A NEW MODEL-FREE FEATURE SCREENING PROCEDURE FOR ULTRAHIGH-DIMENSIONAL INTERVAL-CENSORED FAILURE TIME DATA

Jing Zhang, Mingyue Du, Yanyan Liu* and Jianguo Sun

Zhongnan University of Economics and Law

The Hong Kong Polytechnic University

Wuhan University and University of Missouri

Supplementary Material

In this supplementary material, we will sketch the proof of the sure screening and ranking consistency properties of the proposed screening procedure described in Theorems 1 and 2. Throughout the proof, we will use C_i as a generic notation for some positive constant with an appropriate subscript i.

Proof of Theorem 1

Proof. To prove Theorem 1, we need to show the uniform consistency of the denominator and numerator of $\widehat{\omega}_k$ respectively. Since they have similar form, we only deal with its numerator $\widehat{\mathrm{dcov}}^2(Z_k, \boldsymbol{Y})$ below. For simplicity, we use V_k^2 and $\widehat{V}_k^2$ to represent $\mathrm{dcov}^2(Z_k, \boldsymbol{Y})$ and $\widehat{\mathrm{dcov}}^2(Z_k, \boldsymbol{Y})$.

By the definitions of distance covariance and sample distance covariance, we have

$$V_k^2 = S_{k1} + S_{k2} - 2S_{k3}, \quad \widehat{V}_k^2 = \widehat{S}_{k1} + \widehat{S}_{k2} - 2\widehat{S}_{k3}.$$

We further define

$$\widetilde{V}_k^2 \;\; = \;\; \widetilde{S}_{k1} + \widetilde{S}_{k2} - 2\widetilde{S}_{k3},$$

where $\widetilde{S}_{k1}$, $\widetilde{S}_{k2}$ and $\widetilde{S}_{k3}$ are obtained by replacing $\widehat{\boldsymbol{Y}}_i$ $(i = 1, \ldots, n)$ in $\widehat{S}_{k1}$, $\widehat{S}_{k2}$ and $\widehat{S}_{k3}$ with $\boldsymbol{Y}_i$. By Theorem 1 in Li, Zhong, and Zhu (2012), for any $\epsilon > 0$ and $0 < \gamma < 1/2 - \kappa$, it holds

$$P\left(\left|V_k^2 - \widetilde{V}_k^2\right| \geq \epsilon\right) \leq O\left(\exp(-C_1\epsilon^2 n^{1-2\gamma}) + n\exp(-C_2 n^{\gamma})\right). \tag{S0.1}$$

In order to obtain the exponential tail probability bound of $P(|V_k^2 - \widehat{V}_k^2| \geq \epsilon)$, it is sufficient to compute $P(|\widehat{V}_k^2 - \widetilde{V}_k^2| \geq \epsilon)$. It follows from the definition of $\widehat{V}_k^2$ and $\widetilde{V}_k^2$ that

$$\left|\widehat{V}_k^2 - \widetilde{V}_k^2\right| \leq \left|\widehat{S}_{k1} - \widetilde{S}_{k1}\right| + \left|\widehat{S}_{k2} - \widetilde{S}_{k2}\right| + 2\left|\widehat{S}_{k3} - \widetilde{S}_{k3}\right|. \tag{S0.2}$$

We first deal with $|\widehat{S}_{k1} - \widetilde{S}_{k1}|$. Some straightforward calculations entail

that

$$\begin{aligned}
\left|\widehat{S}_{k1}-\widetilde{S}_{k1}\right| &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|Z_{ki}-Z_{kj}|\left(\|\widehat{\boldsymbol{Y}}_i-\widehat{\boldsymbol{Y}}_j\|_2-\|\boldsymbol{Y}_i-\boldsymbol{Y}_j\|_2\right)\\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(|Z_{ki}|+|Z_{kj}|)\Big(\Psi_{ij}(\widehat{\sigma}_1,\widehat{\sigma}_1;\widehat{\sigma}_2,\widehat{\sigma}_2)-\Psi_{ij}(\sigma_1,\sigma_1;\sigma_2,\sigma_2)\Big)\\
&\leq \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(|Z_{ki}|+|Z_{kj}|)\Big(\Psi_{ii}(\widehat{\sigma}_1,\sigma_1;\widehat{\sigma}_2,\sigma_2)+\Psi_{jj}(\widehat{\sigma}_1,\sigma_1;\widehat{\sigma}_2,\sigma_2)\Big),
\end{aligned}$$

where

$$\Psi_{ij}(\widehat{\sigma}_1,\sigma_1;\widehat{\sigma}_2,\sigma_2)=\sqrt{\left(\frac{L_i}{\widehat{\sigma}_1}-\frac{L_j}{\sigma_1}\right)^2+\left(\frac{H_i}{\widehat{\sigma}_2}-\frac{H_j}{\sigma_2}\right)^2},$$

and the last inequality is obtained by using the triangle inequality. As we all know, for any events Ω_1 and Ω_2, it holds

$$P(\Omega_1)=P(\Omega_1|\Omega_2)P(\Omega_2)+P(\Omega_1|\Omega_2^{\mathrm{c}})P(\Omega_2^{\mathrm{c}})\leq P(\Omega_1|\Omega_2)+P(\Omega_2^{\mathrm{c}}). \quad \text{(S0.3)}$$

Define events $A_i=\{|Z_{ki}|<n^{\gamma}\}$ $(i=1,\ldots,n)$, by (S0.3), condition C1, Bonferroni's inequality and Markov's inequality, we obtain that

$$\begin{aligned}
& P\left(|\widehat{S}_{k1}-\widetilde{S}_{k1}|\geq\epsilon\right)\\
\leq\ & P\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(|Z_{ki}|+|Z_{kj}|)\big(\Psi_{ii}(\widehat{\sigma}_1,\sigma_1;\widehat{\sigma}_2,\sigma_2)+\Psi_{jj}(\widehat{\sigma}_1,\sigma_1;\widehat{\sigma}_2,\sigma_2)\big)\geq\epsilon\,\middle|\bigcap_{i=1}^{n}A_i\right)\\
& +P\left(\overline{\bigcap_{i=1}^{n}A_i}\right)\\
\leq\ & P\left(4n^{\gamma}\cdot\frac{1}{n}\sum_{i=1}^{n}\Psi_{ii}(\widehat{\sigma}_1,\sigma_1;\widehat{\sigma}_2,\sigma_2)\geq\epsilon\right)+C_3n\exp(-n^{\gamma})\\
\leq\ & P\left(\frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{L_i}{\widehat{\sigma}_1}-\frac{L_i}{\sigma_1}\right|+\left|\frac{H_i}{\widehat{\sigma}_2}-\frac{H_i}{\sigma_2}\right|\right)\geq\frac{\epsilon}{4n^{\gamma}}\right)+C_3n\exp(-n^{\gamma}). \qquad \text{(S0.4)}
\end{aligned}$$

Straightforward calculations yield

$$\begin{aligned}
& P\left(\frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{L_i}{\widehat{\sigma}_1}-\frac{L_i}{\sigma_1}\right|+\left|\frac{H_i}{\widehat{\sigma}_2}-\frac{H_i}{\sigma_2}\right|\right)\geq\frac{\epsilon}{4n^{\gamma}}\right) \\
\leq\ & P\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{L_i}{\widehat{\sigma}_1}-\frac{L_i}{\sigma_1}\right|\geq\frac{\epsilon}{8n^{\gamma}}\right)+P\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{H_i}{\widehat{\sigma}_2}-\frac{H_i}{\sigma_2}\right|\geq\frac{\epsilon}{8n^{\gamma}}\right) \\
\leq\ & P\left(\left|\frac{1}{\widehat{\sigma}_1}-\frac{1}{\sigma_1}\right|\cdot\frac{1}{n}\sum_{i=1}^{n}L_i\geq\frac{\epsilon}{8n^{\gamma}}\right)+P\left(\left|\frac{1}{\widehat{\sigma}_2}-\frac{1}{\sigma_2}\right|\cdot\frac{1}{n}\sum_{i=1}^{n}H_i\geq\frac{\epsilon}{8n^{\gamma}}\right) \\
\equiv\ & G_1+G_2. \qquad \text{(S0.5)}
\end{aligned}$$

We first consider G_1. Defining events $B_1=\{|\widehat{\sigma}_1-\sigma_1|\leq\sigma_1/2\}$ and $B_2=\{|n^{-1}\sum_{i=1}^{n}L_i-\mu_1|\leq\mu_1/2\}$, then we have

$$\begin{aligned}
G_1 &= P\left(\left|\frac{1}{\widehat{\sigma}_1}-\frac{1}{\sigma_1}\right|\cdot\frac{1}{n}\sum_{i=1}^{n}L_i\geq\frac{\epsilon}{8n^{\gamma}}\right) \\
&\leq P\left(\left|\frac{1}{\widehat{\sigma}_1}-\frac{1}{\sigma_1}\right|\cdot\frac{1}{n}\sum_{i=1}^{n}L_i\geq\frac{\epsilon}{8n^{\gamma}}\middle|B_1,B_2\right)+P(\overline{B_1B_2}) \\
&\leq P\left(|\widehat{\sigma}_1^2-\sigma_1^2|\geq C_4n^{-\gamma}\epsilon\right)+P\left(\overline{B_1}\right)+P\left(\overline{B_2}\right) \\
&\equiv G_3+P\left(\overline{B_1}\right)+P\left(\overline{B_2}\right).
\end{aligned}$$

Since

$$|\widehat{\sigma}_1^2-\sigma_1^2|=\left|\frac{1}{n}\sum_{i=1}^{n}(L_i-\widehat{\mu}_1)^2-\sigma_1^2\right|\leq\left|\frac{1}{n}\sum_{i=1}^{n}(L_i-\mu_1)^2-\sigma_1^2\right|+|\widehat{\mu}_1-\mu_1|^2,$$

G_3 can be deduced as

$$\begin{aligned}
G_3 &\leq P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(L_i-\mu_1)^2-\sigma_1^2\right|\geq\frac{C_4n^{-\gamma}\epsilon}{2}\right)+P\left(|\widehat{\mu}_1-\mu_1|^2\geq\frac{C_4n^{-\gamma}\epsilon}{2}\right) \\
&\leq 2\exp(-C_5\epsilon^2n^{1-2\gamma})+2\exp(-C_6\epsilon n^{1-\gamma}) \\
&\leq 2\exp(-C_7\epsilon^2n^{1-2\gamma}).
\end{aligned}$$

Combing the fact

$$\begin{aligned} P(B_1^c) &= P\left(|\widehat{\sigma}_1 - \sigma_1| \geq \sigma/2\right) \leq P\left(|\widehat{\sigma}_1^2 - \sigma_1^2| \geq \sigma^2/2\right) \leq 2\exp(-C_8 n), \\ P(B_2^c) &= P\left(\left|\frac{1}{n}\sum_{i=1}^{n} L_i - \mu_1\right| \geq \mu_1/2\right) \leq 2\exp(-C_8 n), \end{aligned}$$

we have

$$\begin{aligned} G_1 &\leq 2\exp(-C_7\epsilon^2 n^{1-2\gamma}) + 2\exp(-C_8 n) + 2\exp(-C_8 n) \\ &= O\left(\exp(-C_9\epsilon^2 n^{1-2\gamma})\right). \end{aligned} \tag{S0.6}$$

Using the similar arguments, we can also conclude that

$$G_2 = O\left(\exp(-C_{10}\epsilon^2 n^{1-2\gamma})\right). \tag{S0.7}$$

Following (S0.4)–(S0.7), we have

$$\begin{aligned} & P\left(\left|\widehat{S}_{k1} - \widetilde{S}_{k1}\right| \geq \epsilon\right) \\ \leq\ & O\left(\exp(-C_9\epsilon^2 n^{1-2\gamma})\right) + O\left(\exp(-C_{10}\epsilon^2 n^{1-2\gamma})\right) + C_3 n\exp(-n^\gamma) \\ =\ & O\left(\exp(-C_{11}\epsilon^2 n^{1-2\gamma}) + n\exp(-n^\gamma)\right). \end{aligned}$$

Utilizing the analogous arguments, along with some tedious calculations, we can draw the similar conclusions of $P(|\widehat{S}_{k2} - \widetilde{S}_{k2}| \geq \epsilon)$ and $P(|\widehat{S}_{k3} - \widetilde{S}_{k3}| \geq \epsilon)$. Consequently,

$$\begin{aligned} & P\left(\left|\widehat{V}_k^2 - \widetilde{V}_k^2\right| \geq \epsilon\right) \\ \leq\ & P\left(\left|\widehat{S}_{k1} - \widetilde{S}_{k1}\right| \geq \epsilon/4\right) + P\left(\left|\widehat{S}_{k2} - \widetilde{S}_{k2}\right| \geq \epsilon/4\right) + P\left(\left|\widehat{S}_{k3} - \widetilde{S}_{k3}\right| \geq \epsilon/4\right) \\ =\ & O\left(\exp(-C_{12}\epsilon^2 n^{1-2\gamma}) + n\exp(-n^\gamma)\right). \end{aligned} \tag{S0.8}$$

Set $\epsilon = cn^{-\kappa}$, where κ satisfies $0 < \kappa + \gamma < 1/2$. By (S0.8), there exist some positive constants c_1 and c_2 such that

$$\begin{aligned}
& P\left(\left|V_k^2 - \widehat{V}_k^2\right| \geq cn^{-\kappa}\right) \\
\leq\ & P\left(\left|V_k^2 - \widetilde{V}_k^2\right| \geq cn^{-\kappa}/2\right) + P\left(\left|\widetilde{V}_k^2 - \widehat{V}_k^2\right| \geq cn^{-\kappa}/2\right) \\
\leq\ & O\left(\exp\left\{-C_{13}n^{1-2(\kappa+\gamma)}\right\} + n\exp(-C_{14}n^{\gamma})\right) + O\left(\exp\left\{-C_{15}n^{1-2(\kappa+\gamma)}\right\} + n\exp(-n^{\gamma})\right) \\
=\ & O\left(\exp\left\{-c_1 n^{1-2(\kappa+\gamma)}\right\} + n\exp(-c_2 n^{\gamma})\right). \qquad \text{(S0.9)}
\end{aligned}$$

The convergence rate of the numerator of $\widehat{\omega}_k$ is now achieved. Following similar arguments, we can obtain the same convergence rate of the denominator. Therefore,

$$\begin{aligned}
& P\left(\max_{1\leq k\leq p} |\widehat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right) \\
\leq\ & p \max_{1\leq k\leq p} P\left(|\widehat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right) \\
=\ & O\left(p\left[\exp\left\{-c_1 n^{1-2(\kappa+\gamma)}\right\} + n\exp(-c_2 n^{\gamma})\right]\right). \qquad \text{(S0.10)}
\end{aligned}$$

The first part of Theorem 1 is proved.

We now consider the second part of Theorem 1. If $\mathcal{A} \not\subseteq \widehat{\mathcal{A}}$, then there must exist some $k \in \mathcal{A}$ such that $\widehat{\omega}_k < cn^{-\kappa}$. It follows from condition C2 that $|\widehat{\omega}_k - \omega_k| > cn^{-\kappa}$ for some $k \in \mathcal{A}$, which implies that $\{\mathcal{A} \not\subseteq \widehat{\mathcal{A}}\} \subseteq \{|\widehat{\omega}_k - \omega_k| > cn^{-\kappa}$ for some $k \in \mathcal{A}\}$. As a result, $\{\max_{k\in\mathcal{A}} |\widehat{\omega}_k - \omega_k| \leq$

$cn^{-\kappa}\} \subseteq \{\mathcal{A} \subseteq \widehat{\mathcal{A}}\}$. Using (S0.10), we have

$$\begin{aligned} P\left(\mathcal{A} \subseteq \widehat{\mathcal{A}}\right) &\geq P\left(\max_{k\in\mathcal{A}} \left|\widehat{\omega}_k - \omega_k\right| \leq cn^{-\kappa}\right) \\ &\geq 1 - O\left(a\left[\exp\left\{-c_1 n^{1-2(\kappa+\gamma)}\right\} + n\exp(-c_2 n^{\gamma})\right]\right), \end{aligned}$$

where $a = |\mathcal{A}|$. Thus, the proof of Theorem 1 is completed. □

Proof of Theorem 2

Proof. Under condition C2, noting that $\min_{k\in\mathcal{A}} \omega_k \geq 2cn^{-\kappa}$ and combining (S0.10), we have

$$\begin{aligned} P\left(\min_{k\in\mathcal{A}} \widehat{\omega}_k \leq \max_{k\notin\mathcal{A}} \widehat{\omega}_k\right) &= P\left(\max_{k\notin\mathcal{A}} \widehat{\omega}_k - \max_{k\notin\mathcal{A}} \omega_k - \min_{k\in\mathcal{A}} \widehat{\omega}_k + \min_{k\in\mathcal{A}} \omega_k \geq \min_{k\in\mathcal{A}} \omega_k\right) \\ &\leq P\left(\max_{k\notin\mathcal{A}} |\widehat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right) + P\left(\max_{k\in\mathcal{A}} |\widehat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right) \\ &\leq 2P\left(\max_{1\leq k\leq p} |\widehat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right) \\ &= O\left(p\left[\exp\left\{-c_1 n^{1-2(\kappa+\gamma)}\right\} + n\exp(-c_2 n^{\gamma})\right]\right), \end{aligned}$$

leads to

$$P\left(\max_{k\notin\mathcal{A}} \widehat{\omega}_k < \min_{k\in\mathcal{A}} \widehat{\omega}_k\right) \geq 1 - O\left(p\left[\exp\left\{-c_1 n^{1-2(\kappa+\gamma)}\right\} + n\exp(-c_2 n^{\gamma})\right]\right),$$

which completes the proof of Theorem 2. □

References

Li, R., Zhong, W. and Zhu, L. (2012). Feature screening via distance correlation learning. *J. Am. Statist. Assoc.* **107**, 1129–1139.