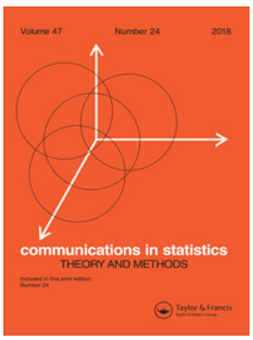

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

A new robust model-free feature screening method for ultra-high dimensional right censored data

Yi Liu & Xiaolin Chen

To cite this article: Yi Liu & Xiaolin Chen (2020): A new robust model-free feature screening method for ultra-high dimensional right censored data, Communications in Statistics - Theory and Methods, DOI: 10.1080/03610926.2020.1769672

To link to this article: https://doi.org/10.1080/03610926.2020.1769672

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2020.1769672

A new robust model-free feature screening method for ultra-high dimensional right censored data

Yi Liu[a] and Xiaolin Chen[b]

[a]School of Mathematical Sciences, Ocean University of China, Qingdao, China; [b]School of Statistics, Qufu Normal University, Qufu, China

ABSTRACT

This paper is concerned with the robust feature screening method for ultra-high dimensional right censored data. A new robust and model-free feature screening approach is built upon a screening index constructed from a fresh measure of dependence between the survival time and a single covariate. One attractive property of this newly introduced index is that it equals zero if the survival time and covariate are independent. In addition, it is invariant to monotonic transformation of the response, and robust against heavy-tailed distributions and outliers for the response. These appealing properties render the derived feature screening approach to be a competitive one among feature screening methods for ultra-high dimensional right censored data. We establish the sure screening property for the suggested feature screening procedure under some regularity conditions, and evaluate its performance through numerical studies. Numerical comparisons with several main competitors show the advantages of the newly proposed means over its main competitors. The well-known diffuse large-B-cell lymphoma (DLBCL) data is utilized to illustrate our methodology.

ARTICLE HISTORY
Received 5 July 2019
Accepted 11 May 2020

KEYWORDS
Censored quantile correlation; right censored data; sure screening property; ultra-high; dimensional data

1. Introduction

Ultra-high dimensional right censored data have appeared frequently in various fields such as genomics, imaging, economics and so on. Because the dimensionality of the predictors is much larger than the sample size, the traditional penalized variable selection methods such as LASSO (Tibshirani 1996), adaptive LASSO (Zou 2006), SCAD (Fan and Li 2001) become ineffective, due to the simultaneous challenges of computational expediency, statistical accuracy and algorithmic stability (Fan, Samworth, and Wu 2009). To tackle these problems, Fan and Lv (2008) proposed the sure independence screening (SIS) method which filtered features by ranking marginal absolute Pearson correlation coefficients of response and predictors. This method has the so-called sure screening property, i.e., all of the important features could be identified with probability very close to one. Due to the computational simplicity of this kind of marginal screening methods, inspired by Fan and Lv (2008), much research effort has been made to the

CONTACT Xiaolin Chen xlchen@amss.ac.cn School of Statistics, Qufu Normal University Qufu 273165, China.

study of variable screening for ultrahigh dimensional data, see Liu, Zhong, and Li (2015) for an overview.

Feature screening approaches for ultra-high dimensional right censored data have been investigated extensively in recent years. For the Cox proportional hazard model, Fan, Feng, and Wu (2010) proposed to screen variables by the maximum of the partial likelihood of each covariate, while Zhao and Li (2012) suggested to screen variables by a standardized marginal maximum partial likelihood estimator and established the sure screening property theoretically. Gorst-Rasmussen and Scheike (2013) developed a marginal screening method for a general class of single-index hazard rate models, which include the Cox proportional hazard model as a special case. Zhou and Zhu (2017) extended the SIRS of Zhu et al. (2011) for complete data to right censored data by the inverse probability censoring weighted idea under the multiple-index models. To relax the model assumption completely, many authors studied the model-free feature screening procedures. Song et al. (2014) proposed a censored rank independence screening (CRIS) method through an inverse probability censoring weighted Kendall's tau. Wu and Yin (2015) proposed a weight-adjusted censored conditional quantile screening method. Zhang, Liu, and Wu (2017) developed a correlation rank screening based on the correlation of cumulative distribution function transformed response and each feature. Yan, Tang, and Zhao (2017) suggested to screen variables by the Spearman rank correlation. Zhang et al. (2018) proposed a censored cumulative residual independent screening method. Hao et al. (2019) proposed a method based on Pearson correlation of cumulative distribution function transformed response and covariates.

In this article, under model-free context, we develop a new robust and model-free screening method for right censored data based on a screening index constructed from a fresh measure of dependence between the survival time and a single covariate. This new index is designed mainly by combing the quantile correlation (Li, Li, and Tsai 2015) and the redistribution-of-mass idea (Portnoy 2003). One attractive property of this newly introduced index is that it equals zero if the survival time and covariate are independent. Thus it could be used as the marginal utility of feature screening for ultra-high dimensional right censored data. Moreover, this index possesses at least three more appealing properties: firstly, it is invariant to monotonic transformation of the response, and robust against heavy-tailed distributions and outliers for the response; secondly, it does not rely on the model assumption, and can effectively capture both of linear and nonlinear relationships between the response and predictors; thirdly, by treating the quantile level as a standard uniform random variable, this new index could fuse information from all quantile levels and make full use of the data information. These appealing properties render the derived feature screening to be a competitive approach among feature screening methods for ultra-high dimensional right censored data.

The rest of this article is organized as follows. In Section 2, we introduce the definition of the new screening index, present the methodology of the corresponding screening method and its sure screening property. Numerical evaluations of the new screening procedure and comparisons with several main competitors are given in Section 3. A real example is provided to demonstrate the proposed procedure's usefulness in Section 4. We present a short discussion in Section 5. All the technical proofs for the theoretical properties are deferred to the appendix.

2. Censored quantile correlation based screening approach

For two random variables U and V and quantile level $\tau \in (0,1)$, the quantile correlation (Li, Li, and Tsai 2015) is defined as

$$\mathrm{qcor}_\tau(U,V) = \frac{E(\psi_\tau(U-Q_\tau)(V-EV))}{\sqrt{(\tau-\tau^2)Var(V)}} \tag{1}$$

where Q_τ is the τ-th quantile of U and $\psi_\tau(u) = \tau - I(u<0)$ with $I(\cdot)$ being the indicator function. It was proved in Li, Li, and Tsai (2015) that $\mathrm{qcor}_\tau(U,V)$ is a rescaled slope of the linear quantile regression model. In addition, if U and V are independent, then $\mathrm{qcor}_\tau(U,V) = 0$. As pointed out by Ma and Zhang (2016), quantile correlation is robust against outliers, heavy tailed distributions and influence points, and could effectively measure the dependent relationship between two random variables. Thus Ma and Zhang (2016) applied it for the purpose of feature screening for the complete data. However, their method could not be used for feature screening of right censored data.

We extended the work of Ma and Zhang (2016) to the right censored data via the redistribution-of-mass idea (Portnoy 2003). Denote by T and C the survival time and censoring time, respectively. Let $Y = T \wedge C$ and $\delta = I(T \le C)$. Denote by $X = (X_1, X_2, ..., X_p)^T$ the p-dimensional covariate vector. For simplicity, it is supposed that C is independent of the covariates vector X. According to the sparsity principle, we assume that only a small number of the covariates are truly predictive of the survival time. Define

$$\mathcal{M} = \{k: \ F(t|X) \ \text{ functionally depends on } X_k\}$$

where $F(t|X)$ is the conditional cumulative distribution function of T given X. If $k \in \mathcal{M}$, X_k is regarded as active predictor; otherwise, as inactive predictor. Furthermore, let $X_{\mathcal{M}}$ be the vector consisting of all variables X_k with $k \in \mathcal{M}$. Similarly, $X_{\mathcal{M}^c}$ represents the vector consisting of all variables X_k with $k \in \mathcal{M}^c$, where $\mathcal{M}^c$ is the complementary set of $\mathcal{M}$.

To recover $\mathcal{M}$, we need to obtain an index to measure the dependence of the survival time and each covariate as many other marginal screening methods. Here, we consider the extension of quantile correlation for complete data to right censored data. With a little confusion with $F(\cdot|X)$, we write the cumulative distribution function of the survival time T as $F(\cdot)$. To evaluate the dependence of X_k on T' τth quantile, we could consider the Pearson correlation between the weight-adjusted check function $\phi_\tau(Y-Q_\tau)$ and X_k:

$$\mathrm{cqcor}_\tau(T,X_k) = \frac{E(\phi_\tau(Y-Q_\tau)(X_k-EX_k))}{\sqrt{Var(\phi_\tau(Y-Q_\tau))Var(X_k)}} \tag{2}$$

where $\phi_\tau(u) = \tau - \varpi(F)I(u<0), Q_\tau$ is the τ-th quantile of T and

$$\varpi(F) = \begin{cases} 1, & \delta = 1 \ \text{ or } \ F(C) > \tau, \\ \dfrac{\tau - F(C)}{1-F(C)}, & \delta = 0, \ F(C) \le \tau \end{cases}$$

When the right censoring is absent, i.e., $\delta = 1$, $\varpi(F)$ always equals 1. Then $\mathrm{cqcor}_\tau(T,X_k)$ degrades into the quantile correlation defined in Li, Li, and Tsai (2015) for complete

data. Here the notation of "$\mathrm{cqcor}_\tau(\cdot,\cdot)$" is modified from the quantile correlation "$\mathrm{qcor}_\tau(\cdot,\cdot)$" used in Li, Li, and Tsai (2015), and "c" represents censoring. It is easy to see that $\mathrm{cqcor}_\tau(T, X_k) = 0$ if X_k has no influence on the τth quantile of T.

To completely capture the dependence information provided by all quantile levels, we view quantile level as a standard uniform random variable, and then take expectation with regard to it for the fusion of information. Specifically, for each k, we suggest to use

$$\omega_k = E_\tau\left[\mathrm{cqcor}^2_\tau(T, X_k)\right] \tag{3}$$

as the marginal screening utility. It is easy to note that $\omega_k = 0$ if X_k is independent of T. Otherwise, it is reasonable to anticipate that ω_k is greater than 0. Thus ω_k could be served as the screening utility. From the construction, it is easy to see that ω_k is model-free, invariant to monotonic transformation of the response, and robust against heavy-tailed distributions and outliers for the response. Without loss of generality, it is assumed that $E(X_k) = 0$ and $Var(X_k) = 1$ for $k = 1, ..., p$. This will simplify the subsequent presentation in this article.

Let $\{Y_i, \delta_i, X_i\}, i = 1, ..., n$ be n independent and identically distributed random sample from $\{Y, \delta, X\}$. Recall that τ is regarded as a standard uniform random variable factitiously, and there is no random sample for it. Here we suppose that there is a hypothetical sample $\tau_i, i = 1, ..., n$, while we specify it by equally spaced points on an interval in simulation studies. Without loss of generality, it is assumed that these τ_is are sorted, and $0 < \tau_1 < \tau_2 < \cdots < \tau_n < 1$. Then ω_k could be easily estimated by

$$\hat{\omega}_k = \frac{1}{n}\sum_{j=1}^{n}\left[\frac{1}{\sqrt{\widehat{Var}(\phi_{\tau_j}(Y - Q_{\tau_j}))}}\frac{1}{n}\sum_{i=1}^{n}\hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j})X_{ik}\right]^2, \quad k = 1, ..., p \tag{4}$$

where $\hat{Q}_\tau = \inf\{y : \hat{F}(y) \geq \tau\}, \hat{\phi}_\tau(u) = \tau - \varpi(\hat{F})I(u < 0)$ with $\hat{F}$ being the Kaplan-Meier estimate of the survival time, and $\widehat{Var}(\phi_{\tau_j}(Y - Q_{\tau_j})) = 1/n\sum_{i=1}^{n}\hat{\phi}^2_{\tau_j}(Y_i - \hat{Q}_{\tau_j}) - (1/n\sum_{i=1}^{n}\hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j}))^2$. We select the covariates with large $\hat{\omega}_k$ s as the active predictors. Specifically, the index set of identified features are

$$\hat{\mathcal{M}} = \{1 \leq k \leq p : \hat{\omega}_k \geq cn^{-\kappa}\}$$

where c is a positive constant and $0 \leq \kappa < 1/2$. For practical use, it is more common to estimate $\mathcal{M}$ by

$$\tilde{\mathcal{M}} = \{1 \leq k \leq p : \hat{\omega}_k \text{ is among the first } d_n \text{ largest of all }\}$$

where d_n is a pre-determined sample-dependent sequence. Hereafter, we refer the proposed procedure as quantile correlation sure independence screening for censored data (cQC-SIS for short).

Before introducing cQC-SIS's the theoretical properties, we list the following assumptions:

(C1) Given $\beta^\top X_{\mathcal{M}}$, T and X are conditional independent;

(C2) The following linear condition holds:

$$E(X|\beta^\top X_{\mathcal{M}}) = \mathrm{cov}(X, X_{\mathcal{M}}^\top)\beta\{\mathrm{cov}(\beta^\top X_{\mathcal{M}})\}^{-1}\beta^\top X_{\mathcal{M}};$$

(C3) It holds uniformly for p that

$$\frac{q^2\lambda_{\max}(E_\tau\{(Var(\phi_\tau(Y-Q_\tau)))^{-1}E^{\otimes 2}[X_{\mathcal{M}}\varpi(F)I(Y-Q_\tau<0)]\})}{\lambda_{\min^2}\{\mathrm{cov}(X_{\mathcal{M}}, X_{\mathcal{M}}^\top)\}} < \frac{\min_{k\in\mathcal{M}}\omega_k}{\lambda_{\max}(\mathrm{cov}(X_{\mathcal{M}}, X_{\mathcal{M}^c}^\top)\mathrm{cov}(X_{\mathcal{M}^c}, X_{\mathcal{M}}^\top))}$$

where $\otimes$ is the Kronecker product, q is the cardinality of $\mathcal{M}$, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the largest and smallest eigenvalues of matrix A, respectively;

(C4) The covariates X_k's are uniformly bounded for $k = 1, ..., p$;

(C5) The cumulative distribution function of T, $F(\cdot)$, is continuously differentiable in a small neighborhood of Q_{τ_j}, say $[Q_{\tau_j} - \delta_0, Q_{\tau_j} + \delta_0], \delta_0 > 0, j = 1, ..., n$. Let $G_{1j}(\delta_0) = \inf_{t\in[Q_{\tau_j}-\delta_0, Q_{\tau_j}+\delta_0]} f(t)$ and $G_{2j}(\delta_0) = \sup_{t\in[Q_{\tau_j}-\delta_0, Q_{\tau_j}+\delta_0]} f(t)$, where $f(t)$ is the density function of T. Assume that $0 < G_{1j}(\delta_0) < G_{2j}(\delta_0) < \infty$.

(C6) There exists a positive constant m_0 such that $Var(\phi_{\tau_j}(Y - Q_{\tau_j})) \geq m_0, j = 1, ..., n$.

Condition (C1) states that T depends on X only through $\beta^\top X_{\mathcal{M}}$, which could allow for a rich class of parametric and semiparametric models for them (Zhu et al. 2011). However, we want to emphasize that this condition is just imposed here to ensure the theoretical property of our cQC-SIS. Our cQC-SIS still has satisfactory performance when this condition is violated in numerical studies. Conditions (C2) and (C3) have been used in Zhu et al. (2011) and Zhou and Zhu (2017) respectively, and are key technical conditions for the success of our suggested screening approach. For more detailed discussion of these two conditions, the reader could refer to these two articles. Condition (C4) is relatively restrictive and only used to simplify our theoretical proofs. It is common to impose Condition (C5) in problems concerned with quantile regression, and it can be satisfied by a large variety of continuous distributions. Condition (C6) is required to guarantee that the proposed index is well defined.

Our first theoretical property is presented in the following Theorem 1, which is concerned with the ranking property of ω_ks.

Theorem 1. *Under the regularity Conditions (C1) to (C3), we have*

$$\max_{k\in\mathcal{M}^c}\omega_k < \min_{k\in\mathcal{M}}\omega_k.$$

This theorem declares that the values of our marginal screening utility ω_ks with indexes in the active set are larger than those not in it.

The following theorem describes the sure screening property of the proposed approach.

Theorem 2. *Under the regularity conditions (C1) to (C5), we have*

$$P\left(\max_{1\leq k\leq p}|\hat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right) \leq O(np\exp(-Cn^{1-2\kappa}))$$

where C is a positive constant and κ is defined in Condition (C1). Furthermore, if we assume that $\min_{k\in\mathcal{M}}\omega_k \geq 2cn^{-\kappa}$, then

$$P(\mathcal{M} \subset \hat{\mathcal{M}}) \geq 1 - O(nq \exp(-Cn^{1-2\kappa}))$$

For ultra-high dimensional right censored data, Theorem 2 indicates that the dimensionality could be $p = o(\exp(n^{1-2\kappa}))$, i.e., it could grow at an exponential rate of the sample size n.

3. Numerical studies

In this section, we examine the finite-sample performances of the proposed method and make comparisons with several main existing methods through Monte Carlo simulation studies, including FAST (Gorst-Rasmussen and Scheike 2013), CRIS (Song et al. 2014), CR-SIS (Zhang, Liu, and Wu 2017), cSIRS (Zhou and Zhu 2017), SRCS (Yan, Tang, and Zhao 2017), and CCRIS (Zhang et al. 2018). A majority of our simulation elements in this section are borrowed from Zhang, Liu, and Wu (2017). Survival time T follows the distribution specified in each example, while censoring times are generated from the uniform distribution on $(0, c_0)$, where different c_0's are set to produce censoring rates (CR) 20% and 40%. In all the three examples, the covariate vector X follows the multivariate normal distribution with mean zero and covariance matrix $\Sigma = (0.6^{|i-j|})_{p\times p}$. We set the sample size n to be 200 and take the number of features $p = 1000$ and 3000.

To assess the performances of various methods, we independently generate 500 simulated datasets for each example. To summarize these results, we consider the following performance measures: the minimum model size, denoted by S (only 5%, 25%, 50%, 75% and 95% quantiles of 500 Ss will be displayed); the proportion of including a single active variable P_k, where k is the index of truly important predictor; the proportion of including all active variables P_{All}. We choose estimated model size d_n to be $[n/\log(n)]$, where $[x]$ denotes the integer part of x. In addition, we take τ_i, $i = 1, \ldots, n$ as n equally spaced points of the interval $[c_1, c_2]$ in our specific estimation, where $c_1 = 0.05$ and c_2 is the largest possible value such that $\hat{Q}_{c_2}$ is well defined.

Example 1. Survival times are generated from the following high dimensional Cox proportional hazards model:

$$\lambda(t|X) = \lambda_0(t) \exp(X^\top \beta_0)$$

where the baseline hazard function is set to be $\lambda_0(t) = (t - 0.5)^2$ and the regression coefficient vector $\beta_0 = (0.35, 0.35, 0.35, 0.35, 0.35, 0, \ldots, 0)^\top$. It is easy to see that only the first five covariates are truly predictive of T. This model has been used in Example 2 of Zhang, Liu, and Wu (2017).

Example 2. Different from Example 1, we consider the following nonlinear accelerated failure time model:

$$\log(T) = X_1^2 + (2 + \sin X_2)^2 + (1 + X_3)^{-3} + (X_4^2 + X_4 - 1)^{-1} + X_5 + \epsilon$$

where the error follows the standard normal distribution. This model has been exploited in Example 3 of Zhang, Liu, and Wu (2017).

Table 1. The quantiles of S for the three examples with $p = 1000$.

	CR	Method	5%	25%	50%	75%	95%
Example 1	0.20	cQC-SIS	5.00	5.00	5.00	5.00	5.00
		CRIS	5.00	5.00	6.00	11.50	84.00
		FAST	5.00	5.00	5.00	5.00	6.00
		CR-SIS	5.00	5.00	5.00	5.00	9.00
		SRCS	5.00	5.00	5.00	5.00	7.00
		cSIRS	5.00	5.00	5.00	5.00	7.00
		CCRIS	5.00	5.00	7.00	14.00	62.00
	0.40	cQC-SIS	5.00	5.00	5.00	5.00	6.00
		CRIS	5.00	5.00	7.00	23.50	191.50
		FAST	5.00	5.00	5.00	5.00	6.00
		CR-SIS	5.00	5.00	6.00	10.00	54.50
		SRCS	5.00	5.00	5.00	6.00	13.50
		cSIRS	5.00	5.00	5.00	6.00	18.50
		CCRIS	15.00	67.50	169.00	368.50	750.50
Example 2	0.20	cQC-SIS	5.00	5.00	7.00	20.00	90.50
		CRIS	37.00	188.50	547.50	820.50	977.50
		FAST	79.50	339.50	630.00	859.00	976.50
		CR-SIS	5.00	9.00	27.00	83.00	384.00
		SRCS	5.00	7.00	19.00	70.00	368.50
		cSIRS	5.00	5.00	9.00	22.00	94.50
		CCRIS	5.00	6.50	16.00	67.50	364.50
	0.40	cQC-SIS	5.00	5.00	8.00	20.00	111.00
		CRIS	52.00	263.50	639.50	889.50	982.50
		FAST	10.00	44.00	150.50	397.00	859.50
		CR-SIS	5.00	11.00	36.50	125.50	517.00
		SRCS	5.00	6.00	12.00	33.50	179.50
		cSIRS	5.00	7.00	13.00	35.00	143.00
		CCRIS	7.00	24.50	86.50	220.00	564.00
Example 3	0.20	cQC-SIS	5.00	5.00	8.00	19.00	90.00
		CRIS	32.50	259.00	577.50	849.00	980.00
		FAST	88.00	382.50	659.00	855.00	976.00
		CR-SIS	5.00	9.00	25.50	90.00	335.00
		SRCS	5.00	9.00	24.50	75.50	380.00
		cSIRS	5.00	6.00	9.00	22.00	105.00
		CCRIS	5.00	6.00	17.00	60.00	309.50
	0.40	cQC-SIS	5.00	5.00	8.00	24.00	103.00
		CRIS	45.50	368.00	693.50	898.00	986.00
		FAST	11.50	61.00	180.00	462.00	849.00
		CR-SIS	5.00	11.00	32.00	131.00	488.00
		SRCS	5.00	6.00	13.00	37.00	187.00
		cSIRS	5.00	6.00	12.00	36.00	146.00
		CCRIS	6.00	25.00	73.50	211.50	576.50

cQC-SIS, the proposed method; CRIS, Song et al. (2014); FAST, Gorst-Rasmussen and Scheike (2013); CR-SIS, Zhang, Liu, and Wu (2017); cSIRS, Zhou and Zhu (2017); SRCS, Yan, Tang, and Zhao (2017); CCRIS, Zhang et al. (2018).

Example 3. All the elements are the same with Example 2 except that the errors ϵ s are generated from the type-I extreme value distribution EV(0,1).

Tables 1 and 2 exhibit the simulation results for each example with $p = 1000$, while Tables 3 and 4 summarize those for all the examples with $p = 3000$. From these tables, it could be concluded that our proposed cQC-SIS is consistently superior to the other six methods, especially when the dimension of predictors is high and survival models are complex. As seen from the results for Example 1, the performance of FAST is very satisfactory and almost the same as that of cQC-SIS. The reason may be that FAST is specially designed for the single-index hazards model, and thus it naturally behaves well for the Cox's model, which is almost the simplest single-index hazards model. However, the existence of nonlinear covariate effects on the survival time could easily lead to the

Table 2. The proportions of P_j and P_{All} for the three examples with $p = 1000$.

			P_j					P_{All}
	CR	Method	X_1	X_2	X_3	X_4	X_5	All
Example 1	0.20	cQC-SIS	1.00	1.00	1.00	1.00	1.00	1.00
		CRIS	0.95	0.99	0.99	0.99	0.93	0.89
		FAST	1.00	1.00	1.00	1.00	1.00	1.00
		CR-SIS	1.00	1.00	1.00	1.00	0.99	0.99
		SRCS	1.00	1.00	1.00	1.00	1.00	1.00
		cSIRS	1.00	1.00	1.00	1.00	1.00	1.00
		CCRIS	0.94	1.00	1.00	0.99	0.96	0.92
	0.40	cQC-SIS	1.00	1.00	1.00	1.00	1.00	1.00
		CRIS	0.91	0.96	0.99	0.97	0.88	0.81
		FAST	1.00	1.00	1.00	1.00	1.00	1.00
		CR-SIS	0.97	1.00	1.00	0.99	0.96	0.93
		SRCS	1.00	1.00	1.00	1.00	0.99	0.99
		cSIRS	0.99	1.00	1.00	1.00	0.99	0.99
		CCRIS	0.35	0.55	0.63	0.56	0.36	0.15
Example 2	0.20	cQC-SIS	0.92	1.00	1.00	0.99	0.94	0.88
		CRIS	0.43	0.82	0.10	0.35	0.34	0.05
		FAST	0.40	0.74	0.05	0.20	0.36	0.02
		CR-SIS	0.80	1.00	0.86	0.85	0.79	0.57
		SRCS	0.91	1.00	0.75	0.90	0.88	0.64
		cSIRS	0.93	1.00	1.00	0.99	0.92	0.86
		CCRIS	0.84	0.97	0.93	0.90	0.77	0.64
	0.40	cQC-SIS	0.90	1.00	1.00	0.98	0.93	0.84
		CRIS	0.29	0.70	0.09	0.30	0.27	0.03
		FAST	0.66	0.96	0.43	0.59	0.59	0.22
		CR-SIS	0.71	0.99	0.91	0.83	0.76	0.51
		SRCS	0.92	1.00	0.91	0.96	0.90	0.77
		cSIRS	0.85	1.00	1.00	0.97	0.89	0.76
		CCRIS	0.62	0.94	0.89	0.73	0.56	0.34
Example 3	0.20	cQC-SIS	0.91	1.00	1.00	0.99	0.95	0.86
		CRIS	0.41	0.80	0.10	0.37	0.36	0.06
		FAST	0.38	0.68	0.05	0.23	0.35	0.01
		CR-SIS	0.79	1.00	0.87	0.87	0.83	0.57
		SRCS	0.88	1.00	0.74	0.93	0.88	0.61
		cSIRS	0.90	1.00	1.00	0.99	0.94	0.85
		CCRIS	0.83	0.98	0.93	0.89	0.81	0.67
	0.40	cQC-SIS	0.89	1.00	1.00	0.98	0.94	0.83
		CRIS	0.32	0.72	0.09	0.28	0.29	0.04
		FAST	0.64	0.95	0.36	0.54	0.59	0.19
		CR-SIS	0.73	0.99	0.89	0.84	0.76	0.53
		SRCS	0.91	1.00	0.89	0.95	0.92	0.75
		cSIRS	0.86	1.00	1.00	0.97	0.89	0.76
		CCRIS	0.61	0.93	0.87	0.75	0.57	0.35

cQC-SIS, the proposed method; CRIS, Song et al. (2014); FAST, Gorst-Rasmussen and Scheike (2013); CR-SIS, Zhang, Liu, and Wu (2017); cSIRS, Zhou and Zhu (2017); SRCS, Yan, Tang, and Zhao (2017); CCRIS, Zhang et al. (2018).

complete failure of FAST, as demonstrated in the results for Examples 2 and 3. In addition, although the performance of SRCS and cSIRS is close to our suggested cQC-SIS sometimes, cQC-SIS is more stable under different censoring rates and performs uniformly better than them. It also should be noted that Condition (C1) is violated in Examples 2 and 3. But the performance of cQC-SIS is still very well. This indicates that cQC-SIS is insensitive to the violation of Condition (C1).

In the previous simulation studies, the censoring time is independent of the covariates, while it is not necessarily the case in practical problems. As suggested by one reviewer, it is meaningful to examine the situations with censoring time being

Table 3. The quantiles of S for the three examples with $p = 3000$.

	CR	Method	5%	25%	50%	75%	95%
Example 1	0.20	cQC-SIS	5.00	5.00	5.00	5.00	6.00
		CRIS	5.00	5.00	7.00	23.00	219.50
		FAST	5.00	5.00	5.00	5.00	6.00
		CR-SIS	5.00	5.00	5.00	6.00	15.50
		SRCS	5.00	5.00	5.00	5.00	10.00
		cSIRS	5.00	5.00	5.00	5.00	8.00
		CCRIS	5.00	6.00	10.00	30.50	220.50
	0.40	cQC-SIS	5.00	5.00	5.00	5.00	6.00
		CRIS	5.00	6.00	12.00	62.00	584.50
		FAST	5.00	5.00	5.00	5.00	7.00
		CR-SIS	5.00	5.00	7.50	21.00	125.00
		SRCS	5.00	5.00	5.00	7.00	32.50
		cSIRS	5.00	5.00	5.00	7.00	25.00
		CCRIS	37.50	217.00	543.00	1147.50	2234.50
Example 2	0.20	cQC-SIS	5.00	7.00	15.00	58.50	341.50
		CRIS	78.50	662.00	1683.50	2588.50	2938.00
		FAST	206.00	1110.50	1909.50	2585.50	2904.00
		CR-SIS	5.00	22.00	87.50	293.00	1346.00
		SRCS	5.00	16.00	64.50	209.00	1231.00
		cSIRS	5.00	8.00	20.00	68.00	366.50
		CCRIS	5.00	12.00	50.50	228.00	995.00
	0.40	cQC-SIS	5.00	7.00	16.00	61.50	346.00
		CRIS	120.00	854.50	1930.00	2667.50	2960.50
		FAST	16.50	122.50	433.50	1054.50	2457.50
		CR-SIS	6.00	26.50	122.50	473.00	1391.00
		SRCS	5.00	8.00	27.00	94.00	515.50
		cSIRS	5.00	9.50	31.00	110.00	424.00
		CCRIS	13.00	81.00	253.00	794.50	1941.50
Example 3	0.20	cQC-SIS	5.00	7.00	18.00	55.50	348.50
		CRIS	96.50	712.00	1749.50	2525.00	2917.50
		FAST	349.50	1153.00	1973.00	2557.00	2928.50
		CR-SIS	6.00	23.00	102.50	298.00	1360.00
		SRCS	6.00	19.00	69.50	250.50	1384.50
		cSIRS	5.00	8.00	23.00	58.50	331.00
		CCRIS	5.00	10.50	43.00	197.50	905.50
	0.40	cQC-SIS	5.00	6.00	14.00	69.00	452.00
		CRIS	94.00	912.00	1961.50	2658.50	2960.00
		FAST	20.00	157.50	571.50	1377.00	2589.50
		CR-SIS	6.00	20.50	97.50	392.50	1174.50
		SRCS	5.00	8.00	27.00	121.00	761.50
		cSIRS	5.00	9.00	28.50	111.00	566.50
		CCRIS	10.50	60.00	189.00	485.50	1515.50

cQC-SIS, the proposed method; CRIS, Song et al. (2014); FAST, Gorst-Rasmussen and Scheike (2013); CR-SIS, Zhang, Liu, and Wu (2017); cSIRS, Zhou and Zhu (2017); SRCS, Yan, Tang, and Zhao (2017); CCRIS, Zhang et al. (2018).

dependent on covariates. Here, we reexamine Example 2 by generating censoring times from the model $C = 20\exp(X^\top\gamma) + \varepsilon$, where $\gamma = (3, 2.5, 0, ..., 0)$ is the p dimensional vector of regression coefficients and the error ε follows a standard normal distribution. The other settings are the same as those used previously. Under these settings, the censoring rate is 30% in Example 2. The results are summarized in Tables 5 and 6, which indicate that the proposed method performed very well under the scenario that censoring time is dependent on covariates. In addition, this setting obviously violates the assumption that C is dependent of X. The results in Tables 5 and 6 show that our methodology is insensitive to this assumption.

Table 4. The proportion of P_j and P_{All} for the three examples with $p = 3000$.

			P_j					P_{All}
	CR	Method	X_1	X_2	X_3	X_4	X_5	All
Example 1	0.20	cQC-SIS	1.00	1.00	1.00	1.00	1.00	1.00
		CRIS	0.89	0.98	0.99	0.99	0.88	0.79
		FAST	1.00	1.00	1.00	1.00	1.00	1.00
		CR-SIS	0.99	1.00	1.00	1.00	1.00	0.99
		SRCS	1.00	1.00	1.00	1.00	0.99	0.99
		cSIRS	1.00	1.00	1.00	1.00	0.99	0.99
		CCRIS	0.87	0.97	0.99	0.98	0.89	0.79
	0.40	cQC-SIS	1.00	1.00	1.00	1.00	1.00	1.00
		CRIS	0.79	0.94	0.97	0.95	0.83	0.69
		FAST	1.00	1.00	1.00	1.00	1.00	1.00
		CR-SIS	0.91	0.99	1.00	0.99	0.92	0.83
		SRCS	0.98	1.00	1.00	1.00	0.99	0.96
		cSIRS	0.99	1.00	1.00	1.00	0.98	0.97
		CCRIS	0.20	0.43	0.51	0.38	0.22	0.05
Example 2	0.20	cQC-SIS	0.80	1.00	1.00	0.96	0.83	0.68
		CRIS	0.31	0.75	0.07	0.25	0.25	0.03
		FAST	0.22	0.56	0.02	0.12	0.21	0.00
		CR-SIS	0.63	0.98	0.76	0.74	0.66	0.35
		SRCS	0.77	1.00	0.61	0.83	0.75	0.39
		cSIRS	0.79	1.00	1.00	0.95	0.80	0.65
		CCRIS	0.72	0.98	0.87	0.79	0.63	0.46
	0.40	cQC-SIS	0.79	1.00	1.00	0.97	0.85	0.67
		CRIS	0.24	0.63	0.03	0.20	0.22	0.01
		FAST	0.49	0.91	0.28	0.44	0.48	0.11
		CR-SIS	0.54	0.97	0.81	0.72	0.61	0.30
		SRCS	0.83	1.00	0.82	0.91	0.81	0.57
		cSIRS	0.70	1.00	1.00	0.93	0.77	0.54
		CCRIS	0.41	0.90	0.79	0.57	0.39	0.15
Example 3	0.20	cQC-SIS	0.80	1.00	1.00	0.96	0.86	0.68
		CRIS	0.25	0.71	0.05	0.20	0.23	0.02
		FAST	0.23	0.54	0.02	0.11	0.15	0.00
		CR-SIS	0.63	0.99	0.71	0.71	0.65	0.35
		SRCS	0.74	1.00	0.58	0.82	0.77	0.36
		cSIRS	0.78	1.00	1.00	0.94	0.82	0.63
		CCRIS	0.75	0.98	0.85	0.79	0.62	0.47
	0.40	cQC-SIS	0.78	1.00	1.00	0.96	0.85	0.66
		CRIS	0.22	0.62	0.05	0.20	0.16	0.02
		FAST	0.48	0.90	0.22	0.39	0.41	0.09
		CR-SIS	0.56	0.99	0.79	0.72	0.60	0.32
		SRCS	0.81	1.00	0.76	0.90	0.83	0.56
		cSIRS	0.72	1.00	0.99	0.92	0.77	0.55
		CCRIS	0.49	0.95	0.85	0.61	0.44	0.19

cQC-SIS, the proposed method; CRIS, Song et al. (2014); FAST, Gorst-Rasmussen and Scheike (2013); CR-SIS, Zhang, Liu, and Wu (2017); cSIRS, Zhou and Zhu (2017); SRCS, Yan, Tang, and Zhao (2017); CCRIS, Zhang et al. (2018).

4. Real data analysis

As an illustration, we apply the proposed cQC-SIS along with approaches examined in Section 3 to the diffuse large-B-cell lymphoma (DLBCL) data (Rosenwald et al. 2002). This dataset contains $n = 240$ patients, for each of whom $p = 7399$ genes are measured in addition to survival time with diffuse large B-cell lymphoma after chemotherapy. Among 240 patients, survival times of 102 patients are right censored, causing 42.5% censoring rate. In such a large p small n setting, feature screening would be indispensable before performing any sophisticated statistical modeling and analysis. We apply cQC-SIS, CRIS, FAST, CR-SIS, SRCS, cSIRS and CCRIS to pick out important genes

Table 5. The quantiles of S for Example 2 with censoring time being dependent on covariates.

p	Method	5%	25%	50%	75%	95%
1000	cQC-SIS	5.00	5.00	8.00	21.00	110.00
	CRIS	15.00	101.00	345.50	716.00	965.50
	FAST	35.00	172.00	421.50	731.50	953.00
	CR-SIS	5.00	5.00	6.00	16.00	129.00
	SRCS	5.00	6.00	12.00	36.50	134.00
	cSIRS	5.00	6.00	8.00	19.50	95.00
	CCRIS	5.00	5.00	8.00	35.50	273.50
3000	cQC-SIS	5.00	6.00	13.00	44.50	281.00
	CRIS	32.00	253.00	1032.50	2180.50	2893.50
	FAST	69.50	375.50	994.00	1936.00	2811.00
	CR-SIS	5.00	5.00	6.00	18.00	202.00
	SRCS	5.00	8.00	19.50	67.50	402.00
	cSIRS	5.00	7.00	13.50	46.50	225.50
	CCRIS	5.00	5.00	11.00	72.00	546.00

cQC-SIS, the proposed method; CRIS, Song et al. (2014); FAST, Gorst-Rasmussen and Scheike (2013); CR-SIS, Zhang, Liu, and Wu (2017); cSIRS, Zhou and Zhu (2017); SRCS, Yan, Tang, and Zhao (2017); CCRIS, Zhang et al. (2018).

Table 6. The proportions of P_j and P_{All} for for Example 2 with censoring time being dependent on covariates.

		P_j					P_{All}
p	Method	X_1	X_2	X_3	X_4	X_5	All
1000	cQC-SIS	0.90	1.00	1.00	0.99	0.94	0.85
	CRIS	0.57	0.92	0.23	0.45	0.47	0.13
	FAST	0.74	0.95	0.11	0.36	0.47	0.06
	CR-SIS	1.00	1.00	1.00	0.97	0.88	0.86
	SRCS	0.92	1.00	0.89	0.97	0.93	0.76
	cSIRS	1.00	1.00	1.00	0.99	0.87	0.86
	CCRIS	1.00	1.00	1.00	0.95	0.76	0.76
3000	cQC-SIS	0.81	1.00	1.00	0.98	0.89	0.73
	CRIS	0.43	0.85	0.13	0.32	0.34	0.06
	FAST	0.58	0.86	0.08	0.26	0.34	0.03
	CR-SIS	1.00	1.00	1.00	0.96	0.84	0.83
	SRCS	0.85	1.00	0.85	0.93	0.87	0.63
	cSIRS	1.00	1.00	1.00	0.95	0.73	0.72
	CCRIS	1.00	1.00	1.00	0.94	0.67	0.66

cQC-SIS, the proposed method; CRIS, Song et al. (2014); FAST, Gorst-Rasmussen and Scheike (2013); CR-SIS, Zhang, Liu, and Wu (2017); cSIRS, Zhou and Zhu (2017); SRCS, Yan, Tang, and Zhao (2017); CCRIS, Zhang et al. (2018).

related to the survival time. All predictors are standardized with mean zero and variance 1. We choose $d_n = [n/\log n] = 43$ for various screening procedures. The screened genes by these methods are listed in Table 7, from which we find that the number of overlapping genes between cQC-SIS and CRIS, FAST, CR-SIS, SRCS, cSIRS are 2, 21, 33, 27, 22, 0, respectively. Except for CRIS and CCSRIS, there are 15 genes being all selected by the examined screening procedures, which have been confirmed in some previous studies. For example, 4131, BF129543, indicates ESTs, weakly similar to A47224 thyroxine-binding globulin precursor, has been found in Ando et al. (2003); Annest et al. (2009); Gui and Li (2005); Li and Luan (2005); 1825, BC012161, indicates septin 1, has been found in Annest et al. (2009); Li and Luan (2005); 3799, 3822, 3825, indicate major histocompatibility complex, have been found in Gui and Li (2005). These foundings indicate that these genes could be strongly associated with patients' survival risk.

Table 7. Screened genes for the DLBCL data by various approaches.

cQC-SIS	CRIS	FAST	CR-SIS	SRCS	cSIRS	CCRIS
1188	69	394	1188	1188	80	24
1456	958	1072	1439	1456	1072	150
1671	1096	1188	1456	1671	1188	522
1825	1097	1439	1825	1825	1439	561
1994	1314	1456	1841	2579	1456	567
2576	1383	1660	1994	2902	1660	572
2672	1825	1662	2576	3787	1662	625
3787	1830	1664	2672	3795	1664	635
3795	2636	1671	2902	3799	1671	676
3796	3387	1672	3787	3810	1672	677
3799	4523	1678	3795	3811	1678	758
3806	4574	1680	3799	3813	1681	806
3807	4576	1681	3801	3822	1682	837
3808	4578	1682	3806	3824	1825	843
3810	4680	1825	3807	3825	1831	846
3811	4681	1841	3808	4131	1841	986
3812	4682	1871	3810	4148	1854	1114
3813	4683	1986	3811	4162	1871	1116
3815	4684	2579	3813	4190	1985	1153
3818	4733	2672	3814	4202	2437	1286
3820	4808	3248	3815	4232	2579	1297
3821	4827	3799	3818	5023	3248	2935
3822	4886	3810	3819	5024	3799	3281
3824	4887	3811	3820	5025	3810	3307
3825	4916	3812	3821	5027	3811	3308
3828	5068	3813	3822	5054	3812	3488
4131	5179	3822	3824	5055	3813	3493
4202	5404	3824	3825	5260	3820	3591
5023	5614	3825	4131	5296	3821	3592
5024	5621	4131	4574	5297	3822	3593
5025	5767	5027	5024	5298	3824	3601
5027	5802	5054	5025	5301	3825	4061
5055	5951	5055	5027	5442	4131	4321
5254	6166	5297	5055	5451	5027	4392
5260	6352	5301	5254	5456	5055	4506
5296	6411	5614	5260	5459	5297	6485
5297	6439	5621	5296	5475	5301	6686
5301	6442	5950	5301	5476	5614	6871
5364	6747	6365	5614	5614	5950	7102
5459	6790	6411	5975	6134	6365	7129
5614	6844	7069	6365	6896	6411	7279
6365	6956	7343	6508	7343	7343	7384
7357	7062	7357	6896	7357	7357	7385

cQC-SIS, the proposed method; CRIS, Song et al. (2014); FAST, Gorst-Rasmussen and Scheike (2013); CR-SIS, Zhang, Liu, and Wu (2017); cSIRS, Zhou and Zhu (2017); SRCS, Yan, Tang, and Zhao (2017); CCRIS, Zhang et al. (2018).

5. Discussion

Our proposed cQC-SIS procedure is a natural extension of the work Ma and Zhang (2016) for the complete data to the right censored data. This extension is mainly based on the redistribution-of-mass idea (Portnoy 2003), which has been used widely for a variety of problems involving quantiles for right censored data. In addition, the marginal screening utility is constructed in such a manner like the composite quatile regression (Zou and Yuan 2008; Chen, Chen, and Liu 2019).

Most recently, some authors investigated the conditional screening problems for ultrahigh dimensional right censored data, see Hong, Kang, and Li (2018), Liu and Chen (2018) and Chen (2018). In such kind of problems, some predictors are known to

be predictive in advance. This information should be used in the process of screening. We believe that it is meaningful to extend our suggested approaches to this context, and will investigate this in the near future.

Funding

We are deeply grateful to the Editor, the Associate Editor and two referees for their valuable comments which help to improve quality of this paper substantially. Yi Liu's research was supported by the National Natural Science Foundation of China (11801567). Xiaolin Chen's research was supported by the National Natural Science Foundation of China (11771250).

References

Ando, T., M. Suguro, T. Kobayashi, M. Seto, and H. Honda. 2003. Multiple fuzzy neural network system for outcome prediction and classification of 220 lymphoma patients on the basis of molecular profiling. *Cancer Science* 94 (10):906–13. doi:10.1111/j.1349-7006.2003.tb01374.x.

Annest, A.,. R. E. Bumgarner, A. E. Raftery, and K. Y. Yeung. 2009. Iterative bayesian model averaging: A method for the application of survival analysis to high-dimensional microarray data. *BMC Bioinformatics* 10 (1):72. doi:10.1186/1471-2105-10-72.

Chen, X. 2018. Model-free conditional feature screening for ultrahigh dimensional right censored data. *Journal of Statistical Computation and Simulation* 88 (12):2425–46. doi:10.1080/00949655.2018.1466142.

Chen, X., X. Chen, and Y. Liu. 2019. A note on quantile feature screening via distance correlation. *Statistical Papers* 60 (5):1741–2. doi:10.1007/s00362-017-0894-8.

Chen, X., X. Chen, and H. Wang. 2018. Robust feature screening for ultra-high dimensional right censored data via distance correlation. *Computational Statistics & Data Analysis* 119:118–38. doi:10.1016/j.csda.2017.10.004.

Fan, J., Y. Feng, and Y. Wu. 2010. High-dimensional variable selection for Cox's proportional hazards model. In *Borrowing strength: Theory powering applications–A Festschrift for Lawrence D. Brown*, ed. J. O. Berger, T. Tony Cai and I. M. Johnstone, 70–86. Beachwood, OH: Institute of Mathematical Statistics.

Fan, J., and R. Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96 (456):1348–60. doi:10.1198/016214501753382273.

Fan, J., and J. Lv. 2008. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70 (5):849–911. doi:10.1111/j.1467-9868.2008.00674.x.

Fan, J., R. Samworth, and Y. Wu. 2009. Ultrahigh dimensional feature selection: Beyond the linear model. *The Journal of Machine Learning Research* 10:2013–38.

Gorst-Rasmussen, A., and T. Scheike. 2013. Independent screening for single-index hazard rate models with ultrahigh dimensional features. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75 (2):217–45. doi:10.1111/j.1467-9868.2012.01039.x.

Gui, J., and H. Li. 2005. Penalized cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics (Oxford, England)* 21 (13):3001–8. doi:10.1093/bioinformatics/bti422.

Hao, M., Y. Lin, X. Liu, and W. Tang. 2019. Robust feature screening for high-dimensional survival data. *Journal of Applied Statistics* 46 (6):979–94. doi:10.1080/02664763.2018.1529151.

Hong, H.,. J. Kang, and Y. Li. 2018. Conditional screening for ultra-high dimensional covariates with survival outcomes. *Lifetime Data Analysis* 24 (1):45–71. doi:10.1007/s10985-016-9387-7.

Li, G., Y. Li, and C.-L. Tsai. 2015. Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association* 110 (509):246–61. doi:10.1080/01621459.2014.892007.

Li, H., and Y. Luan. 2005. Boosting proportional hazards models using smoothing splines, with applications to high-dimensional microarray data. *Bioinformatics (Oxford, England)* 21 (10): 2403–9. doi:10.1093/bioinformatics/bti324.

Liu, Y., and X. Chen. 2018. Quantile screening for ultra-high-dimensional heterogeneous data conditional on some variables. *Journal of Statistical Computation and Simulation* 88 (2): 329–42. doi:10.1080/00949655.2017.1389944.

Liu, J., W. Zhong, and R. Li. 2015. A selective overview of feature screening for ultrahigh-dimensional data. *Science China Mathematics* 58 (10):1–2054. doi:10.1007/s11425-015-5062-9.

Li, R., W. Zhong, and L. Zhu. 2012. Feature screening via distance correlation learning. *Journal of the American Statistical Association* 107 (499):1129–39. doi:10.1080/01621459.2012.695654.

Ma, X., and J. Zhang. 2016. Robust model-free feature screening via quantile correlation. *Journal* of Multivariate Analysis. 143:472–80. doi:10.1016/j.jmva.2015.10.010.

Portnoy, S. 2003. Censored regression quantiles. *Journal of the American Statistical Association* 98 (464):1001–12. doi:10.1198/016214503000000954.

Rosenwald, A., G. Wright, W. C. Chan, J. M. Connors, E. Campo, R. I. Fisher, R. D. Gascoyne, H. K. Muller-Hermelink, E. B. Smeland, J. M. Giltnane, and E. M. Hurt. 2002. The use of molecular profiling to predict survival after chemotherapy for diffuse large-b-cell lymphoma. *The New England Journal of Medicine* 346 (25):1937–47. doi:10.1056/NEJMoa012914.

Serfling, R. J. 1980. *Approximation theorems of mathematical statistics*, vol. 162. New York: John Wiley & Sons, Inc.

Song, R., W. Lu, S. Ma, and X. J. Jeng. 2014. Censored rank independence screening for high-dimensional survival data. *Biometrika* 101 (4):799–814. doi:10.1093/biomet/asu047.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1):267–88. doi:10.1111/j.2517-6161.1996.tb02080.x.

Wu, Y., and G. Yin. 2015. Conditional quantile screening in ultrahigh-dimensional heterogeneous data. *Biometrika* 102 (1):65–76. doi:10.1093/biomet/asu068.

Xu, K., and X. Huang. 2018. Conditional-quantile screening for ultrahigh-dimensional survival data via martingale difference correlation. *Science China Mathematics* 61 (10):1907–22. doi:10.1007/s11425-016-9208-6.

Yan, X., N. Tang, and X. Zhao. 2017. The spearman rank correlation screening for ultrahigh-dimensional censored data. arXiv:1702.02708.

Zhang, J., Y. Liu, and Y. Wu. 2017. Correlation rank screening for ultrahigh-dimensional survival data. *Computational Statistics & Data Analysis* 108:121–32. doi:10.1016/j.csda.2016.11.005.

Zhang, J., G. Yin, Y. Liu, and Y. Wu. 2018. Censored cumulative residual independent screening for ultrahigh-dimensional survival data. *Lifetime Data Analysis* 24 (2):273–92. doi:10.1007/s10985-017-9395-2.

Zhao, S. D., and Y. Li. 2012. Principled sure independence screening for cox models with ultra-high-dimensional covariates. *Journal of Multivariate Analysis* 105 (1):397–411. doi:10.1016/j.jmva.2011.08.002.

Zhou, T., and L. Zhu. 2017. Model-free feature screening for ultrahigh dimensional censored regression. *Statistics and Computing* 27 (4):947–61. doi:10.1007/s11222-016-9664-z.

Zhu, L.-P., L. Li, R. Li, and L.-X. Zhu. 2011. Model-free feature screening for Ultrahigh Dimensional Data. *Journal of the American Statistical Association* 106 (496):1464–75. doi:10.1198/jasa.2011.tm10563.

Zou, H. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101 (476):1418–29. doi:10.1198/016214506000000735.

Zou, H., and M. Yuan. 2008. Composite quantile regression and the oracle model selection theory. *Annals of Statistics* 36:1008–126.

Appendix: Proofs of the theorems

Proof of Theorem 1. By the definition of $\mathrm{cqcor}_\tau(T, X_k)$ in (2) and the assumption that $\beta^\top \mathrm{cov}(X_\mathcal{M})\beta = I_q$, where I_q is the $q \times q$ identity matrix. We have

$$\begin{aligned}
&\mathrm{cqcor}_\tau(T, X_k)\\
&= (Var(\phi_\tau(Y - Q_\tau)))^{-1/2} E(\phi_\tau(Y - Q_\tau)X_k)\\
&= (Var(\phi_\tau(Y - Q_\tau)))^{-1/2} E\{[\tau - \varpi(F)I(Y - Q_\tau < 0)]X_k\}\\
&= -(Var(\phi_\tau(Y - Q_\tau)))^{-1/2} E\{\varpi(F)I(Y - Q_\tau < 0)X_k\}\\
&= -(Var(\phi_\tau(Y - Q_\tau)))^{-1/2} E\{E\big[\varpi(F)I(Y - Q_\tau < 0)|\beta^\top X_\mathcal{M}\big]E\big[X_k|\beta^\top X_\mathcal{M}\big]\}\\
&= -(Var(\phi_\tau(Y - Q_\tau)))^{-1/2} E\{\mathrm{cov}(X_k, X_\mathcal{M}^\top\beta)\beta^\top X_\mathcal{M} E\big[\varpi(F)I(Y - Q_\tau < 0)|\beta^\top X_\mathcal{M}\big]\}\\
&= -(Var(\phi_\tau(Y - Q_\tau)))^{-1/2} \mathrm{cov}(X_k, X_\mathcal{M}^\top\beta)E\{\beta^\top X_\mathcal{M}\varpi(F)I(Y - Q_\tau < 0)\},
\end{aligned}$$

where the fourth equality is derived from Condition (C1) and the fifth equality uses the linear condition (C2). Thus, by (3), we have

$$\begin{aligned}
&w_k\\
&= E_\tau \mathrm{cqcor}_\tau^2(T, X_k)\\
&= \mathrm{cov}(X_k, X_\mathcal{M}^\top)\beta E_\tau\{(Var(\phi_\tau(Y - Q_\tau)))^{-1}E^{\otimes 2}\big[\beta^\top X_\mathcal{M}\varpi(F)I(Y - Q_\tau < 0)\big]\}\beta^\top \mathrm{cov}(X_\mathcal{M}, X_k).
\end{aligned}$$

By the fact that $\lambda_{\max}(C^\top BC) \le \lambda_{\max}(B)\lambda_{\max}(C^\top C)$ for any matrix $B \ge 0$, it follows

$$\begin{aligned}
\max_{k\in\mathcal{M}^c} w_k \le\ & \lambda_{\max}(E_\tau\{(Var(\phi_\tau(Y - Q_\tau)))^{-1}E^{\otimes 2}\big[\beta^\top X_\mathcal{M}\varpi(F)I(Y - Q_\tau < 0)\big]\})\\
&\times \max_{k\in\mathcal{M}^c}(\mathrm{cov}(X_k, X_\mathcal{M}^\top)\beta\beta^\top \mathrm{cov}(X_\mathcal{M}, X_k)).
\end{aligned} \tag{5}$$

On one hand,

$$\begin{aligned}
&\lambda_{\max}(E_\tau\{(Var(\phi_\tau(Y - Q_\tau)))^{-1}E^{\otimes 2}\big[\beta^\top X_\mathcal{M}\varpi(F)I(Y - Q_\tau < 0)\big]\})\\
&\le \sum_{j=1}^{q} E_\tau\{(Var(\phi_\tau(Y - Q_\tau)))^{-1}E^{\otimes 2}\Big[\beta_j^\top X_\mathcal{M}\varpi(F)I(Y - Q_\tau < 0)\Big]\}\\
&\le \sum_{j=1}^{q} \lambda_{\max}(\mathrm{cov}(X_\mathcal{M}, X_\mathcal{M}^\top)^{-1/2}E_\tau\{(Var(\phi_\tau(Y - Q_\tau)))^{-1}E^{\otimes 2}[X_\mathcal{M}\varpi(F)I(Y - Q_\tau < 0)]\}\\
&\qquad \times \mathrm{cov}(X_\mathcal{M}, X_\mathcal{M}^\top)^{-1/2})\\
&\le q\lambda_{\max}\{\mathrm{cov}(X_\mathcal{M}, X_\mathcal{M}^\top)^{-1}\}\\
&\quad \times \lambda_{\max}(E_\tau\{(Var(\phi_\tau(Y - Q_\tau)))^{-1}E^{\otimes 2}[X_\mathcal{M}\varpi(F)I(Y - Q_\tau < 0)]\})\\
&= q\lambda_{\max}(E_\tau\{(Var(\phi_\tau(Y - Q_\tau)))^{-1}E^{\otimes 2}[X_\mathcal{M}\varpi(F)I(Y - Q_\tau < 0)]\})/\lambda_{\min}\{\mathrm{cov}(X_\mathcal{M}, X_\mathcal{M}^\top)\},
\end{aligned} \tag{6}$$

where the second inequality follows by $\beta^\top \mathrm{cov}(X_\mathcal{M})\beta = I_q$.

On the other hand,

$$\begin{aligned}
&\max_{k\in\mathcal{M}^c}(\mathrm{cov}(X_k, X_\mathcal{M}^\top)\beta\beta^\top \mathrm{cov}(X_\mathcal{M}, X_k))\\
&\le \sum_{j=1}^{q} \max_{k\in\mathcal{M}^c}(\mathrm{cov}(\beta_j^\top X_\mathcal{M}, X_k)\mathrm{cov}(X_k, X_\mathcal{M}^\top\beta_j))\\
&\le \sum_{j=1}^{q} \beta_j^\top(\mathrm{cov}(X_\mathcal{M}, X_{\mathcal{M}^c}^\top)\mathrm{cov}(X_{\mathcal{M}^c}, X_\mathcal{M}^\top))\beta_j\\
&\le q\lambda_{\max}\{\mathrm{cov}(X_\mathcal{M}, X_{\mathcal{M}^c}^\top)\mathrm{cov}(X_{\mathcal{M}^c}, X_\mathcal{M}^\top)\}/\lambda_{\min}\{\mathrm{cov}(X_\mathcal{M}, X_\mathcal{M}^\top)\}.
\end{aligned} \tag{7}$$

By the above inequalities (5)-(7) and condition (C3), we have

$$\begin{aligned}
&\max_{k\in\mathcal{M}^c} w_k \\
&\le q^2\lambda_{\max}(E_\tau\{(Var(\phi_\tau(Y-Q_\tau)))^{-1}E^{\otimes 2}[X_{\mathcal{M}}\varpi(F)I(Y-Q_\tau<0)]\}) \\
&\times\lambda_{\max}\{\mathrm{cov}(X_{\mathcal{M}},X_{\mathcal{M}^c}^\top)\mathrm{cov}(X_{\mathcal{M}^c},X_{\mathcal{M}}^\top)\}/\lambda_{\min^2}\{\mathrm{cov}(X_{\mathcal{M}},X_{\mathcal{M}}^\top)\} \\
&\le \min_{k\in\mathcal{M}} w_k.
\end{aligned}$$

This completes the proof of Theorem 1.

Proof of Theorem 2. We first consider the following $\hat{w}_k$.

$$\begin{aligned}
\hat{w}_k =& \frac{2}{n(n-1)(n-2)}\sum_{j<i<l}^{n}\Bigg\{\frac{1}{\widehat{Var}(\phi_{\tau_j}(Y-Q_{\tau_j}))}X_{ik}X_{lk}\hat{\phi}_{\tau_j}(Y_i-\hat{Q}_{\tau_j})\hat{\phi}_{\tau_j}(Y_l-\hat{Q}_{\tau_j}). \\
&+\frac{1}{\widehat{Var}(\phi_{\tau_i}(Y-Q_{\tau_i}))}X_{jk}X_{lk}\hat{\phi}_{\tau_i}(Y_j-\hat{Q}_{\tau_i})\hat{\phi}_{\tau_i}(Y_l-\hat{Q}_{\tau_i}) \\
&+\frac{1}{\widehat{Var}(\phi_{\tau_l}(Y-Q_{\tau_l}))}X_{ik}X_{jk}\hat{\phi}_{\tau_l}(Y_i-\hat{Q}_{\tau_l})\hat{\phi}_{\tau_l}(Y_j-\hat{Q}_{\tau_l})\Bigg\}.
\end{aligned}$$

In a similar way, we define $\tilde{w}_k$ with $\hat{Q}_{\tau_i}$ and $\hat{F}(\cdot)$ in $\hat{w}_k$ being replaced by Q_{τ_i} and $F(\cdot)$, i.e.,

$$\begin{aligned}
\tilde{w}_k =& \frac{2}{n(n-1)(n-2)}\sum_{j<i<l}^{n}\Bigg\{\frac{1}{Var(\phi_{\tau_j}(Y-Q_{\tau_j}))}X_{ik}X_{lk}\phi_{\tau_j}(Y_i-Q_{\tau_j})\phi_{\tau_j}(Y_l-Q_{\tau_j}) \\
&+\frac{1}{Var(\phi_{\tau_i}(Y-Q_{\tau_i}))}X_{jk}X_{lk}\phi_{\tau_i}(Y_j-Q_{\tau_i})\phi_{\tau_i}(Y_l-Q_{\tau_i}) \\
&+\frac{1}{Var(\phi_{\tau_l}(Y-Q_{\tau_l}))}X_{ik}X_{jk}\phi_{\tau_l}(Y_i-Q_{\tau_l})\phi_{\tau_l}(Y_j-Q_{\tau_l})\Bigg\}.
\end{aligned}$$

Noting that $\omega_k = E(\tilde{w}_k)$, by employing the Markov inequality, we obtain that for any $s>0$,

$$P(\tilde{w}_k - w_k \ge cn^{-\kappa}) \le \exp(-scn^{-\kappa})\exp(-sw_k)E(\exp(s\tilde{w}_k)).$$

Following Serfling (1980), the U statistics $\tilde{w}_k$ can be represented as an average of iid random variables. That is,

$$\tilde{w}_k = (n!)^{-1}\sum_{n!}\varphi(X_{1k},Y_1,\ldots,X_{nk},Y_n)$$

where $\sum_{n!}$ denotes the summation over all possible permutations of $(1,2,\ldots,n)$, and each $\varphi(X_{1k},Y_1,\ldots,X_{nk},Y_n)$ is an average of $m=[n/3]$ iid random variables. Thus, by the Jensen's inequality, we have

$$\begin{aligned}
E(\exp(s\tilde{w}_k)) &= E\left(\exp\left[s\frac{1}{n!}\sum_{n!}\varphi(X_{1k},Y_1,\ldots,X_{nk},Y_n)\right]\right) \\
&\le \frac{1}{n!}\sum_{n!}E(\exp[s\varphi(X_{1k},Y_1,\ldots,X_{nk},Y_n)]) \\
&= E^m\exp(\frac{1}{m}s\varphi(X_{1k},Y_1,\ldots,X_{nk},Y_n)).
\end{aligned}$$

As a result, by Lemma 1 in Li, Zhong, and Zhu (2012), it could be obtained

$$\begin{aligned}
&P(\tilde{w}_k - w_k \ge cn^{-\kappa}) \\
&\le \exp(-scn^{-\kappa})E^m\exp(\frac{1}{m}s(\varphi(X_{1k},Y_1,\ldots,X_{nk},Y_n)-w_k)) \\
&\le \exp(-scn^{-\kappa})\left[\exp((m^{-1}s)^2\tfrac{4M^2}{8})\right]^m \\
&= \exp\left(-scn^{-\kappa}+\frac{M^2s^2}{2m}\right) \\
&= O(\exp(-c'n^{1-2\kappa})),
\end{aligned} \tag{8}$$

where M is an upper bound of the function φ, and the last equation is obtained by choosing $s = cn^{-\kappa}m/M^2$.

It is easy to see that

$$
\begin{aligned}
&P(\hat{w}_k - \tilde{w}_k \geq cn^{-\kappa}) \\
&\leq P\Bigg(\frac{2}{n(n-1)(n-2)} \sum_{j<i<l}^{n} X_{ik}X_{lk} \\
&\quad \times \left[\frac{\hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j})\hat{\phi}_{\tau_j}(Y_l - \hat{Q}_{\tau_j})}{\widehat{Var}(\phi_{\tau_j}(Y - Q_{\tau_j}))} - \frac{\phi_{\tau_j}(Y_i - Q_{\tau_j})\phi_{\tau_j}(Y_l - Q_{\tau_j})}{Var(\phi_{\tau_j}(Y - Q_{\tau_j}))} \right] \geq cn^{-\kappa}/3 \Bigg) \\
&+P\Bigg(\frac{2}{n(n-1)(n-2)} \sum_{j<i<l}^{n} X_{jk}X_{lk} \\
&\quad \times \left[\frac{\hat{\phi}_{\tau_i}(Y_j - \hat{Q}_{\tau_i})\hat{\phi}_{\tau_i}(Y_l - \hat{Q}_{\tau_i})}{\widehat{Var}(\phi_{\tau_i}(Y - Q_{\tau_i}))} - \frac{\phi_{\tau_i}(Y_j - Q_{\tau_i})\phi_{\tau_i}(Y_l - Q_{\tau_i})}{Var(\phi_{\tau_i}(Y - Q_{\tau_i}))} \right] \geq cn^{-\kappa}/3 \Bigg) \\
&+P\Bigg(\frac{2}{n(n-1)(n-2)} \sum_{j<i<l}^{n} X_{ik}X_{jk} \\
&\quad \times \left[\frac{\hat{\phi}_{\tau_l}(Y_i - \hat{Q}_{\tau_l})\hat{\phi}_{\tau_l}(Y_j - \hat{Q}_{\tau_l})}{\widehat{Var}(\phi_{\tau_l}(Y - Q_{\tau_l}))} - \frac{\phi_{\tau_l}(Y_i - Q_{\tau_l})\phi_{\tau_l}(Y_j - Q_{\tau_l})}{Var(\phi_{\tau_l}(Y - Q_{\tau_l}))} \right] \geq cn^{-\kappa}/3 \Bigg) \\
&:= I_1 + I_2 + I_3.
\end{aligned}
$$

For I_1, we have

$$
\begin{aligned}
I_1 &= P\Bigg(\frac{2}{n(n-1)(n-2)} \sum_{j<i<l}^{n} X_{ik}X_{lk} \\
&\quad \times [\frac{\hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j})\hat{\phi}_{\tau_j}(Y_l - \hat{Q}_{\tau_j})}{\widehat{Var}(\phi_{\tau_j}(Y - Q_{\tau_j}))} - \frac{\phi_{\tau_j}(Y_i - Q_{\tau_j})\phi_{\tau_j}(Y_l - Q_{\tau_j})}{Var(\phi_{\tau_j}(Y - Q_{\tau_j}))}] \geq cn^{-\kappa}/3 \Bigg) \\
&= P\Bigg(\frac{2}{n(n-1)(n-2)} \sum_{j<i<l}^{n} X_{ik}X_{lk} \\
&\quad \times \Bigg[\frac{1}{Var(\phi_{\tau_j}(Y - Q_{\tau_j}))} [\hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j})\hat{\phi}_{\tau_j}(Y_l - \hat{Q}_{\tau_j}) - \phi_{\tau_j}(Y_i - Q_{\tau_j})\phi_{\tau_j}(Y_l - Q_{\tau_j})] \\
&\qquad + [\frac{1}{\widehat{Var}(\phi_{\tau_j}(Y - Q_{\tau_j}))} - \frac{1}{Var(\phi_{\tau_j}(Y - Q_{\tau_j}))}]\hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j})\hat{\phi}_{\tau_j}(Y_l - \hat{Q}_{\tau_j})] \Bigg] \geq cn^{-\kappa}/3 \Bigg) \\
&= P\Bigg(\frac{2}{n(n-1)(n-2)} \sum_{j<i<l}^{n} X_{ik}X_{lk} \\
&\quad \times \Bigg[\frac{1}{Var(\phi_{\tau_j}(Y - Q_{\tau_j}))} \hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j})[\hat{\phi}_{\tau_j}(Y_l - \hat{Q}_{\tau_j}) - \phi_{\tau_j}(Y_l - Q_{\tau_j})] \\
&+ \frac{1}{Var(\phi_{\tau_j}(Y - Q_{\tau_j}))} \phi_{\tau_j}(Y_l - Q_{\tau_j})[\hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j}) - \phi_{\tau_j}(Y_i - Q_{\tau_j})] \Bigg] \\
&+ \left[\frac{1}{\widehat{Var}(\phi_{\tau_j}(Y - Q_{\tau_j}))} - \frac{1}{Var(\phi_{\tau_j}(Y - Q_{\tau_j}))}]\hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j})\hat{\phi}_{\tau_j}(Y_l - \hat{Q}_{\tau_j})] \right] \geq cn^{-\kappa}/3 \Bigg) \\
&\leq I_{11} + I_{12} + I_{13},
\end{aligned}
$$

where

$$
\begin{aligned}
&I_{11}\\
&= P\Bigg(\frac{2}{n(n-1)(n-2)} \sum_{j<i<l}^{n} X_{ik}X_{lk} \frac{1}{Var(\phi_{\tau_j}(Y-Q_{\tau_j}))} \\
&\qquad \times \hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j}) \Big[\hat{\phi}_{\tau_j}(Y_l - \hat{Q}_{\tau_j}) - \phi_{\tau_j}(Y_l - Q_{\tau_j}) \Big] \geq cn^{-\kappa}/9 \Bigg) \\
&\leq P\Bigg(\frac{2C}{n(n-1)} \sum_{j<l}^{n} |\hat{\phi}_{\tau_j}(Y_l - \hat{Q}_{\tau_j}) - \phi_{\tau_j}(Y_l - Q_{\tau_j})| \geq cn^{-\kappa}/9 \Bigg) \\
&\leq P\Bigg(\frac{2C}{n(n-1)} \sum_{j<l}^{n} |\varpi(\hat{F}) I(Y_l - \hat{Q}_{\tau_j} < 0) - \varpi(F) I(Y_l - Q_{\tau_j} < 0)| \geq cn^{-\kappa}/9 \Bigg) \\
&= O(n \exp(-c'' n^{1-2\kappa}))
\end{aligned}
$$

with a similar discussion of Lemma 3 in Chen, Chen, and Wang (2018), where C is a positive constant. Similarly, $I_{12} \leq O(n \exp(-c'' n^{1-2\kappa}))$.

Note that

$$
\begin{aligned}
&I_{13}\\
&= P\Bigg(\frac{2}{n(n-1)(n-2)} \sum_{j<i<l}^{n} X_{ik}X_{lk} \\
&\qquad \times \left[\frac{1}{\widehat{Var}(\phi_{\tau_j}(Y-Q_{\tau_j}))} - \frac{1}{Var(\phi_{\tau_j}(Y-Q_{\tau_j}))} \right] \hat{\phi}_{\tau_j}(Y_i - \hat{Q}_{\tau_j}) \hat{\phi}_{\tau_j}(Y_l - \hat{Q}_{\tau_j}) \geq cn^{-\kappa}/9 \Bigg) \\
&\leq P\Bigg(\frac{2C}{n} \sum_{j=1}^{n} \left[\frac{1}{\widehat{Var}(\phi_{\tau_j}(Y-Q_{\tau_j}))} - \frac{1}{Var(\phi_{\tau_j}(Y-Q_{\tau_j}))} \right] \geq cn^{-\kappa}/9 \Bigg) \\
&\leq \sum_{j=1}^{n} P\Bigg(\left[\frac{1}{\widehat{Var}(\phi_{\tau_j}(Y-Q_{\tau_j}))} - \frac{1}{Var(\phi_{\tau_j}(Y-Q_{\tau_j}))} \right] \geq cn^{-\kappa}/9 \Bigg) \\
&\leq O(n \exp(-c'' n^{1-2\kappa}))
\end{aligned}
$$

with a similar discussion with Xu and Huang (2018).

Then, we have $I_1 \leq O(n \exp(-c'' n^{1-2\kappa}))$. Consequently,

$$
P(\hat{w}_k - \tilde{w}_k \geq cn^{-k}) \leq O(n \exp(-c'' n^{1-2\kappa})). \tag{9}
$$

Then, combing (8) and (9), we have

$$
P(\hat{w}_k - w_k \geq cn^{-\kappa}) \leq P(\hat{w}_k - \tilde{w}_k \geq cn^{-\kappa}) + P(\tilde{w}_k - w_k \geq cn^{-\kappa}) \leq O(n \exp(-Cn^{1-2\kappa})).
$$

Similarly, we can prove that

$$
P(\hat{w}_k - w_k \leq -cn^{-\kappa}) \leq O(n \exp(-Cn^{1-2\kappa})).
$$

Therefore, we get

$$
P(\max_{1\leq k\leq p} |\hat{\omega}_k - \omega_k| \geq cn^{-\kappa}) \leq O(np \exp(-Cn^{1-2\kappa})).
$$

This finishes the proof of the first part of Theorem 2. Let's turn to the second part. If $\mathcal{M} \not\subset \hat{\mathcal{M}}$, then there exist some $k \in \mathcal{M}$ such that $\hat{\omega}_k < cn^{-\kappa}$ due to $\min_{k\in\mathcal{M}} \omega_k \geq 2cn^{-\kappa}$. That is

$$
\{\mathcal{M} \not\subset \hat{\mathcal{M}}\} \subset \{|\hat{\omega}_k - \omega_k| \geq cn^{-\kappa} \quad \text{for some } k \in \mathcal{M}\}.
$$

We finally arrive at

$$
\begin{aligned}
&P\{\mathcal{M} \subset \hat{\mathcal{M}}\} \\
&\geq P\left\{ \max_{j \in \mathcal{M}} |\hat{\omega}_k - \omega_k| < cn^{-\kappa} \right\} \\
&= 1 - P\left\{ \max_{j \in \mathcal{M}} |\hat{\omega}_k - \omega_k| \geq cn^{-\kappa} \right\} \\
&\geq 1 - qP\{|\hat{\omega}_k - \omega_k| \geq cn^{-\kappa}\} \\
&\geq 1 - O(nq \exp(-Cn^{1-2\kappa})).
\end{aligned}
$$