Gene-Environment Interaction Analysis Under the Cox Model

Kuangnan Fang[1], Jingmao Li[1], Yaqing Xu[2], Shuangge Ma[3], Qingzhao Zhang[1,4]

[1] *Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen, China*

[2] *School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, China*

[3] *Department of Biostatistics, Yale School of Public Health, New Haven, U.S.A.*

[4] *The Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China*

Supplementary Material

This file contains additional information for calculation (Section S1), the proof of Theorem 1 (Section S2), and additional numerical results (Section S3).

S1 Additional information for calculation

S1.1 Calculation for proximal mapping $\mathbf{Prox}(\mathbf{g}; \lambda_1, \lambda_2, \kappa)$

Recall that the proximal mapping problem is defined as:

$$\text{prox}(\mathbf{g}; \lambda_1, \lambda_2, \kappa) = \arg\min_{\mathbf{d}} \frac{1}{2}\|\mathbf{g} - \mathbf{d}\|_2^2 + P(\mathbf{d}; \lambda_1, \lambda_2, \kappa), \tag{13}$$

where $\mathbf{d}$ and $\mathbf{g}$ denote $(q+1)$-dimensional vectors, and the penalty $P(\mathbf{d}; \lambda_1, \lambda_2, \kappa) = \rho(\|\mathbf{d}\|_2; \sqrt{q+1}\lambda_1, \kappa) + \sum_{k=2}^{q+1} \rho(d_k; \lambda_2, \kappa)$.

We can characterize the optimal solution of (13) using the subgradient equation:

$$\mathbf{d} - \mathbf{g} + \mathbf{v} + \mathbf{w} = \mathbf{0}, \tag{14}$$

where $\mathbf{v}$ denotes the subgradient of $\rho(\|\mathbf{d}\|_2; \sqrt{q+1}\lambda_1, \kappa)$ with respect to $\mathbf{d}$. $\mathbf{w} = (0, w_2 \ldots, w_{q+1})^\top$ with w_k $(k = 2, \ldots, q+1)$ denoting the subgradient of $\rho(d_k; \lambda, \kappa)$ with respect to d_k. Specif-

ically, we have:

$$\mathbf{v} = \begin{cases} \rho'(\|\mathbf{d}\|_2; \sqrt{q+1}\lambda_1, \kappa)\mathbf{d}/\|\mathbf{d}\|_2 & \mathbf{d} \neq \mathbf{0} \\ \in \{\mathbf{v} : \|\mathbf{v}\|_2 \leq \sqrt{q+1}\lambda_1\} & \mathbf{d} = \mathbf{0} \end{cases},$$

$$w_k = \begin{cases} \rho'(d_k; \lambda_2, \kappa) & d_k \neq 0 \\ \in [-\lambda_2, \lambda_2] & d_k = 0 \end{cases},$$

where $\rho'(x; \lambda, \kappa) = \text{sign}(x)\,\text{I}(|x| < \kappa\lambda)(\lambda - |x|/\kappa)$ denotes the gradient function for the MCP penalty.

We can derive that subgradient equation (14) is satisfied with $\widehat{\mathbf{d}} = \mathbf{0}$ if the following inequality holds:

$$\|\,\text{ST}_{-1}(\mathbf{g}; \lambda)\|_2 \leq \sqrt{q+1}\lambda_1, \tag{15}$$

where $\text{ST}_{-1}(\mathbf{x}; \lambda) = (x_1, [sgn(\mathbf{x}_{-1}) \circ (|\mathbf{x}_{-1}| - \lambda)_+]^\top)^\top$.

Specifically, since inequality (15) holds, we can find a suitable $\mathbf{w}$ satisfying that $w_1 = 0$, $w_k \in [-\lambda_2, \lambda_2]$ $(k \in 2, \ldots, (q+1))$, and $\|\mathbf{g} - \mathbf{w}\|_2 \leq \sqrt{q+1}\lambda_1$, or equivalently $\mathbf{g} - \mathbf{w} = \mathbf{v}$ for some $\|\mathbf{v}\|_2 \leq \sqrt{q+1}\lambda_1$. This indicates that $\mathbf{d} = \mathbf{0}$ is a solution to equation (14).

Otherwise, when inequality (15) does not holds and $\mathbf{d} \neq \mathbf{0}$, subgradient function (14) becomes:

$$\mathbf{d} - \mathbf{g} + \rho'(\|\mathbf{d}\|_2; \sqrt{q+1}\lambda_1, \kappa)\frac{\mathbf{d}}{\|\mathbf{d}\|_2} + \mathbf{w} = \mathbf{0},$$

or equivalently

$$\mathbf{d} = \frac{\mathbf{g} - \mathbf{w}}{1 + \rho'(\|\mathbf{d}\|_2; \sqrt{q+1}\lambda_1, \kappa)/\|\mathbf{d}\|_2}. \tag{16}$$

Denote $\Psi = \{k = 1, \ldots, (q+1) : k = 1 \quad \text{or} \quad |g_k| > \lambda_2\}$ and its complement $\Psi^c = \{1, \ldots, (q+1)\} \setminus \Psi$. Then we have $d_k \neq 0$ for $k \in \Psi$ (since $|g_k - w_k| > |g_k| - \lambda_2 > 0$ for $k > 2$ and $k \in \Psi$) and $d_k = 0$ for $k \in \Psi^c$ (since $|g_k| \leq \lambda_2$, we let $d_k = 0$ and the subgradient $w_k = g_k$, and then equation (16) holds in its kth component). Based on these two conclusions, the solution

follows that:

$$\widehat{\mathbf{d}}_{\Psi^c} = \mathbf{0}, \tag{17}$$

$$\widehat{\mathbf{d}}_{\Psi} = \frac{\mathbf{g}_{\Psi} - \rho'(\widehat{\mathbf{d}}_{\Psi}; \lambda_2, \kappa)}{1 + \rho'(\|\widehat{\mathbf{d}}_{\Psi}\|_2; \sqrt{q+1}\lambda_1, \kappa)/\|\widehat{\mathbf{d}}_{\Psi}\|_2}. \tag{18}$$

For $\widehat{\mathbf{d}}_{\Psi}$, we first initialize $\widehat{\mathbf{d}}_{\Psi} = \mathbf{g}_{\Psi}$ and then iterate equation (18) until convergence. Note that, in all iteration steps, $\widehat{d}_k \neq 0$ for $k \in \Psi$, and thus the subgradient $\rho'(\widehat{\mathbf{d}}_{\Psi}; \lambda_2, \kappa)$ is in fact a gradient vector.

The overall procedure for calculating the proximal mapping is presented in Algorithm 1 in the main text.

S1.2 Convergence of the ADMM algorithm

We examine convergence properties of the ADMM algorithm. In Figure 3, for one simulation replicate, we provide the trace plots of the primal residual $\sum_{j=1}^{p} \|\mathbf{b}_j^{(m+1)} - \mathbf{d}_j^{(m+1)}\|_2^2$ and dual residual $\sum_{j=1}^{p} \|\mathbf{d}_j^{(m+1)} - \mathbf{d}_j^{(m)}\|_2^2$ over iterations. We see that the ADMM algorithm converges with a moderate number of iterations. The trend is similar for other simulated cases. The following Theorem 1 shows that the primal feasibility and dual feasibility are achieved by the algorithm, and the resulting solution can be a local minimum of the objective function.

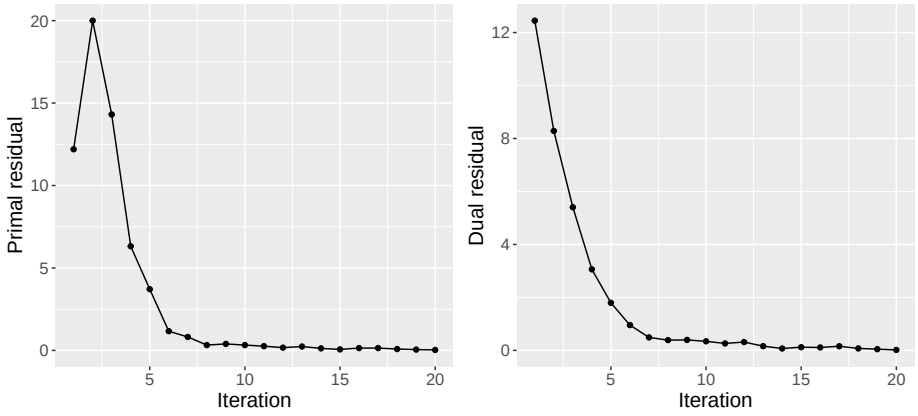

Figure 3: Trace plots of primal residual (left) and dual residual (right) over iterations for one simulation replicate. Data is generated under Example 1 and the AR(0.3) correlation structure. $\lambda_1 = 0.05$ and $\lambda_2 = 0.1$.

Theorem 1. *The primal residual* $\mathbf{r}^{(m+1)} = vec\left[\mathbf{B}^{(m+1)}\right] - vec\left[\mathbf{D}^{(m+1)}\right]$ *and the dual residual* $\mathbf{s}^{(m+1)} = vec\left[\mathbf{D}^{(m+1)}\right] - vec\left[\mathbf{D}^{(m)}\right]$ *of the ADMM algorithm satisfy* $\lim_{m\to\infty} \|\mathbf{r}^{(m+1)}\|_2^2 = 0$ *and* $\lim_{m\to\infty} \|\mathbf{s}^{(m+1)}\|_2^2 = 0$.

Proof. Recall $\tilde{Q}(\boldsymbol{\phi}, \mathbf{D})$ in (5) and the corresponding Lagrangian $\tilde{Q}_\psi(\boldsymbol{\phi}, \mathbf{D}, \boldsymbol{\Upsilon})$ in (6). By the definition $\mathbf{D}^{(m+1)} = \arg\min_{\mathbf{D}} \tilde{Q}_\psi(\boldsymbol{\phi}^{(m+1)}, \mathbf{D}, \boldsymbol{\Upsilon}^{(m)})$, for any $(q+1)\times p$- dimensional matrix $\mathbf{D}$:

$$\tilde{Q}_\psi(\boldsymbol{\phi}^{(m+1)}, \mathbf{D}^{(m+1)}, \boldsymbol{\Upsilon}^{(m)}) \leq \tilde{Q}_\psi(\boldsymbol{\phi}^{(m+1)}, \mathbf{D}, \boldsymbol{\Upsilon}^{(m)}). \tag{19}$$

Let $\mathbf{D} = \mathbf{B}^{(m+1)}$, and we have:

$$f^{(m+1)} \doteq \tilde{Q}_\psi(\boldsymbol{\phi}^{(m+1)}, \mathbf{B}^{(m+1)}, \boldsymbol{\Upsilon}^{(m)}) = \tilde{Q}(\boldsymbol{\phi}^{(m+1)}, \mathbf{B}^{(m+1)}) \geq \tilde{Q}_\psi(\boldsymbol{\phi}^{(m+1)}, \mathbf{D}^{(m+1)}, \boldsymbol{\Upsilon}^{(m)}). \tag{20}$$

It is noted that for any integer $t \geq 2$, the scaled dual variable $\boldsymbol{\upsilon}_j$ $(j = 1, \ldots, p)$ can be represented as $\boldsymbol{\upsilon}_j^{(m+t-1)} = \boldsymbol{\upsilon}_j^{(m)} + \sum_{i=1}^{t-1}\left(\mathbf{b}_j^{(m+i)} - \mathbf{d}_j^{(m+i)}\right)$. Thus we have:

$$\begin{aligned}
&\tilde{Q}_\psi(\boldsymbol{\phi}^{(m+t)}, \mathbf{D}^{(m+t)}, \boldsymbol{\Upsilon}^{(m+t-1)}) \\
=&\tilde{Q}(\boldsymbol{\phi}^{(m+t)}, \mathbf{D}^{(m+t)}) + \frac{\psi}{2}\sum_{j=1}^{p}\left(\|\mathbf{b}_j^{(m+t)} - \mathbf{d}_j^{(m+t)} + \boldsymbol{\upsilon}_j^{(m+t-1)}\|_2^2 - \|\boldsymbol{\upsilon}_j^{(m+t-1)}\|_2^2\right) \\
=&\tilde{Q}(\boldsymbol{\phi}^{(m+t)}, \mathbf{D}^{(m+t)}) + \frac{\psi}{2}\sum_{j=1}^{p}\left\{\|\mathbf{b}_j^{(m+t)} - \mathbf{d}_j^{(m+t)}\|_2^2 + 2\boldsymbol{\upsilon}_j^{(m)\top}\left[\mathbf{b}_j^{(m+t)} - \mathbf{d}_j^{(m+t)}\right]\right\} \\
&+ \psi\sum_{j=1}^{p}\sum_{i=1}^{t-1}\left[\mathbf{b}_j^{(m+t)} - \mathbf{d}_j^{(m+t)}\right]^\top\left[\mathbf{b}_j^{(m+i)} - \mathbf{d}_j^{(m+i)}\right] \leq f^{(m+t)}.
\end{aligned} \tag{21}$$

It is noted that function $\tilde{Q}_\psi(\boldsymbol{\phi}, \mathbf{D}, \boldsymbol{\Upsilon})$ is convex and differentiable with respect to $\boldsymbol{\phi}$. When ψ is sufficiently large, it is also convex with respect to $\mathbf{D}$. Thus by Theorem 4.1 in Tseng (2001), we conclude that the sequence $\{\boldsymbol{\phi}^{(m)}, \mathbf{D}^{(m)}\}$ has a limit point $\{\boldsymbol{\phi}^\dagger, \mathbf{D}^\dagger\}$. Then,

$$f^\dagger \doteq \lim_{m\to\infty} f^{(m+1)} = \lim_{m\to\infty} f^{(m+t)} = \tilde{Q}(\boldsymbol{\phi}^\dagger, \mathbf{B}^\dagger). \tag{22}$$

Meanwhile, we have:

$$\begin{aligned}&\lim_{m\to\infty}\tilde{Q}_{\psi}(\boldsymbol{\phi}^{(m+t)},\mathbf{D}^{(m+t)},\boldsymbol{\Upsilon}^{(m+t-1)})\\ =\ &\tilde{Q}(\boldsymbol{\phi}^{\dagger},\mathbf{D}^{\dagger})+\lim_{m\to\infty}\sum_{j=1}^{p}\psi\left\{\boldsymbol{\upsilon}_j^{(m)\top}\left[\mathbf{b}_j^{\dagger}-\mathbf{d}_j^{\dagger}\right]+(t-\frac{1}{2})\|\mathbf{b}_j^{\dagger}-\mathbf{d}_j^{\dagger}\|_2^2)\right\}\leq f^{\dagger}.\end{aligned}\tag{23}$$

To ensure that (23) holds for all integer $t>0$, we have that $\|\mathbf{b}_j^{\dagger}-\mathbf{d}_j^{\dagger}\|_2^2=0$ $(j=1,\ldots,p)$. Thus we can conclude that the primal residual satisfies $\lim_{m\to\infty}\|\mathrm{vec}(\mathbf{B}^{(m+1)})-\mathrm{vec}(\mathbf{D}^{(m+1)})\|_2^2=\lim_{m\to\infty}\sum_{j=1}^{p}\|\mathbf{b}_j^{(m+1)}-\mathbf{d}_j^{(m+1)}\|_2^2=\sum_{j=1}^{m}\|\mathbf{b}_j^{\dagger}-\mathbf{d}_j^{\dagger}\|_2^2=0$. Thus the primal feasibility is proved.

Moreover, $\boldsymbol{\phi}^{(m+1)}=\arg\min_{\boldsymbol{\phi}}\tilde{Q}_{\psi}(\boldsymbol{\phi},\mathbf{D}^{(m)},\Upsilon^{(m)})$ indidates that the following sub-gradient equation holds:

$$\begin{aligned}\mathbf{0}&=\frac{\partial\tilde{Q}_{\psi}(\boldsymbol{\phi}^{(m+1)},\mathbf{D}^{(m)},\Upsilon^{(m)})}{\partial\left[\mathrm{vec}(\mathbf{B}^{(m+1)})\right]}\\ &=\frac{\partial\tilde{Q}(\boldsymbol{\phi}^{(m+1)},\mathbf{D}^{(m)})}{\partial\left[\mathrm{vec}(\mathbf{B}^{(m+1)})\right]}+\psi\left[\mathrm{vec}(\mathbf{B}^{(m+1)})-\mathrm{vec}(\mathbf{D}^{(m)})+\mathrm{vec}(\Upsilon^{(m)})\right]\\ &=\frac{\partial\tilde{Q}(\boldsymbol{\phi}^{(m+1)},\mathbf{D}^{(m)})}{\partial\left[\mathrm{vec}(\mathbf{B}^{(m+1)})\right]}+\psi\mathrm{vec}(\Upsilon^{(m+1)})+\psi\left[\mathrm{vec}(\mathbf{D}^{(m+1)})-\mathrm{vec}(\mathbf{D}^{(m)})\right].\end{aligned}\tag{24}$$

Therefore, we have:

$$\mathbf{s}^{(m+1)}=\mathrm{vec}(\mathbf{D}^{(m+1)})-\mathrm{vec}(\mathbf{D}^{(m)})=-\frac{1}{\psi}\frac{\partial\tilde{Q}(\boldsymbol{\phi}^{(m+1)},\mathbf{D}^{(m)})}{\partial\left[\mathrm{vec}(\mathbf{B}^{(m+1)})\right]}-\mathrm{vec}(\Upsilon^{(m+1)}).\tag{25}$$

It is noted that, at the limiting point, $\|\mathbf{b}_j^{\dagger}-\mathbf{d}_j^{\dagger}\|_2^2=0$ $(j=1,\ldots,p)$. Thus we also have:

$$\mathbf{0}=\lim_{m\to\infty}\frac{\partial\tilde{Q}_{\psi}(\boldsymbol{\phi}^{(m+1)},\mathbf{D}^{(m)},\Upsilon^{(m)})}{\partial\left[\mathrm{vec}(\mathbf{B}^{\dagger})\right]}=\frac{\partial\tilde{Q}(\boldsymbol{\phi}^{\dagger},\mathbf{D}^{\dagger})}{\partial\left[\mathrm{vec}(\mathbf{B}^{\dagger})\right]}+\psi\mathrm{vec}(\Upsilon^{\dagger}).\tag{26}$$

This indicates that $\lim_{m\to\infty}\mathbf{s}^{(m+1)}=\mathbf{0}$. This concludes the proof. □

S2 Proof of Theorem 1

S2.1 Proof of results (a) and (b)

Proof. Define the oracle estimator $\widehat{\boldsymbol{\phi}}^o$ as $\widehat{\boldsymbol{\phi}}_2^o = \mathbf{0}$ and $\widehat{\boldsymbol{\phi}}_1^o = \arg\max_{\boldsymbol{\phi}_1} \mathcal{L}_n^o(\boldsymbol{\phi}_1)$, where

$$\mathcal{L}_n^o(\boldsymbol{\phi}_1) = n^{-1} \sum_{i=1}^{n} \int_0^\tau \left\{ \mathbf{A}_{i,\Psi} \boldsymbol{\phi}_1 - \log \left(\sum_{i'=1}^{n} \mathcal{Y}_{i'}(t) \exp(\mathbf{A}_{i',\Psi} \boldsymbol{\phi}_1) \right) \right\} \mathrm{d}N_i(t). \tag{27}$$

Our proof contains two steps. In the first step, we establish estimation consistency of $\widehat{\boldsymbol{\phi}}^o$. And in the second step, we show that with probability tending to 1, $\widehat{\boldsymbol{\phi}}^o$ is a local minimizer of objective function $\mathcal{Q}_n(\boldsymbol{\phi})$.

Step 1. We show that $\left\| \widehat{\boldsymbol{\phi}}^o - \boldsymbol{\phi}^* \right\|_2 = \left\| \widehat{\boldsymbol{\phi}}_1^o - \boldsymbol{\phi}_1^* \right\|_2 = O_p\left(\sqrt{s/n} \right)$. It is sufficient to show that, for any $\epsilon > 0$, there exists a large constant B such that:

$$\Pr\left(\sup_{\boldsymbol{\phi}_1 \in \mathcal{I}} \mathcal{L}_n^o(\boldsymbol{\phi}_1) \geq \mathcal{L}_n^o(\boldsymbol{\phi}_1^*) \right) \geq 1 - \epsilon, \tag{28}$$

where $\mathcal{I} = \{\boldsymbol{\phi}_1 : \|\boldsymbol{\phi}_1 - \boldsymbol{\phi}_1^*\|_2 = B\sqrt{s/n}\}$. This indicates that, with probability tending to 1, function $\mathcal{L}_n^o(\boldsymbol{\phi}_1)$ has a local maximizer $\widehat{\boldsymbol{\phi}}_1^o$ within the ball $\{\boldsymbol{\phi}_1 : \|\boldsymbol{\phi}_1 - \boldsymbol{\phi}_1^*\|_2 \leq B\sqrt{s/n}\}$. That is, $\widehat{\boldsymbol{\phi}}_1^o$ satisfies $\|\widehat{\boldsymbol{\phi}}_1^o - \boldsymbol{\phi}_1^*\|_2 = O_p(\sqrt{s/n})$. The proof of Step 1 follows the strategy of Bradic et al (2011). The main steps are as follows.

For $\boldsymbol{\phi}_1 \in \mathcal{I}$, we rewrite $\boldsymbol{\phi}_1 = \boldsymbol{\phi}_1^* + \gamma_n \mathbf{u}$, where $\gamma_n = B\sqrt{s/n}$ and $\|\mathbf{u}\|_2 = 1$. The first and second order derivatives of $\mathcal{L}_n^o(\boldsymbol{\phi}_1)$ are:

$$\mathcal{U}_n(\boldsymbol{\phi}_1) = \frac{\partial \mathcal{L}_n^o(\boldsymbol{\phi}_1)}{\partial \boldsymbol{\phi}_1} = n^{-1} \sum_{i=1}^{n} \int_0^\tau \left\{ \mathbf{A}_{i,\Psi}^\top - \mathbf{H}_{n1}(t, \boldsymbol{\phi}_1) \right\} \mathrm{d}N_i(t),$$

$$\partial \mathcal{U}_n(\boldsymbol{\phi}_1) = \frac{\partial^2 \mathcal{L}_n^o(\boldsymbol{\phi}_1)}{\partial \boldsymbol{\phi}_1 \partial \boldsymbol{\phi}_1^\top} = n^{-1} \sum_{i=1}^{n} \int_0^\tau \mathbf{V}_{n11}(t, \boldsymbol{\phi}_1) \mathrm{d}N_i(t).$$

With Taylor expansion, we have:

$$\begin{aligned} \mathcal{L}_n^o(\boldsymbol{\phi}_1) - \mathcal{L}_n^o(\boldsymbol{\phi}_1^*) &= \mathcal{L}_n^o(\boldsymbol{\phi}_1^* + \gamma_n \mathbf{u}) - \mathcal{L}_n^o(\boldsymbol{\phi}_1^*) \\ &= \gamma_n \mathbf{u}^\top \mathcal{U}_n(\boldsymbol{\phi}_1^*) + \frac{1}{2} \gamma_n^2 \mathbf{u}^\top \partial \mathcal{U}_n(\boldsymbol{\phi}_1^*) \mathbf{u} + r_n, \end{aligned}$$

where $r_n = O_p(\gamma_n^3)$.

With Conditions 1, 3, 4 and based on Lemmas 2.2, 2.3, 4.1 in Bradic et al (2011), we have $\|\mathcal{U}_n(\boldsymbol{\phi}_1^*)\|_2 = O_p(\sqrt{s/n})$ and $\|\partial\mathcal{U}_n(\boldsymbol{\phi}_1^*) - \boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)\|_2 = o_p(1)$. It follows that:

$$\mathbf{u}^\top \partial\mathcal{U}_n(\boldsymbol{\phi}_1^*)\mathbf{u} = \mathbf{u}^\top \left\{\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*) + [\partial\mathcal{U}_n(\boldsymbol{\phi}_1^*) - \boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)]\right\}\mathbf{u} \geq \sigma_{\min}\left[\boldsymbol{\Sigma}(\boldsymbol{\phi_1}^*)\right] + o_p(1).$$

Overall, we have:

$$\begin{aligned}\mathcal{L}_n^o(\boldsymbol{\phi}_1) - \mathcal{L}_n^o(\boldsymbol{\phi}_1^*) &= \gamma_n \mathbf{u}^\top \mathcal{U}_n(\boldsymbol{\phi}_1^*) + \frac{1}{2}\gamma_n^2 \mathbf{u}^\top \partial\mathcal{U}_n(\boldsymbol{\phi}_1^*)\mathbf{u} + r_n \\ &\geq \gamma_n \left[\left\{\sigma_{\min}\left[\boldsymbol{\Sigma}(\boldsymbol{\phi_1}^*)\right] + o_p(1)\right\}\gamma_n - O_p(\sqrt{s/n})\right].\end{aligned}$$

Then with probability tending to 1, $\mathcal{L}_n^o(\boldsymbol{\phi}_1) - \mathcal{L}_n^o(\boldsymbol{\phi}_1^*) > 0$ for $\gamma_n = B\sqrt{s/n}$ with a sufficiently large constant B, which implies (28).

Step 2. We show that with probability tending to 1, the oracle estimator $\widehat{\boldsymbol{\phi}}^o$ is a local minimizer of $\mathcal{Q}_n(\boldsymbol{\phi})$. To this end, we need to verify the following first order Karush-Kuhn-Tucher (KKT) conditions with respect to $\widehat{\boldsymbol{\phi}}^o = (\widehat{\mathbf{b}}_1^{o\top}, \ldots, \widehat{\mathbf{b}}_p^{o\top}, \widehat{\boldsymbol{\eta}}^{o\top})^\top$:

$$\left.\frac{\partial\mathcal{Q}_n(\boldsymbol{\phi})}{\partial\boldsymbol{\eta}}\right|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}^o} = -\mathcal{U}_{n,\boldsymbol{\eta}}(\widehat{\boldsymbol{\phi}}_1^o) = \mathbf{0}, \tag{29}$$

$$\left.\frac{\partial\mathcal{Q}_n(\boldsymbol{\phi})}{\partial\mathbf{b}_{j,\Xi_j}}\right|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}^o} = -\mathcal{U}_{n,j}(\widehat{\boldsymbol{\phi}}_1^o) + \frac{\partial P(\widehat{\mathbf{b}}_j^o;\lambda_1,\lambda_2,\kappa)}{\partial\widehat{\mathbf{b}}_{j,\Xi_j}^o} = \mathbf{0}, j \in \Phi, \tag{30}$$

$$\left.\frac{\partial\mathcal{Q}_n(\boldsymbol{\phi})}{\partial\mathbf{b}_{j,\Xi_j^c}}\right|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}^o} = -\mathcal{V}_{n,j}(\widehat{\boldsymbol{\phi}}_1^o) + \frac{\partial P(\widehat{\mathbf{b}}_j^o;\lambda_1,\lambda_2,\kappa)}{\partial\widehat{\mathbf{b}}_{j,\Xi_j^c}^o} = \mathbf{0}, j \in \Phi, \tag{31}$$

$$\left.\frac{\partial\mathcal{Q}_n(\boldsymbol{\phi})}{\partial\mathbf{b}_j}\right|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}^o} = -\mathcal{V}_{n,j}(\widehat{\boldsymbol{\phi}}_1^o) + \frac{\partial P(\widehat{\mathbf{b}}_j^o;\lambda_1,\lambda_2,\kappa)}{\partial\widehat{\mathbf{b}}_j^o} = \mathbf{0}, j \in \Phi^c, \tag{32}$$

where $\mathcal{U}_{n,\boldsymbol{\eta}}(\boldsymbol{\phi}_1)$ denotes the sub-vector of $\mathcal{U}_n(\boldsymbol{\phi}_1)$ corresponding to $\boldsymbol{\eta}$, and $\mathcal{U}_{n,j}(\boldsymbol{\phi}_1)$, $j \in \Phi_1$ denotes the sub-vector of $\mathcal{U}_n(\boldsymbol{\phi}_1)$ corresponding to $\mathbf{b}_{j,\Xi_j}$. We also denote $\mathcal{V}_n(\boldsymbol{\phi}) = \partial\mathcal{L}_n(\boldsymbol{\phi})/\partial\boldsymbol{\phi}_2$ and use $\mathcal{V}_{n,j}(\boldsymbol{\phi})$ to denote the corresponding sub-vector with respect to $\mathbf{b}_{j,\Xi_j^c}$ for $j \in \Phi$ or $\mathbf{b}_j$ for $j \in \Phi^c$.

With the specific definition of the oracle estimate $\widehat{\boldsymbol{\phi}}^o$, equation (29) is already satisfied. For equation (30), considering the minimal signal condition (Condition 6), we can conclude

that for $j \in \Phi$, $\|\widehat{\mathbf{b}}_{j,\Xi_j}\|_2 \geq b_{\min} - \|\widehat{\mathbf{b}}_{j,\Xi_j} - \mathbf{b}^*_{j,\Xi_j}\|_2 \gg \lambda_1$, and $|\widehat{b}_{j,k}| \geq \mu_{\min} - |\widehat{b}^o_{j,k} - b^*_{j,k}| \gg \lambda_2$, $k \in \Xi_j \setminus \{1\}$. Based on these two results and the properties of the folded-concave penalty function (Condition 7) $\rho'(x;\lambda,\kappa) = 0$ for $|x| \geq \kappa\lambda$, we obtain:

$$\frac{\partial P(\widehat{\mathbf{b}}^o_j;\lambda_1,\lambda_2,\kappa)}{\partial \widehat{b}^o_{j,1}} = \frac{\rho'(\|\widehat{\boldsymbol{b}}^o_{j,\Xi_j}\|_2;\sqrt{q+1}\lambda_1,\kappa)\widehat{b}^o_{j,1}}{\|\widehat{\boldsymbol{b}}^o_{j,\Xi_j}\|_2} = 0,$$
$$\frac{\partial P(\widehat{\mathbf{b}}^o_j;\lambda_1,\lambda_2,\kappa)}{\partial \widehat{b}^o_{j,k}} = \frac{\rho'(\|\widehat{\boldsymbol{b}}^o_{j,\Xi_j}\|_2;\sqrt{q+1}\lambda_1,\kappa)\widehat{b}^o_{j,k}}{\|\widehat{\boldsymbol{b}}^o_{j,\Xi_j}\|_2} + \rho'(\widehat{b}^o_{j,k};\lambda_2,\kappa) = 0, \quad k \in \Xi_j \setminus \{1\}.$$

These two equations imply that $\partial P(\widehat{\mathbf{b}}^o_j;\lambda_1,\lambda_2,\kappa)/\partial \widehat{\mathbf{b}}^o_{j,\Xi_j} = \mathbf{0}$ $(j \in \Phi)$. Additionally, we have $\mathcal{U}_{n,j}(\widehat{\boldsymbol{\phi}}^o_1) = \mathbf{0}$ $(j \in \Phi)$ by the definition of the oracle estimate. Then we conclude that equation (30) holds.

For equation (31), we have that for $j \in \Phi$, $\partial P(\widehat{\mathbf{b}}^o_j;\lambda_1,\lambda_2,\kappa)/\partial \widehat{\mathbf{b}}^o_{j,\Xi^c_j} = \lambda_2 \mathbf{w}_j$ with $\|\mathbf{w}_j\|_\infty \leq 1$. Then we can express equation (31) as $\mathcal{V}_{n,j}(\widehat{\boldsymbol{\phi}}^o_1) = \lambda_2 w_j$ or equivalently:

$$\|\mathcal{V}_{n,j}(\widehat{\boldsymbol{\phi}}^o_1)\|_\infty \leq \lambda_2, \quad j \in \Phi, \tag{33}$$

where $\|\cdot\|_\infty$ denotes the infinite norm.

For equation (32), we have that, for $j \in \Phi^c$, the subgradient $\partial P(\widehat{\mathbf{b}}^o_j;\lambda_1,\lambda_2,\kappa)/\partial \widehat{\mathbf{b}}^o_j = \sqrt{q+1}\lambda_1 \mathbf{v}_j + \lambda_2 \bar{\mathbf{w}}_j$ with $\|\mathbf{v}_j\|_2 \leq 1$, $\|\bar{\mathbf{w}}_j\|_\infty \leq 1$, and $\bar{w}_{j,1} = 0$. Hence, we can reformulate equation (32) as $\mathcal{V}_{n,j}(\widehat{\boldsymbol{\phi}}^o_1) - \lambda_2 \bar{\mathbf{w}}_j = \sqrt{q+1}\lambda_1 \mathbf{v}_j$, which is equivalent to:

$$\|\mathcal{V}_{n,j}(\widehat{\boldsymbol{\phi}}^o_1) - \lambda_2 \bar{\mathbf{w}}_j\|_2 \leq \sqrt{q+1}\lambda_1, \quad j \in \Phi^c.$$

Additionally, by considering the range of $\bar{\mathbf{w}}_j$, we can further express as:

$$\|\mathrm{ST}_{-1}(\mathcal{V}_{n,j}(\widehat{\boldsymbol{\phi}}^o_1);\lambda_2)\|_2 \leq \sqrt{q+1}\lambda_1, \quad j \in \Phi^c, \tag{34}$$

where $\mathrm{ST}_{-1}(\mathbf{x};\lambda) = (x_1, [sgn(\mathbf{x}_{-1}) \circ (|\mathbf{x}_{-1}| - \lambda)_+]^\top)^\top$ is a special type of soft-thresholding function in that soft-thresholding shrinkage is applied to all elements except for the first one

of vector $\mathbf{x}$. For inequality (34), it is sufficient to verify that:

$$\|\mathcal{V}_{n,j}(\widehat{\phi}_1^o)\|_\infty \leq \lambda_1, \quad j \in \Phi^c. \tag{35}$$

Based on the above, we are left to verify (33) and (35), for which it is sufficient to show that $\|\mathcal{V}_n(\widehat{\phi}^o)\|_\infty \leq \min\{\lambda_1, \lambda_2\}$.

We define the sequence $u_n = n^{0.5\zeta+\alpha_1}$ and event $\mathcal{M} = \{\|\mathcal{V}_n(\boldsymbol{\phi}_1^*)\|_\infty \leq u_n/\sqrt{n}\}$. Based on Conditions 1, 2 and the assumption $\max_j(\sigma_j^2) = O(n^{0.5\zeta+\alpha_1}) = O(u_n)$, we have that Theorem 3.1 of Bradic et al (2011) holds, which implies that there exist positive constants c_0 and c_1 such that:

$$\Pr(|\xi_{j'}| > u_n) \leq c_0 e^{-c_1 u_n}, \quad j' = 1, \ldots, [p(q+1)+q-s],$$

where $\xi_{j'}$ denotes the j'th element of $\mathcal{V}_n(\boldsymbol{\phi}_1^*)$. In this case, by using the Bonferroni union bound formula, we can derive that:

$$\Pr(\mathcal{M}) = \Pr\left(\max_{j'} |\xi_{j'}| \leq u_n\right) \geq 1 - c_0[p(q+1)+q-s]e^{-c_1 u_n} \to 1,$$

since $\log p = O(n^{\zeta_0})$, $u_n = n^{0.5\zeta+\alpha_1}$, and $\zeta_0 < 0.5\zeta + \alpha_1$.

Under event $\mathcal{M}$ and Condition 5, we follow the proof of Theorem 4.3 in Bradic et al (2011) and obtain that:

$$\begin{aligned}
\|\mathcal{V}_n(\widehat{\boldsymbol{\phi}}_1^o)\|_\infty &\leq \|\mathcal{V}_n(\boldsymbol{\phi}_1^*)\|_\infty + \|\mathcal{V}_n(\widehat{\boldsymbol{\phi}}_1^o) - \mathcal{V}_n(\boldsymbol{\phi}_1^*)\|_\infty \\
&= O_p(u_n/\sqrt{n}) + O_p\left(\sup_{0\leq t\leq\infty} \sup_{\boldsymbol{\phi}_1 \in \mathcal{B}_1(\boldsymbol{\phi}_1^*, \boldsymbol{\phi}_{\min})} \|\boldsymbol{V}_{n21}(t, \boldsymbol{\phi}_1)\|_{2,\infty} \|\widehat{\boldsymbol{\phi}}_1^o - \boldsymbol{\phi}_1^*\|_2\right) \\
&= O_p(n^{0.5\zeta+\alpha_1-0.5}).
\end{aligned}$$

Under assumptions $\lambda_1 \gg n^{0.5\zeta+\alpha_1-0.5}$ and $\lambda_2 \gg n^{0.5\zeta+\alpha_1-0.5}$, we have that $\|\mathcal{V}_n(\widehat{\boldsymbol{\phi}}^o)\|_\infty \ll \min\{\lambda_1, \lambda_2\}$ holds with probability tending to 1. This concludes the proof. □

S2.2 Proof of result (c)

Proof. Based on results (a) and (b), we have that the estimate $\widehat{\boldsymbol{\phi}}_1 = \arg\max_{\boldsymbol{\phi}_1} \mathcal{L}_n^o(\boldsymbol{\phi}_1)$, and it satisfies $\mathcal{U}_n(\widehat{\boldsymbol{\phi}}_1) = \mathbf{0}$. With Taylor's expansion at the true value $\boldsymbol{\phi}_1^*$, we have:

$$\mathcal{U}_n(\widehat{\boldsymbol{\phi}}_1) = \mathcal{U}_n(\boldsymbol{\phi}_1^*) + \partial\mathcal{U}_n(\boldsymbol{\phi}_1^*)\left(\widehat{\boldsymbol{\phi}}_1 - \boldsymbol{\phi}_1^*\right) + \mathbf{r}'_n, \tag{36}$$

where $\mathbf{r}'_n$ is a s-dimensional residual vector, its jth element is defined as:

$$\boldsymbol{r}'_{n,j} = \left(\widehat{\boldsymbol{\phi}}_1 - \boldsymbol{\phi}_1^*\right)^\top \frac{\partial^2 \mathcal{U}_{n,j}(\tilde{\boldsymbol{\phi}}_1)}{\partial\tilde{\boldsymbol{\phi}}_1 \partial\tilde{\boldsymbol{\phi}}_1^\top}\left(\widehat{\boldsymbol{\phi}}_1 - \boldsymbol{\phi}_1^*\right), \tag{37}$$

with $\tilde{\boldsymbol{\phi}}_1$ being a vector locating between $\widehat{\boldsymbol{\phi}}_1$ and $\boldsymbol{\phi}_1^*$, and $\mathcal{U}_{n,j}(\tilde{\boldsymbol{\phi}}_1)$ denotes the jth element of $\mathcal{U}_n(\tilde{\boldsymbol{\phi}}_1)$. Following (36), we can derive that:

$$\begin{aligned}\sqrt{n}\boldsymbol{\nu}_n^\top \boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{1/2}\left(\widehat{\boldsymbol{\phi}}_1 - \boldsymbol{\phi}_1^*\right) =& \boldsymbol{\nu}_n^\top \boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{1/2}\left[-\partial\mathcal{U}_n(\boldsymbol{\phi}_1^*)\right]^{-1}\left[n^{1/2}\mathcal{U}_n(\boldsymbol{\phi}_1^*)\right] \\ &+ \boldsymbol{\nu}_n^\top \boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{1/2}\left[-\partial\mathcal{U}_n(\boldsymbol{\phi}_1^*)\right]^{-1}\left[n^{1/2}\mathbf{r}'_n\right] \\ \doteq& \omega_{n1} + \omega_{n2}.\end{aligned} \tag{38}$$

We note that $\left\|\frac{\partial^2 \mathcal{U}_{n,j}(\tilde{\boldsymbol{\phi}}_1)}{\partial\tilde{\boldsymbol{\phi}}_1\partial\tilde{\boldsymbol{\phi}}_1^\top}\right\|_2 = O_p(1)$. Thus by adopting the Cauchy-Schwarz inequality in (37), we can derive that:

$$|r'_{n,j}| \leq \|\widehat{\boldsymbol{\phi}}_1 - \boldsymbol{\phi}_1^*\|_2^2 O_p(1) = O_p(s/n), \tag{39}$$

and thus $\|n^{1/2}\mathbf{r}'_n\|_2 = O_p(s^{3/2}/n^{1/2}) = o_p(1)$. Furthermore, based on Condition 4 and Lemmas 2.3 and 4.1 in Bradic et al (2011), we have that $\|\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)\|_2 \asymp 1$ and $\|\partial\mathcal{U}_n(\boldsymbol{\phi}_1^*) - \boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)\|_2 = o_p(1)$. It follows from the Cauchy-Schwarz inequality that:

$$|\omega_{n2}| \leq \|\boldsymbol{\nu}_n\|_2 \|\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{1/2}\|_2 \left[\|\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)\|_2 + \|\partial\mathcal{U}_n(\boldsymbol{\phi}_1^*) - \boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)\|_2\right]^{-1} \|n^{1/2}\mathbf{r}'_n\|_2 = o_p(1).$$

On the other hand, since $\|\partial\mathcal{U}_n(\boldsymbol{\phi}_1^*) - \boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)\|_2 = o_p(1)$, we have that $\omega_{n1} \to_P \omega'_{n1}$ with

$$\omega'_{n1} = \boldsymbol{\nu}_n^\top \boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{-1/2}\left[n^{1/2}\mathcal{U}_n(\boldsymbol{\phi}_1^*)\right]. \tag{40}$$

It is now sufficient to prove the asymptotic normality of ω'_{n1}. Recall that:

$$\begin{aligned}\mathcal{U}_n(\boldsymbol{\phi}_1^*) &= n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{A}_{i,\Psi}^{\top}-\mathbf{H}_{n1}(t,\boldsymbol{\phi}_1^*)\right\}\mathrm{d}N_i(t)\\ &= n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{A}_{i,\Psi}^{\top}-\mathbf{H}_{n1}(t,\boldsymbol{\phi}_1^*)\right\}\mathrm{d}M_i(t),\end{aligned} \tag{41}$$

where $M_i(t)=N_i(t)-\Lambda_i(t)$ is a martingale on the time interval $[0,\tau]$, and the second equality follows from the fact that:

$$\begin{aligned}&\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{A}_{i,\Psi}^{\top}-\mathbf{H}_{n1}(t,\boldsymbol{\phi}_1^*)\right\}\mathrm{d}\Lambda_i(t)\\ =&\int_0^{\tau}\sum_{i=1}^{n}\left[\mathbf{A}_{i,\Psi}^{\top}\mathcal{Y}_i(t)\exp\{\mathbf{A}_{i,\Psi}\boldsymbol{\phi}_1^*\}\right]\mathrm{d}t-\int_0^{\tau}n\mathbf{H}_{n1}(t,\boldsymbol{\phi}_1^*)n^{-1}\sum_{i=1}^{n}\left[\mathcal{Y}_i(t)\exp\{\mathbf{A}_{i,\Psi}\boldsymbol{\phi}_1^*\}\right]\mathrm{d}t\\ =&\int_0^{\tau}nS_{n1}^{(1)}(\tau,\boldsymbol{\phi}_1^*)\mathrm{d}t-\int_0^{\tau}n\mathbf{H}_{n1}(t,\boldsymbol{\phi}_1^*)S_{n1}^{(0)}(\tau,\boldsymbol{\phi}_1^*)\mathrm{d}t=\mathbf{0}.\end{aligned}$$

Based on (41), we have that $E\left[\mathcal{U}_n(\boldsymbol{\phi}_1^*)\right]=\mathbf{0}$, $E(w'_{n1})=0$ and

$$\begin{aligned}\mathrm{Var}(w'_{n1}) &= \mathrm{E}\left[\sum_{i=1}^{n}\int_0^{\tau}\boldsymbol{\nu}_n^{\top}\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{-1/2}\frac{\left\{\mathbf{A}_{i,\Psi}^{\top}-\mathbf{H}_{n1}(t,\boldsymbol{\phi}_1^*)\right\}^{\otimes 2}}{n}\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{-1/2}\boldsymbol{\nu}_n\mathrm{d}\Lambda_i(t)\right]\\ &= \boldsymbol{\nu}_n^{\top}\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{-1/2}\mathrm{E}\left[\int_0^{\tau}\frac{\sum_{i=1}^{n}\left\{\mathbf{A}_{i,\Psi}^{\top}-\mathbf{H}_{n1}(t,\boldsymbol{\phi}_1^*)\right\}^{\otimes 2}\lambda_i(t)}{n}\mathrm{d}t\right]\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{-1/2}\boldsymbol{\nu}_n\\ &= \boldsymbol{\nu}_n^{\top}\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{-1/2}\mathrm{E}\left[\int_0^{\tau}\mathbf{V}_{n11}\left(t,\boldsymbol{\phi}_1^*\right)S_n^{(0)}(t,\boldsymbol{\phi}^*)\mathrm{d}t\right]\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{-1/2}\boldsymbol{\nu}_n.\end{aligned}$$

By Lemma 4.1 in Bradic et al (2011), we can obtain that $\mathrm{E}\left[\int_0^{\tau}\mathbf{V}_{n11}\left(t,\boldsymbol{\phi}_1^*\right)S_n^{(0)}(t,\boldsymbol{\phi}^*)\mathrm{d}t\right]=\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)$, which leads to $\mathrm{Var}(w'_{n1})=1$. Therefore, by using the central limit theorem (Andersen and Gill, 1982; Fleming and Harrington, 2011), we get that $\omega'_{n1}\to_d\mathcal{N}(0,1)$. Combining the above results, we have that $\sqrt{n}\boldsymbol{\nu}_n^{\top}\boldsymbol{\Sigma}(\boldsymbol{\phi}_1^*)^{1/2}\left(\widehat{\boldsymbol{\phi}}_1-\boldsymbol{\phi}_1^*\right)\to_d\mathcal{N}(0,1)$, which concludes the proof.

□

S3 Additional numeric results

S3.1 Additional simulation Results

Table 3: Simulation results under the Band1 and Band2 correlation structures. In each cell, mean (sd) based on 100 replicates.

Example	Correlation	Method	Main effects (M)		Interactions (I)		SSE	C-index
			TP	FP	TP	FP		
1	Band1	Proposed	12.8(2.5)	0.2(0.4)	5.8(2.3)	1.4(1.4)	5.030(1.533)	0.827(0.031)
		MCP	8.2(2.6)	0.1(0.4)	5.9(2.1)	3.5(2.8)	7.082(1.147)	0.788(0.034)
		grMCP	8.2(3.8)	0.0(0.0)	9.4(1.9)	31.8(17.1)	6.887(1.539)	0.784(0.046)
		Marginal	1.0(1.5)	3.5(5.6)	1.9(1.4)	27.1(35.3)	-(-)	-(-)
	Band2	Proposed	13.4(2.2)	1.2(1.2)	5.8(2.8)	2.8(2.3)	5.936(1.342)	0.819(0.031)
		MCP	8.2(1.6)	0.8(0.8)	5.7(2.6)	4.7(3.0)	7.755(1.038)	0.783(0.031)
		grMCP	7.3(3.9)	0.3(1.1)	8.2(2.2)	30.0(21.8)	7.951(1.445)	0.767(0.048)
		Marginal	1.1(1.4)	4.2(5.2)	1.3(1.4)	31.7(34.5)	-(-)	-(-)
2	Band1	Proposed	7.3(0.7)	0.2(0.6)	5.1(1.0)	1.8(1.6)	2.315(0.635)	0.842(0.021)
		MCP	6.0(1.0)	0.2(0.5)	4.7(1.3)	3.3(1.4)	2.738(0.577)	0.810(0.025)
		grMCP	7.1(1.0)	0.1(0.4)	5.9(0.4)	30.4(4.8)	2.773(0.440)	0.837(0.019)
		Marginal	0.8(1.0)	3.4(4.4)	2.4(1.3)	31.0(31.4)	-(-)	-(-)
	Band2	Proposed	7.1(0.6)	0.3(0.5)	4.6(1.1)	2.1(1.5)	2.910(0.733)	0.850(0.019)
		MCP	4.8(1.3)	0.2(0.4)	4.2(1.3)	3.8(1.9)	3.268(0.903)	0.820(0.022)
		grMCP	6.7(0.9)	0.6(0.5)	6.0(0.2)	30.3(4.8)	3.212(0.798)	0.845(0.020)
		Marginal	0.9(1.0)	3.8(5.1)	1.8(1.3)	35.3(38.5)	-(-)	-(-)
3	Band1	Proposed	10.9(1.0)	0.2(0.4)	0.0(0.0)	2.0(1.4)	1.554(0.556)	0.843(0.016)
		MCP	9.3(1.3)	0.2(0.4)	0.0(0.0)	3.9(2.2)	2.534(0.527)	0.807(0.012)
		grMCP	8.9(3.0)	0.0(0.0)	0.0(0.0)	44.8(15.0)	3.202(0.718)	0.807(0.037)
		Marginal	1.2(1.6)	4.3(7.5)	0.0(0.0)	25.9(36.5)	-(-)	-(-)
	Band2	Proposed	11.3(1.0)	0.5(0.8)	0.0(0.0)	2.1(1.7)	1.929(0.610)	0.854(0.017)
		MCP	10.1(1.5)	0.5(0.6)	0.0(0.0)	3.4(2.4)	2.730(0.661)	0.819(0.018)
		grMCP	8.4(3.0)	0.2(0.5)	0.0(0.0)	43.0(16.3)	3.625(0.836)	0.809(0.046)
		Marginal	1.6(1.7)	5.7(8.8)	0.0(0.0)	34.1(43.3)	-(-)	-(-)
4	Band1	Proposed	4.0(0.0)	0.2(0.7)	12.6(3.9)	0.1(0.3)	4.284(1.314)	0.860(0.013)
		MCP	3.6(0.6)	0.0(0.0)	10.1(3.1)	0.3(0.8)	4.824(0.856)	0.845(0.017)
		grMCP	4.0(0.0)	0.0(0.0)	20.0(0.0)	0.0(0.0)	2.407(0.465)	0.878(0.011)
		Marginal	0.2(0.4)	2.8(3.9)	1.6(1.6)	32.2(30.8)	-(-)	-(-)
	Band2	Proposed	4.0(0.0)	0.1(0.3)	13.1(2.8)	0.0(0.0)	4.548(1.451)	0.869(0.013)
		MCP	2.5(1.1)	0.1(0.3)	12.1(2.6)	0.7(1.1)	4.712(0.991)	0.848(0.016)
		grMCP	4.0(0.0)	0.0(0.0)	20.0(0.0)	0.0(0.0)	2.374(0.571)	0.881(0.010)
		Marginal	0.1(0.3)	2.2(3.6)	1.0(1.1)	29.5(36.8)	-(-)	-(-)

Table 4: Simulation results under the CS(0.3) and CS(0.5) correlation structures. In each cell, mean (sd) based on 100 replicates.

Example	Correlation	Method	Main effects (M)		Interactions (I)		SSE	C-index
			TP	FP	TP	FP		
1	CS(0.3)	Proposed	14.8(1.1)	1.8(2.1)	5.6(2.8)	2.4(2.5)	4.884(1.219)	0.922(0.015)
		MCP	13.2(1.4)	2.2(1.9)	5.3(2.3)	3.8(2.0)	6.075(1.203)	0.903(0.024)
		grMCP	13.9(1.9)	0.6(0.8)	11.3(1.0)	61.1(11.2)	5.449(1.228)	0.914(0.021)
		Marginal	2.8(2.3)	36.6(33.7)	0.1(0.2)	0.5(1.1)	-(-)	-(-)
	CS(0.5)	Proposed	14.6(0.9)	6.2(4.2)	6.0(2.2)	4.7(3.8)	5.321(0.786)	0.938(0.011)
		MCP	12.1(1.7)	4.8(2.7)	5.8(2.0)	7.3(3.2)	6.122(0.830)	0.930(0.015)
		grMCP	12.2(2.8)	1.1(1.4)	10.6(1.2)	55.6(16.8)	6.050(1.231)	0.925(0.020)
		Marginal	8.5(4.0)	217.8(127.8)	0.0(0.2)	0.7(1.2)	-(-)	-(-)
2	CS(0.3)	Proposed	7.5(0.9)	0.7(1.6)	4.3(1.5)	1.4(1.2)	2.431(0.639)	0.879(0.030)
		MCP	6.2(1.2)	1.6(1.9)	4.4(1.4)	3.9(2.6)	2.750(0.753)	0.851(0.030)
		grMCP	7.5(0.8)	0.6(0.9)	5.8(0.5)	34.6(5.5)	2.600(0.593)	0.878(0.023)
		Marginal	0.8(1.2)	8.3(13.9)	0.6(0.8)	1.3(2.6)	-(-)	-(-)
	CS(0.5)	Proposed	7.1(1.0)	1.8(2.9)	3.5(1.7)	2.0(2.5)	2.702(0.864)	0.892(0.017)
		MCP	5.8(1.4)	3.4(2.0)	3.9(1.4)	7.7(3.8)	3.079(0.940)	0.871(0.022)
		grMCP	7.2(0.9)	1.6(1.6)	5.8(0.4)	37.6(9.1)	2.850(0.757)	0.890(0.012)
		Marginal	2.8(2.2)	75.7(80.9)	0.4(0.6)	4.7(9.3)	-(-)	-(-)
3	CS(0.3)	Proposed	11.6(0.5)	0.5(0.6)	0.0(0.0)	0.5(0.9)	1.340(0.437)	0.923(0.010)
		MCP	11.7(0.7)	0.8(1.0)	0.0(0.0)	1.4(1.3)	1.481(0.552)	0.900(0.018)
		grMCP	11.3(0.9)	0.5(0.7)	0.0(0.0)	59.2(6.5)	2.076(0.553)	0.906(0.017)
		Marginal	3.4(2.2)	53.3(46.9)	0.0(0.0)	0.2(0.7)	-(-)	-(-)
	CS(0.5)	Proposed	11.2(1.1)	0.9(1.4)	0.0(0.0)	0.1(0.2)	1.452(0.557)	0.935(0.010)
		MCP	11.4(0.8)	3.1(2.6)	0.0(0.0)	1.4(1.1)	1.716(0.599)	0.915(0.013)
		grMCP	10.6(1.4)	0.8(1.1)	0.0(0.0)	57.0(10.4)	2.357(0.625)	0.914(0.016)
		Marginal	8.4(3.0)	282.6(142.6)	0.0(0.0)	0.5(1.3)	-(-)	-(-)
4	CS(0.3)	Proposed	4.0(0.0)	0.7(1.2)	15.4(2.8)	0.4(0.8)	3.161(0.881)	0.888(0.023)
		MCP	1.9(1.2)	0.1(0.3)	13.2(3.4)	0.8(1.2)	4.127(1.002)	0.874(0.026)
		grMCP	4.0(0.0)	0.0(0.0)	20.0(0.0)	0.0(0.0)	1.954(0.498)	0.905(0.015)
		Marginal	0.0(0.0)	0.0(0.0)	0.7(0.9)	5.6(5.7)	-(-)	-(-)
	CS(0.5)	Proposed	4.0(0.0)	2.9(2.0)	17.1(1.9)	2.0(2.0)	2.823(0.755)	0.912(0.015)
		MCP	1.4(0.9)	0.6(0.8)	15.9(2.2)	3.6(2.3)	3.420(0.678)	0.904(0.016)
		grMCP	4.0(0.0)	0.0(0.0)	20.0(0.0)	0.0(0.0)	1.709(0.323)	0.922(0.011)
		Marginal	0.0(0.0)	0.0(0.0)	1.0(1.1)	21.3(15.1)	-(-)	-(-)

S3.2 Additional data analysis results

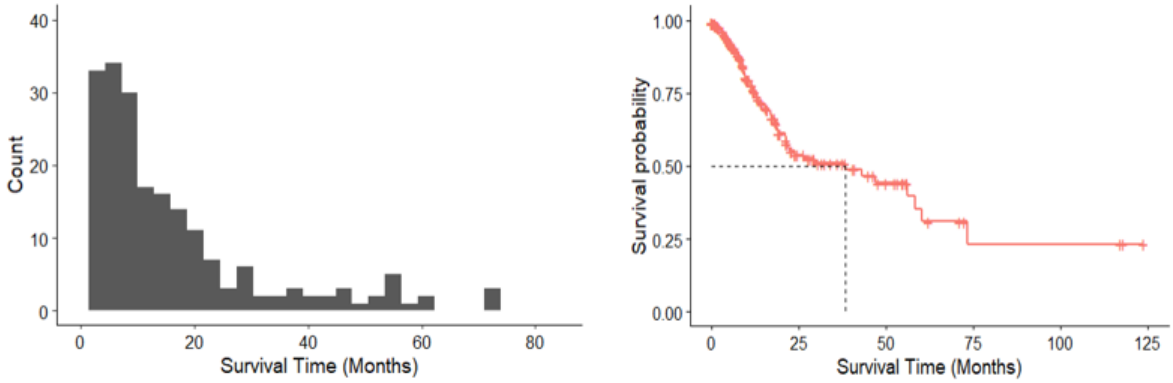

Figure 4: Histogram (left) and Kaplan-Meier curve (right) of the response variable.

Table 5: Data analysis using the proposed and alterative methods: number of overlapping identifications (RV-coefficient) for the main G effects and G-E interactions.

Main G effects	Proposed	MCP	grMCP	Marginal
Proposed	11	3(0.436)	2(0.415)	1(0.236)
MCP		10	1(0.222)	1(0.269)
grMCP			2	1(0.502)
Marginal				2
Interactions	Proposed	MCP	grMCP	Marginal
Proposed	12	9(0.577)	5(0.515)	0(0.057)
MCP		25	4(0.283)	0(0.077)
grMCP			10	0(0.020)
Marginal				9

References

Andersen PK, Gill RD (1982) Cox's regression model for counting processes: A large sample study. Annals of Statistics 10(4):1100 – 1120

Bradic J, Fan J, Jiang J (2011) Regularization for cox's proportional hazards model with np-dimensionality. Annals of Statistics 39 6:3092–3120

Fleming TR, Harrington DP (2011) Counting processes and survival analysis. John Wiley & Sons

Tseng P (2001) Convergence of a block coordinate descent method for nondifferentiable minimization. Journal of optimization theory and applications 109(3):475–494