To appear in the *Journal of Statistical Computation and Simulation*
Vol. 00, No. 00, Month 20XX, 1–9

# Supplementary Material for *Ultrahigh Dimensional Feature Screening for Additive Model with Multivariate Response*

Shishi Liu, Xiangjie Li and Jingxiao Zhang*

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

## 1. Proofs

We show all the proofs in this supplementary material. Recall some notations firstly.

- The multivariate response is $\boldsymbol{y} = (y_1, \ldots, y_q)^\top$, and denote the sample matrix of $\boldsymbol{y}$ as $\mathbf{Y} = (Y_1, \ldots, Y_q)$, where $Y_k = (Y_{1k}, \ldots, Y_{nk})^\top$, $k = 1, \ldots, q$.
- Denote $\mathbb{P}_n y_m y_k = \frac{1}{n} Y_m^\top Y_k$ as the sample average of $\{Y_{im} Y_{ik}\}_{i=1}^n$. Similarly, $\mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top = \frac{1}{n}\mathbf{Y}^\top\mathbf{Y}$ is the sample average of $\boldsymbol{y}\boldsymbol{y}^\top$.
- Correspondingly, $Ey_m y_k$ and $E\boldsymbol{y}\boldsymbol{y}^\top$ are the expectation of $y_m y_k$ and $\boldsymbol{y}\boldsymbol{y}^\top$ respectively.
- For a matrix $\mathbf{A}$, $\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{A}\mathbf{A}^\top)}$, $\|\mathbf{A}\|_\infty = \max_{i,j} |A_{ij}|$.
- $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ denote the minimum and maximum eigenvalue of matrix $\mathbf{A}$.

Lemma 1.1 *(Hoeffding's inequality, [4]) Let $Z_1, \ldots, Z_n$ be independent random variables. Assume that $P(Z_i \in [a_i, b_i]) = 1$ for $1 \le i \le n$, where $a_i$ and $b_i$ are constants. Let $\bar{Z} = \sum_{i=1}^n Z_i/n$. Then for every $x > 0$,*

$$P(|\bar{Z} - E(\bar{Z})| \ge x) \le 2\exp\left(-\frac{2n^2x^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Lemma 1.2 *(Liu et al., [3]) Let $Z$ be a random variable with $E\exp(a|Z|) < \infty$ for some $a > 0$. Then for any $M > 0$, there exist positive constants $b$ and $c$ such that*

$$P(|Z| \ge M) \le b\exp(-cM),$$

*where $b > 0$ such that $E\exp(a|Z|) \le b/2$, and $c = a$.*

Lemma 1.3 *(Li et al., [2]) Under Condition (C8), for $\varepsilon > 0$, there exist some positive constants $0 < \alpha < 1/2$, $m_1$ and $s_1$ such that*

$$P(|\mathbb{P}_n y_m y_k - Ey_m y_k| \ge \varepsilon) \le 2\exp(-\frac{1}{2}n^{1-2\alpha}\varepsilon^2) + nm_1\exp(-s_1 n^\alpha).$$

*Proof.* See proof of Lemma 3 in [2]. ■

*Corresponding author. Email: zhjxiaoruc@163.com

Lemma 1.4 *(Li et al., [2]) Under Condition (C7), for any $\varepsilon > 0$,*

$$P\left(\left|\lambda_{\min}(\mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top) - \lambda_{\min}(E\boldsymbol{y}\boldsymbol{y}^\top)\right| \geq \varepsilon\right) \leq q_n^2 \left[2\exp\left(-\frac{n^{1-2\alpha}\varepsilon^2}{2q_n^2}\right) + nm_1\exp(-s_1 n^\alpha)\right].$$

*In addition, there exist some positive constants $c_1$ and $c_2$ such that*

$$P\left(\left|\|(\mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top)^{-1}\| - \|(E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\|\right| \geq c_1\|(E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\|\right) \leq q_n^2\left[2\exp(-c_2 n^{1-2\alpha}q_n^{-4}) + nm_1\exp(-s_1 n^\alpha)\right].$$

*Proof.* See proof of Lemma 4 in [2]. ■

Lemma 1.5 *(Bernstein's inequality, [4]) For independent random variables $Z_1, \dots, Z_n$ with bounded ranges $[-M, M]$ and zero means,*

$$P(|Z_1 + \cdots + Z_n| > x) \leq 2\exp\{-\frac{x^2}{2(v + Mx/3)}\},$$

*for $v \geq var(Z_1 + \cdots + Z_n)$.*

Lemma 1.6 *(Bernstein's inequality, [4]) Let $Z_1, \dots, Z_n$ be independent random variables with zero mean such that $E|Z_i|^m \leq m!M^{m-2}v_i/2$, for every $m \geq 2$ and some constants $M$ and $v_i$. Then*

$$P(|Z_1 + \cdots + Z_n| > x) \leq 2\exp\{-\frac{x^2}{2(v + Mx)}\},$$

*for $v \geq v_1 + \cdots + v_n$.*

Lemma 1.7 *Under Conditions (C1), (C2), (C5) and (C6), for any $\delta > 0$, there exist some positive constants $c_4$ and $c_5$ such that*

$$P\left(|\mathbb{P}_n \psi_{jm} y_k - E\psi_{jm} y_k| \geq \delta\right) \leq 4\exp\left(-\frac{n\delta^2}{2(c_4 d_n^{-1} + c_5\delta)}\right),$$

*for $m = 1, \dots, d_n$, $j = 1, \dots, p_n$.*

*Proof.* See proof of Lemma 4 in [1]. ■

Lemma 1.8 *Under Conditions (C1) and (C2), for any $\delta > 0$,*

$$P\left(\left|\lambda_{\min}(\mathbb{P}_n \Psi_j \Psi_j^\top) - \lambda_{\min}(E\Psi_j\Psi_j^\top)\right| \geq \delta\right) \leq 2d_n^2\exp\left(-\frac{n\delta^2}{2d_n(D_5 + \delta)}\right).$$

*In addition, for any given constant $c_7$, there exists some positive constant $c_6$ such that*

$$P\left(\left|\|(\mathbb{P}_n \Psi_j\Psi_j^\top)^{-1}\| - \|(E\Psi_j\Psi_j^\top)^{-1}\|\right| \geq c_6\|(E\Psi_j\Psi_j^\top)^{-1}\|\right) \leq 2d_n^2\exp(-c_7 n d_n^{-3}).$$

*Proof.* See proof of Lemma 5 in [1]. ■

Lemma 1.9 *Suppose that Conditions (C1), (C2), (C5) and (C6) hold. For any $\delta > 0$, there exist some positive constants $c_9 \sim c_{13}$, such that*

$$P\left(\left|\mathbb{P}_n \hat{f}_{nkj} y_k - E f_{nkj} y_k\right| \geq c_9 d_n^2 \delta^2 + c_{10} d_n^{3/2} \delta + c_{11} d_n^2 \delta\right)$$
$$\leq (8d_n + 2d_n^2) \exp\left(-\frac{n\delta^2}{2(c_{12} d_n^{-1} + c_{13}\delta)}\right) + 6d_n^2 \exp\left(-c_7 n d_n^{-3}\right).$$

*Proof.* See part (i) of proof of Theorem 1 in [1]. ■

**Proof of Theorem 2.1**:

*Proof.* We first show the first inequality. Recall that

$$\omega_j = (Ef_{n1j}y_1, \ldots, Ef_{nqj}y_q)(E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}(Ef_{n1j}y_1, \ldots, Ef_{nqj}y_q)^\top \triangleq \boldsymbol{\phi}_j^\top (E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\boldsymbol{\phi}_j,$$
$$\hat{\omega}_j = (\mathbb{P}_n\hat{f}_{n1j}y_1, \ldots, \mathbb{P}_n\hat{f}_{nqj}y_q)(\mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top)^{-1}(\mathbb{P}_n\hat{f}_{n1j}y_1, \ldots, \mathbb{P}_n\hat{f}_{nqj}y_q)^\top \triangleq \hat{\boldsymbol{\phi}}_j^\top (\mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\hat{\boldsymbol{\phi}}_j.$$

Similar to the proof of Lemma 1.9, we divide $\hat{\omega}_j - \omega_j$ into three parts,

$$\hat{\omega}_j - \omega_j =: S_1 + S_2 + S_3,$$

where

$$\begin{aligned} S_1 &= (\hat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j)^\top (\mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top)^{-1}(\hat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j), \\ S_2 &= 2(\hat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j)^\top (\mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\boldsymbol{\phi}_j, \\ S_3 &= \boldsymbol{\phi}_j^\top [(\mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top)^{-1} - (E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}]\boldsymbol{\phi}_j \\ &= \boldsymbol{\phi}_j^\top (\mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top)^{-1}[E\boldsymbol{y}\boldsymbol{y}^\top - \mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top](E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\boldsymbol{\phi}_j. \end{aligned}$$

Note that

$$|S_1| \leq \|(\mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\| \cdot \|\hat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j\|^2. \tag{1}$$

By Lemma 1.4, we have

$$P\left(\left|\|(\mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\| - \|(E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\|\right| \geq c_1\|(E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\|\right) \leq q_n^2\left[2\exp(-c_2 n^{1-2\alpha} q_n^{-4}) + nm_1\exp(-s_1 n^\alpha)\right],$$

and by Condition (C7) that $\|(E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\| \leq D_1^{-1} q_n$. Thus,

$$P\left(\|(\mathbb{P}_n\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\| \geq (1+c_1)D_1^{-1}q_n\right) \leq q_n^2\left[2\exp(-c_2 n^{1-2\alpha} q_n^{-4}) + nm_1\exp(-s_1 n^\alpha)\right]. \tag{2}$$

Since by Lemma 1.9,

$$\begin{aligned} &P\left(\|\hat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j\|^2 \geq q_n(c_9 d_n^2\delta^2 + c_{10} d_n^{3/2}\delta + c_{11} d_n^2\delta)^2\right) \\ &\leq q_n(8d_n + 2d_n^2)\exp\left(-\frac{n\delta^2}{2(c_{12}d_n^{-1} + c_{13}\delta)}\right) + 6q_n d_n^2 \exp(-c_7 n d_n^{-3}). \end{aligned} \tag{3}$$

It follows from (1), (2) and (3) that

$$\begin{aligned} & P\left(|S_1| \geq (1+c_1)D_1^{-1}q_n^2(c_9 d_n^2\delta^2 + c_{10}d_n^{3/2}\delta + c_{11}d_n^2\delta)^2\right) \\ & \leq 2q_n^2\exp(-c_2 n^{1-2\alpha}q_n^{-4}) + nq_n^2 m_1\exp(-s_1 n^{\alpha}) \\ & + q_n(8d_n + 2d_n^2)\exp\left(-\frac{n\delta^2}{2(c_{12}d_n^{-1} + c_{13}\delta)}\right) + 6q_n d_n^2\exp(-c_7 n d_n^{-3}). \end{aligned} \tag{4}$$

Now we bound $S_2$, note that

$$|S_2| \leq 2\|\hat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j\| \cdot \|(\mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top)^{-1}\| \cdot \|\boldsymbol{\phi}_j\|. \tag{5}$$

By (3),

$$\begin{aligned} & P\left(\|\hat{\boldsymbol{\phi}}_j - \boldsymbol{\phi}_j\| \geq q_n^{1/2}(c_9 d_n^2\delta^2 + c_{10}d_n^{3/2}\delta + c_{11}d_n^2\delta)\right) \\ & \leq q_n(8d_n + 2d_n^2)\exp\left(-\frac{n\delta^2}{2(c_{12}d_n^{-1} + c_{13}\delta)}\right) + 6q_n d_n^2\exp(-c_7 n d_n^{-3}). \end{aligned} \tag{6}$$

Since by Fact 3,

$$\|(E\Psi_j\Psi_j^\top)^{-1}\| \leq D_3^{-1}d_n.$$

Moreover, it follows from Condition (C5) and Fact 2 that

$$\|E\Psi_j y_k\|^2 = \sum_{m=1}^{d_n}(E\psi_{jm}y_k)^2 = \sum_{m=1}^{d_n}(E\psi_{jm}m_k(\boldsymbol{x}))^2 \leq \sum_{m=1}^{d_n} B_1^2 \cdot E\psi_{jm}^2 \leq D_5 B_1^2.$$

So we have

$$\begin{aligned} \|\boldsymbol{\phi}_j\|^2 &= \sum_{k=1}^{q_n}|\phi_{kj}|^2 = \sum_{k=1}^{q_n}|Ef_{nkj}y_k|^2 \\ &= \sum_{k=1}^{q_n}|(E\Psi_j y_k)^\top(E\Psi_j\Psi_j^\top)^{-1}(E\Psi_j y_k)|^2 \\ &\leq \sum_{k=1}^{q_n}\left(\|(E\Psi_j\Psi_j^\top)^{-1}\| \cdot \|E\Psi_j y_k\|^2\right)^2 \\ &\leq q_n d_n^2 D_3^{-2}D_5^2 B_1^4. \end{aligned} \tag{7}$$

Combining (2), (5), (6) and (7), we have

$$\begin{aligned} & P\left(|S_2| \geq 2q_n^2 d_n(c_9 d_n^2\delta^2 + c_{10}d_n^{3/2}\delta + c_{11}d_n^2\delta)(1+c_1)D_1^{-1}D_3^{-1}D_5 B_1^2\right) \\ & \leq q_n(8d_n + 2d_n^2)\exp\left(-\frac{n\delta^2}{2(c_{12}d_n^{-1} + c_{13}\delta)}\right) + 6q_n d_n^2\exp(-c_7 n d_n^{-3}) \\ & + 2q_n^2\exp(-c_2 n^{1-2\alpha}q_n^{-4}) + nq_n^2 m_1\exp(-s_1 n^{\alpha}). \end{aligned} \tag{8}$$

To bound $S_3$, we note that

$$|S_3| \le \|(\mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top)^{-1}\| \cdot \|E\boldsymbol{y}\boldsymbol{y}^\top - \mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top\| \cdot \|(E\boldsymbol{y}\boldsymbol{y}^\top)^{-1}\| \cdot \|\boldsymbol{\phi}_j\|^2. \tag{9}$$

Since

$$\|\mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top - E\boldsymbol{y}\boldsymbol{y}^\top\| \le q_n \|\mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top - E\boldsymbol{y}\boldsymbol{y}^\top\|_\infty = q_n \max_{1\le m,k\le q_n} |\mathbb{P}_n y_m y_k - E y_m y_k| \,,$$

it follows from Lemma 1.3 that

$$\begin{aligned} &P\left(\|\mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top - E\boldsymbol{y}\boldsymbol{y}^\top\| \ge q_n\delta\right) \le P\left(q_n\|\mathbb{P}_n \boldsymbol{y}\boldsymbol{y}^\top - E\boldsymbol{y}\boldsymbol{y}^\top\|_\infty \ge q_n\delta\right) \\ &\le 2q_n^2 \exp\left(-\frac{n^{1-2\alpha}\delta^2}{2}\right) + nq_n^2 m_1 \exp(-s_1 n^\alpha). \end{aligned} \tag{10}$$

By Condition (C7), (2), (7) and (10), we have

$$\begin{aligned} &P\left(|S_3| \ge q_n^4 d_n^2 \delta(1+c_1) D_1^{-2} D_3^{-2} D_5^2 B_1^4\right) \\ &\le 2q_n^2 \exp(-c_2 n^{1-2\alpha} q_n^{-4}) + 2nq_n^2 m_1 \exp(-s_1 n^\alpha) + 2q_n^2 \exp(-n^{1-2\alpha}\delta^2/2). \end{aligned} \tag{11}$$

It follows from (4), (8) and (11) that there exist some positive constants $b_1$, $b_2$, $b_3$, such that

$$\begin{aligned} P(|\hat{\omega}_j - \omega_j| \ge\ & b_1 q_n^2 (c_9 d_n^2 \delta^2 + c_{10} d_n^{3/2}\delta + c_{11} d_n^2 \delta)^2 \\ &+ b_2 q_n^2 d_n (c_9 d_n^2 \delta^2 + c_{10} d_n^{3/2}\delta + c_{11} d_n^2\delta) + b_3 q_n^4 d_n^2 \delta) \\ &\le 6q_n^2 \exp(-c_2 n^{1-2\alpha} q_n^{-4}) + 4nq_n^2 m_1 \exp(-s_1 n^\alpha) + 12 q_n d_n^2 \exp(-c_7 n d_n^{-3}) \\ &+ 2q_n(8d_n + 2d_n^2)\exp(-\frac{n\delta^2}{2(c_{12} d_n^{-1} + c_{13}\delta)}) + 2q_n^2 \exp(-\frac{n^{1-2\alpha}\delta^2}{2}). \end{aligned}$$

For a given $c > 0$, there exists a $b_4$ such that

$$\begin{aligned} b_1 q_n^2 (c_9 d_n^2\delta^2 + c_{10} d_n^{3/2}\delta + c_{11} d_n^2 \delta)^2 + b_2 q_n^2 d_n (c_9 d_n^2 \delta^2 + c_{10} d_n^{3/2}\delta + c_{11} d_n^2\delta) \\ + b_3 q_n^4 d_n^2 \delta \le b_4 q_n^4 d_n^6 \delta \le cn^{-\tau}. \end{aligned}$$

Combining Condition (C10), and taking $\delta \le cn^{-\tau} q_n^{-4} d_n^{-6} b_4^{-1}$, then there exist some positive constants $b_5$ and $b_6$ such that

$$\begin{aligned} P\left(|\hat{\omega}_j - \omega_j| \ge cn^{-\tau}\right) &\le 6q_n^2 \exp(-c_2 n^{1-2\alpha} q_n^{-4}) + 4nq_n^2 m_1 \exp(-s_1 n^\alpha) \\ &+ 12 q_n d_n^2 \exp(-c_7 n d_n^{-3}) + 2q_n(8d_n + 2d_n^2)\exp(-b_5 n^{1-2\tau} q_n^{-8} d_n^{-11}) \\ &+ 2q_n^2 \exp(-b_6 n^{1-2\alpha-2\tau} q_n^{-8} d_n^{-12}) \\ &\le O(6n^{2\beta} \exp(-c_2 n^{1-2\alpha-4\beta}) + 4n^{1+2\beta} m_1 \exp(-s_1 n^\alpha) \\ &+ 12 n^{\beta+2\gamma} \exp(-c_7 n^{1-3\gamma}) + 2n^{2\beta}\exp(-b_6 n^{1-2\alpha-2\tau-8\beta-12\gamma}) \\ &+ (16 n^{\beta+\gamma} + 4n^{\beta+2\gamma}) \exp(-b_5 n^{1-2\tau-8\beta-11\gamma})). \end{aligned}$$

Therefore, we have

$$\begin{aligned}
P\left(\max_{1\le j\le p_n}|\hat{\omega}_j-\omega_j|\ge cn^{-\tau}\right) &\le O(6p_n n^{2\beta}\exp(-c_2 n^{1-2\alpha-4\beta})\\
&+4p_n n^{1+2\beta}m_1\exp(-s_1 n^{\alpha})+12p_n n^{\beta+2\gamma}\exp(-c_7 n^{1-3\gamma})\\
&+2p_n n^{2\beta}\exp(-b_6 n^{1-2\alpha-2\tau-8\beta-12\gamma})\\
&+p_n(16n^{\beta+\gamma}+4n^{\beta+2\gamma})\exp(-b_5 n^{1-2\tau-8\beta-11\gamma}))
\end{aligned}$$

Now we show the second part. Under Condition (C9), we have

$$\begin{aligned}
&\{\mathcal{M}\not\subseteq\widehat{\mathcal{M}}\}\subseteq\{|\hat{\omega}_j-\omega_j|>cn^{-\tau},\exists j\in\mathcal{M}\}\\
&\Rightarrow S_n=\{\max_{j\in\mathcal{M}}|\hat{\omega}_j-\omega_j|\le cn^{-\tau}\}\subseteq\{\mathcal{M}\subseteq\widehat{\mathcal{M}}\},
\end{aligned}$$

so we can conclude that

$$\begin{aligned}
&P(\mathcal{M}\subseteq\widehat{\mathcal{M}})\ge P(S_n)=1-P(S_n^c)\\
&\ge 1-s_nP(|\hat{\omega}_j-\omega_j|\ge cn^{-\tau})\\
&\ge 1-O(6s_n n^{2\beta}\exp(-c_2n^{1-2\alpha-4\beta})+4s_n n^{1+2\beta}m_1\exp(-s_1n^{\alpha})\\
&+12s_n n^{\beta+2\gamma}\exp(-c_7n^{1-3\gamma})+2s_n n^{2\beta}\exp(-b_6n^{1-2\alpha-2\tau-8\beta-12\gamma})\\
&+s_n(16n^{\beta+\gamma}+4n^{\beta+2\gamma})\exp(-b_5n^{1-2\tau-8\beta-11\gamma}))
\end{aligned}$$

It permits $\log p_n = O(n^a)$, where $a < \min\{\alpha, 1-3\gamma, 1-2\alpha-2\tau-8\beta-12\gamma\}$. This completes the proof. ■

**Proof of Theorem 2.2**:

*Proof.* By Condition (C11),

$$\begin{aligned}
&P\left(\min_{j\in\mathcal{M}}\hat{\omega}_j-\max_{j\in\mathcal{M}^c}\hat{\omega}_j<\frac{\delta}{2}\right)\\
&\le P\left((\min_{j\in\mathcal{M}}\hat{\omega}_j-\max_{j\in\mathcal{M}^c}\hat{\omega}_j)-(\min_{j\in\mathcal{M}}\omega_j-\max_{j\in\mathcal{M}^c}\omega_j)<-\frac{\delta}{2}\right)\\
&\le P\left(\left|(\min_{j\in\mathcal{M}}\hat{\omega}_j-\max_{j\in\mathcal{M}^c}\hat{\omega}_j)-(\min_{j\in\mathcal{M}}\omega_j-\max_{j\in\mathcal{M}^c}\omega_j)\right|>\frac{\delta}{2}\right)\\
&=P\left(\left|(\min_{j\in\mathcal{M}}\hat{\omega}_j-\min_{j\in\mathcal{M}}\omega_j)-(\max_{j\in\mathcal{M}^c}\hat{\omega}_j-\max_{j\in\mathcal{M}^c}\omega_j)\right|>\frac{\delta}{2}\right)\\
&\le P\left(2\max_{1\le j\le p_n}|\hat{\omega}_j-\omega_j|>\frac{\delta}{2}\right),
\end{aligned}$$

According to Theorem 1, there exist some positive constants $\tilde{b}_5$ and $\tilde{b}_6$ such that

$$\begin{aligned}
&P\left(\min_{j\in\mathcal{M}}\hat{\omega}_j - \max_{j\in\mathcal{M}^c}\hat{\omega}_j < \frac{\delta}{2}\right)\\
&\le P\left(2\max_{1\le j\le p_n}|\hat{\omega}_j-\omega_j| > \frac{\delta}{2}\right)\\
&\le O(6p_nn^{2\beta}\exp(-c_2n^{1-2\alpha-4\beta}) + 4p_nn^{1+2\beta}m_1\exp(-s_1n^{\alpha})\\
&+ 12p_nn^{\beta+2\gamma}\exp(-c_7n^{1-3\gamma}) + 2p_nn^{2\beta}\exp(-\tilde{b}_6n^{1-2\alpha-8\beta-12\gamma}\delta^2)\\
&+ p_n(16n^{\beta+\gamma}+4n^{\beta+2\gamma})\exp(-\tilde{b}_5n^{1-8\beta-11\gamma}\delta^2))\\
&\triangleq O(p_n(6n^{2\beta}\exp(-c_2n^{1-2\alpha-4\beta}) + 4n^{1+2\beta}m_1\exp(-s_1n^{\alpha})\\
&+ 12n^{\beta+2\gamma}\exp(-c_7n^{1-3\gamma}) + 2n^{2\beta}\exp(-b_6^*n^{1-2\alpha-8\beta-12\gamma})\\
&+ (16n^{\beta+\gamma}+4n^{\beta+2\gamma})\exp(-b_5^*n^{1-8\beta-11\gamma}))) \triangleq O(g(n)).
\end{aligned}$$

where $b_5^*$ and $b_6^*$ determined by $\delta$, $\tilde{b}_5$ and $\tilde{b}_6$. If $\log p_n = o(n^t)$, where $t < \min\{\alpha, 1-3\gamma, 1-2\alpha-8\beta-12\gamma\}$, then there exists $n_0$ such that

$$\sum_{n=n_0}^{\infty} g(n) \le \sum_{n=n_0}^{\infty}(6+4m_1+12+2+16+4)\exp(-2\log n) = \sum_{n=n_0}^{\infty}(40+4m_1)n^{-2} < \infty,$$

by Borel-Contelli lemma, we have

$$\lim\inf_{n\to\infty}\{\min_{j\in\mathcal{M}}\hat{\omega}_j - \max_{j\in\mathcal{M}^c}\hat{\omega}_j\} \ge \frac{\delta}{2} > 0, a.s.$$

■

## 2. More Simulation Results

**Example 7.** Let $g(x) = \sin(2x)+2\cos(2x)+3\sin^2(2x)+4\cos^3(2x)+5\sin^3(2x)$. Consider the model

$$\begin{aligned}
y_1 &= x_1^2 + 5\sin(x_1x_2^2) + \varepsilon_1,\\
\log(y_2) &= x_3 + 0.5x_4 + 0.1x_5 + \varepsilon_2,\\
y_3 &= g(x_2x_3x_4) + x_5^2 + \varepsilon_3,
\end{aligned}$$

with $(n,p) = (200, 2000)$. The true predictor set is $\mathcal{M} = \{x_1,\ldots,x_5\}$. $\mathbf{X}_i \sim N(\mathbf{0},\Sigma)$, where $\Sigma = (\sigma_{jl})_{p\times p}$, $\sigma_{jj} = 1$; $\sigma_{jl} = 0.2$, $j \ne l$. $\varepsilon_k$ is from (a) $N(0,1)$ and (b) $t(3)$. We run 100 replications for each situation. Example 7 is designed to see whether the proposed methods still work well if additive structure is violated. The screening performances are shown in Figure 1 and Tables 1-2.

As shown in Figure 1, none of these methods perform well. The performance of DC-SIS is the best among these methods, and the proposed methods, Naive-GCS1, Naive-GCS2 and GCPS, perform better than linear methods in this situation. Moreover, when noise increases as error is drawn from $t(3)$ distribution, the performance of all methods deteriorates further. Tables 1-2 also show the similar conclusions.

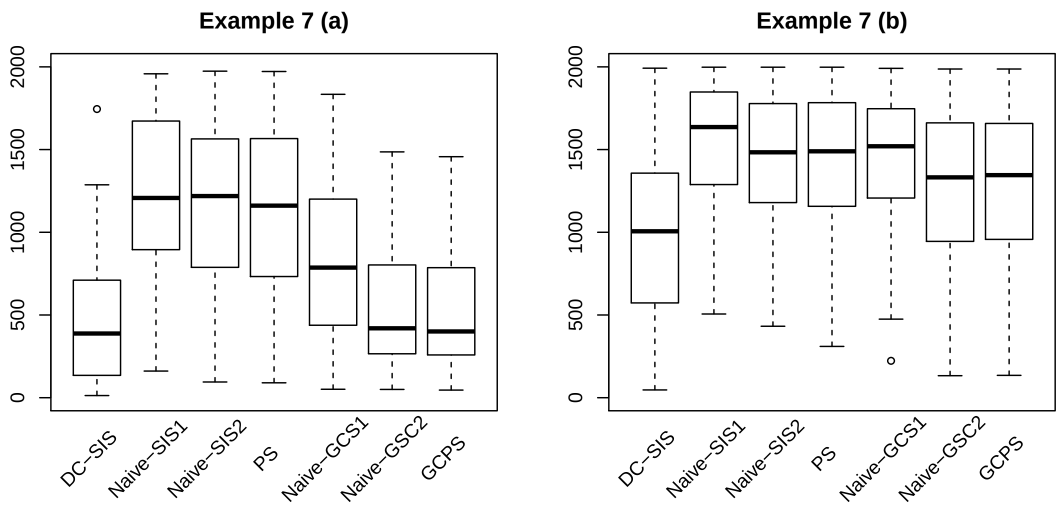


Figure 1. The boxplots of minimum model size (MMS) for Example 7 (a) $\varepsilon \sim N(0,1)$ and (b) $\varepsilon \sim t(3)$.

Since the proposed methods are based on additive models structure and make use of the structure information, it is reasonable that they outperform DC-SIS under additive models. However, once the additive models are misspecified, our methods will lose their edge in screening. This is inherent in all model-based screening approaches.

Table 1. The average rank $R_j$ of true predictor $x_j$ for Example 7 (a) $\varepsilon \sim N(0,1)$ and (b) $\varepsilon \sim t(3)$.

| Example | Method | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ |
|---|---|---|---|---|---|---|
| Example 7 (a) | DC-SIS | 11.19 | 290.4 | 1.040 | 13.51 | 294.9 |
| | Naive-SIS1 | 399.4 | 1002 | 1.320 | 88.78 | 674.7 |
| | Naive-SIS2 | 10.45 | 914.4 | 2.250 | 130.7 | 741.9 |
| | PS | 10.25 | 913.1 | 2.010 | 127.5 | 748.1 |
| | Naive-GCS1 | 320.3 | 676.3 | 13.77 | 121.8 | 480.0 |
| | Naive-GCS2 | 56.05 | 425.9 | 13.50 | 91.26 | 276.3 |
| | GCPS | 58.31 | 419.2 | 13.54 | 89.08 | 267.0 |
| Example 7 (b) | DC-SIS | 294.6 | 686.9 | 186.0 | 361.4 | 718.0 |
| | Naive-SIS1 | 828.6 | 1046 | 431.5 | 720.5 | 1050 |
| | Naive-SIS2 | 53.94 | 945.0 | 532.7 | 645.4 | 981.0 |
| | PS | 54.14 | 946.6 | 537.0 | 648.5 | 982.9 |
| | Naive-GCS1 | 836.3 | 967.3 | 485.3 | 712.0 | 901.6 |
| | Naive-GCS2 | 528.8 | 801.6 | 466.1 | 667.9 | 775.7 |
| | GCPS | 532.2 | 799.0 | 465.9 | 667.0 | 773.1 |

Table 2. The selecting proportion $P_j$'s of true predictors and $P_a$ for Example 7 (a) $\varepsilon \sim N(0,1)$ and (b) $\varepsilon \sim t(3)$.

| Example | Method | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_a$ |
|---|---|---|---|---|---|---|---|
| Example 7 (a) | DC-SIS | 0.96 | 0.17 | 1 | 0.93 | 0.21 | 0.04 |
| | Naive-SIS1 | 0.19 | 0.02 | 1 | 0.71 | 0.05 | 0 |
| | Naive-SIS2 | 0.94 | 0.02 | 1 | 0.56 | 0.04 | 0 |
| | PS | 0.94 | 0.02 | 1 | 0.58 | 0.02 | 0 |
| | Naive-GCS1 | 0.22 | 0.04 | 0.89 | 0.34 | 0.07 | 0 |
| | Naive-GCS2 | 0.64 | 0.05 | 0.89 | 0.39 | 0.11 | 0 |
| | GCPS | 0.64 | 0.05 | 0.89 | 0.40 | 0.11 | 0 |
| Example 7 (b) | DC-SIS | 0.36 | 0.02 | 0.64 | 0.38 | 0.06 | 0 |
| | Naive-SIS1 | 0.01 | 0 | 0.33 | 0.17 | 0.02 | 0 |
| | Naive-SIS2 | 0.88 | 0.03 | 0.30 | 0.13 | 0.02 | 0 |
| | PS | 0.88 | 0.04 | 0.31 | 0.14 | 0.02 | 0 |
| | Naive-GCS1 | 0.03 | 0.01 | 0.18 | 0.04 | 0.04 | 0 |
| | Naive-GCS2 | 0.11 | 0.01 | 0.18 | 0.04 | 0.04 | 0 |
| | GCPS | 0.10 | 0.01 | 0.18 | 0.04 | 0.04 | 0 |

## References

[1] Fan, J., Feng, Y., & Song, R. (2011). Nonparametric independence screening in sparse ultra-high-dimensional additive models. *Journal of the American Statistical Association*, 106(494), 544-557.

[2] Li, X., Cheng, G., Wang, L., Lai, P., & Song, F. (2017). Ultrahigh dimensional feature screening via projection. *Computational Statistics and Data Analysis*, 114, 88-104.

[3] Liu, J., Li, R., & Wu, R. (2014). Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *Journal of the American Statistical Association*, 109(505), 266-274.

[4] Van Der Vaart, A. W., & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.