrm{E}\{\rho_\iota(X_j - \mathbf{Z}^T\boldsymbol{\theta}_j)\}$. As in (3.2), a sample estimate for $\varrho_{Y,X_j|\mathbf{Z}}(\tau,\iota)$ can be given as

$$\widehat{\varrho}_{Y,X_j|\mathbf{Z}}(\tau,\iota) = \frac{n^{-1}\sum_{i=1}^n \psi_\tau(Y_i - \mathbf{Z}_i^T\widehat{\boldsymbol{\alpha}})\psi_\iota(X_{ij} - \mathbf{Z}_i^T\widehat{\boldsymbol{\theta}}_j)}{\sqrt{\tau(1-\tau)\iota(1-\iota)}}, \tag{3.4}$$

where $\widehat{\boldsymbol{\alpha}} = \operatorname{argmin}_{\boldsymbol{\alpha}} n^{-1}\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{Z}_i^T\boldsymbol{\alpha})$ and $\widehat{\boldsymbol{\theta}}_j = \operatorname{argmin}_{\boldsymbol{\theta}_j} n^{-1}\sum_{i=1}^n \rho_\iota (X_{ij} - \mathbf{Z}_i^T\boldsymbol{\theta}_j)$. The CPC screening yields the following empirical active set:

$$\widehat{\mathcal{M}}_b = \{j : |\widehat{\varrho}_{Y,X_j|\mathbf{Z}}(\tau,\iota)| \ge v_n, 1 \le j \le p_n\}, \tag{3.5}$$

where v_n is a user-specified threshold parameter. We refer to this sure independence screening procedure as the CPC-SIS. Clearly, the CPC-SIS extends earlier conditional sure independence screening methods, such as that of Barut, Fan and Verhasselt (2016).

Let $\mathcal{M}_b^* = \{j : |\varrho_{Y,X_j|\mathbf{Z}}(\tau,\iota)| > 0, j = 1, \ldots, p\}$ be the true active set. We write $F_{Y|\mathbf{X},\mathbf{W}}^{-1}(\tau) = \inf\{y : P(Y \le y|\mathbf{X},\mathbf{W}) \ge \tau\}$. For simplicity, we still use $u_j = |\varrho_{Y,X_j|\mathbf{Z}}(\tau,\iota)|$ and $\widehat{u}_j = |\widehat{\varrho}_{Y,X_j|\mathbf{Z}}(\tau,\iota)|$ to denote the underlying and empirical CPC utilities, respectively. To establish the sure screening property, we need the following conditions, which are very mild and similarly imposed in Ma, Li and Tsai (2017).

(D1) (i) The conditional density $f_{Y|\mathbf{Z}=\mathbf{z}}(y)$ of Y, given $\mathbf{Z} = \mathbf{z}$, satisfies the Lipschitz condition of order one, and $f_{Y|\mathbf{Z}=\mathbf{z}}(y) > 0$ for any y in a neighborhood

of $\mathbf{Z}^T\boldsymbol{\alpha}^0 = \mathbf{z}^T\boldsymbol{\alpha}^0$. (ii) For every $1 \le j \le p_n$, the conditional density $f_{X_j|\mathbf{Z}=\mathbf{z}}(x)$ of X_j, given $\mathbf{Z} = \mathbf{z}$, satisfies the Lipschitz condition of order one, and $f_{X_j|\mathbf{Z}=\mathbf{z}}(x) > 0$ for any x in a neighborhood of $\mathbf{Z}^T\boldsymbol{\theta}^0 = \mathbf{z}^T\boldsymbol{\theta}^0$.

(D2) (i) There exist some finite constants m_1, m_2, and m_3, such that

$$\max_{i,j} |Z_{ij}| \le m_1, \quad \max_i |\mathbf{Z}_i^T\boldsymbol{\alpha}^0| \le m_2, \quad \max_{i,j} |\mathbf{Z}_i^T\boldsymbol{\theta}_j^0| \le m_3.$$

(ii) There exist two positive-finite constants $c_{\min}$ and $c_{\max}$, such that

$$c_{\min} \le \lambda_{\min}(\mathrm{E}(\mathbf{Z}\mathbf{Z}^T)) \le \lambda_{\max}(\mathrm{E}(\mathbf{Z}\mathbf{Z}^T)) \le c_{\max},$$

where $\lambda_{\min}(\mathrm{E}(\mathbf{Z}\mathbf{Z}^T))$ and $\lambda_{\max}(\mathrm{E}(\mathbf{Z}\mathbf{Z}^T))$ denote the minimum and maximum eigenvalues of $\mathrm{E}(\mathbf{Z}\mathbf{Z}^T)$, respectively.

(D3) $\min_{j\in\mathcal{M}_b^*} u_j \ge C_0^* n^{-\kappa}$, for some $\kappa > 0$ and $C_0^* > 0$.

Theorem 6. *(Screening Property for CPC-SIS) Suppose conditions* (D1) *and* (D2) *hold. Then:*
(i) For any constant $C > 0$, there exists some positive constant $\tilde{c}_1^*$, such that for sufficiently large n,

$$P\Big(\max_{1\le j\le p_n} \big|\widehat{u}_j - u_j\big| \ge Cn^{-\kappa}\Big) \le 12p_n \exp(-\tilde{c}_1^* q_n^{-1} n^{1-2\kappa}).$$

(ii) In addition, if condition (D3) is satisfied, by choosing $v_n = C_2 n^{-\kappa}$ with $C_2 \le C_0^*/2$, we have

$$P\big(\mathcal{M}_b^* \subset \widehat{\mathcal{M}}_b\big) \ge 1 - 12s_n \exp(-\tilde{c}_1^* q_n^{-1} n^{1-2\kappa})$$

for sufficiently large n, where $s_n = |\mathcal{M}_b^*|$.

When conditional variables are available, our proposed CPC-SIS method can handle dimensionality of order $p_n = o(\exp(q_n^{-1}n^{1-2\kappa}))$. If $q_n = O(1)$, then the dimension can be as high as $o(n^{1-2\kappa})$, the same order as that of the CC-SIS. Moreover, the proposed CPC-SIS can be readily used for ultrahigh-dimensional data, as long as $q_n = o(n^{1-2\kappa})$.

As in Section 2.2, we can determine a proper v_n by controlling the FDR. From Theorem 4, for covariate j, such that $\varrho_{Y,X_j|\mathbf{Z}}(\tau,\iota) = 0$, we have $\sqrt{n}\widehat{\Omega}_2^{-1/2}\widehat{\varrho}_{Y,X_j|\mathbf{Z}}(\tau,\iota) \sim N(0,1)$ asymptotically. Then, we select variables $\widehat{\mathcal{M}}_{b,\delta} = \{j : \sqrt{n}\widehat{\Omega}_2^{-1/2}|\widehat{\varrho}_{Y,X_j|\mathbf{Z}}(\tau,\iota)| \ge \delta\}$ for a small $\delta > 0$, which controls the FDR $\mathrm{E}\{|\widehat{\mathcal{M}}_{b,\delta} \cap (\mathcal{M}_b^*)^c|/|(\mathcal{M}_b^*)^c|\}$.

Proposition 2. *(FDR Property) Assume conditions* (D1)–(D3) *and the condition of Theorem* 4 *hold. Then, if we choose* $\delta = \Phi^{-1}(1 - \bar{d}_n/(2p_n))$, *where* $\Phi(\cdot)$ *and* $\bar{d}_n$ *are defined in Proposition* 1, *then for some constant* $c_b > 0$, *we have*

$$\mathrm{E}\left\{ \frac{|\widehat{\mathcal{M}}_{b,\delta} \cap (\mathcal{M}_b^*)^c|}{|(\mathcal{M}_b^*)^c|} \right\} \leq \frac{\bar{d}_n}{p_n} + \frac{c_b}{\sqrt{n}}.$$

4. Implementation of the CPC-SIS

For the implementation of the CPC-SIS, we consider three practical types of conditional variables.

Case 1. If $\mathbf{W}$ is not available, we consider the conditional variables from $\mathbf{X}$ itself for each X_j; that is, $\mathbf{Z} = \mathbf{X}_{\mathcal{S}_j}$, for $j = 1, \ldots, p_n$. We start with an empty active set $\mathcal{A}^{(0)} = \emptyset$.

- Step 1. For $j = 1, \ldots, p_n$, select the confounding sets $\mathcal{S}_j^\nu$ using the partial correlation-based consequential test (Ma, Li and Tsai (2017)).

- Step 2. In the kth iteration, where $k = 1, \ldots, d^*$ and $d^* = \lfloor 2(n/\log n)^{1/2} \rfloor$, for given $\mathcal{A}^{(k-1)}$, update $\mathcal{S}_j = \mathcal{A}^{(k-1)} \cup \mathcal{S}_j^\nu$ and determine the variable index j^*, such that $j^* = \mathrm{argmax}_{j \notin \mathcal{A}^{(k-1)}} |\widehat{\varrho}_{Y,X_j|\mathbf{Z}}(\tau, \iota)|$. Update $\mathcal{A}^{(k)} = \mathcal{A}^{(k-1)} \cup \{j^*\}$.

- Step 3. In the kth iteration, where $k = d^*+1, \ldots, d_n$, set $\mathcal{S}_j = \mathcal{A}^{(d^*)} \cup \mathcal{S}_j^\nu$ and determine $j^* = \mathrm{argmax}_{j \notin \mathcal{A}^{(k-1)}} |\widehat{\varrho}_{Y,X_j|\mathbf{Z}}(\tau, \iota)|$. Update $\mathcal{A}^{(k)} = \mathcal{A}^{(k-1)} \cup \{j^*\}$. Use $\mathcal{A}^{(d_n)} \equiv \widehat{\mathcal{M}}_b$ as the final set of selected covariates.

Note that the main difference between Steps 2 and 3 is that, in Step 2, the conditional set is updated gradually by adding one selected index variable in the first d^* iterations; in Step 3, the conditional set remains intact in the final $d_n - d^*$ iterations.

Case 2. If $\mathbf{W}$ is available, we consider the same conditional variables for each target X_j; that is, $\mathbf{Z} = \mathbf{W}$, for $j = 1, \ldots, p_n$.

- Step 1. For $j = 1, \ldots, p_n$, compute the CPC utility statistics $\widehat{u}_j = |\widehat{\varrho}_{Y,X_j|\mathbf{Z}}(\tau, \iota)|$.

- Step 2. Rank the covariates in terms of $\widehat{u}_j$ in decreasing order, and then select the top d_n covariates as the final set of selected covariates.

Case 3. If $\mathbf{W}$ is available, we slightly modify the algorithm in Case 1. The

steps are the same as those in Case 1, except that we consider the conditional variables $\mathbf{Z} = (\mathbf{W}^T, \mathbf{X}_{\mathcal{S}_j}^T)^T$ in each iteration, for $1 \le k \le d_n$, for each step.

Note that Case 1 uses only confounding information from the covariates $\mathbf{X}$, whereas Case 2 incorporates exogenous conditional information, but ignores the confounding effect from $\mathbf{X}$ itself. Case 3 is the most flexible version, incorporating all types of covariate information. We implement Case 3 in the real-data analysis in Section 5.

5. Numerical Studies

5.1. Simulation studies

In this section, we present simulation studies to illustrate the finite-sample performance of the proposed screening procedure, CPC-SIS. Owing to space constraints, simulation studies related to the CC and CPC estimates and their corresponding asymptotic variance estimates for small, moderate, and large sample sizes can be found in Appendix C of the online Supplementary Material. The results from Examples S1–S4 in Appendix C reflect the effectiveness of Theorems 1, 2, 4, and 5.

Throughout this subsection, we set the sample size $n = 200$, the covariate dimension $p_n = 1{,}000$, and the number of simulations $N = 200$ for each simulation setup. Moreover, for the purpose of comparison, we use three criteria for evaluation: (i) the minimum model size (MMS), that is, the smallest number of the selected covariates that contain all active covariates, and its robust standard deviation (RSD); (ii) the rank for each active covariate (R_j); and (iii) the proportion of all active covariates selected ($\mathcal{P}$) with the screening threshold specified as $\lfloor n/\log n \rfloor$ over N simulations. We report the median of MMS and R_j.

In Example S5 in Appendix C, we compare our CC-SIS method with the following existing methods: SIS (Fan and Lv (2008)), SIRS (Zhu et al. (2011)), DC-SIS (Li, Zhong and Zhu (2012)), Kendall-SIS (Li et al. (2012)), QC-SIS (Li, Li and Tsai (2015)), and CQC-SIS (Ma and Zhang (2016)). Moreover, in Example S6 in Appendix C, we compare our CPC-SIS procedure with these marginal screening methods, and with the QPC-SIS method of Ma, Li and Tsai (2017), where confounding effects arise from the covariates $\mathbf{X}$; for the latter comparison, we employ the algorithm in Case 1, given in Section 4.

In what follows, we examine the case $\mathbf{Z} = \mathbf{W}$. Because the conditional variables selected for each X_j are the same, we can apply the CPC-SIS with the algorithm in Case 2, and compare the results with those from the QPC-SIS of

Ma, Li and Tsai (2017).

Example 1. We generate the response from the model $Y = 2X_1 + 2X_2 - 4X_3 + 3X_4 + \varepsilon$, where $X_j = \mathbf{W}^T\mathbf{b} + U_j$. Here, $\mathbf{W}$ is distributed as $N(\mathbf{0}_4, \Sigma)$, where $\Sigma = (\rho^{|j-k|})_{1\le j,k\le 4}$; in addition, $\mathbf{b} = (2, 4/3, 2, 4/3)^T$ and $U_j \sim (1/3)Cauchy(0, 1)$, for $j = 1, \ldots, p_n$. The model error ε is simulated as $N(0, 1)$ or $(1/3)Cauchy(0, 1)$. The simulation results are given in Table 1. As expected, none of the marginal screening procedures work, because they are unable to identify the covariate X_3. Our proposed CPC-SIS method outperforms the QPC-SIS in terms of the MMS, and both outperform the marginal procedures.

Furthermore, following a reviewer's suggestion, we can consider feature screening in terms of hypothesis testing. According to Chang, Tang and Wu (2013) and Chang, Tang and Wu (2016), viewing a feature screening problem as a hypothesis testing problem can efficiently avoid the effect of heteroscedasticity in the estimators of the correlations. This is very important when the sample size n is small. For a further comparison, we also implement the testing-based screening procedure (T-SIS), which is given in Appendix B of the online Supplementary Material. Because our simulation results show that the proposed CC-SIS outperforms many existing marginal variable screening methods, we compare our CC-SIS with the T-SIS procedure and the maximum CC-based sure independence screening (mCC-SIS) procedure described in Section 6; see Examples S7 and S8 given in Appendix C in the online Supplementary Material. The simulation results show that our proposed CC-SIS performs best at the median level of (τ, ι) for a small sample size, and that the mCC-SIS dominates other methods for large sample sizes.

5.2. Real-data applications

5.2.1. Rats data

We illustrate the CC-SIS and CPC-SIS using gene expression data on 120 12-weeks-old male rats, including expression measurements from 31,099 gene probes. This data set was analyzed in Scheetz et al. (2006) to investigate gene regulation in mammals; the data set is available at `ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE5nnn/GSE5680/matrix`. We follow Ma, Li and Tsai (2017) and consider the expression of gene TRIM32 (probe 1389163_at) as the response variable Y, because this gene has been identified as the cause of Bardet–Biedl syndrome, which is closely associated with a human hereditary disease of the retina Chiang et al. (2006). The other gene probes are treated as the covariates $\mathbf{X}$. We first

Table 1. Simulation results for Example 1, where R_j indicates the median of the rank of the relevant predictors; and MMS stands for the median of the minimum model size; robust standard deviations (RSD) are given in parentheses.

		$\varepsilon \sim N(0,1)$					$\varepsilon \sim (1/3)Cauchy(0,1)$				
ρ	Method(τ,ι)	R_1	R_2	R_3	R_4	MMS (RSD)	R_1	R_2	R_3	R_4	MMS (RSD)
0.5	SIS	12	12	377	4	488 (518)	14	14	396	5	532 (494)
	SIRS	192	215	910	188	910 (96)	236	222	938	193	938 (98)
	DC-SIS	336	305	511	320	744 (162)	305	272	546	324	746 (154)
	Kendall-SIS	2	2	997	1	997 (15)	2	2	998	1	998 (10)
	CC-SIS$_{(0.25,0.25)}$	3	3	841	2	841 (235)	5	3	830	2	830 (263)
	CC-SIS$_{(0.5,0.5)}$	3	4	886	2	886 (175)	3	3	895	2	895 (207)
	CC-SIS$_{(0.75,0.75)}$	3	3	817	2	817 (240)	4	5	865	2	865 (184)
	QC-SIS$_{(0.25)}$	183	249	713	208	796 (172)	178	148	748	212	806 (170)
	QC-SIS$_{(0.5)}$	269	241	672	307	823 (160)	232	231	687	296	824 (172)
	QC-SIS$_{(0.75)}$	223	174	701	259	829 (179)	154	191	731	209	822 (174)
	QPC-SIS$_{(0.25)}$	6	7	3	3	107 (167)	7	9	3	4	109 (187)
	QPC-SIS$_{(0.5)}$	4	5	2	3	53 (95)	5	5	3	5	77 (118)
	QPC-SIS$_{(0.75)}$	5	7	3	3	75 (148)	8	6	3	4	94 (165)
	CPC-SIS$_{(0.25,0.25)}$	5	4	5	1	28 (58)	7	8	8	2	62 (110)
	CPC-SIS$_{(0.5,0.5)}$	5	6	1	2	14 (27)	6	6	1	2	19 (34)
	CPC-SIS$_{(0.75,0.75)}$	5	8	6	1	41 (88)	7	9	7	2	57 (75)
0.95	SIS	5	9	538	4	581 (378)	17	11	525	4	586 (445)
	SIRS	190	231	894	199	894 (95)	246	220	895	186	895 (103)
	DC-SIS	441	303	508	239	771 (171)	189	273	558	348	776 (174)
	Kendall-SIS	2	3	991	1	991 (32)	2	2	990	1	990 (46)
	CC-SIS$_{(0.25,0.25)}$	3	4	806	2	806 (247)	4	3	825	2	825 (254)
	CC-SIS$_{(0.5,0.5)}$	3	4	831	2	831 (250)	3	4	836	2	836 (257)
	CC-SIS$_{(0.75,0.75)}$	3	4	804	2	804 (314)	4	5	758	2	758 (315)
	QC-SIS$_{(0.25)}$	326	225	624	157	795 (188)	95	202	682	252	805 (192)
	QC-SIS$_{(0.5)}$	400	244	597	226	804 (189)	135	235	661	318	812 (193)
	QC-SIS$_{(0.75)}$	312	168	650	143	785 (185)	73	154	695	272	815 (178)
	QPC-SIS$_{(0.25)}$	5	6	3	3	73 (165)	12	6	3	4	128 (214)
	QPC-SIS$_{(0.5)}$	4	5	3	3	47 (112)	4	8	3	3	78 (115)
	QPC-SIS$_{(0.75)}$	5	5	2	3	52 (146)	8	6	3	4	105 (182)
	CPC-SIS$_{(0.25,0.25)}$	6	5	6	1	31 (47)	6	5	7	2	55 (90)
	CPC-SIS$_{(0.5,0.5)}$	5	6	1	2	20 (40)	6	5	1	2	18 (42)
	CPC-SIS$_{(0.75,0.75)}$	6	6	5	1	41 (80)	6	7	7	2	46 (89)

apply the approach of Iglewicz and Hoaglin (1993) (IH) to check for outliers. IH construct a Z-score, $Z_i = 0.6745(x_i - \tilde{x})/\text{MAD}$, where MAD denotes the median absolute deviation, and $\tilde{x}$ stands for the median. They recommend labeling any i where $Z_i > 3.5$ as an outlier. The IH method is popular in real applications, such as engineering. Following the recommendation of IH, we find that more

than 60% of the gene probes have one or more outliers. Figure 1 displays box plots for the two selected genes and the response. Here, applying a conventional screening method, which ignores outliers, would lead to inappropriate results. The copula-based methods may thus be more robust in this situation. In this data analysis, we use the sample $\{(Y_i, \mathbf{X}_i \in \mathbb{R}^{p_n}), 1 \leq i \leq n\}$, with $n = 120$ and $p_n = 31{,}098$.

We report the overlaps of the top $\lfloor n/\log n \rfloor = 25$ selected genes using various methods in Table 2. We can see that different methods select quite different genes, and such low level of agreement should not be overlooked in practice. Robust and joint screening methods, such as that proposed here lead to entirely different sets of genes that would otherwise be screened out by conventional non-robust marginal screening approaches. Note that the CPC-SIS and QPC-SIS do overlap, partly because both are conditional screening procedures and can adjust the confounder effects.

Table 3 summarized the top 10 gene probes using different methods, and includes the p-value from a marginal Wald-test. We use these 10 genes as regressors and build a joint statistical model to predict Y. Linear and quantile regressions are both considered for this purpose. We provide the mean of their prediction errors (PE1 and PE2) over 500 random partitions, where the partition ratio of the training sample to the test sample is 4 : 1 for each partition. The PE is computed as the average of $\{(Y_i - \widehat{Y}_i)^2, i \in \text{testing set}\}$, and $\widehat{Y}_i$ is the predicted value at the ith test data point, using the model constructed from the training sample of the 10 genes in Table 3. We can see that our proposed copula-based partial correlation screening performs best with the smallest prediction error. This may be because the CPC selects appropriate markers for joint modeling after addressing the distribution heterogeneity and the conditional effects. The heterogeneity problem typically inflates the variance, whereas a purely marginal screener could introduce bias. The prediction error, consisting of the variance and the bias components, is thus much smaller after employing the CPC screening method.

5.2.2. Breast cancer data

The second data set we use to illustrate our proposed method contains a breast cancer data. Breast cancer has become the second most common cancer in the world, and the leading cause in women. Nearly 1.7 million new cases were diagnosed in 2012 (cf., `http://www.wcrf.org/int/cancer-facts-figures/worldwide-data`); see DeSantis et al. (2017) for a discussion on recent trends. Although major progress has been made in terms of breast cancer treatment, our

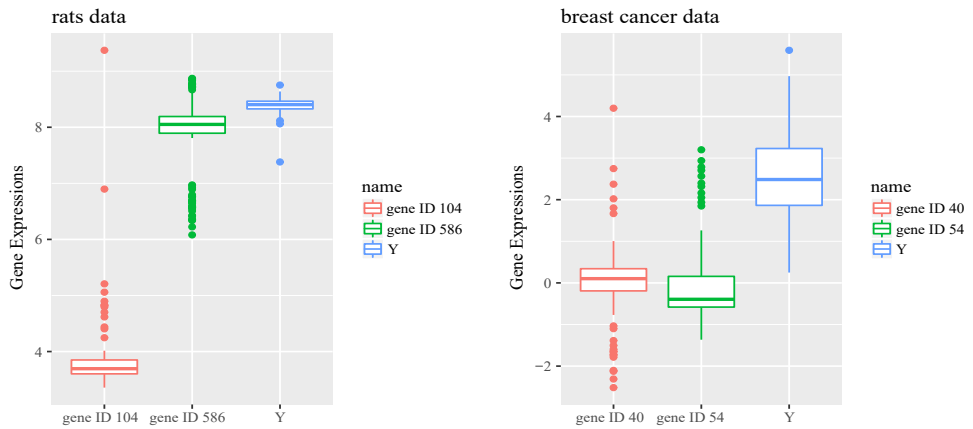

Figure 1. Box plots for the response and two randomly selected genes for the two data sets. The left panel shows the rat data and the right panel shows the breast cancer data.

Table 2. The overlaps of selected genes using various approaches for the rat data, where the screening threshold parameter is set to $\lfloor n/\log n \rfloor = 25$ for each method, and the CPC-SIS applies the algorithm in Case 1.

					QC-SIS(τ)			QPC-SIS(τ)			CC-SIS(τ,ι)			CPC-SIS(τ,ι)		
	SIS	SIRS	DC-SIS	Kendall-SIS	0.25	0.5	0.75	0.25	0.5	0.75	(0.25,0.25)	(0.5,0.5)	(0.75,0.75)	(0.25,0.25)	(0.5,0.5)	(0.75,0.75)
SIS	25	0	1	3	1	1	0	0	0	0	1	3	3	0	0	0
SIRS	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DC-SIS	1	0	25	1	2	2	1	0	0	0	0	0	1	0	0	0
Kendall	3	0	1	25	5	12	3	0	0	0	2	5	3	0	0	0
QC-SIS$_{(0.25)}$	1	0	2	5	25	5	0	0	0	0	3	2	1	0	0	0
QC-SIS$_{(0.5)}$	1	0	2	12	5	25	3	0	0	0	1	7	1	0	0	0
QC-SIS$_{(0.75)}$	0	0	1	3	0	3	25	0	0	0	0	1	1	0	0	0
QPC-SIS$_{(0.25)}$	0	0	0	0	0	0	0	25	3	2	0	0	1	2	1	0
QPC-SIS$_{(0.5)}$	0	0	0	0	0	0	0	3	25	0	0	0	0	1	1	0
QPC-SIS$_{(0.75)}$	0	0	0	0	0	0	0	2	0	25	0	0	0	0	1	0
CC-SIS$_{(0.25,0.25)}$	1	0	0	2	3	1	0	0	0	0	25	1	0	0	0	0
CC-SIS$_{(0.5,0.5)}$	3	0	0	5	2	7	1	0	0	0	1	25	1	0	0	0
CC-SIS$_{(0.75,0.75)}$	3	0	1	3	1	1	1	1	0	0	0	1	25	0	0	0
CPC-SIS$_{(0.25,0.25)}$	0	0	0	0	0	0	0	2	1	0	0	0	0	25	0	0
CPC-SIS$_{(0.5,0.5)}$	0	0	0	0	0	0	0	1	1	1	0	0	0	0	25	1
CPC-SIS$_{(0.75,0.75)}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	25

ability to predict the metastatic behavior of a tumor remains limited. Van't Veer et al. (2002) were the first to study breast cancer using expression data. Their data involved 97 lymph node-negative breast cancer patients, 55 years old or younger, of whom 46 developed distant metastases within five years (metastatic outcome coded as one), and 51 remained metastasis free for at least five years (metastatic outcome coded as zero). This expression data set with clinical variables has been well analyzed in many papers (Boulesteix, Porzelius and Daumer

Table 3. Summary of top 10 gene probes selected by different screening methods for the rat data. ID means the selected gene ID, and p-values are computed as $2(1-\Phi(|\sqrt{n}\widehat{\Omega}_1^{-1/2}\widehat{\varrho}_{Y,X}(0.5,0.5)|))$, where Φ is the CDF of the standard normal random variable. PE is the mean prediction error over 500 random partitions, where the partition ratio of the training sample to the testing sample is $4:1$, where the prediction error is defined as the average of $\{(Y_i-\widehat{Y}_i)^2, i \in \text{testing set}\}$. PE1 and PE2 indicate that $\widehat{Y}_i$ is the predicted value after fitting a median regression model and a linear model, respectively, using the top 10 genes selected.

	SIS		SIRS		DC-SIS		Kendall-SIS		QC-SIS(0.5)		CC-SIS(0.5,0.5)		QPC-SIS(0.5)		CPC-SIS(0.5, 0.5)	
Rank	ID	p-value	ID	p-value	ID	p-value	ID	p-value	ID	p-value	ID	p-value	ID	p-value	ID	p-value
1	14770	2.1E-05	2828	1.000	146	6.4E-04	6083	8.3E-10	22641	5.2E-11	14726	4.4E-16	18602	0.473	1621	0.469
2	21977	1.8E-05	20503	0.480	260	7.2E-04	5002	4.5E-10	14810	6.4E-12	6889	6.6E-14	4101	0.003	11288	0.001
3	6436	2.0E-08	233	0.152	30768	2.3E-06	14726	4.4E-16	22339	8.0E-13	14701	6.2E-14	12365	0.141	12480	1.000
4	4797	1.7E-08	3962	0.716	30745	1.6E-05	14810	6.4E-12	5002	4.5E-10	20898	7.2E-14	8399	1.000	4398	0.271
5	21150	2.3E-07	7656	0.063	285	1.2E-04	25297	1.5E-11	20898	7.2E-14	22339	8.0E-13	5063	0.026	29604	0.467
6	25573	4.5E-10	20453	0.468	30791	1.6E-07	5259	6.5E-12	31008	1.5E-07	23278	6.2E-13	9223	0.467	22679	1.000
7	12127	9.4E-09	22023	0.047	3849	1.1E-04	5223	2.9E-10	26828	1.8E-08	25117	8.8E-14	21746	0.716	22267	0.065
8	9235	1.6E-07	157	0.208	4626	1.1E-04	31008	1.5E-07	24529	4.2E-10	30548	6.9E-14	14019	0.717	17039	0.148
9	3682	2.5E-06	2575	0.153	4490	2.8E-10	22339	8.0E-13	14414	1.5E-07	4512	8.0E-12	30361	0.277	11796	0.720
10	8670	4.5E-11	2841	0.284	3967	2.1E-07	6021	1.7E-07	20724	2.8E-10	4712	8.4E-12	24759	0.010	20967	0.026
PE1	0.0394		0.0252		0.0290		0.0310		0.0283		0.0330		0.0269		0.0247	
PE2	0.0377		0.0269		0.0349		0.0342		0.0344		0.0360		0.0307		0.0257	

(2008), Yu, Li and Ma (2012), among others).

After removing genes with missing values, our sample comprises expression levels of 24,188 gene probes. In addition to gene expression measurements, data on several clinical factors are available as well. Our interest is to identify which gene probes affect the tumor size, given other clinical factors ($\mathbf{W}$), including age, histological grade, angioinvasion, lymphocytic infiltration, estrogen receptor, and progesterone receiptor status. Therefore, we have the data set $\{(Y_i, \mathbf{X}_i \in \mathbb{R}^{p_n}, \mathbf{W}_i \in \mathbb{R}^r), 1 \le i \le n\}$, with $n = 97$, $p_n = 24,188$, and $r = 6$.

Using the IH method for outlier detection, we find that 18,098 gene probes have at least one and at most 29 outliers, suggesting that approximately three-quarters of the gene probes contain extremely large values. The right panel of Figure 1 displays the empirical distributions of the response and the two typical covariates. Thus, it is more suitable to apply a robust joint screening approach, such as the proposed CPC-SIS. We consider the three cases discussed in Section 4, and denote the methods as CPC-SISa_1, CPC-SISa_2, and CPC-SISa_3, respectively. The overlaps of the selected genes from the various methods can be found in Table S13 in the online Supplementary Material. A similar conclusion to that for the rat data analysis is apparent. Furthermore, Table 4 presents a summary of the top 10 gene probes selected using the various methods. The results for

Table 4. Summary of top 10 gene probes selected using different screening methods for the breast cancer data. ID means the selected gene ID and p-values are computed as $2(1-\Phi(|\sqrt{n}\widehat{\Omega}_1^{-1/2}\widehat{\varrho}_{Y,X}(0.5,0.5)|))$, where Φ is the CDF of the standard normal random variable. PE is the mean prediction error over 500 random partitions, where the partition ratio of the training sample to the testing sample is 4 : 1, where the prediction error is defined as the average of $\{(Y_i-\widehat{Y}_i)^2, i \in \text{testing set}\}$. PE1 and PE2 indicate that $\widehat{Y}_i$ is the predicted value after fitting a median regression model and a linear model, respectively, using the top 10 genes selected.

	SIS		SIRS		DC-SIS		Kendall-SIS		QC-SIS(0.5)	
Rank	ID	p-value	ID	p-value	ID	p-value	ID	p-value	ID	p-value
1	24032	3.0E-06	24032	0.000	8349	3.1E-05	17679	6.2E-02	24032	3.0E-06
2	11913	1.2E-07	6841	0.000	24032	3.0E-06	20238	2.3E-02	22705	2.8E-02
3	11870	2.9E-06	9164	0.001	13025	6.7E-04	10408	1.9E-03	6841	6.8E-08
4	17439	6.9E-06	13025	0.001	23670	5.9E-03	1644	6.9E-06	14466	1.8E-03
5	6841	6.8E-08	2172	0.013	20121	1.2E-07	8339	2.6E-03	4767	1.2E-05
6	20938	2.3E-02	17439	0.000	6841	6.8E-08	14028	8.9E-05	5644	2.2E-04
7	10692	2.0E-01	20121	0.000	15674	1.2E-06	23670	5.9E-03	20121	1.2E-07
8	19897	1.5E-03	11870	0.000	1644	6.9E-06	12305	7.3E-04	23670	5.9E-03
9	9164	1.5E-03	22705	0.028	5644	2.2E-04	3929	1.8E-03	13742	1.5E-05
10	17050	2.2E-02	10408	0.002	20238	2.3E-02	14466	1.8E-03	17439	6.9E-06
PE1	1.566		1.483		1.419		1.399		1.409	
PE2	1.550		1.398		1.378		1.366		1.367	
	CC-SIS(0.5,0.5)		QPC-SIS(0.5)		CPC-SISa_1(0.5, 0.5)		CPC-SISa_2(0.5, 0.5)		CPC-SISa_3(0.5, 0.5)	
Rank	ID	p-value	ID	p-value	ID	p-value	ID	p-value	ID	p-value
1	12801	1.5E-06	11696	0.001	301	0.005	20121	0.000	4132	0.136
2	13742	1.5E-05	672	0.005	18678	0.000	4356	0.001	17568	0.620
3	402	6.2E-05	21944	0.021	3524	0.603	13084	0.008	5459	0.482
4	4862	3.4E-04	6466	0.024	5422	0.021	13191	0.035	1079	0.352
5	8349	3.1E-05	518	0.758	14782	0.023	6436	0.299	23942	0.002
6	9158	1.9E-03	12635	0.022	21431	0.922	10179	0.192	14	0.922
7	12074	6.8E-06	12567	0.609	5239	0.295	20102	0.185	1847	0.169
8	14466	1.8E-03	7160	0.483	777	0.352	1299	0.179	3392	0.505
9	18903	2.5E-02	21188	0.007	20958	0.132	1830	0.381	20369	0.367
10	19774	8.1E-06	11916	0.495	4849	0.460	6025	0.467	390	0.920
PE1	1.404		1.466		1.399		1.436		1.372	
PE2	1.290		1.437		1.345		1.304		1.289	

PE1 and PE2 in Table 4 empirically verify that our proposed CPC-SIS method in Case 3 demonstrates the most satisfactory performance in terms of out-of-sample prediction.

6. Choice of Parameters (τ, ι)

In this section, we provide guidance on the choice of the parameters (τ, ι) in the proposed screeners. As in the existing screening literature on quantile correlation, the parameters (τ, ι) play a crucial role in our proposed robust and jointly independent screening procedures. In general, the specification for choosing a suitable (τ, ι) is usually decided by the user, where results are interpreted according to the chosen value. For example, in financial studies, we can choose small or large (τ, ι) depending on whether we are interested in the high or low tail dependence of asset prices. Usually, the median is used in a quantile regression. From our limited simulation experience, using the median quantile level in our CC-SIS and CPC-SIS procedures works better than using other low or high quantile levels; thus, we recommend specifying a median quantile level. Moreover, for the screeners CC-SIS or CPC-SIS, choosing different τ and ι yields different sets of screened covariates. To combine the results from various (τ, ι), we may pursue a global screener over a continuous range of quantiles (e.g., Zheng, Peng and He (2015); Ma and Zhang (2016)). Specifically, we consider the maximum absolute copula-based correlation for variable screening, as suggested by one reviewer as well, in which we let the tuning parameters τ and ι be taken over two intervals. Specifically, we define the following empirical utility as a new screener:

$$\widehat{u}_j = \max_{\tau \in \mathcal{I}_1} \max_{\iota \in \mathcal{I}_2} |\widehat{\varrho}_{Y,X_j}(\tau, \iota)|,$$

where $\mathcal{I}_1 = \mathcal{I}_2 = (0, 1)$. In the implementation, we maximize the absolute correlation with respect to (τ, ι) over a set of discrete points $\{(\tau_k, \iota_l)\}$, where $\tau_k = k/N$, for $1 \le k \le N-1$ and $\iota_l = l/N, 1 \le l \le N-1$ using a pre-specified integer N, rather than maximizing over a continuous range (interval). We call this screening procedure the maximum CC-based sure independence screening (denoted as mCC-SIS). In Examples S7 and S8 in the online supplementary material, we set $N = 10$ for a simple comparison of this method.

7. Conclusion

We propose a copula-based correlation and partial correlation to facilitate robust marginal and joint screening for ultrahigh-dimensional data sets. The large sample properties for the estimated correlation and the sure screening properties for CC and CPC screeners are provided. Empirical studies, including simulations and two data applications, show that our proposed CC-SIS and CPC-SIS outper-

form existing variable screening approaches when outliers are present in both the covariates and the response. Therefore, our current proposals are more applicable to ultrahigh-dimensional heterogeneous data. We provide the following guideline for performing variable screening. If the response and predictors are all normal, without heteroscedastic variance, and the predictors have low correlation, any marginal screening methods (SIS, SIRS, DC-SIS) can be applied. If the response contains outliers, or follows a heavy-tail distribution and the covariates are normal, then robust screening methods (Kendall SIS, QC-SIS, CQC-SIS, CC-SIS) can be employed. If the covariates are highly correlated and conditional variables are available, conditional screening procedures (CSIS, QPC-SIS) can be used. If the data are heteroscedastic for both the response and the covariates, and the covariates may be highly correlated, then only the CPC-SIS is recommended.

The copula formulation suggests several possible extensions to our methodology. First, we may consider a censored survival-time outcome in this framework. See Yue and Li (2017), Hong and Li (2018), and Huang, McKeague and Qian (2019) for recent reviews on feature selection and screening for survival analyses. The estimations for the copula-based correlation and the partial correlation need to incorporate random censoring for such data. In addition, we need to invoke more complicated empirical process theories to argue the weak convergence results. Second, we may even allow the predictors to be censored; see Cheng and Fine (2008) and Cheng and Li (2015) for an earlier discussion. Third, we may consider more pairs of (τ, ι) over a candidate set or an interval in order to incorporate additional information on the quantiles of the response and the covariates. In future work, the relevant theoretical results discussed here can be generalized further to include ultrahigh-dimensional data.

Supplementary Material

The online Supplementary Material provides technical proofs of all theoretical results stated in the manuscript, as well as extensive numerical simulations.

Acknowledgements

We thank the Associate Editor and two anonymous referees for their constructive comments and suggestions. We are grateful to Jon Wellner for his helpful discussions. This work was supported by National Natural Science Foundation of China (Grant No. 11801202) and Academic Research Fundings from Ministry of Eduction (MOE) in Singapore: R-155-000-197-112, R-155-000-195-114, and R-155-000-205-114.

References

Barut, E., Fan, J. and Verhasselt, A. (2016). Conditional sure independence screening. *Journal of American Statistical Association* **111**, 1266–1277.

Billingsley, P. (1999). *Convergence of Probability Measures*. 2nd Edition Wiley, New York.

Boulesteix, A. L., Porzelius, C. and Daumer, M. (2008). Microarray-based classification and clinical predictors: On combined classifiers and additional predictive value. *Bioinformatics* **24**, 1698–1706.

Chang, J., Tang, C. Y. and Wu, Y. (2013). Marginal empirical likelihood and sure independence feature screening. *The Annals of Statistics* **41**, 2123–2148.

Chang, J., Tang, C. Y. and Wu, Y. (2016). Local independence feature screening for nonparametric and semiparametric models by marginal empirical likelihood. *The Annals of Statistics* **44**, 515–539.

Cheng, Y. and Fine, J. P. (2008). Nonparametric estimation of cause-specific cross hazard ratio with bivariate competing risks data. *Biometrika* **95**, 233–240.

Cheng, Y. and Li, J. (2015). Time-dependent diagnostic accuracy analysis with censored outcome and censored predictor. *Journal of Statistical Planning and Inference* **156**, 90–102.

Cheng, M., Honda, T., Li, J. and Peng, H. (2014). Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal data. *The Annals of Statistics* **42**, 1819–1849.

Chiang, A. P., Beck, J. S., Yen, H. J., et al. (2006). Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a BardetBiedl syndrome gene (BBS11). *Proceedings of the National Academy of Sciences* **103**, 6287–6292.

Chu, W., Li, R. and Reimherr, M. (2016). Feature screening for time-varying coefficient models with ultrahigh dimensional longitudinal data. *Annals of Applied Statistics* **10**, 596–617.

DeSantis, C. E., Ma, J., Sauer, A. G., Newman, L. A. and Jemal, A. (2017). Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: A Cancer Journal for Clinicians* **67**, 439–448. Harvard.

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

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

Fan, J., Ma, Y. and Dai, W. (2014). Nonparametric independent screening in sparse ultra-high dimensional varying coefficient models. *Journal of the American Statistical Association*, **109**, 1270–1284.

Fan, J. and Song, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics* **38**, 3567–3604.

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

Huang, T., McKeague, I. and Qian, M. (2019). Marginal screening for high-dimensional predictors of survival outcomes. *Statistica Sinica* **29**, 2105–2139.

Hong, H. and Li, Y. (2018). Feature selection of ultrahigh-dimensional covariates with survival outcomes: a selective review. *Applied Mathematics-A Journal of Chinese Universities* **32**, 379–396.

Iglewicz, B. and Hoaglin, D. (1993). *How to Detect and Handle Outliers. The ASQC Basic Reference in Quality Control: Statistical Techniques Volume 16.* (Edited by E. F. Mykytka, Ph.D.) ASQC Quality Press, Milwaukee.

Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, New York.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

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

Li, G., Peng, H., Zhang, J. and Zhu, L. (2012). Robust rank correlation based screening. *The Annals of Statistics* **40**, 1846–1877.

Li, J., Zheng, Q., Peng, L. and Huang, Z. (2016). Survival impact index and ultrahigh-dimensional model-free screening with survival outcomes. *Biometrics* **72**, 1145–1154.

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

Liu, J., Li, R. and Wu, R. (2014). Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *Journal of the American Statistical Association* **109**, 266–274.

Ma, X. and Zhang, J. (2016). Robust model-free feature screening via quantile correlation. *Journal of Multivariate Analysis* **143**, 472–480.

Ma, S., Li, R. and Tsai, C. (2017). Variable screening via quantile partial correlation. *Journal of the American Statistical Association* **112**, 650–663.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications* **9**, 141–142.

Nelsen, R. B. (2007). *An Introduction to Copulas.* Springer Science & Business Media.

Scheetz, T. E., Kim, K. Y. A., Swiderski, R. E., et al. (2006). Regulation of gene expression in the mammalian eye and its relevance to eye disease. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 14429–14434.

Song, R., Lu, W., Ma, S. and Jeng, X. J. (2014). Censored rank independence screening for high-dimensional survival data. *Biometrika* **101**, 799–814.

Van't Veer, L. J., Dai, H., van de Vijver, M. J., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A* **26**, 359–372.

Wu, Y. and Yin, G. (2015). Conditional qunatile screening in ultrahigh-dimensional heterogeneous data. *Biometrika* **102**, 65–76.

Xia, X., Li, J. and Fu, B. (2019). Conditional quantile correlation learning for ultrahigh dimensional varying coefficient models and its application in survival analysis. *Statistica Sinica* **29**, 645–669.

Xia, X., Jiang, B., Li, J. and Zhang, W. (2016). Low-dimensional confounder adjustment and high-dimensional penalized estimation for survival analysis. *Lifetime Data Analysis* **22**, 547–569.

Xia, X., Yang, H. and Li, J. (2016). Feature screening for generalized varying coefficient models with application to dichotomous responses. *Computational Statistics and Data Analysis*

102, 85–97.

Yu, T., Li, J. and Ma, S. (2012). Adjusting confounders in ranking biomarkers: a model-based ROC approach. *Briefings in Bioinformatics* **13**, 513–523.

Yue, M. and Li, J. (2017). Improvement screening for ultra-high dimensional data with censored survival outcomes and varying coefficients. *International Journal of Biostatistics* **13**, 16.

Zhao, D. S. and Li, Y. (2012). Principled sure independence screening for Cox models with ultra-high-dimensional covariates. *Journal of Multivariate Analysis* **105**, 397–411.

Zheng, Q., Peng, L. and He, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *The Annals of Statistics* **43**, 2225–2258.

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

Xiaochao Xia

College of Science, Huazhong Agricultural University, Yifu Building A303, No.1, Shizishan Street, Hongshan District, Wuhan, Hubei Province, China, 430070.

E-mail: xia_xiao_chao@126.com

Jialiang Li

Department of Statistics & Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore, 117546.

E-mail: stalj@nus.edu.sg

(Received June 2018; accepted April 2019)