Web-based Supplementary Materials for "Ultrahigh Dimensional Time Course Feature Selection"

Peirong Xu[1], Lixing Zhu[2] and Yi Li[3]

[1]Department of Mathematics, Southeast University, Nanjing, China

[2]Department of Mathematics, Hong Kong Baptist University, Hong Kong, China

[3]Department of Biostatistics, University of Michigan, Ann Arbor, USA

1 Web Appendix A - Proofs

To prove Theorems 1 and 2, we will need the Bernstein's inequality (see, e.g. van der Vaart and Wellner, 1996) and a lemma of Wang (2011) (Lemma C.1). We re-state the results.

LEMMA 1.1. *(Bernstein's inequality) Let $Z_1, \ldots, Z_n$ be independent random variables with mean zero and satisfy*

$$E|Z_i|^l \leq l! M^{l-2} V_i/2$$

for every $l \geq 2$ and all i and some positive constants M and V_i. Then

$$P(|Z_1 + \ldots + Z_n| > t) \leq 2 \exp\left(-\frac{1}{2} \frac{t^2}{V + Mt}\right),$$

for $V > V_1 + \ldots + V_n$.

LEMMA 1.2. *(Wang, 2011) Let $\bar{G}(\beta) = n^{-1} \sum_{i=1}^{n} X_i^\tau A_i^{1/2}(\beta) \bar{R}^{-1} A_i^{-1/2}(\beta)(Y_i - \mu_i(\beta))$ and $\nabla(\beta) = -\partial \bar{G}(\beta)/\partial\beta$. Then, we have*

$$\nabla(\beta) = \bar{H}(\beta) + \bar{E}(\beta) + \bar{S}(\beta),$$

where

$$\bar{H}(\beta) = \frac{1}{n} \sum_{i=1}^{n} X_i^\tau A_i^{1/2}(\beta) \bar{R}^{-1} A_i^{1/2}(\beta) X_i,$$
$$\bar{E}(\beta) = \frac{1}{2n} \sum_{i=1}^{n} X_i^\tau A_i^{1/2}(\beta) \bar{R}^{-1} A_i^{-3/2}(\beta) D_i(\beta) F_i(\beta) X_i,$$
$$\bar{S}(\beta) = \frac{1}{2n} \sum_{i=1}^{n} X_i^\tau A_i^{1/2}(\beta) F_i(\beta) J_i(\beta) X_i,$$

with

$$D_i(\beta) = \mathrm{diag}(Y_{i1} - \mu_{i1}(\beta), \ldots, Y_{im} - \mu_{im}(\beta)),$$
$$F_i(\beta) = \mathrm{diag}(\ddot{a}(X_{i1}^\tau \beta), \ldots, \ddot{a}(X_{im}^\tau \beta)),$$
$$J_i(\beta) = \mathrm{diag}(\bar{R}^{-1} A_i^{-1/2}(\beta)(Y_i - \mu_i(\beta))).$$

Proof of Theorem 1. According to the definition of $\widehat{\mathcal{M}}_{\gamma_n}$, we know that $\{\mathcal{M}_0 \subset \widehat{\mathcal{M}}_{\gamma_n}\}$ is equivalent to $\{\min_{j\in\mathcal{M}_0} |\widehat{G}_j(0)| \geq \gamma_n\}$. Then, it is easy to see that

$$P(\mathcal{M}_0 \subset \widehat{\mathcal{M}}_{\gamma_n}) \geq 1 - \sum_{j\in\mathcal{M}_0} P(|\widehat{G}_j(0)| < \gamma_n).$$

Let $\bar{G}(0) = n^{-1}\sum_{i=1}^n X_i^\tau A_i^{1/2}(0)\bar{R}^{-1}A_i^{-1/2}(0)(Y_i - \mu_i(0))$ and $\bar{G}_j(0)$ be the jth element of $\bar{G}(0)$. Then for each $j \in \mathcal{M}_0$, we have

$$\begin{aligned} P(|\widehat{G}_j(0)| < \gamma_n) &\leq P(|\bar{G}_j(0)| - |\bar{G}_j(0) - \widehat{G}_j(0)| < \gamma_n) \\ &\leq P(|\bar{G}_j(0)| < 2\gamma_n) + P(|\bar{G}_j(0) - \widehat{G}_j(0)| > \gamma_n). \end{aligned}$$

We first consider the term $P(|\bar{G}_j(0)| < 2\gamma_n)$, $j \in \mathcal{M}_0$. Under conditions (C2), (C4) and (C5), we have that

$$\begin{aligned} P(|\bar{G}_j(0)| < 2\gamma_n) &\leq P(|\bar{g}_j(0)| - |\bar{G}_j(0) - \bar{g}_j(0)| < 2\gamma_n) \\ &\leq P(|\bar{G}_j(0) - \bar{g}_j(0)| > c_3 n^{-\kappa} - 2\gamma_n) \\ &= P(|\bar{G}_j(0) - \bar{g}_j(0)| > c_3 n^{-\kappa}/2) \\ &\leq 2\exp\left\{-\frac{c_3^2 n^{1-2\kappa}/4}{2c + c_3 n^{-\kappa}}\right\}, \end{aligned}$$

for every $j \in \mathcal{M}_0$, where the second inequality is due to the bound $\min_{j\in\mathcal{M}_0} |\bar{g}_j(0)| \geq c_3 n^{-\kappa}$ in condition (C5), the last inequality follows from Lemma 1.1, and c is a positive constant depending on c_2. Hereafter, we use c to denote a generic positive constant which may vary for every appearance.

Next, let e_j be a p-dimensional basis vector with the jth element being one and all the other elements being zero, $1 \le j \le p$. Then,

$$
\begin{aligned}
& P(|\bar{G}_j(0) - \widehat{G}_j(0)| > \gamma_n) \\
\le\ & P(n^{-1} \sum_{i=1}^{n} |e_j^\tau X_i^\tau A_i^{1/2}(0)(\widehat{R}^{-1} - \bar{R}^{-1}) A_i^{-1/2}(0)(Y_i - \mu_i(0))| > \gamma_n) \\
\le\ & P(n^{-1} \sum_{i=1}^{n} \|e_j^\tau X_i^\tau A_i^{1/2}(0)\|_2 \cdot \|\widehat{R}^{-1} - \bar{R}^{-1}\|_F \cdot \|A_i^{-1/2}(0)(Y_i - \mu_i(0))\|_2 > \gamma_n) \\
\le\ & P(n^{-1} \sum_{i=1}^{n} \|A_i^{-1/2}(0)(Y_i - \mu_i(0))\|_2 > c\gamma_n / \|\widehat{R}^{-1} - \bar{R}^{-1}\|_F) \\
\le\ & c(s_n/n^{1-2\kappa})^{1/2}, \qquad (1.1)
\end{aligned}
$$

where the third inequality follows from condition (C3), and the last inequality follows from conditions (C2), (C4) and the Markov's inequality.

Therefore, under condition (C6), we have

$$
\begin{aligned}
P(\mathcal{M}_0 \subset \widehat{\mathcal{M}}_{\gamma_n}) &\ge 1 - \sum_{j \in \mathcal{M}_0} \left\{ P(|\bar{G}_j(0)| < 2\gamma_n) + P(|\bar{G}_j(0) - \widehat{G}_j(0)| > \gamma_n) \right\} \\
&\ge 1 - 2s_n \exp\left\{ -\frac{c_3^2 n^{1-2\kappa}/4}{2c + c_3 n^{-\kappa}} \right\} - \frac{c s_n^{3/2}}{n^{1/2-\kappa}} \\
&\to 1.
\end{aligned}
$$

□

Proof of Theorem 2. Note that $\gamma_n \le |\widehat{G}_j(0)| \le |\bar{G}_j(0)| + |\bar{G}_j(0) -$

$\widehat{G}_j(0)|$ for every $j \in \widehat{\mathcal{M}}_{\gamma_n}$. Thus, we have

$$\begin{aligned}
|\widehat{\mathcal{M}}_{\gamma_n}| &\leq |\{1 \leq j \leq p : |\bar{G}_j(0)| \geq \gamma_n/2 \text{ or } |\bar{G}_j(0) - \widehat{G}_j(0)| \geq \gamma_n/2\}| \\
&\leq |\{1 \leq j \leq p : |\bar{G}_j(0)| \geq \gamma_n/2\}| \\
&\quad + |\{1 \leq j \leq p : |\bar{G}_j(0)| < \gamma_n/2 \text{ and } |\bar{G}_j(0) - \widehat{G}_j(0)| \geq \gamma_n/2\}| \\
&\triangleq I_1 + I_2.
\end{aligned}$$

Consequently, it is sufficient to provide upper bounds on I_1 and I_2 that hold with a high probability, respectively. Now suppose that $\|\bar{G}(0) - \bar{g}(0)\|_\infty \leq \gamma_n/4$. Then $|\bar{G}_j(0)| \geq \gamma_n/2$ implies that $|\bar{g}_j(0)| \geq \gamma_n/4$. Hence, under $\|\bar{G}(0) - \bar{g}(0)\|_\infty \leq \gamma_n/4$, we have

$$I_1 \leq |\{1 \leq j \leq p_n : |\bar{g}_j(0)| \geq \gamma_n/4\}| \leq 16\|\bar{g}(0)\|_2^2/\gamma_n^2.$$

Consequently, it follows that

$$|\widehat{\mathcal{M}}_{\gamma_n}| \leq 16\|\bar{g}(0)\|_2^2/\gamma_n^2$$

under $\|\bar{G}(0) - \bar{g}(0)\|_\infty \leq \gamma_n/4$ and $\|\widehat{G}(0) - \bar{G}(0)\|_\infty < \gamma_n/2$, which implies that we only need to provide an upper bound on $\|\bar{g}(0)\|_2^2$ when $\|\bar{G}(0) - \bar{g}(0)\|_\infty \leq \gamma_n/4$ and $\|\widehat{G}(0) - \bar{G}(0)\|_\infty < \gamma_n/2$ hold with a high probability.

Let $\beta_0^* = \Sigma^{1/2}\beta_0$. Note that $\bar{g}(\beta_0) = EX^\tau A^{1/2}(\beta_0)\bar{R}^{-1}A^{-1/2}(\beta_0)(Y -$

$\mu(\beta_0)) = 0$. Thus, we have

$$\begin{aligned}
\|\bar{g}(0)\|_2^2 &= \|\bar{g}(\beta_0) - \bar{g}(0)\|_2^2 \\
&= \|E\{\bar{G}(\beta_0) - \bar{G}(0)\}\|_2^2 \\
&= \| - E\{\nabla(\tilde{\beta})\}\beta_0\|_2^2 \\
&\leq \lambda_{\max}(MM^\tau)\|\beta_0^*\|_2^2,
\end{aligned}$$

where $\tilde{\beta}$ lies on the line segment between β_0 and 0 so that $\tilde{\beta} \in \mathcal{B}$ and $M = E\{\nabla(\tilde{\beta})\}\Sigma^{-1/2}$. Since

$$\begin{aligned}
MM^\tau &= E\{\nabla(\tilde{\beta})\}\Sigma^{-1}E^\tau\{\nabla(\tilde{\beta})\} \\
&\leq \lambda_{\min}^{-1}(\Sigma)E\{\nabla(\tilde{\beta})\}E^\tau\{\nabla(\tilde{\beta})\},
\end{aligned}$$

we have $\lambda_{\max}(MM^\tau) \leq \lambda_{\min}^{-1}(\Sigma)\lambda_{\max}^2\{E(\nabla(\tilde{\beta}))\}$. Now, we only need to provide an upper bound on $\lambda_{\max}\{E(\nabla(\tilde{\beta}))\}$. Lemma 1.2 implies that

$$\lambda_{\max}\{E(\nabla(\tilde{\beta}))\} \leq \lambda_{\max}\{E(\bar{H}(\tilde{\beta}))\} + \lambda_{\max}\{E(\bar{E}(\tilde{\beta}))\} + \lambda_{\max}\{E(\bar{S}(\tilde{\beta}))\}.$$

We first consider term $\lambda_{\max}\{E(\bar{H}(\beta))\}$, $\beta \in \mathcal{B}$. Under conditions (C2) and (C7), for any unit length p_n-dimensional vector r, we have

$$\begin{aligned}
r^\tau \bar{H}(\beta)r &\leq \lambda_{\max}(\bar{R}^{-1})r^\tau \left(n^{-1}\sum_{i=1}^n X_i^\tau A_i(\beta)X_i\right) r \\
&\leq \lambda_{\min}^{-1}(\bar{R}) \cdot \max_{1\leq k\leq m} \dot{a}(X_{ik}^\tau\beta) \cdot r^\tau \left(n^{-1}\sum_{i=1}^n X_i^\tau X_i\right) r \\
&\leq cr^\tau \left(n^{-1}\sum_{i=1}^n X_i^\tau X_i\right) r.
\end{aligned}$$

Therefore,

$$\lambda_{\max}\{E(\bar{H}(\beta))\} \leq E\{\lambda_{\max}(\bar{H}(\beta))\} \leq cE\{\lambda_{\max}(n^{-1}\sum_{i=1}^n X_i^\tau X_i)\},$$

for any $\beta \in \mathcal{B}$. Next, we consider term $E\{\bar{E}(\beta)\}$. Let $\bar{D}_i(\beta) = \mathrm{diag}(\mu_{i1}(\beta_0) - \mu_{i1}(\beta), \ldots, \mu_{im}(\beta_0) - \mu_{im}(\beta))$. Then, we have

$$E\{\bar{E}(\beta)\} = E\left\{\frac{1}{2n}\sum_{i=1}^{n} X_i^\tau A_i^{1/2}(\beta)\bar{R}^{-1}A_i^{-3/2}(\beta)\bar{D}_i(\beta)F_i(\beta)X_i\right\},$$

which can be decomposed as

$$E\{\bar{E}(\beta)\} = E\{\bar{E}_{11}(\beta)\} + E\{\bar{E}_{12}(\beta)\} + E\{\bar{E}_{13}(\beta)\} + E\{\bar{E}_{14}(\beta)\},$$

where

$$\begin{aligned}
\bar{E}_{11}(\beta) &= \frac{1}{2n}\sum_{i=1}^{n} X_i^\tau [A_i^{1/2}(\beta) - A_i^{1/2}(\beta_0)]\bar{R}^{-1}A_i^{-3/2}(\beta_0)\bar{D}_i(\beta)F_i(\beta_0)X_i \\
\bar{E}_{12}(\beta) &= \frac{1}{2n}\sum_{i=1}^{n} X_i^\tau A_i^{1/2}(\beta)\bar{R}^{-1}[A_i^{-3/2}(\beta) - A_i^{-3/2}(\beta_0)]\bar{D}_i(\beta)F_i(\beta_0)X_i \\
\bar{E}_{13}(\beta) &= \frac{1}{2n}\sum_{i=1}^{n} X_i^\tau A_i^{1/2}(\beta)\bar{R}^{-1}A_i^{-3/2}(\beta)\bar{D}_i(\beta)[F_i(\beta) - F_i(\beta_0)]X_i \\
\bar{E}_{14}(\beta) &= \frac{1}{2n}\sum_{i=1}^{n} X_i^\tau A_i^{1/2}(\beta_0)\bar{R}^{-1}A_i^{-3/2}(\beta_0)\bar{D}_i(\beta)F_i(\beta_0)X_i.
\end{aligned}$$

For any $r \in R^{p_n}$ with $\|r\|_2 = 1$,

$$\begin{aligned}
& |r^\tau \bar{E}_{11}(\beta) r| \\
= \ & \frac{1}{2n}\left|\sum_{i=1}^{n}\sum_{k=1}^{m}(\mu_{ik}(\beta_0) - \mu_{ik}(\beta)) r^\tau X_i^\tau [A_i^{1/2}(\beta) - A_i^{1/2}(\beta_0)]\bar{R}^{-1}A_i^{-3/2}(\beta_0)e_k e_k^\tau F_i(\beta_0)X_i r\right| \\
\leq \ & \frac{1}{2}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}(\mu_{ik}(\beta_0) - \mu_{ik}(\beta))\right\}^{1/2} \\
& \cdot \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}(r^\tau X_i^\tau (A_i^{1/2}(\beta) - A_i^{1/2}(\beta_0))\bar{R}^{-1}A_i^{-3/2}(\beta_0)e_k e_k^\tau F_i(\beta_0)X_i r)^2\right\}^{1/2}.
\end{aligned}$$

The application of Taylor expansion yields that

$$
\begin{aligned}
& \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}(\mu_{ik}(\beta_0)-\mu_{ik}(\beta))\right\}^{1/2} \\
= & \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}\dot{a}^2(X_{ik}^{\tau}\beta^*)(\beta-\beta_0)^{\tau}X_{ik}X_{ik}^{\tau}(\beta-\beta_0)\right\}^{1/2} \\
\leq & \left(\sup_{\tilde{\beta}\in\mathcal{B}}\dot{a}^2(X_{ik}^{\tau}\tilde{\beta})\right)^{1/2}\cdot\left\{(\beta-\beta_0)^{\tau}\frac{1}{n}\sum_{i=1}^{n}X_i^{\tau}X_i(\beta-\beta_0)\right\}^{1/2} \\
\leq & \; c\lambda_{\max}^{1/2}(n^{-1}\sum_{i=1}^{n}X_i^{\tau}X_i)\|\beta-\beta_0\|_2,
\end{aligned}
$$

where β^* lies on the line segment between β_0 and β. Under conditions (C2), (C3), and (C7), we have

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}\{r^{\tau}X_i^{\tau}(A_i^{1/2}(\beta)-A_i^{1/2}(\beta_0))\bar{R}^{-1}A_i^{-3/2}(\beta_0)e_ke_k^{\tau}F_i(\beta_0)X_ir\}^2 \\
\leq & \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}\{r^{\tau}X_i^{\tau}(A_i^{1/2}(\beta)-A_i^{1/2}(\beta_0))\bar{R}^{-1}A_i^{-3/2}(\beta_0)e_k\}^2\cdot(e_k^{\tau}F_i(\beta_0)X_ir)^2 \\
\leq & \; c\|\beta-\beta_0\|_2^2\frac{1}{n}\sum_{i=1}^{n}\|X_ir\|_2^2 \\
\leq & \; c\lambda_{\max}(n^{-1}\sum_{i=1}^{n}X_i^{\tau}X_i)\|\beta-\beta_0\|_2^2.
\end{aligned}
$$

Hence, for any β satisfying $\|\beta\|_2\leq c_\beta$,

$$
\begin{aligned}
|r^{\tau}\bar{E}_{11}(\beta)r| & \leq \frac{c}{2}\lambda_{\max}(n^{-1}\sum_{i=1}^{n}X_i^{\tau}X_i)\|\beta-\beta_0\|_2^2 \\
& \leq c\lambda_{\max}(n^{-1}\sum_{i=1}^{n}X_i^{\tau}X_i),
\end{aligned}
$$

which implies that

$$
\lambda_{\max}\{E(\bar{E}_{11}(\beta))\}\leq E\{\lambda_{\max}(\bar{E}_{11}(\beta))\}\leq cE\{\lambda_{\max}(n^{-1}\sum_{i=1}^{n}X_i^{\tau}X_i)\}.
$$

Similarly, we can show that $\lambda_{\max}\{E(\bar{E}_{1s}(\beta))\} \leq cE\{\lambda_{\max}(n^{-1}\sum_{i=1}^n X_i^\tau X_i)\}$ for $s = 2, 3, 4$. Thus,

$$\lambda_{\max}\{E(\bar{E}(\beta))\} \leq \sum_{s=1}^{4} \lambda_{\max}\{E(\bar{E}_{1s}(\beta))\} \leq cE\{\lambda_{\max}(n^{-1}\sum_{i=1}^{n} X_i^\tau X_i)\}.$$

We can also have $\lambda_{\max}\{E(\bar{S}(\beta))\} \leq cE\{\lambda_{\max}(n^{-1}\sum_{i=1}^n X_i^\tau X_i)\}$, and then $\lambda_{\max}\{E(\nabla(\beta))\} \leq cE\{\lambda_{\max}(n^{-1}\sum_{i=1}^n X_i^\tau X_i)\}$ for $\beta \in \mathcal{B}$. Consequently, under condition (C7), we have

$$\begin{aligned} \|\bar{g}(0)\|_2 &\leq \lambda_{\min}^{-1/2}(\Sigma)\lambda_{\max}\{E(\nabla(\tilde{\beta}))\}\|\beta_0^*\|_2 \\ &\leq c\lambda_{\min}^{-1/2}(\Sigma)E\{\lambda_{\max}(n^{-1}\sum_{i=1}^{n} X_i^\tau X_i)\}\|\beta_0^*\|_2. \end{aligned}$$

Further, note that $\lambda_{\min}(\Sigma) \geq E\{\lambda_{\min}(n^{-1}\sum_{i=1}^n X_i^\tau X_i)\}$. Thus, we have

$$\|\bar{g}(0)\|_2 \leq c\frac{E\{\lambda_{\max}(n^{-1}\sum_{i=1}^n X_i^\tau X_i)\}}{[E\{\lambda_{\min}(n^{-1}\sum_{i=1}^n X_i^\tau X_i)\}]^{1/2}}\|\beta_0^*\|_2,$$

which results in, combining $\|\bar{G}(0) - \bar{g}(0)\|_\infty \leq \gamma_n/4$ and $\|\widehat{G}(0) - \bar{G}(0)\|_\infty < \gamma_n/2$,

$$|\widehat{\mathcal{M}}_{\gamma_n}| \leq 16\|\bar{g}(0)\|_2^2/\gamma_n^2 \leq O(n^{2\kappa}\sigma_n).$$

On the other hand, invoking Lemma 1.1, we have

$$\begin{aligned} & P(\|\bar{G}(0) - \bar{g}(0)\|_\infty > \gamma_n/4) \\ \leq & \sum_{j=1}^{p} P(|\bar{G}_j(0) - \bar{g}_j(0)| > \gamma_n/4) \\ \leq & 2p_n \exp\left\{\frac{c_3^2 n^{1-2\kappa}/16^2}{2c + c_3 n^{-\kappa}}\right\} \\ \to & 0 \end{aligned}$$

when $\log p_n = o(n^{1-2\kappa})$. Similar to the inequality (1.1) in the proof of Theorem 1, we have

$$\begin{aligned} & P(\|\widehat{G}(0) - \bar{G}(0)\|_\infty \geq \gamma_n/2) \\ = \quad & P(\max_{1\leq j\leq p} |\widehat{G}_j(0) - \bar{G}_j(0)| \geq \gamma_n/2) \\ \leq \quad & c(s_n/n^{1-2\kappa})^{1/2} \\ \to \quad & 0. \end{aligned}$$

Therefore,

$$\begin{aligned} & P(|\widehat{\mathcal{M}}_{\gamma_n}| \leq O(n^{2\kappa}\sigma_n)) \\ \geq \quad & P(\|\bar{G}(0) - \bar{g}(0)\|_\infty \leq \gamma_n/4 \text{ and } \|\widehat{G}(0) - \bar{G}(0)\|_\infty < \gamma_n/2) \\ \geq \quad & 1 - P(\|\bar{G}(0) - \bar{g}(0)\|_\infty > \gamma_n/4) - P(\|\widehat{G}(0) - \bar{G}(0)\|_\infty \geq \gamma_n/2) \\ \geq \quad & 1 - 2p_n \exp\left\{\frac{c_3^2 n^{1-2\kappa}/16^2}{2c + c_3 n^{-\kappa}}\right\} - \frac{c s_n^{1/2}}{n^{1/2-\kappa}} \\ \to \quad & 1, \end{aligned}$$

which concludes the proof. □

2 Web Appendix B - Additional Simulation and Data Analysis Results

[Table 1 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

[Table 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

3 Web Appendix C - Additional Comparisons with DC-SIS

To our knowledge, the DC-SIS method proposed by Li et al. (2012) is the first screening procedure for correlated responses. It is of substantial interest to compare the proposed GEES method with the DC-SIS method. The simulations reported in the main text hint that our GEES method performs better than the DC-SIS under generalized linear models (GLM) for ultrahigh time course data. However, unlike the GEES method, the DC-SIS method is model-free. As suggested by one referee, we conjecture that the DC-SIS method may work better than the GEES method when the GLM assumptions do not hold. For this purpose, we consider a simulation study.

We generate the response from the following heteroscedastic model

$$Y_{ik} = c_1\beta_1 X_{ik1} + c_2\beta_2 X_{ik2} + c_2\beta_3 \mathcal{I}(X_{ik3} < 0) + \exp(c_4 X_{ik4})\epsilon,$$

where $\mathcal{I}(\cdot)$ is an indicator function, $i = 1, \ldots, 200$, and $k = 1, \ldots, 10$. Following Li et al. (2012), we set $\beta_j = (-1)^u(a + |z|)$ for j =1, 2, and 3, where $a = 4\log n/\sqrt{n}$, u is from Bernoulli(0.4), and z is from the standard normal distribution. We set $(c_1, c_2, c_3, c_4) = (2, 0.5, 3, 2)$. We generate $X_{ik} = (X_{ik1}, \ldots, X_{ikp})^\tau$ from the multivariate normal distribution with mean 0 and an AR(1) covariance matrix with marginal variance 1 and autocorrelation coefficient 0.8. The random errors $(\epsilon_{i1}, \ldots, \epsilon_{i10})^\tau$ are generated from the multivariate normal distribution with marginal mean 0, marginal variance 1 and an CS correlation matrix with 0.5. And we set $p = 1000$ and repeat the experiment 400 times.

As in Examples 1 and 2, we report the 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size (MMS) and the average computing time in seconds to assess the sure screening property and computational efficiency. Table 3 summarizes the results by the GEES method and the DC-SIS method. It indicates that the DC-SIS needs somewhat smaller MMS to ensure the inclusion of all truly active covariates, at the expense of 10 times computing time.

[Table 3 about here.]

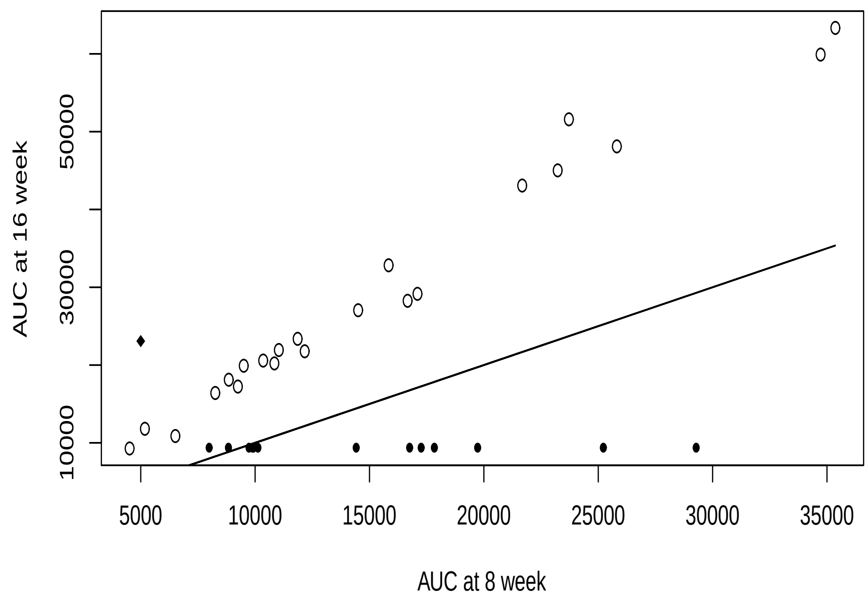

Figure 1: A scatter plot of CCI-779 cumulative AUCs against 8 and 16 weeks. The line is the 45 degree line. The solid circles correspond to patients who only had AUC at 8 week, while the solid diamond corresponds to the patient who only had AUC at 16 week

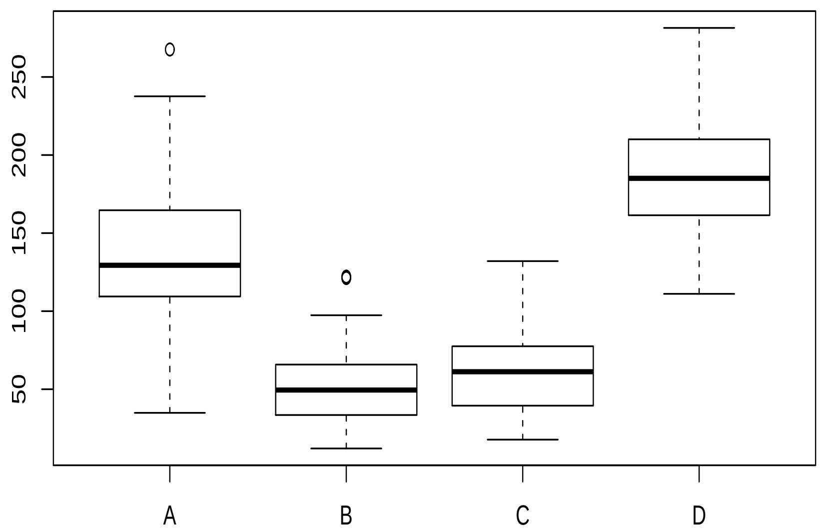

Figure 2: Prediction error results by 100 random splits of the advanced renal cancer data set. The procedures from A to D are GEES_IND, GEES_CS, GEES_AR1 and SIS

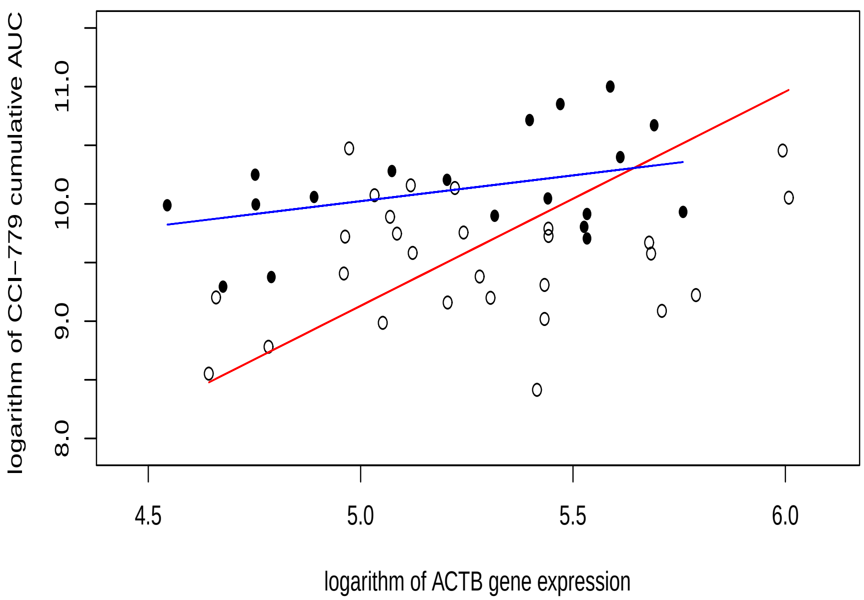

Figure 3: CCI-779 cumulative AUC versus ACTB gene expression level. The unfilled circles correspond to data at week 8, while the filled circles correspond to data at week 16. The red and blue lines denote the estimated regression lines for data points at 8 and 16 weeks, respectively

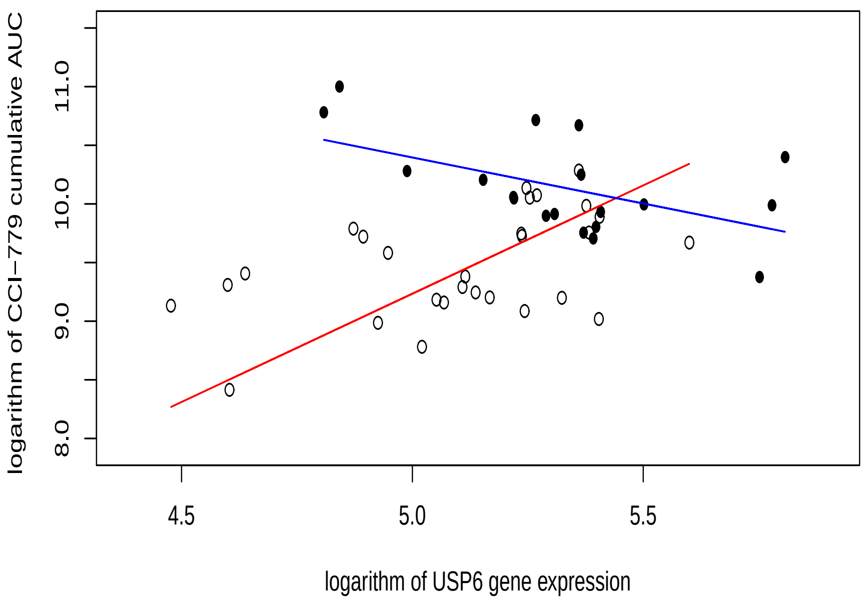

Figure 4: CCI-779 cumulative AUC versus USP6 gene expression level. The unfilled circles correspond to data at week 8, while the filled circles correspond to data at week 16. The red and blue lines denote the estimated regression lines for data points at 8 and 16 weeks, respectively

Table 1: The 5%, 25%, 50%, 75%, and 95% percentiles of the minimum model size and the average runtime in seconds (standard deviation) in Example 1 (with Xeon X5670 2.93 GHz CPU) when $SNR = 80\%$

p	ρ	Method	5%	25%	50%	75%	95%	TIME
1000	0.5	GEES_IND	4	4	6	17.25	174.05	0.04(0.01)
		GEES_CS	4	4	6	18	167.25	0.11(0.01)
		GEES_AR1	4	5	11.50	52.75	464.20	0.13(0.01)
		SIS	4	4	6	17.25	174.05	0.04(0.01)
		DC-SIS	30.80	136.75	300.50	559.50	883.10	1.12(0.04)
	0.8	GEES_IND	4	4	6	17	109.65	0.04(0.01)
		GEES_CS	4	4	5	11	68.05	0.10(0.01)
		GEES_AR1	4	5	11	38	222.40	0.12(0.01)
		SIS	4	4	6	17	109.65	0.04(0.01)
		DC-SIS	30.95	123.75	301	571	859.10	1.16(0.02)
6000	0.5	GEES_IND	4	5	15	97.25	924.40	0.28(0.02)
		GEES_CS	4	4	14	80.25	619.45	0.37(0.02)
		GEES_AR1	4	10	49.50	294.25	1762.10	0.39(0.02)
		SIS	4	5	15	97.25	924.40	0.28(0.02)
		DC-SIS	156.75	812.25	1910	3505.75	5408.40	6.86(0.11)
	0.8	GEES_IND	4	5	16	92.75	1231.55	0.30(0.02)
		GEES_CS	4	4	8	39	491.50	0.36(0.01)
		GEES_AR1	4	8	34	187.50	1206.95	0.38(0.02)
		SIS	4	5	16	92.75	1231.55	0.30(0.02)
		DC-SIS	118.85	734.25	1792	3355	5205.50	6.92(0.04)
20000	0.5	GEES_IND	4	9	41.50	301.50	2862.10	1.16(0.01)
		GEES_CS	4	7.75	35.50	186.50	1926.55	1.31(0.05)
		GEES_AR1	5	29	139	704.50	5524.50	1.30(0.02)
		SIS	4	9	41.50	301.50	2862.10	1.16(0.01)
		DC-SIS	343.45	2168	5653	10816.25	17624.50	23.11(0.16)
	0.8	GEES_IND	4	6	32.50	298.75	2996.25	1.13(0.02)
		GEES_CS	4	5	18	126.25	1850.45	1.36(0.03)
		GEES_AR1	4	13	106	780.50	4502.45	1.32(0.05)
		SIS	4	6	32.50	298.75	2996.25	1.13(0.02)
		DC-SIS	646.35	2974.50	6248	11632.50	17498.30	25.54(0.12)

Table 2: The number of selected informative genes (labeled "Model size") and the median of prediction errors ("PE") from 100 random splits for procedures in the advanced renal cancer data set. "IGEES" stands for the IGEES screening procedure with the PWLS variable selection method. "ISIS" stands for the ISIS procedure in Fan and Lv (2008), in which the SCAD method is used to refine the results

	Model size	PE
IGEES_IND	5	128.98
IGEES_CS	5	37.94
IGEES_AR1	6	56.75
ISIS	10	185.07

Table 3: The 5%, 25%, 50%, 75%, and 95% percentiles of the minimum model size and the average runtime in seconds (standard deviation) in heteroscedastic model (with Xeon X5670 2.93 GHz CPU)

Method	5%	25%	50%	75%	95%	TIME
GEES_IND	3	4	10	358	858.65	0.06(0.01)
GEES_CS	3	4	10	357.25	883.15	0.13(0.01)
GEES_AR1	3	4	10	358.25	862.20	0.14(0.01)
DC-SIS	3	4	5	12	153.45	1.18(0.03)