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 $\hat{\omega}_k$ respectively. Since they have similar form, we only deal with its numerator $\text{dco}v(Z_k, Y)$ below. For simplicity,
we use $V^2_k$ and $\tilde{V}^2_k$ to represent $\text{dcov}^2(Z_k, Y)$ and $\hat{\text{dcov}}^2(Z_k, Y)$.

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

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

We further define

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

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

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

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

$$\left| \tilde{V}^2_k - \hat{V}^2_k \right| \leq \left| \tilde{S}_{k1} - \hat{S}_{k1} \right| + \left| \tilde{S}_{k2} - \hat{S}_{k2} \right| + 2 \left| \tilde{S}_{k3} - \hat{S}_{k3} \right|. \quad (S0.2)$$
We first deal with $|\tilde{S}_{k1} - \tilde{S}_{k1}|$. Some straightforward calculations entail that

$$\left| \tilde{S}_{k1} - \tilde{S}_{k1} \right| = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |Z_{ki} - Z_{kj}| \left( \| \hat{Y}_i - \hat{Y}_j \|_2 - \| Y_i - Y_j \|_2 \right)$$

$$\leq \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (|Z_{ki}| + |Z_{kj}|) \left( \Psi_{ij}(\sigma_1, \sigma_1; \sigma_2, \sigma_2) - \Psi_{ij}(\sigma_1, \sigma_1; \sigma_2, \sigma_2) \right)$$

$$\leq \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (|Z_{ki}| + |Z_{kj}|) \left( \Psi_{ii}(\sigma_1, \sigma_1; \sigma_2, \sigma_2) + \Psi_{jj}(\sigma_1, \sigma_1; \sigma_2, \sigma_2) \right),$$

where

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

and the last inequality is obtained by using the triangle inequality. As we all know, for any events $\Omega_1$ and $\Omega_2$, it holds

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

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

\[
P\left( |\tilde{S}_{k1} - \tilde{S}_{k1}| \geq \epsilon \right) \\
\leq P\left( \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (|Z_{ki}| + |Z_{kj}|) \left( \Psi_{ii}(\tilde{\sigma}_1, \sigma_1; \tilde{\sigma}_2, \sigma_2) + \Psi_{jj}(\tilde{\sigma}_1, \sigma_1; \tilde{\sigma}_2, \sigma_2) \right) \geq \epsilon \right) \\
\quad + P\left( \bigcap_{i=1}^{n} A_i \right) \\
\leq P\left( 4n^\gamma \cdot \frac{1}{n} \sum_{i=1}^{n} \Psi_{ii}(\tilde{\sigma}_1, \sigma_1; \tilde{\sigma}_2, \sigma_2) \geq \epsilon \right) + C_3 n \exp(-n^\gamma) \\
\leq P\left( \frac{1}{n} \sum_{i=1}^{n} \left( \frac{L_i}{\tilde{\sigma}_1} - \frac{L_i}{\sigma_1} + \frac{H_i}{\tilde{\sigma}_2} - \frac{H_i}{\sigma_2} \right) \geq \frac{\epsilon}{4n^\gamma} \right) + C_3 n \exp(-n^\gamma).
\]

(S0.4)

Straightforward calculations yield

\[
P\left( \frac{1}{n} \sum_{i=1}^{n} \left( \frac{L_i}{\tilde{\sigma}_1} - \frac{L_i}{\sigma_1} + \frac{H_i}{\tilde{\sigma}_2} - \frac{H_i}{\sigma_2} \right) \geq \frac{\epsilon}{4n^\gamma} \right) \\
\leq P\left( \frac{1}{n} \sum_{i=1}^{n} \left| \frac{L_i}{\tilde{\sigma}_1} - \frac{L_i}{\sigma_1} \right| \geq \frac{\epsilon}{8n^\gamma} \right) + P\left( \frac{1}{n} \sum_{i=1}^{n} \left| \frac{H_i}{\tilde{\sigma}_2} - \frac{H_i}{\sigma_2} \right| \geq \frac{\epsilon}{8n^\gamma} \right) \\
\leq P\left( \left| \frac{1}{\tilde{\sigma}_1} - \frac{1}{\sigma_1} \right| \cdot \frac{1}{n} \sum_{i=1}^{n} L_i \geq \frac{\epsilon}{8n^\gamma} \right) + P\left( \left| \frac{1}{\tilde{\sigma}_2} - \frac{1}{\sigma_2} \right| \cdot \frac{1}{n} \sum_{i=1}^{n} H_i \geq \frac{\epsilon}{8n^\gamma} \right) \\
\equiv G_1 + G_2.
\]

(S0.5)

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

\[
G_1 = P \left( \left| \frac{1}{\hat{\sigma}_1} - \frac{1}{\sigma_1} \right| \cdot \frac{1}{n} \sum_{i=1}^{n} L_i \geq \frac{\epsilon}{8n} \right) \\
\leq P \left( \left| \frac{1}{\hat{\sigma}_1} - \frac{1}{\sigma_1} \right| \cdot \frac{1}{n} \sum_{i=1}^{n} L_i \geq \frac{\epsilon}{8n} \mid B_1, B_2 \right) + P(B_1 \cap B_2) \\
\leq P \left( |\hat{\sigma}_1^2 - \sigma_1^2| \geq C_4 n^{-\gamma} \epsilon \right) + P(B_1) + P(B_2) \\
\equiv G_3 + P(B_1) + P(B_2).
\]

Since

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

\(G_3\) can be deduced as

\[
G_3 \leq P \left( \left| \frac{1}{n} \sum_{i=1}^{n} (L_i - \mu_1)^2 - \sigma_1^2 \right| \geq C_4 n^{-\gamma} \epsilon \right) + P \left( |\hat{\mu}_1 - \mu_1|^2 \geq C_4 n^{-\gamma} \epsilon \right) \\
\leq 2 \exp(-C_5 \epsilon^2 n^{1-2\gamma}) + 2 \exp(-C_6 \epsilon n^{1-\gamma}) \\
\leq 2 \exp(-C_7 \epsilon^2 n^{1-2\gamma}).
\]
Combing the fact

\[ P(B^c_1) = P(|\hat{\sigma}_1 - \sigma_1| \geq \sigma/2) \leq P(|\hat{\sigma}_1^2 - \sigma_1^2| \geq \sigma^2/2) \leq 2 \exp(-C_8 n), \]

\[ P(B_2^c) = P \left( \left| \frac{1}{n} \sum_{i=1}^{n} L_i - \mu_1 \right| \geq \mu_1/2 \right) \leq 2 \exp(-C_8 n), \]

we have

\[ G_1 \leq 2 \exp(-C_7 \epsilon^2 n^{1-2\gamma}) + 2 \exp(-C_8 n) + 2 \exp(-C_8 n) \]

\[ = O \left( \exp(-C_9 \epsilon^2 n^{1-2\gamma}) \right). \tag{S0.6} \]

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

\[ P \left( \left| \tilde{S}_{k_1} - \tilde{S}_{k_1} \right| \geq \epsilon \right) \]

\[ \leq O \left( \exp(-C_9 \epsilon^2 n^{1-2\gamma}) \right) + O \left( \exp(-C_{10} \epsilon^2 n^{1-2\gamma}) \right) + C_3 n \exp(-n^\gamma) \]

\[ = O \left( \exp(-C_{11} \epsilon^2 n^{1-2\gamma}) + n \exp(-n^\gamma) \right). \]

Utilizing the analogous arguments, along with some tedious calculations, we
FEATURE SCREENING FOR ULTRAHIGH-DIMENSIONAL IC DATA

can draw the similar conclusions of $P(\left| \widehat{S}_{k2} - \widetilde{S}_{k2} \right| \geq \epsilon)$ and $P(\left| \widehat{S}_{k3} - \widetilde{S}_{k3} \right| \geq \epsilon)$.

Consequently,

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

(S0.8)

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

$$
P \left( \left| V_k^2 - \widehat{V}_k^2 \right| \geq cn^{-\kappa} \right) \\
\leq P \left( \left| V_k^2 - \widehat{V}_k^2 \right| \geq cn^{-\kappa}/2 \right) + P \left( \left| \widehat{V}_k^2 - \widetilde{V}_k^2 \right| \geq cn^{-\kappa}/2 \right) \\
\leq O \left( \exp \left\{-C_{13}n^{1-2(\kappa+\gamma)}\right\} + n \exp(-C_{14}n^{\gamma}) \right) + O \left( \exp \left\{-C_{15}n^{1-2(\kappa+\gamma)}\right\} + n \exp(-n^{\gamma}) \right) \\
= O \left( \exp \left\{-c_1n^{1-2(\kappa+\gamma)}\right\} + n \exp(-c_2n^{\gamma}) \right).
$$

(S0.9)

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 denom-
inator. Therefore,

\[ P \left( \max_{1 \leq k \leq p} |\hat{\omega}_k - \omega_k| \geq cn^{-\kappa} \right) \]
\[ \leq \ p \ \max_{1 \leq k \leq p} P \left( |\hat{\omega}_k - \omega_k| \geq cn^{-\kappa} \right) \]
\[ = \ O \left( p \left[ \exp \left\{ -c_1 n^{1-2(\kappa+\gamma)} \right\} + n \exp(-c_2 n^\gamma) \right]\right). \quad (S0.10) \]

The first part of Theorem 1 is proved.

We now consider the second part of Theorem 1. If \( \mathcal{A} \not\subseteq \hat{\mathcal{A}} \), then there must exist some \( k \in \mathcal{A} \) such that \( \hat{\omega}_k < cn^{-\kappa} \). It follows from condition C2 that \( |\hat{\omega}_k - \omega_k| > cn^{-\kappa} \) for some \( k \in \mathcal{A} \), which implies that \( \{ \mathcal{A} \not\subseteq \hat{\mathcal{A}} \} \subseteq \{ |\hat{\omega}_k - \omega_k| > cn^{-\kappa} \ \text{for some} \ k \in \mathcal{A} \} \). As a result, \( \{ \max_{k \in \mathcal{A}} |\hat{\omega}_k - \omega_k| \leq cn^{-\kappa} \} \subseteq \{ \mathcal{A} \subseteq \hat{\mathcal{A}} \} \). Using \((S0.10)\), we have

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

where \( a = |\mathcal{A}| \). Thus, the proof of Theorem 1 is completed. \( \square \)
**Proof of Theorem 2**

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

\[
P(\min_{k \in A} \hat{\omega}_k \leq \max_{k \notin A} \hat{\omega}_k) = P\left(\max_{k \notin A} \hat{\omega}_k - \max_{k \in A} \omega_k - \min_{k \in A} \hat{\omega}_k + \min_{k \in A} \omega_k \geq \min_{k \in A} \omega_k\right)
\]
\[
\leq P\left(\max_{k \notin A} |\hat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right) + P\left(\max_{k \in A} |\hat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right)
\]
\[
\leq 2P\left(\max_{1 \leq k \leq p} |\hat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right)
\]
\[
= O\left(p \left[\exp\{-c_1n^{1-2(\kappa+\gamma)}\} + n \exp(-c_2n^\gamma)\}\right],
\]

leads to

\[
P\left(\max_{k \notin A} \hat{\omega}_k < \min_{k \in A} \hat{\omega}_k\right) \geq 1 - O\left(p \left[\exp\{-c_1n^{1-2(\kappa+\gamma)}\} + n \exp(-c_2n^\gamma)\}\right),
\]

which completes the proof of Theorem 2.

\[\square\]

**References**

School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, Hubei 430073, China

E-mail: jing66@zuel.edu.cn

Department of Mathematics and Statistics, The Hong Kong Polytechnic University, Hong Kong, China

E-mail: dummoon@163.com

School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China

E-mail: liuyy@whu.edu.cn

Department of Statistics, University of Missouri, Columbia, Missouri 65211, U.S.A

E-mail: sunj@missouri.edu