Robust control experiments for multivariate tests with covariates and network information

Shaohua Xu and Yongdao Zhou

Nankai University

Supplementary Materials

In Section S1, we provide the proofs for all the theoretical results. In Section S2, we demonstrate the applications of our proposed experimental schemes in both A/B testing and sequential experiments. Finally, we present supplementary simulation results in Section S3.

S1 Theoretical Proofs

Lemma S1. *For any $n \times n$ positive definite matrices $\mathbf{A}$, $\mathbf{B}$, and $n \times p$ column full-rank matrix $\mathbf{X}$, we have:*

a. If $\mathbf{A} \succeq \mathbf{B}$, then for $1 \leq s \leq \infty$, $\mathrm{tr}((\mathbf{A}-\mathbf{B})^s)^{1/s} \geq \mathrm{tr}(\mathbf{A}^s)^{1/s} - \mathrm{tr}(\mathbf{B}^s)^{1/s}$.

b. For $-\infty < s \leq r < \infty$, $\lambda_{min}(\mathbf{A}) \leq \mathrm{tr}(\mathbf{A}^s/n)^{1/s} \leq \mathrm{tr}(\mathbf{A}^r/n)^{1/r} \leq \lambda_{max}(\mathbf{A})$, where $\lambda_{min}(\mathbf{A})$ and $\lambda_{max}(\mathbf{A})$ denote the minimum and maximum eigenvalues of matrix $\mathbf{A}$, respectively.

c. $\kappa(\mathbf{X}^T\mathbf{A}\mathbf{X}) \leq \kappa(\mathbf{X}^T\mathbf{X})\kappa(\mathbf{A})$, where $\kappa(\mathbf{A})$ is the condition number of $\mathbf{A}$.

d. $\lambda_{min}(\mathbf{X}^T\mathbf{X})\lambda_{max}(\mathbf{A}) \leq \lambda_{max}(\mathbf{X}^T\mathbf{A}\mathbf{X}) \leq \lambda_{max}(\mathbf{X}^T\mathbf{X})\lambda_{max}(\mathbf{A})$.

Lemma S1 presents some fundamental results for positive definite matrices, with proofs omitted. The following lemma establishes upper and lower bounds for all ϕ_s-optimality criteria of any positive definite matrix.

Lemma S2. *For any $p \times p$ positive definite matrix $\mathbf{M}$, we have*

$$C_1 \mathrm{tr}(\mathbf{M}/p)^{-1} \leq \mathrm{tr}((\mathbf{M})^{-s}/p)^{1/s} \leq C_2 \mathrm{tr}(\mathbf{M}/p)^{-1}, \ \textit{for } 0 \leq s \leq \infty,$$

where when $0 \leq s < 1$, $C_1 = \kappa(\mathbf{M})^{-1}$; when $s \geq 1$, $C_1 = \max\{p^{-(1-1/s)}, \kappa(\mathbf{M})^{-1}\}$, and $C_2 = \kappa(\mathbf{M})$.

Proof of Lemma S2. Let $\lambda_1, \lambda_2, \cdots, \lambda_p$ denote the eigenvalues of matrix $\mathbf{M}$

sorted in descending order. Note that

$$\mathrm{tr}(\mathbf{M}^{-s}) = \sum_{i=1}^{p} \lambda_i^{-s} \begin{cases} \leq p\lambda_p^{-s} = p\kappa(\mathbf{M})^s \lambda_1^{-s} \leq p^{s+1}\kappa(\mathbf{M})^s \mathrm{tr}(\mathbf{M})^{-s}, \\ \geq p\lambda_1^{-s} = p\kappa(\mathbf{M})^{-s}\lambda_p^{-s} \geq p^{s+1}\kappa(\mathbf{M})^{-s}\mathrm{tr}(\mathbf{M})^{-s}, \end{cases} \tag{S1-1}$$

which holds for any $s > 0$. When $s \geq 1$, we have a tighter lower bound:

$$\mathrm{tr}(\mathbf{M}^{-s}) = p\left(\frac{p}{\sum_{i=1}^{p}\lambda_i^{-s}}\right)^{-1} \geq p\left(\frac{\mathrm{tr}(\mathbf{M}^s)}{p}\right)^{-1} \geq \frac{p^2}{\mathrm{tr}(\mathbf{M})^s}, \tag{S1-2}$$

where the first inequality follows from Lemma $S1$, and the second inequality holds because for $s \geq 1$, $\mathrm{tr}(\mathbf{M}^s) = \sum_{i=1}^{p}\lambda_i^s \leq \left(\sum_{i=1}^{p}\lambda_i\right)^s = \mathrm{tr}(\mathbf{M})^s$. Combining (S1-1) and (S1-2) yields the conclusion of Lemma $S2$. □

Proof of Proposition 1. For simplicity, denote $[\mathbf{A}]_s = \mathrm{tr}(\mathbf{A}^s/p)^{1/s}$ for a $p \times p$ positive definite matrix $\mathbf{A}$. Clearly, $\|\mathbf{A}\| = \lambda_{\max}(\mathbf{A}) = \lim_{s\to\infty}[\mathbf{A}]_s$ and $\lambda_{\min}(\mathbf{A}) = \lim_{s\to-\infty}[\mathbf{A}]_s$. Note that

$$\begin{aligned} \frac{[\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0,\mathbf{T}))]_s - [\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T}))]_s}{[\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T}))]_s} &\overset{(i)}{\leq} \frac{[\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0,\mathbf{T})) - \mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T}))]_s}{[\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T}))]_s} \\ &\overset{(ii)}{\leq} \frac{\|\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T}))\|}{[\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T}))]_s} \frac{\|\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0,\mathbf{T})) - \mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T}))\|}{\|\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T}))\|} \\ &\overset{(ii)}{\leq} \kappa(\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T})))\mathrm{RT}(\mathbf{R},\mathbf{R}_0,\mathbf{T}) \\ &\overset{(iii)}{\leq} \kappa(\mathbf{X}^T\mathbf{X})\kappa(\mathbf{R})\mathrm{RT}(\mathbf{R},\mathbf{R}_0,\mathbf{T}), \end{aligned}$$

where inequalities (i), (ii), and (iii) follow from conclusions of Lemma $S1$ a, b, and c, respectively. Additionally, if the covariates matrix $\mathbf{Z}$ satisfies

Assumption 4, we have

$$\mathbf{T}^T\mathbf{R}^{-1}\mathbf{T} \succeq \mathbf{T}^T\Sigma(\mathbf{R})\mathbf{T} \succeq \mu\mathbf{T}^T\mathbf{R}^{-1}\mathbf{T}.$$

Thus, we have $\kappa(\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R},\mathbf{T}))) \leq \mu^{-1}\kappa(\mathbf{T}^T\mathbf{T})\kappa(\mathbf{R})$. □

Proof of Theorem 1. To derive the upper bound of the loss, we need to provide upper bounds for the spectral norms of $\mathbf{D}^T\mathbf{R}\mathbf{D}$ and $\mathbf{L}(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{L}^T$. For simplicity, let $\mathbf{V}_0 = \mathbf{X}^T\mathbf{R}_0^{-1}\mathbf{X}$ and $\mathbf{V} = \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}$.

On the one hand, according to the properties of the spectral norm, we have

$$\begin{aligned}
\|\mathbf{D}^T\| &= \|\mathbf{L}\mathbf{V}^{-1}\mathbf{X}^T\mathbf{R}^{-1} - \mathbf{L}\mathbf{V}_0^{-1}\mathbf{X}^T\mathbf{R}_0^{-1}\| \\
&\leq \|\mathbf{L}\mathbf{V}^{-1}\mathbf{X}^T(\mathbf{R}^{-1} - \mathbf{R}_0^{-1})\| + \|\mathbf{L}(\mathbf{V}^{-1} - \mathbf{V}_0^{-1})\mathbf{X}^T\mathbf{R}_0^{-1}\| \\
&\leq (1 + \|\mathbf{X}\mathbf{V}_0^{-1}\mathbf{X}^T\mathbf{R}_0^{-1}\|)\|\mathbf{L}\mathbf{V}^{-1}\mathbf{X}^T\|\|\mathbf{R}_0^{-1} - \mathbf{R}^{-1}\| \\
&\leq (1 + \|\mathbf{X}\mathbf{V}_0^{-1}\mathbf{X}^T\mathbf{R}_0^{-1}\|)\|\mathbf{L}\mathbf{V}^{-1}\mathbf{X}^T\|\|\mathbf{R}^{-1}\|\|\mathbf{R}_0^{-1}\|\|\mathbf{R} - \mathbf{R}_0\|,
\end{aligned} \tag{S1-3}$$

where the last two inequalities hold because

$$\mathbf{V}^{-1}-\mathbf{V}_0^{-1} = \mathbf{V}^{-1}\mathbf{X}^T(\mathbf{R}_0^{-1}-\mathbf{R}^{-1})\mathbf{X}\mathbf{V}_0^{-1} \quad \text{and} \quad \mathbf{R}_0^{-1}-\mathbf{R}^{-1} = \mathbf{R}^{-1}(\mathbf{R}-\mathbf{R}_0)\mathbf{R}_0^{-1}.$$

Denote

$$\mathbf{V}^{-1} = \begin{pmatrix} \mathbf{V}^{11} & \mathbf{V}^{12} \\ \mathbf{V}^{21} & \mathbf{V}^{22} \end{pmatrix},$$

By the inverse formula for block matrices, we have

$$\mathbf{V}^{11} = (\mathbf{T}^T\Sigma(\mathbf{R})\mathbf{T})^{-1} \quad \text{and} \quad \mathbf{V}^{12} = -V^{11}\mathbf{T}^T\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z})^{-1}.$$

Therefore,

$$
\begin{aligned}
\|\mathbf{L}\mathbf{V}^{-1}\mathbf{X}^T\| &= \|\mathbf{V}^{11}\mathbf{T}^T\| + \|\mathbf{V}^{12}\mathbf{Z}^T\| \\
&= \|\mathbf{V}^{11}\mathbf{T}^T\| + \|\mathbf{V}^{11}\mathbf{T}^T\mathbf{R}^{-1/2}\mathbf{R}^{-1/2}\mathbf{Z}(\mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{R}^{-1/2}\mathbf{R}^{1/2}\| \\
&\le (1+\kappa(\mathbf{R})^{1/2}\|\mathbf{R}^{-1/2}\mathbf{Z}(\mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{R}^{-1/2}\|)\|\mathbf{V}^{11}\|\|\mathbf{T}^T\| \\
&= (1+\kappa(\mathbf{R})^{1/2})\|\mathbf{V}^{11}\|\|\mathbf{T}^T\|,
\end{aligned}
\tag{S1-4}
$$

where the last equality holds because $\mathbf{R}^{-1/2}\mathbf{Z}(\mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{R}^{-1/2}$ is an orthogonal projection matrix, so its spectral norm equals 1. Similarly, we have

$$
\|\mathbf{X}\mathbf{V}_0^{-1}\mathbf{X}^T\mathbf{R}_0^{-1}\| = \|\mathbf{R}_0^{1/2}\mathbf{R}_0^{-1/2}\mathbf{X}\mathbf{V}_0^{-1}\mathbf{X}^T\mathbf{R}_0^{-1/2}\mathbf{R}_0^{-1/2}\| \le \kappa(\mathbf{R}_0)^{1/2}. \tag{S1-5}
$$

Substituting (S1-4) and (S1-5) back into (S1-3), we obtain

$$
\|\mathbf{D}\| \le (1+\kappa(\mathbf{R}_0)^{1/2})(1+\kappa(\mathbf{R})^{1/2})\frac{\|\mathbf{V}^{11}\|\|\mathbf{T}^T\|\|\mathbf{R}-\mathbf{R}_0\|}{\lambda_{\min}(\mathbf{R})\lambda_{\min}(\mathbf{R}_0)}. \tag{S1-6}
$$

On the other hand, under Assumption 4, note that

$$
\|\mathbf{L}(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{L}^T\| = \|\mathbf{V}^{11}\| \le (\mu\lambda_{\min}(\mathbf{T}^T\mathbf{T})\lambda_{\min}(\mathbf{R}^{-1}))^{-1}. \tag{S1-7}
$$

Combining (S1-6) and (S1-7), we have

$$
\begin{aligned}
\mathrm{RT}(\mathbf{R},\mathbf{R}_0,\mathbf{T}) &= \frac{\|\mathbf{D}^T\mathbf{R}\mathbf{D}\|}{\|\mathbf{L}(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{L}^T\|} \\
&\le \frac{(1+\kappa(\mathbf{R}_0)^{1/2})^2(1+\kappa(\mathbf{R})^{1/2})^2\kappa(\mathbf{T}^T\mathbf{T})\kappa(\mathbf{R})^2}{\mu\lambda_{\min}(\mathbf{R}_0)^2}\|\mathbf{R}-\mathbf{R}_0\|^2.
\end{aligned}
$$

Note that $\lambda_{\min}(\mathbf{R}_0) \geq \delta\lambda_{\min}(\Sigma_0)$ and $\kappa(\mathbf{R}) \leq \kappa_0 + \delta^{-1}r$. Therefore, for any $(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) \in \Omega_r \times \Omega_0 \times \Theta$, we have

$$\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) \leq C_1 \kappa(\mathbf{T}^T\mathbf{T}) \|\mathbf{R} - \mathbf{R}_0\|^2,$$

where $C_1 = \mu^{-1}(\delta\lambda_{\min}(\Sigma_0))^{-2}(1 + \kappa_0^{1/2})^2(1 + (\kappa_0 + \delta^{-1}r)^{1/2})^2(\kappa_0 + \delta^{-1}r)^2$. Additionally, $\|\mathbf{R} - \mathbf{R}_0\|^2 \leq (\|\mathbf{R} - \mathbf{R}^*\| + \sum_{t=0}^{k} \|\Sigma_t\| |\sigma_t^{2*} - \sigma_{t0}^2|)^2 \leq (r + C_2\|\boldsymbol{\sigma}^{*2} - \boldsymbol{\sigma}_0^2\|_1)^2$, where $C_2 = \max_{0 \leq t \leq k} \|\Sigma_t\|$. □

Proof of Lemma 1. Let $\delta = \left(\kappa_0 - \kappa(\Sigma_0)\lambda_{\min}(\Sigma_0) / \sum_{t=1}^{k} \lambda_{\max}(\Sigma_t)\right)$. For any $\mathbf{R} \in \cup_{\mathbf{j} \in J}[\Omega_{\mathbf{j}}^*]$, we have

$$\kappa(\mathbf{R}) = \frac{\lambda_{\max}(\mathbf{R})}{\lambda_{\min}(\mathbf{R})} \leq \kappa(\Sigma_0) + \frac{\sum_{t=1}^{k} \lambda_{\max}(\Sigma_t)}{\delta\lambda_{\min}(\Sigma_0)} = \kappa_0,$$

i.e., $\mathbf{R} \in \Omega_0$. On the other hand, for some $\mathbf{j} \in J$ with $j_0 = 0$, let $\mathbf{R}_0 = \sigma_0^2\Sigma_0 + \Sigma_1 \in \Omega_{\mathbf{j}}^*$ and $s_0 = \mathbf{x}_0^T\Sigma_0\mathbf{x}_0$, where $\mathbf{x}_0$ is a unit vector such that $\mathbf{x}_0^T\Sigma_1\mathbf{x}_0 = 0$. When $\sigma_0^2 \to 0^+$, we have

$$\kappa(\mathbf{R}_0) \geq \frac{\lambda_{\max}(\Sigma_1)}{\sigma_0^2 s_0} \to \infty.$$

This contradicts the assumption that the condition number is bounded, thus proving $\mathbf{R} \in \cup_{\mathbf{j} \in J}[\Omega_{\mathbf{j}}^*]$. □

Proof of Theorem 2. By Theorem 1 and Lemma 1, for any $\mathbf{j} \in J$, Problem (3.8) is equivalent to

$$\arg \min_{\mathbf{a}_0 \in A_\delta} \max_{\mathbf{a} \in A_{\mathbf{j}}} \|\mathbf{a} - \mathbf{a}_0\|_1, \tag{S1-8}$$

where $A_\delta = \cup_{\mathbf{j}\in J} A_{\mathbf{j}}$.

On the one hand, for any $\mathbf{j} \in J$, if $j_0 = 0$, i.e., $\mathbf{a} \in A_{\mathbf{j}} = \{\mathbf{a} \mid a_0 \in [\delta, 1);\ a_t = 1 \text{ if } j_t = 1;\ a_t \in [0,1) \text{ if } j_t = 0, t \neq 0\}$, we have

$$
\begin{aligned}
\max_{\mathbf{a}\in A_{\mathbf{j}}} \|\mathbf{a}-\mathbf{a}_0\|_1^2 &= \max_{\mathbf{a}\in A_{\mathbf{j}}} \left(\sum_{t=0}^{k} |a_t - a_{t0}|\right)^2 \\
&= \max_{\mathbf{a}\in A_{\mathbf{j}}} \left(|a_0 - a_{00}| + \sum_{t=1, j_t=0}^{k} |a_t - a_{t0}| + \sum_{t=1, j_t=1}^{k} |1 - a_{t0}|\right)^2 \\
&= \left(\max_{a_0\in[\delta,1)} |a_0 - a_{00}| + \max_{\mathbf{a}\in A_{\mathbf{j}}} \sum_{t=1, j_t=0}^{k} |a_t - a_{t0}| + \sum_{t=1, j_t=1}^{k} |1 - a_{t0}|\right)^2 \\
&= \left(\max\{1 - a_{00}, a_{00} - \delta\} + \sum_{t=1, j_t=0}^{k} \max\{1 - a_{t0}, a_{t0}\} + \sum_{t=1, j_t=1}^{k} |1 - a_{t0}|\right)^2 .
\end{aligned}
$$

Therefore, $\max_{\mathbf{a}\in A_{\mathbf{j}}} \|\mathbf{a}-\mathbf{a}_0\|_1^2 \geq \left((1-\delta)/2 + \sum_{t=1,j_t=0}^{k} 1/2\right)^2$, with equality if and only if for $t = 0, 1, \cdots, k$,

$$
a_{t0} = \begin{cases} (1+\delta)/2, & t = 0; \\ 1/2, & j_t = 0, t \neq 0; \\ 1, & j_t = 1, \end{cases} \tag{S1-9}
$$

i.e., $\mathbf{R}_0 = (1/2)\sum_{t=1,j_t=0}^{k} \Sigma_t + \sum_{t=1,j_t=1}^{k} \Sigma_t + [(1+\delta)/2]\Sigma_0$.

If $j_0 = 1$, i.e., $\mathbf{a} \in A_{\mathbf{j}} = \{\mathbf{a} \mid a_t = 1 \text{ if } j_t = 1;\ a_t \in [0,1) \text{ if } j_t = 0\}$, we have

$$
\begin{aligned}
\max_{\mathbf{a}\in A_{\mathbf{j}}} \|\mathbf{a}-\mathbf{a}_0\|_1^2 &= \sup_{\mathbf{a}\in A_{\mathbf{j}}} \left(\sum_{t=0}^{k} |a_t - a_{t0}| \right)^2 \\
&= \max_{\mathbf{a}\in A_{\mathbf{j}}} \left(|1-a_{00}| + \sum_{t=1,j_t=0}^{k} |a_t - a_{t0}| + \sum_{t=1,j_t=1}^{k} |1-a_{t0}| \right)^2 \\
&= \left(|1-a_{00}| + \max_{\mathbf{a}\in A_{\mathbf{j}}} \sum_{t=1,j_t=0}^{k} |a_t - a_{t0}| + \sum_{t=1,j_t=1}^{k} |1-a_{t0}| \right)^2 \\
&= \left(|1-a_{00}| + \sum_{t=1,j_t=0}^{k} \max\{1-a_{t0}, a_{t0}\} + \sum_{t=1,j_t=1}^{k} |1-a_{t0}| \right)^2 .
\end{aligned}
$$

Therefore, $\max_{\mathbf{a}\in A_{\mathbf{j}}} \|\mathbf{a}-\mathbf{a}_0\|_1^2 \geq \left(\sum_{t=1,j_t=0}^{k} 1/2\right)^2$, with equality if and only if for $t = 0, 1, \cdots, k$,

$$
a_{t0} = \begin{cases} 1/2, & j_t = 0; \\ 1, & j_t = 1, \end{cases} \tag{S1-10}
$$

i.e., $\mathbf{R}_0 = (1/2)\sum_{t=1,j_t=0}^{k} \Sigma_t + \sum_{t=1,j_t=1}^{k} \Sigma_t + \Sigma_0$.

On the other hand, for any $\mathbf{T} \in \Theta$, we have

$$
\kappa(\mathbf{T}^T\mathbf{T}) \geq 1, \tag{S1-11}
$$

with equality if and only if $\mathbf{T}^T\mathbf{T} = n\mathbf{I}_n$, i.e., $\mathbf{T}$ is a column-orthogonal matrix. Combining (S1-8) – (S1-11), we obtain the conclusion of the theorem. $\square$

Proof of Theorem 3. By Theorem 1 and Lemma 1, Problem (3.9) is equivalent to

$$\arg\min_{\mathbf{a}_0 \in A_\delta} \max_{\mathbf{a} \in A_\delta} \|\mathbf{a} - \mathbf{a}_0\|_1, \tag{S1-12}$$

where $A_\delta = \cup_{\mathbf{j} \in J} A_{\mathbf{j}}$. When $k = 1$, we have

$$\begin{aligned}
\max_{\mathbf{a} \in A_\delta} \|\mathbf{a} - \mathbf{a}_0\|_1^2 &= \max\left\{ \max_{\mathbf{a} \in A_{(1,0)}} \|\mathbf{a} - \mathbf{a}_0\|_1^2, \sup_{\mathbf{a} \in A_{(0,1)}} \|\mathbf{a} - \mathbf{a}_0\|_1^2, \max_{\mathbf{a} \in A_{(1,1)}} \|\mathbf{a} - \mathbf{a}_0\|_1^2 \right\} \\
&= \max\{A_1(\mathbf{a}_0), A_2(\mathbf{a}_0)\} \\
&= \begin{cases} A_1(\mathbf{a}_0), & \text{if } a_{00} = 1,\ a_{10} \geq 1 - \delta/2 \text{ or } a_{10} = 1; \\ A_2(\mathbf{a}_0), & \text{if } a_{00} = 1,\ a_{10} \leq 1 - \delta/2, \end{cases}
\end{aligned}$$

where $A_1(\mathbf{a}_0) = (|1 - a_{00}| + \max\{1 - a_{10}, a_{10}\})^2$ and $A_2(\mathbf{a}_0) = (\max\{1 - a_{00}, a_{00} - \delta\} + |1 - a_{10}|)^2$. Therefore, $\max_{\mathbf{a} \in A_\delta} \|\mathbf{a} - \mathbf{a}_0\|_1^2 \geq (1 - \delta/2)^2$, with equality if and only if $a_{00} = 1$ and $a_{10} = 1 - \delta/2$, i.e.,

$$\mathbf{R}_0 = \Sigma_0 + (1 - \delta/2)\Sigma_1. \tag{S1-13}$$

If $k \geq 2$, let $J_0 = \{\mathbf{j} \mid j_0 = 0, \mathbf{j} \in J\}$ and $J_1 = \{\mathbf{j} \mid j_0 = 1, \mathbf{j} \in J\}$. For $\mathbf{j} \in J_0$, we have

$$\begin{aligned}
\max_{\mathbf{a} \in A_{\mathbf{j}}} \|\mathbf{a} - \mathbf{a}_0\|_1 &= \max\{1 - a_{00}, a_{00} - \delta\} + \sum_{t=1, j_t=0}^{k} \max\{1 - a_{t0}, a_{t0}\} + \sum_{t=1, j_t=1}^{k} |1 - a_{t0}| \\
&\leq \max\{1 - a_{00}, a_{00} - \delta\} + \sum_{t=1, t \neq t^*}^{k} \max\{1 - a_{t0}, a_{t0}\} + |1 - a_{t^*0}|,
\end{aligned}$$

where $t^* \in \arg\min_{t \in [k]}\{a_{t0}\}$, and equality holds if and only if $j_t = 0$ for

$t \neq t^*$ and $j_t = 1$ for $t = t^*$. Similarly, for $\mathbf{j} \in J_1$, we have

$$\begin{aligned}\max_{\mathbf{a}\in A_{\mathbf{j}}} \|\mathbf{a} - \mathbf{a}_0\|_1 &= \sum_{t=1, j_t=0}^{k} \max\{1 - a_{t0}, a_{t0}\} + \sum_{t=0, j_t=1}^{k} |1 - a_{t0}| \\ &\leq \sum_{t=1}^{k} \max\{1 - a_{t0}, a_{t0}\} + |1 - a_{00}|,\end{aligned}$$

with equality if and only if $\mathbf{j} = (1, 0, \cdots, 0)$.

Define $B_1(\mathbf{a}_0) = |1-a_{00}|+\max\{1-a_{t^*0}, a_{t^*0}\}+A(a_{t0})$, $B_2(\mathbf{a}_0) = \max\{1-a_{00}, a_{00} - \delta\} + |1 - a_{t^*0}| + A(a_{t0})$, and $A(a_{t0}) = \sum_{t=1,t\neq t^*}^{k} \max\{1 - a_{t0}, a_{t0}\}$. Thus, we obtain

$$\begin{aligned}\max_{\mathbf{a}\in A_\delta} \|\mathbf{a} - \mathbf{a}_0\|_1 &= \max\left\{\max_{\mathbf{j}\in J_0}\left\{\max_{\mathbf{a}\in A_{\mathbf{j}}} \|\mathbf{a} - \mathbf{a}_0\|_1\right\}, \max_{\mathbf{j}\in J_1}\left\{\max_{\mathbf{a}\in A_{\mathbf{j}}} \|\mathbf{a} - \mathbf{a}_0\|_1\right\}\right\} \\ &= \max\{B_1(\mathbf{a}_0), B_2(\mathbf{a}_0)\} \\ &= \begin{cases} B_1(\mathbf{a}_0), & \text{if } a_{00} = 1,\ a_{t0} \geq a_{t^*0} \geq 1 - \delta/2 \text{ or } a_{t0} = 1,\ t \in [k]; \\ B_2(\mathbf{a}_0), & \text{if } a_{00} = 1,\ a_{t^*0} \leq 1 - \delta/2,\ a_{t^*0} \leq a_{t0},\ t \in [k]. \end{cases}\end{aligned}$$

Therefore, $\max_{\mathbf{a}\in A_\delta} \|\mathbf{a} - \mathbf{a}_0\|_1^2 \geq (1 + k/2 - \delta)^2$, with equality if and only if, for $k > 2$, $a_{00} = 1$ and $a_{t0} = 1/2$; for $k = 2$, $a_{00} = 1$, $a_{10} = a_{20} = 1/2$ or $a_{00} = 1$, $a_{10} = a_{20} = 1 - \delta/2$, i.e.,

$$\mathbf{R}_0 = \begin{cases} \Sigma_0 + (1 - \delta/2)(\Sigma_1 + \Sigma_2) \text{ or } \Sigma_0 + (1/2)(\Sigma_1 + \Sigma_2), & k = 2; \\ \Sigma_0 + (1/2)\sum_{t=1}^{k} \Sigma_t, & k > 2. \end{cases} \tag{S1-14}$$

Combining (S1-11) – (S1-14), we derive the conclusion of the theorem. □

Proof of Theorem 4. For any positive definite matrices $\mathbf{R}$ and $\mathbf{R}_0$, our goal is to prove that when the response variable $\mathbf{Y}$ satisfies $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\theta}$ and $\mathrm{cov}(\mathbf{Y}) = \mathbf{R}$, the estimators

$$\tilde{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{Y} \text{ and } \tilde{\boldsymbol{\theta}}_0 = (\mathbf{X}^T\mathbf{R}_0^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{R}_0^{-1}\mathbf{Y}$$

are equivalent if and only if $\mathbf{R} \in [\Omega_{\mathbf{R}_0}]$.

Let $\tilde{\mathbf{Y}} = \mathbf{R}^{-1/2}\mathbf{Y}$, $\tilde{\mathbf{X}} = \mathbf{R}^{-1/2}\mathbf{X}$, and $\tilde{\mathbf{R}}_0 = \mathbf{R}^{-1/2}\mathbf{R}_0\mathbf{R}^{-1/2}$. Then, $\tilde{\boldsymbol{\theta}} = (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\tilde{\mathbf{Y}}$, $\tilde{\boldsymbol{\theta}}_0 = (\tilde{\mathbf{X}}^T\tilde{\mathbf{R}}_0^{-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\tilde{\mathbf{R}}_0^{-1}\tilde{\mathbf{Y}}$, $E(\tilde{\mathbf{Y}}) = \tilde{\mathbf{X}}\boldsymbol{\theta}$, and $\mathrm{cov}(\tilde{\mathbf{Y}}) = \mathbf{I}_n$. By Lemma 5$a$ in Rao (1967), $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}_0$ are equivalent if and only if $\tilde{\mathbf{R}} = \mathbf{I}_n + \tilde{\mathbf{X}}\mathbf{D}_1\tilde{\mathbf{X}}^T + \tilde{\mathbf{V}}\mathbf{D}_2\tilde{\mathbf{V}}^T$, where $\mathbf{D}_1, \mathbf{D}_2$ are arbitrary symmetric matrices such that $\tilde{\mathbf{R}}$ is positive definite, and $\tilde{\mathbf{V}}$ is any full-rank matrix satisfying $\mathbf{X}^T\mathbf{R}_1^{-1}\mathbf{V} = \mathbf{0}$ and $\tilde{\mathbf{X}}^T\tilde{\mathbf{V}} = \mathbf{0}$. Equivalently, we have $\mathbf{R} = \mathbf{R}_0 + \mathbf{X}\mathbf{D}_1\mathbf{X}^T + \mathbf{V}\mathbf{D}_2\mathbf{V}^T$. □

Proof of Proposition 2. By the construction of Algorithm 1 and Theorem 3.4.2 from Ben-Tal and Nemirovski (2001), we know that $E[\mathbf{v}_j\mathbf{v}_j^T] = \frac{2}{\pi}\arcsin[\mathbf{S}^*]$ for $j \in [p]$, where $\arcsin[\mathbf{X}]$ denotes the matrix with entries $\arcsin(X_{ij})$.

On the one hand, it's clear that $\mathbf{T}_r = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p) \in \mathscr{T}_p$ with probability one, meaning $\mathbf{T}_r$ is indeed a feasible solution for Problem (4.11). On the other hand, the properties of expectation and trace imply that

$$
\begin{aligned}
E(\mathrm{tr}(\mathbf{T}_r^T\Sigma(\mathbf{R}_*)\mathbf{T}_r)/p) &= E(\mathrm{tr}(\Sigma(\mathbf{R}_*)\mathbf{T}_r\mathbf{T}_r^T)/p) \\
&= \mathrm{tr}(\Sigma(\mathbf{R}_*)E(\mathbf{T}_r\mathbf{T}_r^T)/p) \\
&= \frac{2}{\pi}\mathrm{tr}(\Sigma(\mathbf{R}_*)\arcsin[\mathbf{S}^*]) \\
&\geq \frac{2}{\pi}\mathrm{tr}(\Sigma(\mathbf{R}_*)\mathbf{S}^*),
\end{aligned}
$$

where the inequality holds because Theorem 3.4.2 from Ben-Tal and Nemirovski (2001) shows that $\arcsin[\mathbf{X}] \succeq \mathbf{X}$ for any semi-positive definite matrix $\mathbf{X}$ with $|X_{ij}| \leq 1$. Furthermore, note that Problem (4.12) is a relaxation of Problem (4.11), and thus $\mathrm{tr}(\Sigma(\mathbf{R}_*)\mathbf{S}^*) \geq \mathrm{tr}(\mathbf{T}_r^T\Sigma(\mathbf{R}_*)\mathbf{T}_r)/p$, which completes the proof. □

Proof of Theorem 5. For simplicity, let $\mathbf{M}_r = \mathbf{T}_r^T\Sigma_*\mathbf{T}_r$, $\mathbf{M}^* = \mathbf{T}^{*T}\Sigma_*\mathbf{T}^*$, $\mathbf{M}_s^* = \mathbf{T}_s^{*T}\Sigma_*\mathbf{T}_s^*$, $e_s = \mathrm{tr}(\mathbf{M}^*)/\mathrm{tr}(\mathbf{M}_s^*)$, and $\kappa_s(\mathbf{M}_s^*) = \kappa(\mathbf{M}_s^*)$ if $s < 1$ and $\kappa_s(\mathbf{M}_s^*) = \min\{p^{1-1/s}, \kappa(\mathbf{M}_s^*)\}$ if $s \geq 1$. Clearly, $\mathbf{M}_r$, $\mathbf{M}^*$, and $\mathbf{M}_s^*$ are all positive semi-definite matrices, and $e_s \geq 1$. By the definition of ϕ_s-efficiency and Lemma $S2$, we have

$$
\begin{aligned}
\frac{\phi_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_s^*))}{\phi_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_r))} &\geq \frac{1}{\kappa_s(\mathbf{M}_s^*)\kappa(\mathbf{M}_r)}\frac{\mathrm{tr}(\mathbf{M}_r)}{\mathrm{tr}(\mathbf{M}_s^*)} \\
&= \frac{e_s}{\kappa_s(\mathbf{M}_s^*)\mathrm{tr}(\mathbf{M}^*)}\frac{\mathrm{tr}(\mathbf{M}_r)}{\kappa(\mathbf{M}_r)}.
\end{aligned} \tag{S1-15}
$$

By Taylor expansion, we obtain

$$
E\left(\frac{\mathrm{tr}(\mathbf{M}_r)}{\kappa(\mathbf{M}_r)}\right) = \frac{E(\mathrm{tr}(\mathbf{M}_r))}{E(\kappa(\mathbf{M}_r))} - \frac{E(\mathrm{tr}(\mathbf{M}_r)(\kappa(\mathbf{M}_r) - E(\kappa(\mathbf{M}_r))))}{[E(\kappa(\mathbf{M}_r))]^2} + O_2,
$$

where $O_2 = E\left(\frac{\mathrm{tr}(\mathbf{M}_r)[\kappa(\mathbf{M}_r)-E(\kappa(\mathbf{M}_r))]^2}{\eta^3}\right) \geq 0$ and η is a random variable between $\kappa(\mathbf{M}_r)$ and $E(\kappa(\mathbf{M}_r))$. Thus, we have

$$\begin{aligned} E\left(\frac{\mathrm{tr}(\mathbf{M}_r)}{\kappa(\mathbf{M}_r)}\right) &\geq \frac{2E(\mathrm{tr}(\mathbf{M}_r))}{E(\kappa(\mathbf{M}_r))} - \frac{E(\mathrm{tr}(\mathbf{M}_r)\kappa(\mathbf{M}_r))}{[E(\kappa(\mathbf{M}_r))]^2} \\ &\geq \frac{(4/\pi-1)\mathrm{tr}(\mathbf{M}^*)}{E(\kappa(\mathbf{M}_r))}, \end{aligned} \tag{S1-16}$$

where the last equality holds because $E(\mathrm{tr}(\mathbf{M}_r)) \geq (2/\pi)\mathrm{tr}(\mathbf{M}^*)$ (Proposition 2) and $\mathrm{tr}(\mathbf{M}_r) \leq \mathrm{tr}(\mathbf{M}^*)$. Taking expectations on both sides of (S1-15) and combing with inequality (S1-16), we get

$$E\left(\frac{\phi_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_s^*))}{\phi_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_r))}\right) \geq \frac{e_s(4/\pi-1)}{\kappa_s(\mathbf{M}_s^*)E(\kappa(\mathbf{M}_r))} > 0,$$

which proves the desired lower bound. □

S2 Two Special Cases

S2.1 A/B Testing

In this subsection, we discuss the case of A/B testing, that is, $p = 1$. In this scenario, we represent the treatment vector as $\mathbf{t} = (t_1, \cdots, t_n)^{\mathbf{T}} \in \{-1, 1\}^n$, and the loss $\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{t})$ defined in (2.3) is simplified to the relative error of the estimated treatment effect variance:

$$\frac{\mathrm{var}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{t})) - \mathrm{var}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{t}))}{\mathrm{var}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{t}))}.$$

Therefore, the upper bounds in Theorem 1 and Corollary 1 provide a method to control these relative errors. At this time, $\overline{\mathrm{RT}}(\mathbf{R}, \mathbf{R}_0, \mathbf{t}) = C_1\kappa(\mathbf{t}^T\mathbf{t})(C_2\|\boldsymbol{\sigma}^{*2} - \boldsymbol{\sigma}_0^2\|_1 + r)^2$ can provide theoretical guidance for finding a robust working covariance structure, such as the minimax robust covariance structures in Theorems 2 and 3. However, for all treatment vectors, $\kappa(\mathbf{t}^T\mathbf{t}) = 1$. In this subsection, our goal is to find a treatment vector that is optimally matched with the robust covariance structure based on the random matching method in Section 4. When $p = 1$, all ϕ_s-criteria are equal to the variance, that is, $\phi_s(\hat{\boldsymbol{\alpha}}) = (\mathbf{t}^T\Sigma(\mathbf{R}_*)\mathbf{t})^{-1} = \mathrm{var}(\hat{\boldsymbol{\alpha}})$. Therefore, Problem (4.11) is equivalent to Problem (4.10), and the output of Algorithm 1 can provide a $2/\pi$ approximation of the optimal solution of Problem (4.10) (this is a lower bound independent of the scale of the problem!). The conclusion in this special case generalizes the result about the independent and identically distributed covariance structure in Bhat et al. (2020) to the general covariance structure.

Corollary S1. *When the number of treatments is $p = 1$, the expected relative efficiency $E(\mathrm{var}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{t}_r))) \geq (2/\pi)\mathrm{var}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{t}^*))$.*

S2.2 Sequential Experiments

Based on the SDRM scheme proposed in this paper, this subsection provides a robust sequential experiment scheme, and its detailed process is summarized in Algorithm $S1$.

In the initial stage of the sequential experiment, based on the minimax robust covariance structure $\mathbf{R}_*$ in Theorem 3, Algorithm 1 is used to generate the initial design $\mathbf{T}^{(0)}$, and the response $\mathbf{Y}^{(0)}$ in the initial stage is

Algorithm S1: Robust Sequential Experiment Scheme

Input: $\mathbf{R}_*$: Minimax robust covariance structure, B: Total number of stages.

Output: $\{\hat{\boldsymbol{\alpha}}^{(b)}\}_{b=0}^{B}$: Estimated treatment effects.

Initial stage: Based on $\mathbf{R}_*$, use Algorithm 1 to generate the initial design $\mathbf{T}_*^{(0)}$;

Observe the response $\mathbf{Y}^{(0)}$ of the subjects in the initial stage;

Obtain the maximum likelihood estimates of $\hat{\boldsymbol{\alpha}}^{(0)}$ and $\hat{\sigma}^{2(0)}$ by solving the equations in (S2-17).

for $b = 1, \cdots, B$ **do**

  Based on the estimated covariance $\mathbf{R}(\hat{\boldsymbol{\sigma}}^{2(b-1)})$, use Algorithm 1 to generate $\mathbf{T}_*^{(b)}$;

  Observe the response $\mathbf{Y}^{(b)}$ of the subjects in the b-th stage;

  Obtain the maximum likelihood estimates of $\hat{\boldsymbol{\alpha}}^{(b)}$ and $\hat{\boldsymbol{\sigma}}^{2(b)}$ by solving the equations in (S2-17).

end

Return the estimated treatment effects $\{\hat{\boldsymbol{\alpha}}^{(b)}\}_{b=0}^{B}$;

observed. Then, the maximum likelihood estimates $\hat{\boldsymbol{\alpha}}^{(0)}$ of the treatment effect and $\hat{\boldsymbol{\sigma}}^{2(0)} = (\hat{\sigma}_0^{2(0)}, \cdots, \hat{\sigma}_k^{2(0)})^T$ of the variance components are obtained by solving the following estimation equations (Searle et al., 2009, Chapter 6):

$$\begin{cases} \mathbf{X}^{(0)T}\mathbf{R}(\boldsymbol{\sigma}^2)^{-1}\mathbf{X}^{(0)}\boldsymbol{\theta} = \mathbf{X}^{(0)T}\mathbf{R}(\boldsymbol{\sigma}^2)^{-1}\mathbf{Y}^{(0)}, \\ \mathrm{tr}(\Sigma_t^{(0)}\mathbf{R}(\boldsymbol{\sigma}^2)^{-1}) = (\mathbf{Y}^{(0)} - \mathbf{X}^{(0)}\boldsymbol{\theta})^T\mathbf{Q}(\mathbf{Y}^{(0)} - \mathbf{X}^{(0)}\boldsymbol{\theta}), t = 0, \cdots k, \end{cases} \tag{S2-17}$$

where $\mathbf{X}^{(0)} = (\mathbf{T}^{(0)}, \mathbf{Z}^{(0)})$, $\mathbf{R}(\boldsymbol{\sigma}^2) = \sum_{t=0}^{k} \sigma_t^2 \Sigma_t^{(0)}$, $\mathbf{Q} = \mathbf{R}(\boldsymbol{\sigma}^2)^{-1}\Sigma_t^{(0)}\mathbf{R}(\boldsymbol{\sigma}^2)^{-1}$, and $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^{\mathbf{T}})^T$. In the $(b+1)$-th stage $(b \geq 0)$, based on the estimated covariance structure $\mathbf{R}(\hat{\boldsymbol{\sigma}}^{2(b)}) = \sum_{t=0}^{k} \hat{\sigma}_t^{2(b)} \Sigma_t$ in the b-th stage, Algorithm 1 is used to generate the treatment matrix $\mathbf{T}^{(b+1)}$ in the b-th stage, and the response $\mathbf{Y}^{(b)}$ in the b-th stage is observed. Subsequently, the estimated treatment effect $\hat{\boldsymbol{\alpha}}^{(b+1)}$ and variance components $\hat{\boldsymbol{\sigma}}^{2(b+1)}$ in the $(b+1)$-th stage are updated using the formula in (S2-17). Repeat the above process until all the experimental trials are used up.

The initial design of Algorithm $S1$ is optimally matched with the minimax robust covariance structure. Therefore, while utilizing the effectiveness of the optimal design, it ensures the robustness against the misspecification of the true covariance structure. In the subsequent experimental stages, Algorithm $S1$ uses the experimental data to update the covariance structure, thus continuously approaching the optimal design in the case where the true covariance structure is known.

S3 Supplementary Simulation Results

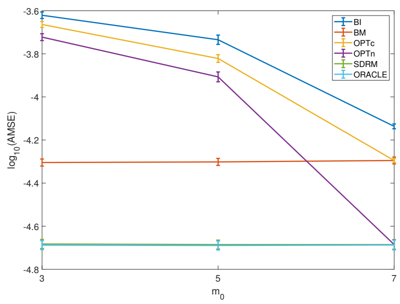
S3.1 Comparison of Experimental Designs under A/B Testing

In this subsection, we compare the performance of different experimental schemes in the case of A/B testing. Bhat et al. (2020) considered the fixed effects of covariates when the random error follows an independent and identically distributed structure, and obtained the optimal experimental design by minimizing the variance of the estimated treatment effect:

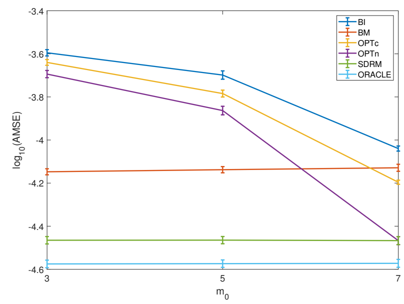
$$\mathbf{t}^* = \arg \min_{\mathbf{t} \in \{-1,1\}^n} \mathbf{t}^T (\mathbf{R}_0^{-1} - \mathbf{R}_0^{-1}\mathbf{Z}(\mathbf{Z}^T\mathbf{R}_0^{-1}\mathbf{Z})^{-1}\mathbf{R}_0^{-1})\mathbf{t}, \tag{S3-18}$$

where the working covariance structure $\mathbf{R}_0 = \sigma_0^2\mathbf{I}_n$. The above experimental scheme is denoted as OPT_c, where the subscript c emphasizes the role of the covariates. Zhang and Kang (2022) further considered the network connections between different subjects (the working covariance structure $\mathbf{R}_0 = \sigma_0^2\Sigma_0$), and obtained the experimental design by solving the optimization problem in (S3-18). This experimental scheme considering network correlations is denoted as OPT_n.

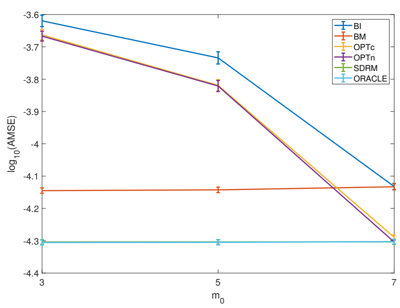
In the simulation, we assume that the relationship between the response and the covariates satisfies the one in (5.13). We set $\mathbf{Z} = (\mathbf{U}_1, \cdots, \mathbf{U}_{m_0})$ with $m_0 \leq k$, that is, only the first m_0 groups of covariates have fixed effects. The true covariance structure is randomly sampled form R1-R4 in Section 5. Figure $S1$ shows the AMSEs of various experimental schemes in 100 repetitions. From this, we can draw the following conclusions: the experimental schemes based on the completely randomized design (BI, BM) are less efficient because they ignore the information such as covariates and network structures. The design in the OPT_c scheme utilizes some covariates information, which significantly reduces the uncertainty of the estimation. The OPT_n scheme further improves its performance by using the information of the network connections between subjects. When the true covariance structure has a parametric form (Figure $S1$ (a) and (c)) and when $m_0 = k$, both the OPT_n and SDRM schemes can achieve the same estimation accuracy as the ORACLE scheme, because when all covariates have fixed effects, the true covariance structure $\mathbf{R} \in \Omega_{\Sigma_0}$ and $\mathbf{R} \in \Omega_{\mathbf{R}_*}$. When m_0 is small, the performance of the OPT_c and OPT_n schemes becomes unacceptable. It is worth noting that the experimental schemes based on the minimax covariance structure (BM, SDRM) are robust to the number of covariates with fixed effects, and the SDRM scheme can achieve the best results in various situations.



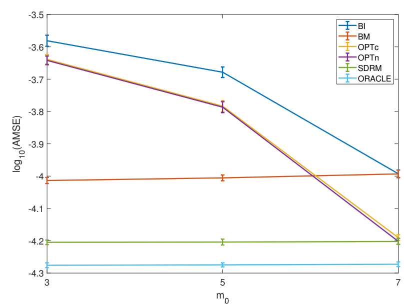
(a) $\mathbf{R} \in \Omega$ and AR Structure



(b) $\mathbf{R} \in \overline{\Omega}$ and AR Structure



(c) $\mathbf{R} \in \Omega$ and CAR Structure.



(d) $\mathbf{R} \in \overline{\Omega}$ and CAR Structure

Figure S1: AMSEs of various experimental schemes under A/B testing.

When the random error follows an AR structure, there is a significant performance gap between the OPT$_c$ and OPT$_n$ schemes, and when $m_0 = k$, the performance of the BM scheme is significantly better than that of the BI scheme, and the performance of the OPT$_n$ scheme is better than that of the OPT$_c$ scheme. However, when the random error follows a CAR structure, the above differences are not obvious. As discussed in Section 5, the difference caused by different structures of the random error is because the CAR structure is closer to the independent and identically distributed structure than the AR structure.

S3.2 Comparison of Experimental Designs under Sequential Experiments

In this subsection, we compare the performance of different experimental schemes in the sequential setting. The baseline scheme uses a completely randomized design to obtain the response in the initial stage, abbreviated to as "Balance". The sequential design in Algorithm $S1$, that is, using the robust experimental scheme $(\mathbf{T}_*, \mathbf{R}_*)$ in Theorem 3 to obtain the response in the initial stage, is referred to as "Robust". For the convenience of comparison, the experimental scheme that uses the true covariance structure as the input of Algorithm 1 to generate the optimal design is referred to as "Optimal". In the subsequent stages, all designs are the same as in Algorithm 1. Therefore, the only difference among the above three schemes is the initial design. In the simulation, we set the total number of stages to $B = 2$, the run size in each stage $n = 100$, and use the model (5.13) to generate the response. For each sequential scheme, the AMSE based on 100 groups of randomly generated true covariance structures from R1, is calculated. Figure $S2$ shows the AMSEs of various schemes in 100 repetitions.

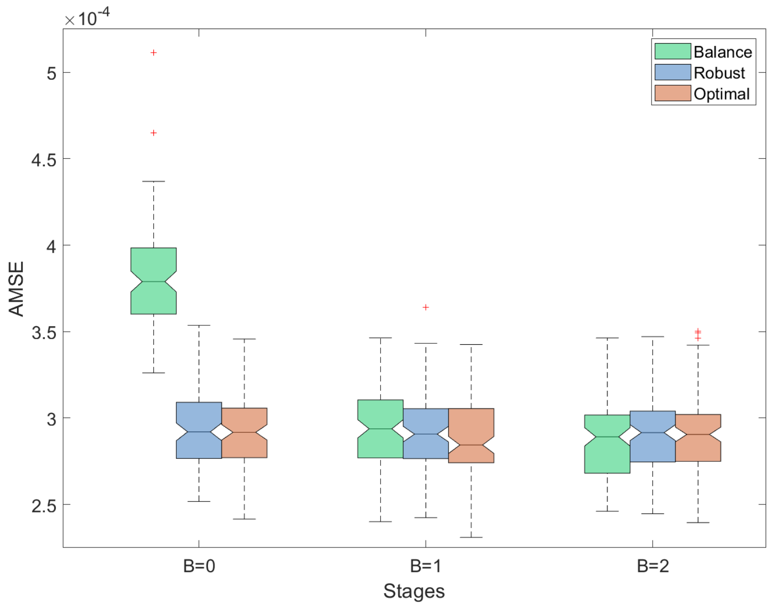


Figure S2: AMSEs of various sequential experimental schemes.

From this, we can draw the following conclusions: in the $B = 1$ and $B = 2$ stages, both the “Robust” and “Balance” schemes can achieve the same estimation accuracy as the “Optimal” scheme. It is worth noting that due to the use of an effective initial design, the “Robust” scheme greatly reduces the uncertainty of the estimation of the treatment effect in the initial stage. This result illustrates the effectiveness of the proposed robust experimental scheme.

S3.3 Supplementary Simulation Results in The Main Text

Following the simulation Section in the main text, we set $p = 5$, $q = m = 34$, $k = 7$, and consider the following covariates distribution: $u_{1i} \equiv 1$, $\mathbf{u}_{2i} \overset{i.i.d.}{\sim} \mathrm{MN}(2; 0.9, 0.1)$, $\mathbf{u}_{3i} \overset{i.i.d.}{\sim} \mathrm{MN}(10; 0.1 \times \mathbf{1}^{\mathbf{T}}_{10\times 1})$, $\mathbf{u}_{4i} \overset{i.i.d.}{\sim} (\chi^2_4, \chi^2_4, \chi^2_4)^T$, $\mathbf{u}_{5i} \overset{i.i.d.}{\sim} \mathbf{t}_3(\mathbf{0}, \mathbf{I}_5)$, $\mathbf{u}_{6i} \overset{i.i.d.}{\sim} N(\mathbf{0}, \mathbf{I}_6)$, $\mathbf{u}_{7i} \overset{i.i.d.}{\sim} N(\mathbf{0}, \mathbf{S})$,$i \in [n]$,where $S_{jl} = (0.7)^{|j-l|}$, $j, l \in [7]$, and χ^2_f represents the chi-squared distribution with degrees of freedom f, $\mathbf{t}_f(\boldsymbol{\mu}, \Sigma)$ represents the multivariate t distribution with location parameter $\boldsymbol{\mu}$, scale matrix Σ and degrees of freedom f, and $N(\boldsymbol{\mu}, \Sigma)$ represents the multivariate normal distribution with location parameter $\boldsymbol{\mu}$ and scale matrix Σ.

The Summary table across various 128-run experimental schemes are listed in Table S1. This results show that when covariates distributions exhibit greater complexity, the advantages of the proposed approaches become more pronounced.

Table S1: Summary table across various 128-run experimental schemes.

	BI	BM	OM	RSM	SDRM
			Average regret (standard deviation)		
R1	2.56(1.98)	1.23(0.45)	1.46(0.62)	0.81(0.24)	0.52(0.17)
R2	2.00(1.37)	1.52(1.32)	1.40(1.10)	1.17(1.23)	1.06(1.52)
R3	0.53(0.12)	0.53(0.12)	0.43(0.08)	0.39(0.10)	0.35(0.07)
R4	0.94(0.54)	0.94(0.54)	0.78(0.57)	0.73(0.49)	0.69(0.49)
			Average ϕ_0-efficiency (standard deviation)		
R1	0.02(0.02)	0.04(0.02)	0.05(0.03)	0.07(0.04)	0.94(0.12)
R2	0.02(0.02)	0.02(0.02)	0.03(0.03)	0.04(0.04)	0.40(0.34)
R3	0.33(0.06)	0.33(0.06)	0.36(0.06)	0.57(0.05)	1.00(0.07)
R4	0.15(0.11)	0.16(0.11)	0.16(0.12)	0.26(0.18)	0.46(0.32)
			Average MSE-efficiency (standard deviation)		
R1	0.54(0.06)	0.61(0.06)	0.65(0.07)	0.72(0.06)	0.97(0.07)
R2	0.54(0.05)	0.57(0.05)	0.62(0.06)	0.66(0.06)	0.88(0.10)
R3	0.80(0.05)	0.51(0.05)	0.84(0.05)	0.92(0.06)	1.00(0.05)
R4	0.72(0.05)	0.72(0.05)	0.74(0.04)	0.82(0.05)	0.87(0.06)

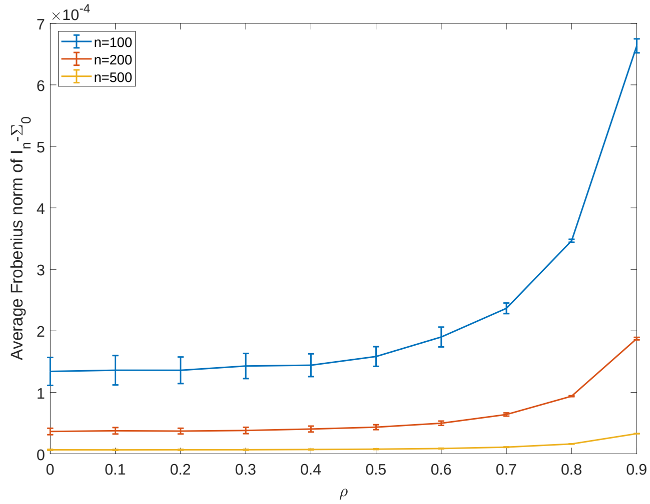


Figure S3: Average Frobenius-norm.

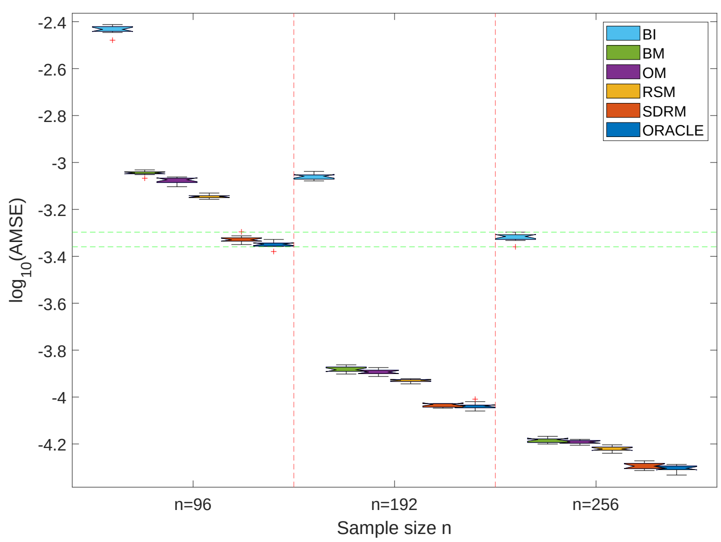


Figure S4: AMSEs in the case study when random errors follow an iid structure.

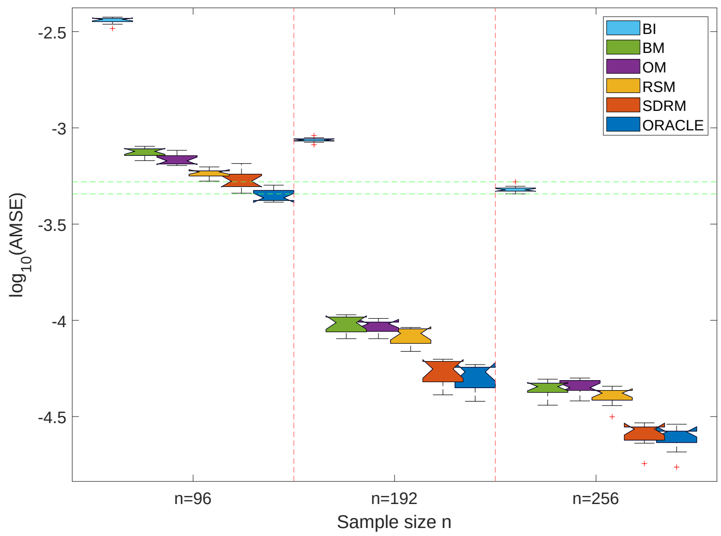


Figure S5: AMSEs in the case study when random errors follow an AR structure.

References

Ben-Tal, A. and A. Nemirovski (2001). *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM.

Bhat, N., V. F. Farias, C. C. Moallemi, and D. Sinha (2020). Near-optimal A-B testing. *Management Science 66*(10), 4477–4495.

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 355–372. Berkeley.

Searle, S. R., G. Casella, and C. E. McCulloch (2009). *Variance Components*. John Wiley & Sons.

Zhang, Q. and L. Kang (2022). Locally optimal design for A/B tests in the presence of covariates and network dependence. *Technometrics 64*(3), 358–369.