The Annals of Statistics
2013, Vol. 41, No. 1, 342–369
DOI: 10.1214/13-AOS1087

QUANTILE-ADAPTIVE MODEL-FREE VARIABLE SCREENING FOR HIGH-DIMENSIONAL HETEROGENEOUS DATA

BY XUMING HE[1], LAN WANG[2] AND HYOKYOUNG GRACE HONG[3]

University of Michigan, University of Minnesota and the City University of New York

We introduce a quantile-adaptive framework for nonlinear variable screening with high-dimensional heterogeneous data. This framework has two distinctive features: (1) it allows the set of active variables to vary across quantiles, thus making it more flexible to accommodate heterogeneity; (2) it is model-free and avoids the difficult task of specifying the form of a statistical model in a high dimensional space. Our nonlinear independence screening procedure employs spline approximations to model the marginal effects at a quantile level of interest. Under appropriate conditions on the quantile functions without requiring the existence of any moments, the new procedure is shown to enjoy the sure screening property in ultra-high dimensions. Furthermore, the quantile-adaptive framework can naturally handle censored data arising in survival analysis. We prove that the sure screening property remains valid when the response variable is subject to random right censoring. Numerical studies confirm the fine performance of the proposed method for various semiparametric models and its effectiveness to extract quantile-specific information from heteroscedastic data.

1. Introduction. We consider the problem of analyzing ultra-high dimensional data, where the number of candidate covariates (or features) may increase at an exponential rate. Many efforts have been devoted to this challenging problem in recent years, motivated by modern applications in genomics, bioinformatics, chemometrics, among others. A practically appealing approach is to first use a fast screening procedure to reduce the dimensionality of the feature space to a moderate scale; and then apply more sophisticated variable selection techniques in the second stage. In this paper, we propose a new quantile-adaptive, model-free variable screening procedure, which is particularly appealing for analyzing high dimensional heterogeneous data and data with censored responses.

Received February 2012; revised January 2013.

[1]Supported in part by NSF Grant DMS-10-07396, NIH Grant R01GM080503 and National Natural Science Foundation of China Grant 11129101.

[2]Supported in part by NSF Grant DMS-10-07603.

[3]Supported in part by a Baruch College Eugene M. Lang Fellowship.

MSC2010 subject classifications. Primary 68Q32, 62G99; secondary 62E99, 62N99.

Key words and phrases. Feature screening, high dimension, polynomial splines, quantile regression, randomly censored data, sure independence screening.

Fan and Lv (2008) proposed the sure independence screening (SIS) methodology for linear regression which screens variables by ranking their marginal correlations with the response variable. They established the desirable sure screening property, that is, some important features are retained with probability approaching one, even if the dimensionality of the features is allowed to grow exponentially fast with the sample size. Fan and Song (2010) further extended the methodology to generalized linear models, see also Fan, Samworth and Wu (2009). The problem of nonlinear features screening was addressed in Hall and Miller (2009) using generalized correlation ranking and more systematically in Fan, Feng and Song (2011) using nonparametric marginal ranking, which extended the scope of applications of sure independence screening. Bühlmann, Kalisch, and Maathuis (2010) introduced the new concept of partial faithfulness and proposed a computationally efficient PC-algorithm for feature screening in linear models.

Zhu et al. (2011) proposed a novel feature screening procedure which avoids the specification of a particular model structure. This model-free screening framework is very appealing because a misspecified model could easily corrupt the performance of a variable selection method. Partly motivated by this interesting piece of work, we propose a new framework called quantile-adaptive model-free screening. We advocate a quantile-adaptive approach which allows the set of active variables to be different when modeling different conditional quantiles. This new framework provides a more complete picture of the conditional distribution of the response given all candidate covariates and is more natural and effective for analyzing high-dimensional data that are characterized by heteroscedasticity.

In the quantile-adaptive model-free screening framework, we estimate marginal quantile regression nonparametrically using B-spline approximation. In this aspect, our technique shares some similarity with that in Fan, Feng and Song (2011). The main technical challenge is to deal with the nonsmooth loss function, because the nonparametric marginal utility we consider does not have a closed form expression as in Fan, Feng and Song (2011). We derive useful exponential bounds using the empirical process theory to establish the sure screening property. When working with marginal quantile regression, the usual sub-Gaussian tail type condition in high-dimensional analysis can be relaxed and replaced by the assumption that the conditional density function of the random error has a positive lower bound around the quantile of interest. Empirically, we also demonstrate that the proposed procedure works well with heavy-tailed error distributions.

Sure independence screening remains challenging and little explored when the response variable is subject to random censoring, a common problem in survival analysis. Fan, Feng and Wu (2010) extended the methodology of sure independence screening using the marginal Cox proportional hazards model and studied its performance empirically. In this paper, we demonstrate that in the quantile-adaptive model-free screening framework, randomly censored responses can be naturally accommodated by ranking a marginal weighted quantile regression utility. We establish the sure screening property for the censored case under some general conditions.

The rest of the paper is organized as follows. In Section 2, we introduce the quantile-adaptive model-free feature screening procedure. In Section 3, we investigate its theoretical properties. Section 4 discusses the extension to survival analysis. In Section 5, we carry out simulation studies to access the performance of the proposed method. The numerical results demonstrate the favorable performance of the proposed method, especially when the errors are heavy-tailed or heteroscedastic. In Section 6, we demonstrate the application on a real data example. Section 7 contains further discussions. The technical details are given in Section 8.

2. Quantile-adaptive model-free feature screening.

2.1. *A general framework.* We consider the problem of nonlinear variable screening in high-dimensional feature space, where we observe a response variable Y and associated covariates $X_1, \dots, X_p$. The goal is to rapidly reduce the dimension of the covariate space p to a moderate scale via a computationally convenient procedure. Since ultra-high dimensional data often display heterogeneity, we advocate a quantile-adaptive feature screening framework. More specifically, we assume that at each quantile level a sparse set of covariates are relevant for modeling Y, but allow this set to be different at different quantiles, see, for instance, Examples 2 and 3 in Section 5. At a given quantile level α $(0 < \alpha < 1)$, we define the set of active variables

$$M_\alpha = \{j : Q_\alpha(Y|\mathbf{X}) \text{ functionally depends on } X_j\},$$

where $Q_\alpha(Y|\mathbf{X})$ is the τth conditional quantile of Y given $\mathbf{X} = (X_1, \dots, X_p)^T$, that is, $Q_\alpha(Y|\mathbf{X}) = \inf\{y : P(Y \leq y|\mathbf{X}) \geq \alpha\}$. Let $S_\alpha = |M_\alpha|$ be the cardinality of M_α. Throughout this paper, we assume S_α, $0 < \alpha < 1$, is smaller than the sample size n.

In practice, we may consider several quantiles to explore the sparsity pattern and the effects of the covariates at different parts of the conditional distribution. We refer to Koenker (2005) for a comprehensive introduction to quantile regression.

In ultrahigh dimensional data analysis, there generally exists little prior information for specifying a statistical model. Given a large number of covariates, it is hard to determine which covariates have linear effects and which have nonlinear effects. In our framework besides the sparsity assumption, we do not impose a specific model structure but allow the covariate effects to be nonlinear.

2.2. *Ranking by marginal quantile utility.* Let $\{(\mathbf{X}_i, Y_i), i = 1, \dots, n\}$ be i.i.d. copies of $(\mathbf{X}, Y)$, where $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$. Note that

$$Y \text{ and } X_j \text{ are independent} \quad \Leftrightarrow \quad Q_\alpha(Y|X_j) - Q_\alpha(Y) = 0 \qquad \forall \alpha \in (0, 1),$$

where $Q_\alpha(Y|X_j)$ is the αth conditional quantile of Y given X_j and $Q_\alpha(Y)$ is the αth unconditional quantile of Y. To estimate the effect of X_j on Y, we consider the marginal quantile regression of Y on X_j. Let $f_j(X_j) = \arg\min_f \mathrm{E}[\rho_\alpha(Y - f(X_j)) - \rho_\alpha(Y)]$, where the inclusion of $\rho_\alpha(Y)$ makes the expectation well defined even when Y has no finite moment, where $\rho_\alpha(u) = u\{\alpha - I(u < 0)\}$ is the quantile loss function (or check function). It is known that $f_j(X_j) = Q_\alpha(Y|X_j)$, the αth conditional quantile of Y given X_j.

Without loss of generality, we assume that each X_j takes values on the interval $[0, 1]$. Let $\mathbb{F}$ be the class of functions defined in condition (C1) in Section 3.1. Let $0 = s_0 < s_1 < \cdots < s_k = 1$ be a partition of the interval. Using the s_i as knots, we construct $N = k + l$ normalized B-spline basis functions of order $l + 1$ which form a basis for $\mathbb{F}$. We write these basis functions as a vector $\boldsymbol{\pi}(t) = (B_1(t), \dots, B_N(t))^T$, where $\|B_k(\cdot)\|_\infty \leq 1$ and $\|\cdot\|_\infty$ denotes the sup norm. Assume that $f_j(t) \in \mathbb{F}$. Then $f_j(t)$ can be well approximated by a linear combination of the basis functions $\boldsymbol{\pi}(t)^T\boldsymbol{\beta}$, for some $\boldsymbol{\beta} \in \mathbb{R}^N$.

Let $\widehat{\boldsymbol{\beta}}_j = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^N} \sum_{i=1}^n \rho_\alpha(Y_i - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta})$, and define

$$\widehat{f}_{nj}(t) = \boldsymbol{\pi}(t)^T\widehat{\boldsymbol{\beta}}_j - F_{Y,n}^{-1}(\alpha),$$

where $F_{Y,n}^{-1}(\alpha)$ is the αth sample quantile function based on $Y_1, \dots, Y_n$. Thus $\widehat{f}_{nj}(t)$ is a nonparametric estimator of $Q_\alpha(Y|X_j) - Q_\alpha(Y)$. We expect $\widehat{f}_{nj}$ to be close to zero if X_j is independent of Y.

The independence screening is based on the magnitude of the estimated marginal components $\|\widehat{f}_{nj}\|_n^2 = n^{-1}\sum_{i=1}^n \widehat{f}_{nj}(X_{ij})^2$. More specifically, we will select the subset of variables

$$\widehat{M}_\alpha = \{1 \leq j \leq p : \|\widehat{f}_{nj}\|_n^2 \geq \nu_n\},$$

where ν_n is a predefined threshold value. In practice, we often rank the features by $\|\widehat{f}_{nj}\|_n^2$ and keep the top $[n/\log(n)]$ features, where $[a]$ denotes the integer part of a.

3. Theoretical properties.

3.1. *Preliminaries.* We impose the following regularity conditions to facilitate our technical derivations.

(C1) The conditional quantile function $Q_\alpha(Y|X_j)$ belongs to $\mathbb{F}$, where $\mathbb{F}$ is the class of functions defined on $[0, 1]$ whose lth derivative satisfies a Lipschitz condition of order c: $|f^{(l)}(s) - f^{(l)}(t)| \leq c_0|s - t|^c$, for some positive constant c_0, $s, t \in [0, 1]$, where l is a nonnegative integer and $c \in (0, 1]$ satisfies $d = l + c > 0.5$.

(C2) $\min_{j\in M_\alpha} E(Q_\alpha(Y|X_j) - Q_\alpha(Y))^2 \geq c_1 n^{-\tau}$ for some $0 \leq \tau < \frac{2d}{2d+1}$ and some positive constant c_1.

(C3) The conditional density $f_{Y|X_j}(t)$ is bounded away from 0 and ∞ on $[Q_\alpha(Y|X_j) - \xi, Q_\alpha(Y|X_j) + \xi]$, for some $\xi > 0$, uniformly in X_j.

(C4) The marginal density function g_j of X_j, $1 \leq j \leq p$, are uniformly bounded away from 0 and ∞.

(C5) The number of basis functions N satisfies $N^{-d}n^{\tau} = o(1)$ and $Nn^{2\tau-1} = o(1)$ as $n \to \infty$.

Condition (C1) assumes that the conditional quantile function $Q_\alpha(Y|X_j)$ belongs to a class of smooth functions. This condition is standard for nonparametric spline approximation. Condition (C2) assumes that the features in the active set at quantile level α have strong enough marginal signals; a smaller τ corresponds to a stronger marginal signal. This condition is important as it guarantees that marginal utilities carries information about the features in the active set. Condition (C3) is a standard condition on random errors in the theory for quantile regression. It relaxes the usual sub-Gaussian assumptions that are needed in the literature on high dimensional inference. Condition (C4) is similar as condition (B) of Fan, Feng and Song (2011). Note that (C4) is not restrictive when X_j is supported on a bounded interval, say [0, 1]. When the distribution of X_j has an unbounded support (e.g., normal), we can view X_j as coming from a truncated distribution. In fact, if there is an outlier in X_j as in the case of a heavy-tailed distribution, we do better by dropping the outlier or transforming X_j to be uniformly distributed on [0, 1]. In our numerical simulations of Section 5, the normally distributed covariates are scaled to the interval [0, 1] and the results would change little if any reasonable truncation is used instead. Condition (C5) describes how fast the number of basis functions is allowed to grow with the sample size.

Given $(Y, \mathbf{X})$, where $\mathbf{X} = (X_1, \dots, X_p)^T$, we define

$$\boldsymbol{\beta}_{0j} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^N} \mathrm{E}\big[\rho_\alpha\big(Y - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}\big) - \rho_\alpha(Y)\big]. \tag{3.1}$$

Let $f_{nj}(t) = \boldsymbol{\pi}(t)^T\boldsymbol{\beta}_{0j} - Q_\alpha(Y)$, whose sample version was defined in Section 2. Furthermore, we let $\|f_{nj}\|^2 = E[f_{nj}(X_j)^2]$. The following lemma shows that the spline approximation error is negligible in the sense that the spline approximation carries the same level of information about the marginal signal.

LEMMA 3.1. *Under condition* (C5), $\min_{j\in M_\alpha}\|f_{nj}\|^2 \geq c_1 n^{-\tau}/8$, *for all n sufficiently large.*

3.2. *Sure screening property.* As covariate screening often serves only as the first step for high dimensional data analysis, the most important property as far as practical application is concerned is the sure screening property. In the quantile-adaptive framework, we require that the sure screening property holds at each quantile level α, that is, the set of selected covariates at quantile level α includes M_α with probability tending to one.

The key step of deriving the sure screening property is to establish exponential probability bounds for $\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{0j}\|$ and $\|\widehat{f}_{nj}\|_n^2 - \|f_{nj}\|^2$. The main technical challenge is that $\widehat{\boldsymbol{\beta}}_j$ is defined by minimizing a nonsmooth objective function, thus does not have a closed-form expression. The exponential bounds are summarized in the following lemma.

LEMMA 3.2. *Assume conditions* (C1)–(C5) *are satisfied.*

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

$$P\Big(\max_{1\le j\le p}\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{0j}\| \ge CN^{1/2}n^{-\tau}\Big) \le 2p\exp(-c_2 n^{1-4\tau}) + p\exp(-c_3 N^{-2}n^{1-2\tau})$$

for all n sufficiently large.

(2) *For any* $C > 0$, *there exist some positive constants* δ_1 *and* δ_2 *such that*

$$P\Big(\max_{1\le j\le p}\big|\|\widehat{f}_{nj}\|_n^2 - \|f_{nj}\|^2\big| \ge Cn^{-\tau}\Big) \le p\{11\exp(-\delta_1 n^{1-4\tau}) + 12N^2\exp(-\delta_2 N^{-3}n^{1-2\tau})\},$$

for all n sufficiently large.

REMARK. The results suggest that we can handle the dimensionality $\log p = o(n^{1-4\tau} + N^{-3}n^{1-2\tau})$. This dimensionality depends on the number of basis functions N and the strength of the marginal signals. If we take $N = n^{1/(2d+1)}$ (the optimal rate for spline approximation), then for $\tau < \min(1/4, (d-1)/(2d+1))$, we can handle ultra-high dimensionality, that is, p can grow at the exponential rate.

The following theorem establishes the sure screening property.

THEOREM 3.3 (Sure screening property). *Under the conditions of Lemma* 3.2, *if* $\tau < 1/4$, $N^3 n^{2\tau-1} = o(1)$, *and we take the threshold value* $\nu_n = \delta^* n^{-\tau}$ *with* $\delta^* \le c_1/16$ *for the constant* c_1 *specified in Lemma* 3.1, *then*

$$P(M_\alpha \subset \widehat{M}_\alpha) \ge 1 - S_\alpha\{11\exp(-\delta_1 n^{1-4\tau}) + 12N^2\exp(-\delta_2 N^{-3}n^{1-2\tau})\},$$

for all n sufficiently large. Especially, $P(M_\alpha \subset \widehat{M}_\alpha) \to 1$ *as* $n \to \infty$.

3.3. *Controlling false discovery.* An interesting question is how many variables are retained after the screening. A simple bound is provided below, which extends the results in Fan, Feng and Song (2011).

Let $\boldsymbol{\Pi} = (\boldsymbol{\pi}(X_1), \dots, \boldsymbol{\pi}(X_p))^T$ and $\boldsymbol{\Sigma} = \mathrm{E}(\boldsymbol{\Pi}\boldsymbol{\Pi}^T)$. Let $\|\cdot\|_F$ denote the Frobenius matrix norm and let $\|\cdot\|_2$ denote the spectral matrix norm. Note that

$$\begin{aligned}\sum_{j=1}^p \|f_{nj}\|^2 &= \sum_{j=1}^p \mathrm{E}\big(\boldsymbol{\pi}(X_j)\boldsymbol{\beta}_{0j}\big)^2 \le \sum_{j=1}^p \lambda_{\max}\big(\mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T\big)\|\boldsymbol{\beta}_{0j}\|^2 \\ &\le N\sum_{j=1}^p \mathrm{trace}\big(\mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T\big) \le N\mathrm{E}\Bigg[\sum_{j=1}^p\sum_{k=1}^N B_k^2(X_j)\Bigg] \\ &\le N\mathrm{E}(\|\boldsymbol{\Pi}\|_F^2) \le N^2\mathrm{E}(\|\boldsymbol{\Pi}\|_2^2) = N^2\lambda_{\max}(\boldsymbol{\Sigma}),\end{aligned}$$

where the second inequality uses the result $\|\boldsymbol{\beta}_{0j}\|^2 \le cN$ for some positive constant c (proved in the supplemental material [He, Wang and Hong (2013)]).

For any $\varepsilon > 0$, we define the set

$$D_n = \Big\{ \max_{1\le j\le p} \big| \|\widehat{f}_{nj}\|_n^2 - \|f_{nj}\|^2 \big| \le \varepsilon n^{-\tau} \Big\}.$$

Then on D_n, the cardinality of $\{j : \|\widehat{f}_{nj}\|_n^2 > 2\varepsilon n^{-\tau}\}$ cannot exceed the cardinality of $\{j : \|f_{nj}\|^2 > \varepsilon n^{-\tau}\}$, which is bounded by $\varepsilon^{-1}N^2 n^{\tau}\lambda_{\max}(\boldsymbol{\Sigma})$.

For $\nu_n = \delta^* n^{-\tau}$, we take $\varepsilon = \delta^*/2$, then $P(|\widehat{M}_\alpha| \le \varepsilon^{-1}N^2 n^{\tau}\lambda_{\max}(\boldsymbol{\Sigma})) \ge P(D_n)$. Applying Lemma 3.2(2), we obtain the following theorem which provides a bound on the size of selected variables.

THEOREM 3.4. *Under the conditions of Theorem 3.3, there exist some positive constants δ_1 and δ_2 such that for all n sufficiently large,*

$$\begin{aligned} P\big(|\widehat{M}_\alpha| \le 2N^2 n^{\tau}\lambda_{\max}(\boldsymbol{\Sigma})/\delta^*\big) & \\ \ge 1 - p\big\{11\exp\big(-\delta_1 n^{1-4\tau}\big) + 12N^2 \exp\big(-\delta_2 N^{-3} n^{1-2\tau}\big)\big\}. & \end{aligned}$$

Especially, $P(|\widehat{M}_\alpha| \le 2N^2 n^{\tau}\lambda_{\max}(\boldsymbol{\Sigma})/\delta^*) \to 1$ as $n \to \infty$.

The above theorem suggests that if $\lambda_{\max}(\boldsymbol{\Sigma}) = O(n^{\gamma})$ for some $\gamma > 0$, then the model obtained after screening is of polynomial size with high probability. Similar observation has been reported for the L_2-based screening procedure of Fan, Feng and Song (2011).

4. Quantile-adaptive screening in survival analysis. There exists very limited amount of work on feature screening with censored responses. Fan, Feng and Wu (2010) and Zhao and Li (2012) investigated marginal screening based on the Cox proportional hazards model. As a powerful alternative to the classical Cox model, quantile regression has recently emerged as a useful tool for analyzing censored data, see Ying, Jung and Wei (1995), McKeague, Subramanian and Sun (2001), Portnoy (2003), Peng and Huang (2008), Wang and Wang (2009), among others. The quantile regression approach directly models the survival time and is easy to interpret. Furthermore, it relaxes the proportional hazards assumption of the Cox model and can naturally accommodate heterogeneity in the data. The quantile regression based screening procedure can be naturally extended to survival analysis.

Assume that Y_i is subject to random right censoring. Instead of $\{(\mathbf{X}_i, Y_i), i = 1, \dots, n\}$, we observe $\{(\mathbf{X}_i, Y_i^*, \delta_i), i = 1, \dots, n\}$ where

$$Y_i^* = \min(Y_i, C_i), \qquad \delta_i = I(Y_i \le C_i). \tag{4.1}$$

The random variable C_i, called the censoring variable, is assumed to be conditionally independent of Y_i given $\mathbf{X}_i$. In this section, we assume that the censoring distribution is the same for all covariates, but this assumption will be relaxed in Section 7.1. Let $G(t) = P(C_i > t)$ be the survival function of C_i. Let $\widehat{G}(t)$ be the Kaplan–Meier estimator of $G(t)$, based on $\{Y_i^*, \delta_i\}$, $i = 1, \dots, n$.

Similarly as in the case of complete data, we consider independence screening based on nonparametric marginal regression given X_j. More specifically, we consider inverse probability weighted marginal quantile regression estimator

$$\widehat{\boldsymbol{\beta}}_j^c = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^N} \sum_{i=1}^n \frac{\delta_i}{\widehat{G}(Y_i^*)} \rho_\alpha\bigl(Y_i^* - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}\bigr).$$

Let $\widehat{f}_{nj}^c(t) = \boldsymbol{\pi}(t)^T \widehat{\boldsymbol{\beta}}_j^c - F_{\mathrm{KM},n}^{-1}(\alpha)$ where $F_{\mathrm{KM},n}^{-1}(\alpha)$ is the nonparametric estimator of the αth conditional quantile of Y based on (Y_i^*, C_i, δ_i), $i = 1, \dots, n$. The estimator we use here is the inverse function (the left-continuous version) of the Kaplan–Meier estimator of the distribution function of Y, whose properties have been studied in Lo and Singh (1986). We will select the subset of variables $\widehat{M}_\alpha^c = \{1 \le j \le p : \|\widehat{f}_{nj}^c\|_n^2 \ge \nu_n^c\}$, where ν_n^c is a predefined threshold value. As for the complete data case, in practice we often rank the features by $\|\widehat{f}_{nj}^c\|_n^2$ and keep the top $[n/\log(n)]$ features.

For the random censoring case, in addition to conditions (C1)–(C5), we assume that:

(C6) $P(t \le Y_i \le C_i) \ge \tau_0 > 0$ for some positive constant τ_0 and any $t \in [0, T]$, where T denotes the maximum follow-up time. Furthermore, $\sup\{t : P(Y > t) > 0\} \ge \sup\{t : P(C > t) > 0\}$. The survival function of the censoring variable $G(t)$ has uniformly bounded first derivative.

(C7) There exist $0 < \beta_1 < \beta_2 < 1$ such that $\alpha \in [\beta_1, \beta_2]$ and that the distribution function of Y_i is twice differentiable in $[Q_{\beta_1}(Y_i) - \varepsilon, Q_{\beta_2}(Y_i) + \varepsilon]$ for some $0 < \varepsilon < 1$, with the first derivative bounded away from zero and the second derivative bounded in absolute value.

Condition (C6) and (C7) are common in the survival analysis literature to ensure that the Kaplan–Meier estimator and its inverse function are well behaved. The following theorem states that the sure screening property holds for the random censoring case and an upper bound that controls the size of selected variables can be obtained.

THEOREM 4.1. *Assume conditions* (C1)–(C7) *are satisfied*, $\tau < 1/4$ *and* $N^3 n^{2\tau-1} = o(1)$, *if we take the threshold value* $\nu_n^c = \delta^* n^{-\tau}$ *with* $\delta^* \le c_1/16$, *then*:

(1) *there exist positive constants* δ_3 *and* δ_4 *such that*

$$P\bigl(M_\alpha \subset \widehat{M}_\alpha^c\bigr) \ge 1 - S_\alpha\bigl\{17 \exp\bigl(-\delta_3 n^{1-4\tau}\bigr) + 12N^2 \exp\bigl(-\delta_4 N^{-3} n^{1-2\tau}\bigr)\bigr\},$$

for all n sufficiently large. Especially, $P(M_\alpha \subset \widehat{M}_\alpha^c) \to 1$ *as* $n \to \infty$.

(2) *for all n sufficiently large*,

$$P\big(|\widehat{M}_\alpha^c| \le 2N^2 n^\tau \lambda_{\max}(\boldsymbol{\Sigma})/\delta^*\big)$$
$$\ge 1 - p\{17\exp(-b_7 n^{1-4\tau}) + 12N^2\exp(-b_8 N^{-3} n^{1-2\tau})\}.$$

Especially, $P(|\widehat{M}_\alpha^c| \le 2N^2 n^\tau \lambda_{\max}(\boldsymbol{\Sigma})/\delta^*) \to 1$ *as* $n \to \infty$.

5. Monte Carlo studies. We carry out simulation studies to investigate the performance of the proposed quantile adaptive sure independence screening procedure (to be denoted by QaSIS). We consider two criteria for evaluating the performance as in Zhu et al. (2011). The first criterion is the minimum model size (denoted by $\mathcal{R}$), that is, the smallest number of covariates that we need to include to ensure that all the active variables are selected. The second criterion is the proportion of active variables (denoted by $\mathcal{S}$) selected by the screening procedure when the threshold $\nu_n = [n/\log(n)]$ is adopted. Note that the first criterion does not need to specify a threshold. An effective variable screening procedure is expected to have the value of $\mathcal{R}$ reasonably small comparing to the number of active variables and the value of $\mathcal{S}$ close to one.

We first consider the complete data case and compare the performance of QaSIS with the nonparametric independence screening (NIS) procedure of Fan, Feng and Song (2011) and the sure independent ranking and screening (SIRS) procedure of Zhu et al. (2011). In computing QaSIS and NIS, the number of basis (d_n) is set to be $[n^{1/5}] = 3$. For each example, we report the results based on 500 simulation runs.

EXAMPLE 1 (Additive model, $n = 400$, $p = 1000$). This example is adapted from Fan, Feng and Song (2011). Let $g_1(x) = x$, $g_2(x) = (2x-1)^2$, $g_3(x) = \sin(2\pi x)/(2-\sin(2\pi x))$, and $g_4(x) = 0.1\sin(2\pi x) + 0.2\cos(2\pi x) + 0.3\sin(2\pi x)^2 + 0.4\cos(2\pi x)^3 + 0.5\sin(2\pi x)^3$. The following cases are studied:

- *Case* (1a): $Y = 5g_1(X_1) + 3g_2(X_2) + 4g_3(X_3) + 6g_4(X_4) + \sqrt{1.74}\varepsilon$, where the vector of covariates $\mathbf{X} = (X_1, \dots, X_{1000})^T$ is generated from the multivariate normal distribution with mean $\mathbf{0}$ and the covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{1000\times1000}$ with $\sigma_{ii} = 1$ and $\sigma_{ij} = \rho^{|i-j|}$ for $i \ne j$, $\varepsilon \sim N(0, 1)$ is independent of $\mathbf{X}$. In case (1a), we consider $\rho = 0$.
- *Case* (1b): same as case (1a) except that $\rho = 0.8$.
- *Case* (1c): Same as case (1b) except that ε has the Cauchy distribution.

Note that the models are homoscedastic in Example 1, thus the number of active variables are the same across different quantiles.

EXAMPLE 2 (Index model, $n = 200$, $p = 2000$). This example is adapted from Zhu et al. (2011). The random data are generated from $Y = 2(X_1 + 0.8X_2 + 0.6X_3 + 0.4X_4 + 0.2X_5) + \exp(X_{20} + X_{21} + X_{22}) \cdot \varepsilon$, where $\varepsilon \sim N(0, 1)$, $\mathbf{X} =$

$(X_1, X_2, \dots, X_{2000})^T$ follows the multivariate normal distribution with the correlation structure described in case (1b). Different from the regression models in Example 1, this model is heteroscedastic: the number of active variables is 5 at the median but 8 elsewhere.

EXAMPLE 3 (A more complex structure, $n = 400$, $p = 5000$). We consider a more complex heteroscedastic model for which the conditional distribution of Y does not have a simple additive or index structure.

- *Case* (3a): $Y = 2(X_1^2 + X_2^2) + \exp((X_1 + X_2 + X_{18} + X_{19} + \cdots + X_{30})/10) \cdot \varepsilon$, where $\varepsilon \sim N(0, 1)$, and $\mathbf{X} = (X_1, X_2, \dots, X_{5000})^T$ follows the multivariate normal distribution with the correlation structure described in case (1b). In this case, the number of active variables is 2 at the median but is 15 elsewhere.
- *Case* (3b): same as case (3a), but with $2(X_1^2 + X_2^2)$ replaced by $2((X_1 + 1)^2 + (X_2 + 2)^2)$.

The median value of $\mathcal{R}$ (with IRQ in the parenthesis) and the average value of $\mathcal{S}$ for QaSIS, NIS and SIRS are summarized in Table 1. For QaSIS, we report results for two quantiles $\alpha = 0.5$ and 0.75. We observe the following from Table 1: (i) The L_2 norm based NIS procedure exhibits the best performance when the random error has a normal distribution, but its performance deteriorates substantially for heavy-tailed or heteroscedasitic errors (Examples 1–2). (ii) We observe that in case (1a) where $\rho = 0$, no method works really well in terms of the minimum model size. This is because the independent signals work against the marginal effect estimation as accumulated noise, thus masking the relatively weak signals from X_3 and X_4 in this model. (iii) In Example 3 where the model has a more complex structure, QaSIS is effective in identifying the number of active variables at different quantiles; while the performance of SIRS depends on the functional form. Overall, our simulations for the complete data case demonstrate that the performance of QaSIS is on par with or better than that of NIS and SIRS for a variety of distributions of covariates and errors.

Variable screening with censored responses has received little attention in the literature. In Example 4 below, we compare the quantile-adaptive nonparametric marginal screening procedure proposed in Section 4 with the Cox model based marginal screening procedure [Cox(SIS)] of Fan, Feng and Wu (2010) and a naive procedure treating the censored data as complete and then applying to the QaSIS procedure (denoted by Naive).

EXAMPLE 4 (Censored responses). We consider a case in which the latent response variable Y_i is generated using the same setup as in case (1b). Let $Y_i^* = \min(Y_i, C_i)$, where the censoring time C_i is generated from a 3-component normal mixture distribution $0.4N(-5, 4) + 0.1N(5, 1) + 0.5N(55, 1)$. The censoring probability is about 45%. Due to the high censoring rate, the performance

TABLE 1

Results for Examples 1–3. The numbers reported are the median of $\mathcal{R}$ [with interquartile range (IQR) given in parentheses] and $\mathcal{S}$

Example	**Case**	**Method**	p^*	$\mathcal{R}$**(IQR)**	$\mathcal{S}$
Example 1	(1a)	QaSIS ($\alpha = 0.50$)	4	655 (434)	0.56
		QaSIS ($\alpha = 0.75$)	4	652 (398)	0.56
		NIS	4	660 (415)	0.55
		SIRS	4	689 (365)	0.53
	(1b)	QaSIS ($\alpha = 0.50$)	4	4 (0)	1.00
		QaSIS ($\alpha = 0.75$)	4	4 (0)	1.00
		NIS	4	4 (0)	1.00
		SIRS	4	6 (9)	0.99
	(1c)	QaSIS($\alpha = 0.50$)	4	4 (0)	1.00
		QaSIS($\alpha = 0.75$)	4	4 (0)	1.00
		NIS	4	6 (79)	0.83
		SIRS	4	7 (14)	0.98
Example 2		QaSIS ($\alpha = 0.50$)	5	6 (2)	1.00
		QaSIS ($\alpha = 0.75$)	8	18 (24)	0.96
		NIS	8	1726 (511)	0.22
		SIRS	8	18 (16)	0.97
Example 3	(3a)	QaSIS ($\alpha = 0.50$)	2	3 (2)	1.00
		QaSIS ($\alpha = 0.75$)	15	153 (207)	0.89
		NIS	15	3117 (4071)	0.50
		SIRS	15	698 (1140)	0.89
	(3b)	QaSIS ($\alpha = 0.50$)	2	2 (1)	1.00
		QaSIS ($\alpha = 0.75$)	15	88 (542)	0.92
		NIS	15	4166 (1173)	0.25
		SIRS	15	29 (21)	1.00

p^*: the number of truly active variables.

of the variable screening procedures is investigated at the median and the 0.25 quantile.

Table 2 summarizes the simulations results based on 100 runs. QaSIS substantially outperforms both Naive and Cox(SIS). Under-performance of Cox(SIS) can be attributed to the fact that the proportional hazards assumption is not satisfied in this example. Table 2 also includes the LQaSIS procedure which will be discussed in Section 7.1.

6. Real data analysis. We illustrate the proposed screening method on the diffuse large-B-cell lymphoma (DLBCL) microarray data of Rosenwald et al. (2002). The data set contains the survival times of 240 patients and the gene ex-

TABLE 2

Simulation results for Example 4 ($n = 400$, $p = 1000$). *The numbers reported are the median of* $\mathcal{R}$ [*with interquartile range* (*IQR*) *given in parentheses*] *and* $\mathcal{S}$

Method	p^*	$\mathcal{R}$**(IQR)**	$\mathcal{S}$
QaSIS ($\alpha = 0.50$)	4	4 (2)	0.99
QaSIS ($\alpha = 0.25$)	4	5 (22)	0.96
LQaSIS ($\alpha = 0.50$)	4	4 (6)	0.98
LQaSIS ($\alpha = 0.25$)	4	5 (13)	0.98
Naive ($\alpha = 0.50$)	4	254 (497)	0.74
Naive ($\alpha = 0.25$)	4	792 (351)	0.14
Cox (SIS)	4	190 (655)	0.70

p^*: the number of truly active variables.

pression measurements of 7399 genes for each patient. The gene expression measurements for each gene are standardized to have mean zero and variance one. To assess the predictive performance of the proposed method, we divide the data set into a training set with $n_1 = 160$ patients and a testing set with remaining $n_2 = 80$ patients, in the same way as Bair and Tibshirani (2004) did. The index of the training set is available from http://www-stat.stanford.edu/~tibs/superpc/staudt.html.

Nearly half of the survival time data are censored, so we focus our attention on two quantile levels $\alpha = 0.2$ and 0.4 that represent the effects of gene expression on the sub-population of patients with poor prognosis. We apply the proposed QaSIS method to the training data to select $[n_1/\log(n_1)] = 31$ genes, which is followed up by a variable selection procedure based on additive quantile regression with the SCAD penalty [Fan and Li (2001)] to find two top genes. As with the empirical studies in the simulation study, we use three internal knots of $N = 3$. Because almost all the censoring occurs above the 0.4th quantile, we do not need to reweight the censored observations for the low quantiles we are considering in this example. Based on the two selected genes at each α, we estimate the corresponding quantile function. The estimated quantile function from the training set is then used to calculate risk scores for each patient in the testing data set. If Y_i is the survival time of the ith patient in the training set, with s_i as the predicted risk score, we expect the αth quantile of Y_i given s_i to have a significant (and positive) slope.

For the purpose of comparison, we also follow the same analysis path but replace QaSIS by the sure independence screening for Cox models, SIS(Cox), of Fan, Feng and Wu (2010) and the SCAD-penalized Cox regression to select two genes. The risk scores are then calculated based on the linear index for the Cox model. Table 3 summarizes the slope coefficients of regressing survival times on risk scores in the training set based on the censored quantile regression of Portnoy (2003). It is clear that the analysis based on QaSIS has the desired predictive power, where the 0.2 and 0.4 quantiles of survival time for the testing data set are significantly associated with the predicted risk scores, but the analysis based on SIS(Cox)

TABLE 3

Estimated slope coefficients (and p-values) for survival time versus the risk score at αth quantile obtained by QaSIS and SIS (Cox)

Method	**Estimated coefficients**	***p*-value**
QaSIS ($\alpha = 0.4$)	0.93	**0.02**
QaSIS ($\alpha = 0.2$)	0.53	**0.04**
SIS (Cox) ($\alpha = 0.4$)	0.16	0.62
SIS (Cox) ($\alpha = 0.2$)	0.17	0.62

did not make it. If we regress the survival time on the risk scores on the training data, we would get coefficients of exactly 1.0 under QaSIS, but it would not have validation power.

We examined the two genes selected by QaSIS. At $\alpha = 0.4$, their GeneIDs are 31981 (known as AA262133, septin 1) and 17902 (AA284323, glutathione synthetase). Both genes belong to the known Proliferation signature group in the study of DLBCL by Rosenwald et al. (2002). Gene AA262133 was also ranked very high by Li and Luan (2005) using the partial likelihood-based scores. At $\alpha = 0.2$, the two selected genes have IDs 31585 (known as NM 018518, MCM10 minichromosome maintenance deficient 10) and 33014 (AA769543, Hypothetical protein MGC4189). We find that Gene 31858 also belongs to the known Proliferation signature group. Gene 33014 was identified as an interesting candidate by Li and Luan (2005). We did not find any signature group associated with it, but it seems quite related to the lower tail of the survival time distribution, and is worth further investigation.

7. Discussions.

7.1. *Further extension on screening with censored responses.* The assumption that the censoring distribution does not depend on the covariates is popular in regression analysis of survival data. It can be further relaxed. Assume that Y_i and C_i are conditionally independent given X_{ij}, $j = 1, \dots, p$. Let $G(t|X_{ij}) = P(C_i > t|X_{ij})$ be the conditional survival function of C_i given X_{ij}. Let $\widehat{G}(t|X_{ij})$ be the local Kaplan–Meier estimator of $G(t|X_{ij})$ [e.g., Beran (1981) and Gonzalez-Manteiga and Cadarso-Suarez (1994)]. More specifically,

$$\widehat{G}(t|x) = \prod_{j=1}^{n} \left\{ 1 - \frac{B_{nj}(x)}{\sum_{k=1}^{n} I(Y_k^* \ge Y_j^*) B_{nk}(x)} \right\}^{I(Y_j^* \le t, \delta_j = 0)}, \tag{7.1}$$

where $B_{nk}(x) = K(\frac{x - x_k}{h_n}) / \{\sum_{i=1}^{n} K(\frac{x - x_i}{h_n})\}, k = 1, \dots, n$, are the Nadaraya–Watson weights, h_n is the bandwidth and $K(\cdot)$ is a density function. We consider estimating $\boldsymbol{\beta}_{0j}$ using the locally weighted censored quantile regression, that is,

$$\widetilde{\boldsymbol{\beta}}_j^c = \underset{\boldsymbol{\beta}\in\mathbb{R}^N}{\arg\min} \sum_{i=1}^n \frac{\delta_i}{\widehat{G}(Y_i^*|X_{ij})} \rho_\alpha\big(Y_i^* - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}\big).$$

Let $\widetilde{f}_{nj}^c(t) = \boldsymbol{\pi}(t)^T \widetilde{\boldsymbol{\beta}}_j^c - F_{\mathrm{KM},n}^{-1}(\alpha)$ and define $\widetilde{M}_\alpha^c = \{1 \le j \le p : \|\widetilde{f}_{nj}^c\|_n^2 \ge \eta_n^c\}$ where η_n^c is a predefined threshold value. We refer to this new procedure as LQaSIS, whose numerical performance is reported in Table 2 and two other examples in the supplemental material [He, Wang and Hong (2013)].

We assume, instead of (C6):

(C6$'$) $\inf_x P(t \le Y_i \le C_i|x) \ge \tau_0 > 0$ for some positive constant τ_0 and any $t \in [0, T]$, where T denotes the maximum follow-up time. $G(t|x)$ has first derivatives with respect to t, which is uniformly bounded away from infinity; and $G(t|x)$ has bounded (uniformly in t) second-order partial derivatives with respect to x. Furthermore, $t_0 \le \sup\{t : G(t|x) > 0\} \le t_1$ uniformly in x for some positive constants t_0 and t_1, and $\sup\{t : P(Y > t|x) > 0\} \ge \sup\{t : G(t|x) > 0\}$ almost surely for x.

Then Theorem 4.1 can be extended as follows.

THEOREM 7.1. *Assume conditions* (C1)–(C5), (C6$'$) *and* (C7) *are satisfied*, $\tau < 1/4$, $nh^3 \to \infty$, $N^3 n^{2\tau-1} = o(1)$, $N^2 n^{2\tau-1}(\log n)^2 h^{-1} = o(1)$ *and* $(N + n^\tau)n^\tau h^2 = o(1)$. *If we take* $\nu_n = \delta^* n^{-\tau}$ *with* $\delta^* \le c_1/16$, *then*:

(1) *there exist positive constants* δ_3 *and* δ_4 *such that*

$$P\big(M_\alpha \subset \widehat{M}_\alpha^c\big) \ge 1 - S_\alpha\big\{17\exp\big(-\delta_3 n^{1-4\tau}\big) + 12N^2\exp\big(-\delta_4 N^{-3} n^{1-2\tau}\big)\big\},$$

for all n sufficiently large. Especially, $P(M_\alpha \subset \widehat{M}_\alpha^c) \to 1$ *as* $n \to \infty$.

(2)

$$\begin{aligned} P\big(\big|\widehat{M}_\alpha^c\big| \le 2N^2 n^\tau \lambda_{\max}(\boldsymbol{\Sigma})/\delta^*\big) \\ \ge 1 - p\big\{17\exp\big(-b_7 n^{1-4\tau}\big) + 12N^2\exp\big(-b_8 N^{-3} n^{1-2\tau}\big)\big\}, \end{aligned}$$

for all n sufficiently large. Especially, $P(|\widehat{M}_\alpha^c| \le 2N^2 n^\tau \lambda_{\max}(\boldsymbol{\Sigma})/\delta^*) \to 1$ *as* $n \to \infty$.

The proof of Theorem 7.1 is given in the supplemental material [He, Wang and Hong (2013)].

7.2. *Limitations and other issues*. We have not investigated the problem of adaptively selecting the number of basis functions in this paper for two reasons: (1) although adaptive tuning is possible, it will significantly increase the computational time as it needs to be done for each covariate separately; (2) optimal estimation is not the goal for marginal screening, instead consistent estimation generally suffices. Empirically, we find that 3 or 4 internal knots are generally enough to flexibly approximate many smooth functions typically seen in practice.

Marginally unimportant but jointly important variables may not be preserved in marginal screening. This is a well-recognized weakness of all existing marginal screening procedures. Iterative procedures may be helpful to a certain degree for this problem [Fan and Lv (2008)]. In the same spirit, we find that in practice a slightly modified QaSIS helps in situations where a dominating variable increases the error variance of the marginal regression model for other variables and hence mask the significance of other variables. If the top ranked variable is dominating, then the modified QaSIS removes its effects from **Y** first and screen the remaining variables again.

The way we define the set of active variables can be considered as a nonparametric approach in the sense that we consider directly the conditional quantile function without a specific model structure. In real life high-dimensional data analysis, the knowledge needed for an appropriate model specification is often inadequate. Using a misspecified model to perform variable selection is likely to produce misleading results. We propose to separate variable screening and model building, where a nonparametric approach is applied to screen high-dimensional variables and then followed by sensible model building in the second stage in a lower-dimensional space. We expect that this model-free approach to variable screening to gain momentum in ultra-high dimensional learning, see, for example, the work of Li, Zhong and Zhu (2012) on distance correlation based screening.

8. Technical proofs.

We present the proof for the random censoring case, as this is the more challenging scenario. The proof for the complete data case and that for Lemma 3.1 are given in the supplemental material [He, Wang and Hong (2013)].

To establish the sure independence property, a key step is to obtain an exponential tail probability bound for

$$P\Big(\max_{1\le j\le p}\big|\|\widehat{f}^c_{nj}\|_n^2-\|f_{nj}\|^2\big|\ge Cn^{-\tau}\Big), \tag{8.1}$$

where C is any positive constant. We have

$$\begin{aligned}\|\widehat{f}^c_{nj}\|_n^2 &= \widehat{\boldsymbol\beta}_j^{cT}\big(\mathbb{P}_n\boldsymbol\pi(X_j)\boldsymbol\pi(X_j)^T\big)\widehat{\boldsymbol\beta}_j^c-2F^{-1}_{\mathrm{KM},n}(\alpha)n^{-1}\big(\mathbb{P}_n\boldsymbol\pi(X_j)\big)^T\widehat{\boldsymbol\beta}_j^c\\&\quad+\big(F^{-1}_{\mathrm{KM},n}(\alpha)\big)^2,\\ \|f_{nj}\|^2&=\boldsymbol\beta_{0j}^T\big(\mathrm{E}\boldsymbol\pi(X_j)\boldsymbol\pi(X_j)^T\big)\boldsymbol\beta_{0j}-2Q_\alpha(Y)\big(E\boldsymbol\pi(X_j)\big)^T\boldsymbol\beta_{0j}+\big(Q_\alpha(Y)\big)^2,\end{aligned}$$

where $\mathbb{P}_n\boldsymbol\pi(X_j)\boldsymbol\pi(X_j)^T = n^{-1}\sum_{i=1}^n\boldsymbol\pi(X_{ij})\boldsymbol\pi(X_{ij})^T$, $\mathbb{P}_n\boldsymbol\pi(X_j) = n^{-1}\times\sum_{i=1}^n\boldsymbol\pi(X_{ij})$, and $\mathrm{E}\boldsymbol\pi(X_j)\boldsymbol\pi(X_j)^T$ denotes the expectation of $\boldsymbol\pi(X_j)\boldsymbol\pi(X_j)^T$ under the true distribution of X_j. Note that

$$\begin{aligned}&\|\widehat{f}^c_{nj}\|_n^2-\|f_{nj}\|^2\\&\quad=\big(\widehat{\boldsymbol\beta}_j^c-\boldsymbol\beta_{0j}\big)^T\big(\mathbb{P}_n\boldsymbol\pi(X_j)\boldsymbol\pi(X_j)^T\big)\big(\widehat{\boldsymbol\beta}_j^c-\boldsymbol\beta_{0j}\big)\end{aligned}$$

$$
\begin{aligned}
&+ 2\big(\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\big)^T \big(\mathbb{P}_n \boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T\big)\boldsymbol{\beta}_{0j} \\
&+ \widehat{\boldsymbol{\beta}}_j^c\big(\mathbb{P}_n \boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T - \mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T\big)\boldsymbol{\beta}_{0j} \\
&- 2F_{\mathrm{KM},n}^{-1}(\alpha)\big[\mathbb{P}_n \boldsymbol{\pi}(X_j)^T \widehat{\boldsymbol{\beta}}_j^c - E\boldsymbol{\pi}(X_j)^T \boldsymbol{\beta}_{0j}\big] \\
&+ 2\big[Q_\alpha(Y) - F_{\mathrm{KM},n}^{-1}(\alpha)\big]\big(E\boldsymbol{\pi}(X_j)^T \boldsymbol{\beta}_{0j}\big) \\
&+ \big[\big(F_{\mathrm{KM},n}^{-1}(\alpha)\big)^2 - \big(Q_\alpha(Y)\big)^2\big] \\
\triangleq{}& \sum_{k=1}^{6} S_{jk},
\end{aligned}
$$

where the definition of S_{jk} is clear from the context. From the argument of Lemma 3.1, $E(\boldsymbol{\pi}(X_j)^T \boldsymbol{\beta}_{0j})$ is uniformly bounded in X_j and by Lemma 8.4(4) below, we have $|S_{j5}| = O(n^{-1/2}(\log n)^{1/2}) = o(n^{-\tau})$ almost surely. Similarly, $|S_{j6}| = O(n^{-1/2}(\log n)^{1/2}) = o(n^{-\tau})$ almost surely. Therefore, for all n sufficiently large,

$$
\begin{aligned}
P\Big(\max_{1\le j\le p} \big|\|\widehat{f}_{nj}^c\|_n^2 - \|f_{nj}\|^2\big| \ge Cn^{-\tau}\Big)
&\le P\left(\max_{1\le j\le p} \sum_{k=1}^{4} |S_{jk}| \ge Cn^{-\tau}/2\right) \\
&\le \sum_{k=1}^{4} P\Big(\max_{1\le j\le p} |S_{jk}| \ge Cn^{-\tau}/8\Big).
\end{aligned}
$$

In the following, we shall provide details on deriving exponential tail bound for $P(\max_{1\le j\le p} |S_{jk}| \ge Cn^{-\tau}/8)$.

8.1. *Properties of the spline basis*. First, we recall some useful properties of the basis vector $\boldsymbol{\pi}(t) = (B_1(t), \dots, B_N(t))^T$. Zhou, Shen and Wolfe (1998) established that there exist two positive constants b_1 and b_2 such that

$$
\begin{aligned}
b_1 N^{-1} &\le \lambda_{\min}\big(\mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T\big) \le \lambda_{\max}\big(\mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T\big) \\
&\le b_2 N^{-1} \qquad \forall j,
\end{aligned}
\tag{8.2}
$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest eigenvalue and the largest eigenvalue, respectively.

Stone (1985) established that there exists a positive constant b_3 such that

$$
E\big(B_k^2(X_{ij})\big) \le b_3 N^{-1}, \qquad 1 \le k \le N, 1 \le i \le n, 1 \le j \le p. \tag{8.3}
$$

Similar result can be found in He and Shi (1996).

LEMMA 8.1. *Let* $\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T = n^{-1}\sum_{i=1}^n \boldsymbol{\pi}(X_{ij})\boldsymbol{\pi}(X_{ij})^T$ *and* $\mathbf{D}_j = \mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T - \mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T$.

(1) *There exists a positive constant* c_4 *such that for all n sufficiently large*

$$P\big(\lambda_{\max}(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T) \ge (b_2+1)N^{-1}\big) \le 2N^2\exp(-c_4 nN^{-3}), \tag{8.4}$$

(2) *For any* $c_5 > 0$, *there exists a positive constant* c_6 *such that for all n sufficiently large*

$$\begin{aligned} &P\big(\max(|\lambda_{\max}(\mathbf{D}_j)|, |\lambda_{\min}(\mathbf{D}_j)|) \ge c_5 N^{-1}n^{-\tau}\big) \\ &\quad \le 2N^2\exp(-c_6 N^{-3}n^{1-2\tau}). \end{aligned} \tag{8.5}$$

PROOF. The proof is an extension of that for Lemma 5 in Fan, Feng and Song (2011) which proved similar results for the smallest eigenvalue. First, for any two symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, we have $\lambda_{\max}(\mathbf{A}+\mathbf{B}) \le \lambda_{\max}(\mathbf{A}) + \lambda_{\max}(\mathbf{B})$. This implies that $\lambda_{\max}(\mathbf{A}) - \lambda_{\max}(\mathbf{B}) \le \lambda_{\max}(\mathbf{A}-\mathbf{B})$ and $\lambda_{\max}(\mathbf{B}) - \lambda_{\max}(\mathbf{A}) \le \lambda_{\max}(\mathbf{B}-\mathbf{A})$. Thus

$$|\lambda_{\max}(\mathbf{A}) - \lambda_{\max}(\mathbf{B})| \le \max\{|\lambda_{\max}(\mathbf{A}-\mathbf{B})|, |\lambda_{\max}(\mathbf{B}-\mathbf{A})|\}.$$

Applying the above inequality, we have

$$\begin{aligned} &|\lambda_{\max}(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T) - \lambda_{\max}(\mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)| \\ &\quad \le \max\{|\lambda_{\max}(\mathbf{D}_j)|, |\lambda_{\max}(-\mathbf{D}_j)|\}. \end{aligned} \tag{8.6}$$

For any N-dimensional vector $\mathbf{a} = (a_1,\dots,a_N)^T$ satisfying $\|\mathbf{a}\| = 1$, we have $|\mathbf{a}^T\mathbf{D}_j\mathbf{a}| \le \|\mathbf{D}_j\|_\infty(\sum_{i=1}^N |a_i|)^2 \le N\|\mathbf{D}_j\|_\infty$, where $\|\mathbf{D}_j\|_\infty$ is the sup norm of the matrix $\mathbf{D}_j$. Thus $\lambda_{\max}(\mathbf{D}_j) = \max_{\|\mathbf{a}\|=1}\mathbf{a}^T\mathbf{D}_j\mathbf{a} \le N\|\mathbf{D}_j\|_\infty$. Also $\lambda_{\max}(\mathbf{D}_j) = -\min_{\|\mathbf{a}\|=1}(-\mathbf{a}^T\mathbf{D}_j\mathbf{a}) \ge -N\|\mathbf{D}_j\|_\infty$. Thus $|\lambda_{\max}(\mathbf{D}_j)| \le N\|\mathbf{D}_j\|_\infty$. Similarly, we have $|\lambda_{\max}(-\mathbf{D}_j)| \le N\|\mathbf{D}_j\|_\infty$. Following (8.6) and using the result on the smallest eigenvalue of $\mathbf{D}_j$ [Fan, Feng and Song (2011)], we have

$$\begin{aligned} &|\lambda_{\max}(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T) - \lambda_{\max}(\mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)| \\ &\quad \le \max(|\lambda_{\max}(\mathbf{D}_j)|, |\lambda_{\min}(\mathbf{D}_j)|) \le N\|\mathbf{D}_j\|_\infty. \end{aligned} \tag{8.7}$$

As in Fan, Feng and Song (2011), applying Bernstein's inequality to each entry of $\mathbf{D}_j$, it can be shown that $\forall\delta > 0$,

$$P\big(N\|\mathbf{D}_j\|_\infty \ge N\delta/n\big) \le 2N^2\exp\left\{-\frac{\delta^2}{2(b_3 nN^{-1} + \delta/3)}\right\}. \tag{8.8}$$

To prove (8.4), we use the bound in (8.2), apply the inequality in (8.7) and take $\delta = N^{-2}n$ in (8.8). This gives

$$P\big(\lambda_{\max}(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T) \ge (b_2+1)N^{-1}\big) \le 2N^2\exp(-c_4 N^{-3}n),$$

for some positive constant c_4 for all n sufficiently large.

To prove (8.5), we apply the inequality in (8.7) and take $\delta = c_5 N^{-2} n^{1-\tau}$ in (8.8). This gives

$$P\big(\max\big(\big|\lambda_{\max}(\mathbf{D}_j)\big|, \big|\lambda_{\min}(\mathbf{D}_j)\big|\big) \geq c_5 N^{-1} n^{-\tau}\big) \leq 2N^2 \exp\big(-c_6 N^{-3} n^{1-2\tau}\big),$$

for some positive constant c_6 for all n sufficiently large. □

8.2. *An exponential tail probability bound for* $\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\|$. Let

$$B_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \delta_i \big[\widehat{G}(Y_i^*)\big]^{-1} \big[\rho_\alpha\big(Y_i^* - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}\big) - \rho_\alpha\big(Y_i^*\big)\big],$$

$$B(\boldsymbol{\beta}) = \mathrm{E}\big\{\delta_i \big[G\big(Y_i^*\big)\big]^{-1} \big[\rho_\alpha\big(Y_i^* - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}\big) - \rho_\alpha\big(Y_i^*\big)\big]\big\}.$$

Then $\widehat{\boldsymbol{\beta}}_j^c = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^N} B_n(\boldsymbol{\beta})$. Applying the iterative expectation formula, we have

$$\begin{aligned} B(\boldsymbol{\beta}) &= \mathrm{E}\big\{\mathrm{E}\big\{I(Y_i \leq C_i)\big[G(Y_i)\big]^{-1} \big[\rho_\alpha\big(Y_i - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}\big) - \rho_\alpha(Y_i)\big] | Y_i, X_{ij}\big\}\big\} \\ &= \mathrm{E}\big[\rho_\alpha\big(Y_i - \boldsymbol{\pi}(X_{ij})^T \boldsymbol{\beta}\big) - \rho_\alpha(Y_i)\big]. \end{aligned}$$

Hence, $\boldsymbol{\beta}_{0j} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^N} B(\boldsymbol{\beta})$.

We can bound the difference $\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\|$ by the difference of their respective objective functions.

LEMMA 8.2. *For any* $\delta > 0$,

$$\begin{aligned} &P\big(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\| \geq \delta\big) \\ &\quad \leq P\bigg(\sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_{0j}\| \leq \delta} \big|B_n(\boldsymbol{\beta}) - B(\boldsymbol{\beta})\big| \geq \frac{1}{2} \inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_{0j}\| = \delta} \big(B(\boldsymbol{\beta}) - B(\boldsymbol{\beta}_{0j})\big)\bigg). \end{aligned} \tag{8.9}$$

PROOF. This is a direct application of Lemma 2 of Hjort and Pollard (1993) making use of the convexity of the objective function. □

The lower bound of the right-hand side of (8.9) can be explicitly evaluated for any given $\delta > 0$. This is summarized in the following lemma.

LEMMA 8.3. *Let* $C > 0$ *be an arbitrary constant. Assume that* $N^{-d} n^{\tau} = o(1)$, *then there exists a positive constant* b_4 *such that*

$$\inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_{0j}\| = C N^{1/2} n^{-\tau}} \big(B(\boldsymbol{\beta}) - B(\boldsymbol{\beta}_{0j})\big) \geq b_4 n^{-2\tau}$$

for all n *sufficiently large.*

PROOF. We consider $\boldsymbol{\beta} = \boldsymbol{\beta}_{0j} + CN^{1/2}n^{-\tau}\mathbf{u}$, where $\mathbf{u} \in \mathbb{R}^N$ satisfying $\|\mathbf{u}\| = 1$. Using the identity by Knight [(1998), page 758], we have

$$
\begin{aligned}
&B(\boldsymbol{\beta}) - B(\boldsymbol{\beta}_{0j}) \\
&\quad = \mathrm{E}\{\rho_\alpha(Y - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j} - CN^{1/2}n^{-\tau}\boldsymbol{\pi}(X_j)^T\mathbf{u}) - \rho_\alpha(Y - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j})\} \\
&\quad = CN^{1/2}n^{-\tau}\mathrm{E}\{\boldsymbol{\pi}(X_j)^T\mathbf{u}[I(Y - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j} \le 0) - \tau]\} \\
&\qquad + \mathrm{E}\bigg\{\int_0^{CN^{1/2}n^{-\tau}\boldsymbol{\pi}(X_j)^T\mathbf{u}}[I(Y - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j} \le s) \\
&\qquad\qquad - I(Y - \boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j} \le 0)]\,ds\bigg\} \\
&\quad = CN^{1/2}n^{-\tau}\mathrm{E}\{\boldsymbol{\pi}(X_j)^T\mathbf{u}[F_{Y|X_j}(\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j}) - F_{Y|X_j}(f_j(X_j))]\} \\
&\qquad + \mathrm{E}\bigg\{\int_0^{CN^{1/2}n^{-\tau}\boldsymbol{\pi}(X_j)^T\mathbf{u}}[F_{Y|X_j}(\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j} + s) \\
&\qquad\qquad - F_{Y|X_j}(\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j})]\,ds\bigg\} \\
&\quad \triangleq I_1 + I_2.
\end{aligned}
$$

By Hölder's inequality, we have

$$
\begin{aligned}
|I_1| &\le CN^{1/2}n^{-\tau}(\mathrm{E}(\boldsymbol{\pi}(X_j)^T\mathbf{u})^2)^{1/2} \\
&\quad \times [\mathrm{E}(F_{Y|X_j}(\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j}) - F_{Y|X_j}(f_j(X_j)))^2]^{1/2} \\
&\le CN^{1/2}n^{-\tau}O(N^{-1/2})O(N^{-d}) \\
&= O(N^{-d}n^{-\tau}),
\end{aligned}
$$

where the second inequality uses inequality (B.3) in the supplementary material and (8.2).

Furthermore, for some ξ between $\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j} + s$ and $\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j}$,

$$
\begin{aligned}
I_2 &= \mathrm{E}\bigg\{\int_0^{CN^{1/2}n^{-\tau}\boldsymbol{\pi}(X_j)^T\mathbf{u}} f_{Y|X_j}(\xi)s\,ds\bigg\} = O(1)\mathrm{E}(N^{1/2}n^{-\tau}\boldsymbol{\pi}(X_j)^T\mathbf{u})^2 \\
&= O(n^{-2\tau})
\end{aligned}
$$

by (8.2). Note that I_2 is nonnegative and $I_1 = o(I_2)$. Thus, the conclusion of the lemma holds. $\square$

Lemmas 8.4–8.6 below provide several useful technical results for evaluating the right-hand side of (8.9).

LEMMA 8.4. *Assume conditions* (C6) *and* (C7). *The Kaplan–Meier estimator* $\widehat{G}(t)$ *satisfies*:

(1) $\sup_{0\le t\le T}|\widehat{G}(t)-G(t)| = O(n^{-1/2}(\log n)^{1/2})$ *almost surely.*

(2) $\widehat{G}(t)^{-1}-G(t)^{-1} = n^{-1}\sum_{j=1}^{n}\frac{\xi(Y_j^*,\delta_j,t)}{G^2(t)} + R_n(t)$, *where* $\xi(Y_j^*,\delta_j,t)$ *are independent mean zero random variables whose expression is given in Theorem* 1 *of* Lo and Singh (1986), *and* $\sup_{0\le t\le T}|R_n(t)| = O(n^{-3/4}(\log n)^{3/4})$ *almost surely.*

(3) $\sup_{0\le t\le T}|\frac{1}{\widehat{G}(t)}-\frac{1}{G(t)}| = O(n^{-1/2}(\log n)^{1/2})$ *almost surely.*

(4) $\sup_{\beta_1\le\alpha\le\beta_2}|F_{\mathrm{KM},n}^{-1}(\alpha)-Q_\tau(Y)| = O(n^{-1/2}(\log n)^{1/2})$ *almost surely.*

PROOF. The results in (1) and (4) are given in Lemma 3 of Lo and Singh (1986). The result in (2) follows from the Taylor expansion, Theorem 1 in Lo and Singh (1986) and the result in (1). The proof of (3) follows Taylor expansion and (1). $\square$

LEMMA 8.5 (Massart's concentration theorem, 2000). *Let* $W_1,\dots,W_n$ *be independent random variables and let* $\mathbb{G}$ *be a class of functions satisfying* $a_{i,g}\le g(W_i)\le b_{i,g}$ *for some real numbers* $a_{i,g}$ *and* $b_{i,g}$, *and for all* $1\le i\le n$ *and* $g\in\mathbb{G}$. *Define* $L^2=\sup_{g\in\mathbb{G}}\sum_{i=1}^{n}(b_{i,g}-a_{i,g})^2/n$ *and* $Z=\sup_{g\in\mathbb{G}}n^{-1}|\sum_{i=1}^{n}(g(W_i)-E(g(W_i)))|$. *Then for any positive* t, $P(Z\ge \mathrm{E}Z+t)\le\exp[-\frac{nt^2}{2L^2}]$.

LEMMA 8.6 [Bernstein inequality for U-statistics, Hoeffding (1963)]. *Let* $U_n^2(g)$ *denote the second-order* U*-statistics with kernel function* $g(t_1,t_2)$ *based on the independent random variables* $Z_1,\dots,Z_n$. *Assume that the function* g *is bounded*: $a<g<b$ *for some finite constants* a *and* b. *If* $E(g(Z_i,Z_j))=0,\forall i\ne j$, *then* $\forall t>0$, $P(|U_n^2(g)|>t)\le 2\exp(-\frac{2kt^2}{(b-a)^2})$, *where* k *denotes the integer part of* $n/2$.

LEMMA 8.7. *Assume the conditions of Theorem* 4.1. *For any* $C>0$, *there exist positive constants* c_7 *and* c_8 *such that for all* n *sufficiently large*

$$P\big(\|\widehat{\boldsymbol{\beta}}_j^c-\boldsymbol{\beta}_{0j}\|\ge CN^{1/2}n^{-\tau}\big)\le 4\exp\big(-c_7n^{1-4\tau}\big)+\exp\big(-c_8N^{-2}n^{1-2\tau}\big).$$

PROOF. Following Lemmas 8.2 and 8.3, there exists some $b_4>0$ such that for all n sufficiently large,

$$\begin{aligned}
&P\big(\|\widehat{\boldsymbol{\beta}}_j^c-\boldsymbol{\beta}_{0j}\|\ge CN^{1/2}n^{-\tau}\big)\\
&\quad\le P\Big(\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}}\big|B_n(\boldsymbol{\beta})-B(\boldsymbol{\beta})\big|\ge b_4n^{-2\tau}\Big)\\
&\quad\le P\left(\big|B_n(\boldsymbol{\beta}_{0j})-B(\boldsymbol{\beta}_{0j})\big|\ge\frac{1}{2}b_4n^{-2\tau}\right)
\end{aligned}$$

$$+ P\Big(\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}} \big|B_n(\boldsymbol{\beta}) - B_n(\boldsymbol{\beta}_{0j}) - B(\boldsymbol{\beta}) + B(\boldsymbol{\beta}_{0j})\big| \ge \frac{1}{2}b_4 n^{-2\tau}\Big)$$

$$\triangleq J_1 + J_2.$$

First, we evaluate J_1. Let $W_i = \delta_i[G(Y_i^*)]^{-1}[\rho_\alpha(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j}) - \rho_\alpha(Y_i^*)]$. Then

$$\begin{aligned}
&B_n(\boldsymbol{\beta}_{0j}) - B(\boldsymbol{\beta}_{0j})\\
&\quad = n^{-1}\sum_{i=1}^n (W_i - EW_i)\\
&\qquad + n^{-1}\sum_{i=1}^n \delta_i\big[\big(\widehat{G}(Y_i^*)\big)^{-1} - \big(G(Y_i^*)\big)^{-1}\big]\big[\rho_\alpha\big(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j}\big) - \rho_\alpha\big(Y_i^*\big)\big]\\
&\quad \triangleq I_1 + I_2.
\end{aligned}$$

Then $J_1 \le P(|I_1| \ge b_4 n^{-2\tau}/4) + P(|I_2| \ge b_4 n^{-2\tau}/4)$. Note that $|W_i| \le C|\boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j}|$, for some positive constant C. By the argument of Lemma 3.1, $\sup_t |f_j(t) - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j}| \le c_2 N^{-d}$. Thus, $|W_i|$ are uniformly bounded by a constant M. Applying Bernstein's inequality, there exists a positive constant b_5 such that for all n sufficiently large,

$$P\big(|I_1| \ge b_4 n^{-2\tau}/4\big) \le 2\exp\Big(-\frac{b_4^2 n^{1-4\tau}/16}{2M^2 + Mb_4 n^{-2\tau}/3}\Big) \le 2\exp\big(-b_5 n^{1-4\tau}\big).$$

Furthermore, applying Lemma 8.4,

$$\begin{aligned}
I_2 &= n^{-2}\sum_{i=1}^n\sum_{j=1}^n \delta_i\big[G\big(Y_i^*\big)\big]^{-2}\xi\big(Y_j^*, \delta_j, Y_i^*\big)\big[\rho_\alpha\big(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j}\big) - \rho_\alpha\big(Y_i^*\big)\big]\\
&\quad + n^{-1}\sum_{i=1}^n \delta_i R_n\big(Y_i^*\big)\big[\rho_\alpha\big(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j}\big) - \rho_\alpha\big(Y_i^*\big)\big] \triangleq I_{21} + I_{22},
\end{aligned}$$

where ξ and R_n are defined in Lemma 8.4. By Lemma 8.4, $I_{22} = O(n^{-3/4}(\log n)^{3/4})$ almost surely. By assumptions $n^{-3/4}(\log n)^{3/4} = o(n^{-2\tau})$, and noting that $\delta_i G^{-2}(Y_i^*)\xi(Y_j^*, \delta_j, Y_i^*)[\rho_\alpha(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j}) - \rho_\alpha(Y_i^*)]$ are independent bounded random variables, we have for all n sufficiently large,

$$\begin{aligned}
&P\big(|I_2| \ge b_4 n^{-2\tau}/4\big)\\
&\quad \le P\Big(\frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j=1, j\ne i}^n \delta_i G^{-2}\big(Y_i^*\big)\xi\big(Y_j^*, \delta_j, Y_i^*\big)
\end{aligned}$$

$$\times \left[\rho_\alpha\left(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j}\right) - \rho_\alpha\left(Y_i^*\right)\right] > b_4 n^{-2\tau}/8\Bigg)$$

$$\le 2\exp\left(-b_6 n^{1-4\tau}\right),$$

where b_6 is a positive constant, by Lemma 8.6. Therefore, $J_1 \le 4\exp(-c_7 n^{1-4\tau})$ where $c_7 = \min(b_5, b_6)$.

Next, we evaluate J_2. Let $V_i(\boldsymbol{\beta}) = \rho_\alpha(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}) - \rho_\alpha(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j})$ and let $Z_i = \delta_i[G(Y_i^*)]^{-1}V_i(\boldsymbol{\beta})$. We have

$$J_2 \le P\Bigg(\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}}\left|n^{-1}\sum_{i=1}^n[Z_i - E(Z_i)]\right| + \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}}\left|n^{-1}\sum_{i=1}^n\delta_i\left[\left(\widehat{G}\left(Y_i^*\right)\right)^{-1} - \left(G\left(Y_i^*\right)\right)^{-1}\right]V_i(\boldsymbol{\beta})\right| \ge b_4 n^{-2\tau}/2\Bigg).$$

Applying Knight's identity [(1998), page 758], we have

$$\begin{aligned} V_i(\boldsymbol{\beta}) &= \boldsymbol{\pi}(X_{ij})^T(\boldsymbol{\beta} - \boldsymbol{\beta}_{0j})\left[I\left(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j} \le 0\right) - \tau\right] \\ &\quad + \int_0^{\boldsymbol{\pi}(X_{ij})^T(\boldsymbol{\beta}-\boldsymbol{\beta}_{0j})}\left[I\left(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j} \le s\right) \right. \\ &\qquad \left. - I\left(Y_i^* - \boldsymbol{\pi}(X_{ij})^T\boldsymbol{\beta}_{0j} \le 0\right)\right]ds. \end{aligned}$$

Thus,

$$\begin{aligned} \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}}\left|V_i(\boldsymbol{\beta})\right| &\le 2\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}}\left|\boldsymbol{\pi}(X_{ij})^T(\boldsymbol{\beta} - \boldsymbol{\beta}_{0j})\right| \\ &\le cNn^{-\tau} \end{aligned} \tag{8.10}$$

for some $c > 0$ because $\|B_k(\cdot)\|_\infty \le 1$. Combining (8.10) with Lemma 8.4(3), we have

$$\begin{aligned} &\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}}\left|n^{-1}\sum_{i=1}^n\delta_i\left[\left(\widehat{G}\left(Y_i^*\right)\right)^{-1} - \left(G\left(Y_i^*\right)\right)^{-1}\right]V_i(\boldsymbol{\beta})\right| \\ &\quad = O\left(Nn^{-\tau-1/2}(\log n)^{1/2}\right) \end{aligned}$$

almost surely. Assume $N^2n^{2\tau-1}\log n = o(1)$, then for all n sufficiently large,

$$J_2 \le P\Bigg(\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}}\left|n^{-1}\sum_{i=1}^n[Z_i - E(Z_i)]\right| \ge b_4 n^{-2\tau}/4\Bigg).$$

We use Lemma 8.5 to evaluate the above inequality. First note that, (8.10) implies that $\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}}|Z_i|\le c^*Nn^{-\tau}$ for some positive constant c^*. Next, let $e_1,\dots,e_n$ be a Rademacher sequence (i.e., i.i.d. sequence taking values of ±1 with probability $1/2$) independent of $Z_1,\dots,Z_n$. We have

$$\begin{aligned}
\mathrm{E}&\left\{\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}} n^{-1}\left|\sum_{i=1}^{n}(Z_i-\mathrm{E}(Z_i))\right|\right\}\\
&\le 2\mathrm{E}\left\{\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}} n^{-1}\left|\sum_{i=1}^{n}e_iZ_i\right|\right\}\\
&\le C\mathrm{E}\left\{\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le CN^{1/2}n^{-\tau}} n^{-1}\left|\sum_{i=1}^{n}e_i\pi(X_{ij})^T(\boldsymbol{\beta}-\boldsymbol{\beta}_{0j})\right|\right\}\\
&\le CN^{1/2}n^{-\tau}\mathrm{E}\left\|n^{-1}\sum_{i=1}^{n}e_i\pi(X_{ij})\right\|\le CN^{1/2}n^{-\tau}\left[\mathrm{E}\left\|n^{-1}\sum_{i=1}^{n}e_i\pi(X_{ij})\right\|^2\right]^{1/2}\\
&= CN^{1/2}n^{-\tau}\left[n^{-2}\mathrm{E}\left(\sum_{i=1}^{n}e_i^2\pi(X_{ij})^T\pi(X_{ij})\right)\right]^{1/2}\le CN^{1/2}n^{-\tau-1/2}
\end{aligned}$$

for some generic constant C which may vary from line to line. In the above, the first inequality applies the symmetrization theorem [Lemma 2.3.1, van der Vaart and Wellner (1996)], the second inequality applies the contraction theorem [Ledoux and Talagrad, (1991)] using the Lipschitz property of the quantile objective function, and the last inequality uses (8.3). Now, we apply Lemma 8.5 to evaluate J_2. Let $Z=\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_{0j}\|\le\Delta N^{1/2}n^{-\tau}} n^{-1}|\sum_{i=1}^{n}(Z_i-\mathrm{E}(Z_i))|$. In Lemma 8.5, we take $t=b_4n^{-2\tau}/2-CN^{1/2}n^{-\tau-1/2}$ and $L^2=4c^2N^2n^{-2\tau}$, which gives

$$\begin{aligned}
J_2&=P\big(Z\ge\mathrm{E}Z+\big(b_4n^{-2\tau}/4-\mathrm{E}Z\big)\big)\\
&\le P\big(Z\ge\mathrm{E}Z+\big(b_4n^{-2\tau}/4-CN^{1/2}n^{-\tau-1/2}\big)\big)\\
&\le\exp\left(-\frac{n(b_4n^{-2\tau}/4-CN^{1/2}n^{-\tau-1/2})^2}{8c^2N^2n^{-2\tau}}\right)\le\exp\big(-c_8N^{-2}n^{1-2\tau}\big)
\end{aligned}$$

for some positive constant c_8 and all n sufficiently large. □

8.3. *Proof of the Theorem* 4.1. In this subsection, we establish the exponential tail probability bounds for $P(|S_{jk}|\ge Cn^{-\tau}/8)$, $k=1,\dots,4$, which lead to the result of Theorem 4.1.

An exponential tail probability bound for S_{j1}. Recall that

$$\begin{aligned}
S_{j1}&=\big(\widehat{\boldsymbol{\beta}}_j^c-\boldsymbol{\beta}_{0j}\big)^T\big(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T\big)\big(\widehat{\boldsymbol{\beta}}_j^c-\boldsymbol{\beta}_{0j}\big)\\
&\le\lambda_{\max}\big(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T\big)\big\|\widehat{\boldsymbol{\beta}}_j^c-\boldsymbol{\beta}_{0j}\big\|^2.
\end{aligned}$$

It follows from Lemmas 8.1 and 8.7 that for some $C^* > 0$,

$$
\begin{aligned}
&P(S_{j1} \geq Cn^{-\tau}/8) \\
&\quad \leq P(\lambda_{\max}(\mathbb{P}_n \boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T) \geq (b_2+1)N^{-1}) \\
&\qquad + P(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\|^2 \geq (b_2+1)^{-1}CNn^{-\tau}/8) \\
&\quad \leq 2N^2\exp(-c_4 nN^{-3}) + P(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\| > C^* N^{1/2}n^{-\tau/2}) \\
&\quad \leq 2N^2\exp(-c_4 nN^{-3}) + P(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\| > C^* N^{1/2}n^{-\tau}) \\
&\quad \leq 2N^2\exp(-c_4 nN^{-3}) + 4\exp(-c_7 n^{1-4\tau}) + \exp(-c_8 N^{-2}n^{1-2\tau}).
\end{aligned} \tag{8.11}
$$

An exponential tail probability bound for S_{j2}. We first establish an upper bound for $\|\boldsymbol{\beta}_{0j}\|$. By result (B.3) in the supplemental material [He, Wang and Hong (2013)], $\mathrm{E}[f_j(X_j) - f_{nj}(X_j)]^2 \leq c_3 N^{-2d}$, $\forall j$, for some $c_3 > 0$. It follows that

$$
\begin{aligned}
\mathrm{E}[f_{nj}(X_j)^2] &\leq 2\mathrm{E}[f_j(X_j)^2] + 2\mathrm{E}[(f_j(X_j) - f_{nj}(X_j))^2] \\
&\leq c_9 + 2c_3 N^{-2d},
\end{aligned}
$$

for some positive constant c_9. Also note that

$$
\mathrm{E}[f_{nj}(X_j)^2] \geq \lambda_{\min}(\mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)\|\boldsymbol{\beta}_{0j}\|^2 \geq b_1 N^{-1}\|\boldsymbol{\beta}_{0j}\|^2.
$$

This implies that $\|\boldsymbol{\beta}_{0j}\| \leq c_{10}\sqrt{N}$ for some positive constant c_{10}.

Since $|S_{j2}| \leq 2\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\|\lambda_{\max}(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)\|\boldsymbol{\beta}_{0j}\|$, we have

$$
\begin{aligned}
&P(|S_{j2}| \geq Cn^{-\tau}/8) \\
&\quad \leq P(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\|\lambda_{\max}(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T) \geq CN^{-1/2}n^{-\tau}/(16c_{10})) \\
&\quad \leq P(\lambda_{\max}(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T) > (b_2+1)N^{-1}) \\
&\qquad + P(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\| \geq (b_2+1)^{-1}CN^{1/2}n^{-\tau}/(16c_{10})) \\
&\quad \leq 2N^2\exp(-c_4 nN^{-3}) + 4\exp(-c_7 n^{1-4\tau}) + \exp(-c_8 N^{-2}n^{1-2\tau}).
\end{aligned}
$$

An exponential tail probability bound for S_{j3}. We have

$$
\begin{aligned}
S_{j3} &= \widehat{\boldsymbol{\beta}}_j^c(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T - \mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)\boldsymbol{\beta}_{0j} \\
&= (\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j})(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T - \mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)\boldsymbol{\beta}_{0j} \\
&\quad + \boldsymbol{\beta}_{0j}^T(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T - \mathrm{E}\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)\boldsymbol{\beta}_{0j} \\
&\triangleq S_{j31} + S_{j32}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&P(|S_{j3}| \ge Cn^{-\tau}/8)\\
&\quad \le P(|S_{j31}| \ge Cn^{-\tau}/16) + P(|S_{j32}| \ge Cn^{-\tau}/16)\\
&\quad \le P(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\| \max(|\lambda_{\max}(\mathbf{D}_j)|, |\lambda_{\min}(\mathbf{D}_j)|) \ge CN^{-1/2}n^{-\tau}/(16c_{10}))\\
&\qquad + P(\|\boldsymbol{\beta}_{0j}\|^2 \max(|\lambda_{\max}(\mathbf{D}_j)|, |\lambda_{\min}(\mathbf{D}_j)|) \ge Cn^{-\tau}/16)\\
&\quad \le P(\max(|\lambda_{\max}(\mathbf{D}_j)|, |\lambda_{\min}(\mathbf{D}_j)|) \ge N^{-1}/(16c_{10}))\\
&\qquad + P(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\| \ge CN^{1/2}n^{-\tau})\\
&\qquad + P(\max(|\lambda_{\max}(\mathbf{D}_j)|, |\lambda_{\min}(\mathbf{D}_j)|) \ge CN^{-1}n^{-\tau}/(16c_{10}^2))\\
&\quad \le 2P(\max(|\lambda_{\max}(\mathbf{D}_j)|, |\lambda_{\min}(\mathbf{D}_j)|) \ge C^*N^{-1}n^{-\tau})\\
&\qquad + P(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\| \ge CN^{1/2}n^{-\tau})\\
&\quad \le 2N^2\exp(-c_6N^{-3}n^{1-2\tau}) + 4\exp(-c_7n^{1-4\tau}) + \exp(-c_8N^{-2}n^{1-2\tau})
\end{aligned}
$$

for all n sufficiently large, where the last inequality uses Lemmas 8.1 and 8.7.

An exponential tail probability bound for S_{j4}.

$$
\begin{aligned}
S_{j4} &= -2F_{\mathrm{KM},n}^{-1}(\alpha)n^{-1}\sum_{i=1}^{n}[\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j} - E\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j}]\\
&\qquad - 2F_{\mathrm{KM},n}^{-1}(\alpha)n^{-1}\sum_{i=1}^{n}\boldsymbol{\pi}(X_j)^T(\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j})\\
&\triangleq S_{j41} + S_{j42}.
\end{aligned}
$$

Note that $F_{\mathrm{KM},n}^{-1}(\alpha)$ is uniformly bounded for $\beta_1 \le \alpha \le \beta_2$ almost surely. From the argument of Lemma 3.1, $E(\boldsymbol{\pi}(X_j)^T\boldsymbol{\beta}_{0j})$ is uniformly bounded in X_j. Applying Bernstein's inequality to S_{j41}, there exists a positive constant c_9 such that $P(|S_{j41}| > Cn^{-\tau}/16) \le \exp(-c_9n^{1-2\tau})$ for all n sufficiently large. On the other hand, by the Cauchy–Schwarz inequality, for all n sufficiently large,

$$
\begin{aligned}
&P(|S_{j42}| > Cn^{-\tau}/16)\\
&\quad \le P\left(\left|n^{-1}\sqrt{n}\left[\sum_{i=1}^{n}[\boldsymbol{\pi}(X_j)^T(\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j})]^2\right]^{1/2}\right| > C^*n^{-\tau}\right)\\
&\quad \le P([(\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j})^T(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T)(\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j})]^{1/2} > C^*n^{-\tau})\\
&\quad \le P(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\|\lambda_{\max}^{1/2}(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T) > C^*n^{-\tau})\\
&\quad \le P(\lambda_{\max}(\mathbb{P}_n\boldsymbol{\pi}(X_j)\boldsymbol{\pi}(X_j)^T) > (b_2+1)N^{-1})
\end{aligned}
$$

$$
\begin{aligned}
&+ P\big(\|\widehat{\boldsymbol{\beta}}_j^c - \boldsymbol{\beta}_{0j}\| \geq C^* N^{1/2} n^{-\tau}\big) \\
&\leq 2N^2 \exp(-c_6 N^{-3} n^{1-2\tau}) + 4\exp(-c_7 n^{1-4\tau}) + \exp(-c_8 N^{-2} n^{1-2\tau})
\end{aligned}
$$

for all n sufficiently large, where the last inequality uses Lemmas 8.1 and 8.7 and C^* denotes a generic positive constant which may vary from line to line. Therefore, for all n sufficiently large,

$$
\begin{aligned}
&P\big(|S_{j4}| > Cn^{-\tau}/8\big) \\
&\quad \leq 2N^2 \exp(-c_6 N^{-3} n^{1-2\tau}) + 4\exp(-c_7 n^{1-4\tau}) + \exp(-c_8 N^{-2} n^{1-2\tau}) \\
&\qquad + \exp(-c_9 n^{1-2\tau}) \\
&\quad \leq 2N^2 \exp(-c_6 N^{-3} n^{1-2\tau}) + 5\exp(-c_7 n^{1-4\tau}) + \exp(-c_8 N^{-2} n^{1-2\tau}).
\end{aligned}
$$

PROOF OF THEOREM 4.1.

(1) We have

$$
\begin{aligned}
&P\Big(\max_{1\leq j\leq p} \big|\|\widehat{f}_{nj}^c\|_n^2 - \|f_{nj}\|^2\big| \geq Cn^{-\tau}\Big) \\
&\quad \leq p\big(4N^2 \exp(-c_4 N^{-3} n) + 17\exp(-c_7 n^{1-4\tau}) + 4\exp(-c_8 N^{-2} n^{1-2\tau}) \\
&\qquad\qquad + 4N^2 \exp(-c_6 N^{-3} n^{1-2\tau})\big) \\
&\quad \leq p\big(17\exp(-\delta_3 n^{1-4\tau}) + 12N^2 \exp(-\delta_4 N^{-3} n^{1-2\tau})\big)
\end{aligned}
$$

for all n sufficiently large, for some positive constants δ_3 and δ_4.

(2) The result follows by making use of the bound in (1) and observing that

$$
\begin{aligned}
P\big(M_\alpha \subset \widehat{M}_\alpha^c\big) &\geq P\Big(\min_{j\in M_\alpha} \|\widehat{f}_{nj}^c\|_n^2 \geq \nu_n\Big) \\
&\geq P\Big(\min_{j\in M_\alpha} \|f_{nj}\|^2 - \max_{j\in M_\alpha} \big|\|\widehat{f}_{nj}^c\|_n^2 - \|f_{nj}\|^2\big| \geq \nu_n\Big) \\
&= 1 - P\Big(\max_{j\in M_\alpha} \big|\|\widehat{f}_{nj}^c\|_n^2 - \|f_{nj}\|^2\big| \geq \min_{j\in M_\alpha} \|f_{nj}\|^2 - \nu_n\Big) \\
&\geq 1 - P\Big(\max_{j\in M_\alpha} \big|\|\widehat{f}_{nj}^c\|_n^2 - \|f_{nj}\|^2\big| \geq c_1 n^{-\tau}/16\Big). \qquad \square
\end{aligned}
$$

Acknowledgements. We are grateful to Professor Bühlmann, the Associate Editor and three referees for their encouragement, insightful and constructive comments.

SUPPLEMENTARY MATERIAL

Supplement A: "Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data" (DOI: 10.1214/13-AOS1087SUPP; .pdf). We provide additional technical details and numerical examples in the supplemental material.

REFERENCES

BAIR, E. and TIBSHIRANI, R. (2004). Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol*. **2** 511–522.

BERAN, R. (1981). Nonparametric regression with randomly censored survival data, Technical report. Univ. California, Berkeley.

BÜHLMANN, P., KALISCH, M. and MAATHUIS, M. H. (2010). Variable selection in high-dimensional linear models: Partially faithful distributions and the PC-simple algorithm. *Biometrika* **97** 261–278. MR2650737

FAN, J., FENG, Y. and WU, Y. (2010). Ultrahigh dimensional variable selection for Cox's proportional hazards model. *IMS Collections* **6** 70–86.

FAN, J., FENG, Y. and SONG, R. (2011). Nonparametric independence screening in sparse ultra-high-dimensional additive models. *J. Amer. Statist. Assoc*. **106** 544–557. MR2847969

FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc*. **96** 1348–1360. MR1946581

FAN, J. and LV, J. (2008). Sure independence screening for ultra-high dimensional feature space (with discussion). *J. Roy. Statist. Soc. Ser. B* **70** 849–911.

FAN, J., SAMWORTH, R. and WU, Y. (2009). Ultrahigh dimensional variable selection: Beyond the linear model. *J. Mach. Learn. Res*. **10** 1829–1853.

FAN, J. and SONG, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist*. **38** 3567–3604. MR2766861

GONZALEZ-MANTEIGA, W. and CADARSO-SUAREZ, C. (1994). Asymptotic properties of a generalized Kaplan–Meier estimator with some applications. *J. Nonparametr. Stat*. **4** 65–78. MR1366364

HALL, P. and MILLER, H. (2009). Using generalized correlation to effect variable selection in very high dimensional problems. *J. Comput. Graph. Statist*. **18** 533–550. MR2751640

HE, X. and SHI, P. (1996). Bivariate tensor-product B-splines in a partly linear model. *J. Multivariate Anal*. **58** 162–181. MR1405586

HE, X., WANG, L. and HONG, H. G. (2013). Supplement to "Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data." DOI:10.1214/13-AOS1087SUPP.

HJORT, N. L. and POLLARD, D. (1993). Asymptotics for minimisers of convex processes. Technical report, Dept. Statistics, Yale Univ., New Haven, CT. Available at http://citeseer.ist.psu.edu/hjort93asymptotics.html.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc*. **58** 13–30. MR0144363

KNIGHT, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist*. **26** 755–770. MR1626024

KOENKER, R. (2005). *Quantile Regression*. *Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015

LI, H. and LUAN, Y. (2005). Boosting proportional hazards models using smoothing splines, with applications to high-dimensional microarray data. *Bioinformatics* **21** 2403–2409.

LI, R., ZHONG, W. and ZHU, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc*. **107** 1129–1139.

LO, S.-H. and SINGH, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probab. Theory Related Fields* **71** 455–465. MR0824714

MASSART, P. (2000). Some applications of concentration inequalities to statistics. *Ann. Fac. Sci. Toulouse Math*. (6) **9** 245–303. MR1813803

MCKEAGUE, I. W., SUBRAMANIAN, S. and SUN, Y. (2001). Median regression and the missing information principle. *J. Nonparametr. Stat*. **13** 709–727. MR1931164

PENG, L. and HUANG, Y. (2008). Survival analysis with quantile regression models. *J. Amer. Statist. Assoc*. **103** 637–649. MR2435468

PORTNOY, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc*. **98** 1001–1012. MR2041488

ROSENWALD, A., WRIGHT, G., CHAN, W. C., CONNORS, J. M., HERMELINK, H. K., SMELAND, E. B. and STAUDT, L. M. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *The New England Journal of Medicine* **346** 1937–1947.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist*. **13** 689–705. MR0790566

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

WANG, H. J. and WANG, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc*. **104** 1117–1128. MR2562007

YING, Z., JUNG, S. H. and WEI, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc*. **90** 178–184. MR1325125

ZHAO, S. D. and LI, Y. (2012). Principled sure independence screening for Cox models with ultra-high-dimensional covariates. *J. Multivariate Anal*. **105** 397–411. MR2877525

ZHOU, S., SHEN, X. and WOLFE, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist*. **26** 1760–1782. MR1673277

ZHU, L.-P., LI, L., LI, R. and ZHU, L.-X. (2011). Model-free feature screening for ultrahigh-dimensional data. *J. Amer. Statist. Assoc*. **106** 1464–1475. MR2896849

X. HE
DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109
USA
E-MAIL: xmhe@umich.edu

L. WANG
SCHOOL OF STATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: wangx346@umn.edu

H. HONG
BARUCH COLLEGE
DEPARTMENT OF STATISTICS AND CIS
CITY UNIVERSITY OF NEW YORK
NEW YORK, NEW YORK 10010
USA
E-MAIL: hyokyoung.hong@baruch.cuny.edu