Statistica Sinica Preprint No: SS-2025-0157	
Title	Robust Control Experiments for Multivariate Tests with Covariates and Network Information
Manuscript ID	SS-2025-0157
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0157
Complete List of Authors	Shaohua Xu and Yongdao Zhou
Corresponding Authors	Yongdao Zhou
E-mails	ydzhou@nankai.edu.cn
Notice: Accepted author version.	

Robust control experiments for multivariate tests with covariates and network information

Shaohua Xu and Yongdao Zhou

Nankai University

Abstract: Multivariate testing has recently emerged as a promising technique in scientific decision-making and electronic information fields. Unlike standard A/B/n testing, which evaluates individual variations, multivariate testing aims to identify the best-performing combination of variations from all possible combinations. We address the challenge of robustly allocating treatments to subjects in multivariate testing when treatment effects are confounded by covariates and subjects are interconnected through a network. In this paper, for the first time we introduce the use of a mixed effect model to account for covariate uncertainty and network structure. Based on this model, we propose a criterion to measure the regret of efficiency due to incorrect specification of the covariance structure. We derive minimax robust experimental schemes and introduce a novel scheme that optimally matches the design with the robust covariance structure. Our proposed experimental schemes demonstrate: (a) resilience to various optimality criteria, (b) efficiency against model misspecification, and (c) applicability to complex scenarios. This work extends existing researches in optimal A/B testing designs, offering theoretical foundations and practical implementations that

outperform current approaches in statistical efficiency, as demonstrated through simulations and a case study.

Key words and phrases: A/B testing, minimax risk, mixed effect model.

1. Introduction

Multivariate testing (MVT) is a type of control experiment that has recently emerged as a promising technique in the fields of scientific decision making and electronic information, such as e-commerce, marketing research and clinical trials. Large-scale control experiments are widely adopted by technology giants such as Amazon and Google to optimize algorithms, user interfaces, and advertising strategies. Many tech companies have achieved significant business benefits through controlled experiments (Pokhilko et al., 2019; Kohavi et al., 2020). For a complete review of control experiments, see Kohavi et al. (2009) and Larsen et al. (2024).

Unlike traditional A/B testing, which evaluates only two versions of a single factor, MVT simultaneously tests multiple factors to identify the treatment combination that maximizes impact on key metrics (e.g., website click-through rates, clinical trial survival rates). The full factorial design, a classic approach for MVT, provides a comprehensive estimation of factor effects by equally allocating subjects to all treatment combinations. However,

its sample size grows exponentially with the number of factors, limiting its application in high-dimensional settings. Fractional factorial designs, such as Taguchi methods (Jiang et al., 2020), slice designs (Sadeghi et al., 2020), and sequential designs (Haizler and Steinberg, 2021), enhance experimental feasibility while maintaining statistical power by reducing the number of test combinations.

In practical applications, control experiments with covariates and network information are common and present unique challenges. For example, in coupon distribution experiments, due to concerns about customer churn risk, testing is typically limited to a subset of customers from different consumption tiers, and these customers often exhibit complex social network connections. In these scenarios, in addition to the treatments, the covariates of the subjects and network structure among subjects are usually available to the experimenter and can significantly influence the responses. To address this issue, existing research focuses primarily on the randomization framework (Rubin, 2005), employing rerandomization (Branson et al., 2016; Li and Ding, 2020; Pashley and Bind, 2023), covariate adjustment (Zhao and Ding, 2023; Liu et al., 2024), and other covariate balancing techniques (Zhao and Ding, 2022; Bai et al., 2024) to obtain treatment effect estimates. Within the A/B testing paradigm, parametric response-treatment-covariate

frameworks have recently been proposed as alternative methodologies to model covariates and network effects (Pokhilko et al., 2019; Bhat et al., 2020; Zhang and Kang, 2022). For example, Zhang and Kang (2022) utilized residuals to capture network dependencies among subjects and devised a D_s-optimal design to minimize the variance of estimated treatment effects in the presence of covariates. However, existing methods still have some limitations: traditional randomization methods face substantial uncertainty estimation under complex data structures, while optimal design approaches based on parametric regression models within A/B testing frameworks exhibit high sensitivity to model misspecification (Wiens, 2015).

In this paper, we introduce a linear mixed effect model to address these challenges, modeling covariate effects as random variables to quantify uncertainty and using residuals to capture network dependencies among subjects. We propose a criterion to quantify covariance structure misspecification and derive the corresponding minimax experimental scheme. Additionally, we develop a novel experimental scheme that optimally matches the design with the robust covariance structure. Our experimental schemes are robust against optimality criteria for estimating treatments and the misspecification of the covariance structure, making them applicable to various control experiments with complex covariates and network structures.

The remainder of this paper is organized as follows. In Section 2, we develop a mixed effect model to quantify the uncertainty of covariates and network structures in MVT, and propose a criterion to measure the loss of efficiency due to incorrect specification of the covariance structure. Section 3 derives minimax robust experimental schemes for estimating treatment effects. In Section 4, we propose a novel experimental scheme that optimally matches the treatment with the robust covariance structure. Sections 5 and 6 compare different experimental schemes under various linear mixed-effects models. Finally, Section 7 discusses future works. For clarity, additional simulation results and proofs of theoretical results are included in the Supplementary Materials.

2. Problem Setups

Consider n subjects randomly selected from a population to participate in a MVT, where the goal is to examine differences between different treatments. We are primarily interested in the treatment effects, thus treat them as fixed effects. Considering the complex structure of the experimental units, we divide the covariates of the subjects into two groups to model the fixed effects and random effects of the covariates respectively. Specifically, suppose that the relationship among the experimental result Y_i, the

treatment factor $\mathbf{t}_i$, and the two groups of covariates $\mathbf{z}_i$ and $\mathbf{u}_i$ of the i-th subject can be described by the following linear mixed-effects model (LMM, Verbeke et al., 1997; Searle et al., 2009):

$$Y_i = \mathbf{t}_i^T \boldsymbol{\alpha} + \mathbf{z}_i^T \boldsymbol{\beta} + \mathbf{u}_i^T \boldsymbol{\xi} + \varepsilon_i, i \in [n], \tag{2.1}$$

where $\boldsymbol{\alpha}$ is the p-dimensional treatment effect; $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ are the q-dimensional fixed effect and m-dimensional random effect of the covariates respectively; ε_i is the random error. The treatment effect $\boldsymbol{\alpha}$ can be interpreted as the expected change in the experimental results when subjects with the same covariates are assigned to different treatments.

Let $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$ be the response vector of the subjects, $\mathbf{T} = (\mathbf{t}_1, \cdots, \mathbf{t}_n)^T$ be the $n \times p$ treatment matrix, $\mathbf{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_n)^T$ and $\mathbf{U} = (\mathbf{u}_1, \cdots, \mathbf{u}_n)^T$ be the $n \times q$ and $n \times m$ covariates matrices corresponding to the fixed effect and random effect respectively, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T$ be the random error vector. As done by Verbeke et al. (1997); Searle et al. (2009), we partition the m-dimensional random effect into k groups, that is, $\boldsymbol{\xi} = (\boldsymbol{\xi}_1^T, \cdots, \boldsymbol{\xi}_k^T)^T$ and $\mathbf{U} = (\mathbf{U}_1, \cdots, \mathbf{U}_k)$, where for $t \in [k]$, $\mathbf{U}_t = (\mathbf{u}_{t1}, \cdots, \mathbf{u}_{tn})^T$ is the covariates matrix corresponding to the m_t-dimensional random effect $\boldsymbol{\xi}_t$, and $m = \sum_{t=1}^{k} m_t$. In this paper, we consider the case where $p + q + m < n$.

We make the following common assumptions about the first two mo-

ments for the random effects and random errors.

Assumption 1. $E(\boldsymbol{\xi}) = \mathbf{0}_{m\times 1}$ and $\mathrm{cov}(\boldsymbol{\xi}) = \mathrm{diag}\{\sigma_1^2\mathbf{I}_{m_1}, \cdots, \sigma_k^2\mathbf{I}_{m_k}\}$.

Assumption 2. $E(\boldsymbol{\varepsilon}) = \mathbf{0}_{n\times 1}$ and $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma_0^2\Sigma_0$, where Σ_0 is a known positive-definite matrix.

Assumption 3. $\mathrm{cov}(\boldsymbol{\xi}, \boldsymbol{\varepsilon}) = \mathbf{0}_{m\times n}$.

Assumption 1 requires that the random effect has a zero mean and no correlation, which is commonly used in the literature of LMM (Verbeke et al., 1997; Searle et al., 2009). Assumption 2 requires that the random error has a zero mean, and a general covariance matrix is used to model the correlation among subjects. When there are network connections among subjects, a common correlation assumption in the A/B testing literature is the conditional autoregressive (CAR) distribution (Pokhilko et al., 2019; Zhang and Kang, 2022). Assumption 3 requires that the random effects and random errors are uncorrelated, a sufficient condition that ensures the separability of inter-group and intra-group variations (Searle et al., 2009).

Under the LMM (2.1) and Assumptions 1-3, the responses of the subjects are modeled as an n-dimensional random vector with the following

mean and covariance matrix:

$$\begin{cases} E(\mathbf{Y}) = \mathbf{T}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}; \\ \text{cov}(\mathbf{Y}) = \sum_{t=0}^{k} \sigma_t^2 \Sigma_t \triangleq \mathbf{R}, \end{cases} \tag{2.2}$$

where σ_t^2 represents the t-th variance component, and $\Sigma_t = \mathbf{U}_t\mathbf{U}_t^T$, for $t \in [k]$. Therefore, the fixed effect $\boldsymbol{\beta}$ and the random effect $\boldsymbol{\xi}$ in the model (2.1) quantify the impacts of the covariates on the mean and covariance of the experimental results respectively, and the experimental results of subjects with similar covariates are more consistent than those of subjects with different covariates. Recently, two special LMMs have been studied in the literature of A/B testing. The first case corresponds to $\text{cov}(\mathbf{Y}) = \sigma_0^2\mathbf{I}_n$, where different subjects are isolated, and their covariates only affect the expected outcomes of the subjects (Bhat et al., 2020). The second case corresponds to $\text{cov}(\mathbf{Y}) = \sigma_0^2\Sigma_0$, where different subjects are connected through a network, but their covariates have no impact on the covariance structure of subjects (Pokhilko et al., 2019; Zhang and Kang, 2022). Example 1 illustrates that the covariance structure in practical problems may be rather complex, then the LMM framework (2.2) provides a more suitable modeling approach for controlled experiments.

In this paper, we consider the problem of experimental design in the LMM (2.1), and focus on the estimation of the treatment effect $\boldsymbol{\alpha}$. Through-

out the paper, let $\mathbf{L} = (\mathbf{I}_p, \mathbf{0}_{p\times q})$, $\mathbf{X} = (\mathbf{T}, \mathbf{Z})$, and $\Sigma(\mathbf{R}) = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{R}^{-1}$. We introduce the following assumption to ensure the existence of the Generalized Least Squares (GLS) estimator

$$\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T}) = \mathbf{L}(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{R}^{-1}\mathbf{Y}.$$

Assumption 4. The covariates matrix $\mathbf{Z}$ satisfies $\text{rank}(\mathbf{Z}) = q$, and there exists a constant $\mu \in (0, 1]$ such that $\mathbf{T}^T\Sigma\mathbf{T} \succeq \mu\mathbf{T}^T\mathbf{R}^{-1}\mathbf{T}$.

Specifically, $\mu = 0$ implies that $\text{rank}(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}) < p + q$, $\mu = 1$ implies that $\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}$ is a block diagonal matrix, and $\mu \in (0, 1)$ indicates that $\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X}$ is between a column-full-rank matrix and a block diagonal matrix. If Assumption 4 does not hold, the generalized inverse based least squares estimator can be discussed similarly (Searle et al., 2009, Appendix M.4).

When the true covariance matrix $\mathbf{R}$ is known, the variance-covariance matrix of $\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})$ is given by

$$\text{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})) = \mathbf{L}(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{L}^T = (\mathbf{T}^T\Sigma(\mathbf{R})\mathbf{T})^{-1}.$$

From the perspective of optimal experimental designs, for a given $s \in (0, \infty)$, the ϕ_s-optimal design (Atkinson et al., 2007) minimizes the design criterion: $\phi_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})) = \left(\text{tr}(\text{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T}))^s/p)\right)^{1/s}$, thus providing the best estimate of the treatment effect from different perspectives. Specifically, the commonly used A-, D-, and E-optimal designs correspond to the cases

where the values of s are 1, 0, and ∞ respectively. It is worth noting that the above ϕ_s-optimal design depends on the true covariance matrix $\mathbf{R}$.

Typically, the above ϕ_s-optimal design is not available because the variance components variance components $\sigma_0^2, \ldots, \sigma_k^2$ in $\mathbf{R}$ are not known in advance. Therefore, under a complex covariance structure, the key to experimental design is to effectively utilize the information in the unknown covariance matrix. In this paper, we seek experimental schemes that are robust to the true covariance matrix. Specifically, when the true covariance matrix $\mathbf{R}$ is misspecified as the working covariance matrix $\mathbf{R}_0$, the variance-covariance matrix of the GLS estimator $\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})$ is

$$\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})) = \mathbf{L}(\mathbf{X}^T\mathbf{R}_0^{-1}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{R}_0^{-1}\mathbf{R}\mathbf{R}_0^{-1}\mathbf{X})(\mathbf{X}^T\mathbf{R}_0^{-1}\mathbf{X})^{-1}\mathbf{L}^T.$$

Although both estimators $\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})$ and $\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})$ are unbiased, the Gauss-Markov theorem (Harville, 1976) shows that $\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})$ is inefficient, that is,

$$\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})) - \mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})) = \mathbf{D}^T\mathbf{R}\mathbf{D} \succeq 0,$$

where $\mathbf{D}^T = \mathbf{L}(\mathbf{X}^T\mathbf{R}_0^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{R}_0^{-1} - \mathbf{L}(\mathbf{X}^T\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{R}^{-1}$. Therefore, $\mathbf{D}^T\mathbf{R}\mathbf{D}$ can be regarded as the regret caused by the misspecification of the covariance matrix. We define this regret as

$$\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) = \frac{\|\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})) - \mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T}))\|}{\|\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T}))\|}, \tag{2.3}$$

where $\|\cdot\|$ is the spectral norm operator. A key property of the regret is that $\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) \geq 0$, and it equals zero when $\mathbf{R}_0 = \mathbf{R}$. Therefore, the smaller the regret, the better the performance of the covariance matrix $\mathbf{R}_0$ and the design $\mathbf{T}$. In particular, zero regret means that the working covariance matrix achieves the same effect as the true covariance matrix. Another important property of the regret is that for any $c, c_0 > 0$, we have $\mathrm{RT}(c\mathbf{R}, c_0\mathbf{R}_0, \mathbf{T}) = \mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T})$. This indicates that the regret depends only on the covariance structure, rather than on a specific covariance matrix. The following example demonstrates that we can significantly reduce the regret by selecting an appropriate working covariance structure.

Example 1. In a controlled experiment, there are n subjects. The first n_1 subjects are male, and the last n_1 subjects are married. That is,

$$\mathbf{U}_1^T = \begin{pmatrix} \overbrace{1, \cdots, 1}^{n_1 \text{ males}}, & 0, \cdots, 0 \\ 0, \cdots, 0, & \underbrace{1, \cdots, 1}_{n-n_1 \text{ females}} \end{pmatrix}, \quad \mathbf{U}_2^T = \begin{pmatrix} \overbrace{1, \cdots, 1}^{n-n_1 \text{ unmarried}}, & 0, \cdots, 0 \\ 0, \cdots, 0, & \underbrace{1, \cdots, 1}_{n_1 \text{ married}} \end{pmatrix}.$$

In this situation, a suitable covariance structure is $\mathbf{R} = \mathbf{I}_n + a_1\Sigma_1 + a_2\Sigma_2$, where $a_1, a_2 \geq 0, \Sigma_1 = \mathbf{U}_1\mathbf{U}_1^T, \Sigma_2 = \mathbf{U}_2\mathbf{U}_2^T$. Suppose that the experimenter uses a balanced design $\mathbf{T}_b = \mathbf{1}_{(n/2)\times 1} \otimes (1, -1)^T$ to test the difference between the two treatments. Under each true covariance structure $\mathbf{R} = \mathbf{I}_n + a_1\Sigma_1 + a_2\Sigma_2$ with $a_1 \in [0, 1]$ and $a_2 \in [0, 1]$, we calculate the regrets (2.3) for the

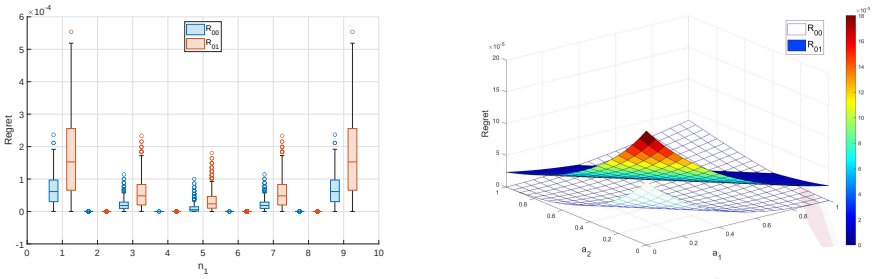
two working covariance structures: $\mathbf{R}_{00} = \mathbf{I}_n + 0.5\Sigma_1 + 0.5\Sigma_2$ and $\mathbf{R}_{01} = \mathbf{I}_n + \Sigma_1 + \Sigma_2$.

As shown in Figure 1(a), the regret is influenced by both the covariates imbalance between treatment and control groups and the chosen working covariance structure. Specifically, when n_1 is odd, the differences in both the number of male/female and married/unmarried subjects between the treatment and control groups are 1; when n_1 is even, those differences are 0. The regret of the experimental scheme is unaffected by the working covariance structure only when the design perfectly balances all covariates. Otherwise, selecting an appropriate working covariance structure can significantly reduce the regret. Figure 1(b) further displays the regret surfaces under $\mathbf{R}_{00}$ and $\mathbf{R}_{01}$ when $n_1 = 5$. Intuitively, the regret surface of $\mathbf{R}_{00}$ is flatter than that of $\mathbf{R}_{01}$. Additionally, the maximum and average regrets of $\mathbf{R}_{01}$ are 1.8 and 3.2 times higher than those of $\mathbf{R}_{00}$, respectively. Therefore, $\mathbf{R}_{00}$ is a more suitable working covariance structure for allocating treatments.

To avoid the situation of ill-conditioned covariance structures, we define the following covariance class:

$$\Omega_0 = \left\{ \mathbf{R} = \sum_{t=0}^{k} \sigma_t^2 \Sigma_t \mid \kappa(\mathbf{R}) \leq \kappa_0, \sigma_t^2 \geq 0 \text{ for } t \in [k] \cup \{0\} \right\}, \qquad (2.4)$$

where $\kappa(\mathbf{R})$ is the condition number of $\mathbf{R}$, and κ_0 controls the condition



(a) Regret under different values of n_1. (b) Regret surface when $n_1 = 5$.

Figure 1: Regrets of working covariance matrices $\mathbf{R}_{00}$ and $\mathbf{R}_{01}$ under various true covariance matrices $\mathbf{R}$.

number of the covariance structures in Ω_0. Note that, for given $\Sigma_0, \cdots, \Sigma_k$, the covariance class in (2.4) is determined by the parameters $\sigma_0^2, \cdots, \sigma_k^2$. Therefore, we call Ω_0 a parametric covariance class. More generally, for any $r \geq 0$, we introduce the following non-parametric covariance class:

$$\Omega_r = \{\mathbf{R} + \mathbf{K} \mid \mathbf{R} \in \Omega_0, \mathbf{R} + \mathbf{K} \succ 0 \text{ and } \|\mathbf{K}\| \leq r\}. \tag{2.5}$$

The covariance class Ω_r contains all covariance matrices whose spectral radius from a certain parametric covariance matrix in (2.4) does not exceed r. Here, the spectral norm $\|\cdot\|$ is chosen for convenience; other matrix norms are also feasible. With this definition, the LMM with Assumptions 1-3 can be extended to responses with $E(\mathbf{Y}) = \mathbf{T}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}$ and $\text{cov}(\mathbf{Y}) \in \Omega_r$. When $r = 0$, the non-parametric covariance class Ω_r degenerates to the parametric covariance class Ω_0. Therefore, this provides a more general

framework than (2.2) for controlled experiments.

Let $\Theta = \{\mathbf{T} \mid \mathbf{T} \in \{-1, 1\}^{n\times p}, \text{rank}(\mathbf{T}) = p\}$ denote the full-rank design space. In this paper, we adopt the minimax robust framework to find a design $\mathbf{T}_* \in \Theta$ and a working covariance matrix $\mathbf{R}_* \in \Omega_0$ such that they have high inference ability for the treatment effect under various true covariance structures in Ω_r, that is,

$$(\mathbf{T}_*, \mathbf{R}_*) = \arg \min_{\mathbf{T}\in\Theta, \mathbf{R}_0\in\Omega_0} \max_{\mathbf{R}\in\Omega_r} \text{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}). \tag{2.6}$$

Since the experimental scheme $(\mathbf{T}_*, \mathbf{R}_*)$ has the minimum regret under the worst-case covariance structure, we call it a minimax robust experimental scheme. Based on the minimax covariance structure $\mathbf{R}_*$, we further provide an approximate solution to the following ϕ_s-optimal design:

$$\mathbf{T}_s^* = \arg \min_{\mathbf{T}\in\Theta} \text{tr}\left((\mathbf{T}^T\Sigma(\mathbf{R}_*)\mathbf{T})^{-1})^{-s}/p\right)^{1/s}, \tag{2.7}$$

which combines the efficiency of the ϕ_s-optimal design and the robustness of $\mathbf{R}_*$ with respect to the true covariance structure.

3. Minimax Experimental Scheme against a Misspecification of Covariance Structure

In this section, we present the minimax robust experimental schemes for problem (2.6) under different scenarios. First, we establish the connection

between the regret in (2.3) and the ϕ_s-criterion. For any GLS estimator $\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})$, its ϕ_s-efficiency relative to $\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})$ is defined as

$$\mathrm{Eff}_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})) \triangleq \frac{\phi_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T}))}{\phi_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T}))}, \quad s \in [0, \infty].$$

Obviously, the ϕ_s-efficiency takes values in the interval $[0, 1]$. The larger the ϕ_s-efficiency of the estimator $\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})$ is, the higher the estimation accuracy of $\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})$ under the ϕ_s-optimality criterion is. In particular, $\mathrm{Eff}_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})) = 1$ means that $\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})$ and $\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})$ are equivalent.

Proposition 1. *For any* $(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) \in \Omega_r \times \Omega_0 \times \Theta$, *and* $s \geq 1$, *the* ϕ_s-*efficiency of* $\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})$ *satisfies*

$$\mathrm{Eff}_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})) \geq \frac{1}{1 + \kappa_0 \kappa(\mathbf{X}^T\mathbf{X})\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T})}.$$

If the covariates matrix $\mathbf{Z}$ *satisfies Assumption 4, then*

$$\mathrm{Eff}_s(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})) \geq \frac{1}{1 + \kappa_0 \mu^{-1} \kappa(\mathbf{T}^T\mathbf{T})\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T})}.$$

The lower bound given by Proposition 1 uniformly holds for all ϕ_s-optimality criteria with $s \geq 1$. For any $\epsilon \in [0, 1)$, if the risk $\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) \leq [(1-\epsilon)^{-1} - 1](\kappa_0 \mu^{-1} \kappa(\mathbf{T}^T\mathbf{T}))^{-1}$, then for all ϕ_s-optimality criteria with $s \geq 1$, the estimation efficiency of $\tilde{\boldsymbol{\alpha}}$ is at least $100(1-\epsilon)\%$. Therefore, the regret in (2.3) provides a way to control all ϕ_s-criteria with $s \geq 1$.

Define the index set $J = \{(j_0, j_1, \cdots, j_k) \mid j_t \in \{0,1\}\} \setminus \mathbf{0}_{(k+1)\times 1}$. For any subclass $C \subseteq \Omega_0$, let $[C] = \{c\mathbf{R} \mid c > 0, \mathbf{R} \in C\}$. The following lemma gives an equivalent characterization of the covariance class Ω_0.

Lemma 1. *For any $\kappa_0 \geq \kappa(\Sigma_0)$, there exists $\delta \in (0,1)$ such that*

$$\Omega_0 = \cup_{\mathbf{j}\in J}[\Omega^*_{\mathbf{j}}],$$

where for any $\mathbf{j} \in J$, $\Omega^*_{\mathbf{j}} = \{\mathbf{R} = \sum_{t=0}^k a_t\Sigma_t \mid a \in A_{\mathbf{j}}\}$, $A_{\mathbf{j}} = \times_{t=0}^k A_{\mathbf{j},t}$, if $j_t = 1$, then $A_{\mathbf{j},t} = \{1\}$; otherwise $A_{\mathbf{j},t} = [\delta\mathbb{I}\{t=0\}, 1)$.

Lemma 1 shows that the covariance class Ω_0 can be partitioned into the union of $2^{k+1}-1$ disjoint subclasses. For example, when $k = 1$, we have $\Omega^*_{(0,1)} = \{a_0\Sigma_0 + \Sigma_1 \mid a_0 \in [\delta, 1)\}$, $\Omega^*_{(1,0)} = \{\Sigma_0 + a_1\Sigma_1 \mid a_1 \in [0,1)\}$, and $\Omega^*_{(1,1)} = \{\Sigma_0 + \Sigma_1\}$. $\Omega^*_{(0,1)}, \Omega^*_{(1,0)}$ and $\Omega^*_{(1,1)}$ correspond to the cases where $\sigma_0^2 > \sigma_1^2, \sigma_0^2 < \sigma_1^2$ and $\sigma_0^2 = \sigma_1^2$ in the true covariance structure, respectively. An intuitive interpretation of the parameter δ is that the smaller δ is, the more ill-conditioned covariance structures are included in Ω_0. In particular, when $\delta \to 0^+$, $\Omega^*_{(0,1)}$ contains covariance structures with the condition number tending to infinity.

The following theorem establishes an upper bound for the regret.

Theorem 1. *For any $\mathbf{R} \in \Omega_r$, let $\mathbf{R}^* = \sum_{t=0}^k \sigma_t^{*2}\Sigma_t$ be the parametric covariance structure in Ω_0 that is closest to $\mathbf{R}$ in the sense of the spectral*

norm. If Assumption 4 holds, then for any $(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) \in \Omega_r \times \Omega_0 \times \Theta$*, the regret* $\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T})$ *is upper bounded by*

$$\overline{\mathrm{RT}}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) = C_1 \kappa(\mathbf{T}^T\mathbf{T})(C_2\|\boldsymbol{\sigma}^{*2} - \boldsymbol{\sigma}_0^2\|_1 + r)^2,$$

where C_1 *and* C_2 *are constants independent of the design and the covariance structure, and* $\|\cdot\|_1$ *denotes the* l_1*-norm.*

Theorem 1 demonstrates that the regret between the nonparametric covariance $\mathbf{R}$ and the parametric covariance $\mathbf{R}_0$ is controlled by the l_1-norm of the difference in the variance components between $\mathbf{R}^*$ and $\mathbf{R}_0$. Therefore, $\max_{\mathbf{R}\in\Omega_r} \overline{\mathrm{RT}}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) = \max_{\mathbf{R}^*\in\Omega_0} \overline{\mathrm{RT}}(\mathbf{R}^*, \mathbf{R}_0, \mathbf{T})$. In addition, the effects of the design $\mathbf{T}$ and the covariance structure $(\mathbf{R}, \mathbf{R}_0)$ on this upper bound are separable. Geometrically, the complex regret surface $\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T})$ can be controlled by a surface $\overline{\mathrm{RT}}(\mathbf{R}, \mathbf{R}_0, \mathbf{T})$ with good properties, and if a experimental design has a smaller upper bound of the regret, then it usually also has a smaller regret, thus ensuring a higher ϕ_s efficiency.

In practical problems, experimenters are also interested in specific linear combinations of treatment effects. For example, the parameter $\alpha_1 - \alpha_2$ measures the difference between the first two treatment effects. Generally speaking, for any $b \times p$ row full-rank matrix Γ, if we focus on the parameter vector in the form of $\boldsymbol{\gamma} = \Gamma\boldsymbol{\alpha}$, then the upper bound in Theorem 1 can still provide effective guidance for the corresponding regret.

Corollary 1. *For any* $(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) \in \Omega_r \times \Omega_0 \times \Theta$, *if Assumption* 4 *holds, then for the estimators* $\hat{\boldsymbol{\gamma}}(\mathbf{R}_0, \mathbf{T}) = \Gamma\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T})$ *and* $\hat{\boldsymbol{\gamma}}(\mathbf{R}, \mathbf{T}) = \Gamma\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})$, *we have*

$$\frac{\|\mathrm{cov}(\hat{\boldsymbol{\gamma}}(\mathbf{R}_0, \mathbf{T})) - \mathrm{cov}(\hat{\boldsymbol{\gamma}}(\mathbf{R}, \mathbf{T}))\|}{\|\mathrm{cov}(\hat{\boldsymbol{\gamma}}(\mathbf{R}, \mathbf{T})\|} \leq \kappa(\Gamma\Gamma^{\mathbf{T}})\overline{\mathrm{RT}}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}).$$

Based on the above representation of Ω_0, we discuss the experimental designs that minimize the maximum regret upper bound in two cases. If the experimenter knows in advance which variance components will dominate, that is, $\mathbf{R} \in [\Omega^*_{\mathbf{j}}]$ for some $\mathbf{j} \in J$, then we consider the following minimax experimental scheme:

$$(\mathbf{T}_*, \mathbf{R}_*) \in \arg \min_{\mathbf{T}\in\Theta, \mathbf{R}_0\in\Omega_0} \max_{\mathbf{R}\in[\Omega^*_{\mathbf{j}}]} \overline{\mathrm{RT}}(\mathbf{R}^*, \mathbf{R}_0, \mathbf{T}). \tag{3.8}$$

We provide a solution to this minimax problem as follows.

Theorem 2. *Suppose that for some* $\mathbf{j} \in J$, $\mathbf{R} \in [\Omega^*_{\mathbf{j}}]$. *If the design* $\mathbf{T}_*$ *satisfies* $\mathbf{T}_*^T\mathbf{T}_* = n\mathbf{I}_p$, *and the working covariance structure is*

$$\mathbf{R}_* = \begin{cases} \Sigma_0 + (1/2)\sum_{t=1,j_t=0}^{k} \Sigma_t + \sum_{t=1,j_t=1}^{k} \Sigma_t, & \text{if } j_0 = 1; \\ [(1+\delta)/2]\Sigma_0 + (1/2)\sum_{t=1,j_t=0}^{k} \Sigma_t + \sum_{t=1,j_t=1}^{k} \Sigma_t, & \text{if } j_0 = 0, \end{cases}$$

then $(\mathbf{T}_*, \mathbf{R}_*) \in \arg\min_{\mathbf{T}\in\Theta, \mathbf{R}_0\in\Omega_0} \max_{\mathbf{R}\in[\Omega^*_{\mathbf{j}}]} \overline{\mathrm{RT}}(\mathbf{R}, \mathbf{R}_0, \mathbf{X})$.

Theorem 2 shows that the column-orthogonal treatment matrix and the parametric covariance structure $\mathbf{R}_*$ can minimize the maximum regret

upper bound, and the variance components with $j_t = 1$ should be larger than the other components. The conclusion in Theorem 2 is applicable when the relative magnitudes of the variance components can be estimated from previous experimental results or determined by experts in the field. In Example 1, we assume that $\mathbf{R} \in \Omega^*_{(1,0,0)}$ and show that the corresponding minimax covariance structure $\mathbf{R}_{00}$ has a smaller regret than $\mathbf{R}_{01}$.

If the experimenter has no prior knowledge about the covariance structure, a reasonable assumption in this case is the nonparametric covariance class Ω_r, and the following minimax experimental design can be considered:

$$(\mathbf{T}_*, \mathbf{R}_*) \in \arg \min_{\mathbf{T}\in\Theta, \mathbf{R}_0\in\Omega_0} \max_{\mathbf{R}\in\Omega_r} \overline{\mathrm{RT}}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}). \tag{3.9}$$

The following theorem provides a solution to the minimax problem (3.9).

Theorem 3. *If the design* $\mathbf{T}_*$ *satisfies* $\mathbf{T}_*^T\mathbf{T}_* = n\mathbf{I}_p$, *and the working covariance structure is*

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

then $(\mathbf{T}_*, \mathbf{R}_*) \in \arg\min_{\mathbf{T}\in\Theta,\mathbf{R}_0\in\Omega_0} \max_{\mathbf{R}\in\Omega_r} \overline{\mathrm{RT}}(\mathbf{R}, \mathbf{R}_0, \mathbf{T})$.

Theorem 3 shows that, in the absence of information about the true covariance structure, the treatment matrix of the minimax experimental

design is still column-orthogonal, but its covariance structure assigns the maximum weight to the components of the random error. It should be noted that when the number of groups of random effects $k \geq 2$, the minimax covariance structure is independent of the parameter δ, which further enhances its practicality. In fact, $\mathbf{R}_{00}$ in Example 1 is the minimax covariance structure in the cases where $\mathbf{R} \in \Omega_r$ or $\mathbf{R} \in \Omega^*_{(1,0,0)}$.

Next, we identify the conditions under which the above minimax experimental schemes achieves zero regret. For any positive definite matrix A, define $\Omega_{\mathbf{A}} = \{\Sigma \mid \Sigma = \mathbf{A} + \mathbf{X}\mathbf{D}_1\mathbf{X}^T + \mathbf{V}\mathbf{D}_2\mathbf{V}^T\}$, where $\mathbf{D}_1, \mathbf{D}_2$ are arbitrary symmetric matrices such that Σ is positive definite, and $\mathbf{V}$ is an arbitrary column full-rank matrix satisfying $\mathbf{X}^T\mathbf{A}^{-1}\mathbf{V} = \mathbf{0}$.

Theorem 4. *For any experimental scheme* $(\mathbf{R}_0, \mathbf{T}) \in \Omega_0 \times \Theta$*, we have* $\mathrm{RT}(\mathbf{R}, \mathbf{R}_0, \mathbf{T}) = 0$ *if and only if* $\mathbf{R} \in [\Omega_{\mathbf{R}_0}]$.

The above conclusion shows that the robust experimental scheme in Theorem 3 has zero regret under a class of parametric covariance structures. Therefore, if $\mathbf{R} \in [\Omega_{\mathbf{R}_*}]$, then the minimax working covariance structure achieves the same effect as the true covariance structure. Note that if there exists a column full-rank matrix $\mathbf{Q} \in \mathbb{R}^{q\times m}$ such that $\mathbf{U} = \mathbf{Z}\mathbf{Q}$, then $\Omega_0 \subset [\Omega_{\mathbf{R}_*}]$, that is, the minimax robust experimental scheme in Theorem 3 achieves zero regret in the parametric covariance class Ω_0.

4. Optimal Matching Based on Minimax Covariance Structure

In this section, we give an approximate solution for any ϕ_s-optimal design $\mathbf{T}_s^*$ in (2.7):

$$\mathbf{T}_s^* = \arg\min_{\mathbf{T}\in\Theta} \mathrm{tr}((\mathbf{T}^T\Sigma(\mathbf{R}_*)\mathbf{T})^{-s}/p)^{1/s}, \tag{4.10}$$

where $\mathbf{R}_*$ is the minimax robust covariance structure given in Theorem 2 or 3. Intuitively, the optimal solution of Problem (4.10) not only preserves the effectiveness of the ϕ_s-optimal design, but also inherits the robustness of the minimax covariance structure $\mathbf{R}_*$.

However, solving the above ϕ_s-optimal design is quite difficult, because even in the simplest case where $s = p = 1$, it is an NP-complete problem (Nesterov, 1998). From Lemma S2 in the Supplementary Materials, any ϕ_s-optimality criterion is controlled by $\mathrm{tr}(\mathbf{T}^T\Sigma(\mathbf{R}_*)\mathbf{T}/p)^{-1}$. An intuitive interpretation is that a larger value of $\mathrm{tr}(\mathbf{T}^T\Sigma(\mathbf{R}_*)\mathbf{T}/p)$ indicates that the performance of the design $\mathbf{T}$ is better under all ϕ_s-optimality criteria. Therefore, we consider the following trace maximization problem:

$$\mathbf{T}^* = \arg\max_{\mathbf{T}\in\Theta} \mathrm{tr}(\mathbf{T}^T\Sigma(\mathbf{R}_*)\mathbf{T}/p). \tag{4.11}$$

Although the solution of Problem (4.11) provides an approximation for all ϕ_s-optimal designs, it is still computationally intractable. To solve this problem, we introduce the semidefinite relaxation (SDR) technique to

approximate the problem effectively. Specifically, consider the following semidefinite programming problem:

$$\begin{aligned} \max_{S} \quad & \mathrm{tr}(\Sigma(\mathbf{R}_*)\mathbf{S}) \\ s.t. \quad & S_{ii} = 1, i \in [n], \\ & \mathbf{S}^T = \mathbf{S} \succeq 0. \end{aligned} \tag{4.12}$$

Given an optimal solution $\mathbf{T}^* \in \Theta$ of Problem (4.11), define $\mathbf{S}^* = \mathbf{T}^*\mathbf{T}^{*T}/p$. It is easy to verify that $\mathbf{S}^*$ is a symmetric positive semi-definite matrix and $S^*_{ii} = 1$ for $i \in [n]$, that is, $\mathbf{S}^*$ satisfies the constraints of Problem (4.12). In addition, $\mathrm{tr}(\Sigma(\mathbf{R}_*)\mathbf{S}^*) = \mathrm{tr}(\mathbf{T}^{*T}\Sigma(\mathbf{R}_*)\mathbf{T}^*/p)$. Therefore, in the sense of achieving a higher objective value, Problem (4.12) is a relaxation of Problem (4.11).

Unless both n and p are very small, it is almost impossible to obtain the optimal solution of Problem (4.11). Fortunately, moderate to large-scale semidefinite programming problems can be solved in polynomial time. For example, the simplest subgradient method can obtain an ϵ-approximate solution of Problem (4.12) after $O(n^2 \log n/\epsilon^2)$ operations (Nesterov, 2007).

Based on the $\pi/2$ theorem in Nesterov (1998), we propose the following random allocation Algorithm 1. This method transforms the optimal solution of Problem (4.12) into a feasible solution of Problem (4.11) by independent sampling from a multivariate Gaussian distribution. The following

Algorithm 1: SDR based Randomized Allocation Algorithm

Input: Sample size n, number of treatments p, and working covariance matrix $\mathbf{R}_*$.

Output: $\mathbf{T}_r$, the random allocation of Problem (4.11).

1 Calculate an optimal solution $\mathbf{S}^*$ of Problem (4.12) based on the covariance structure $\mathbf{R}_*$.

2 **for** $j = 1, \cdots, p$ **do**

3 Independently generate $\mathbf{u}_j \sim N(\mathbf{0}_{n\times 1}, \mathbf{S}^*)$;

4 Set $\mathbf{v}_j = \text{sign}[\mathbf{u}_j]$, where the symbol $\text{sign}[\mathbf{u}_j]$ represents the sign of the elements in $\mathbf{u}_j$.

5 **end**

6 Return the random allocation $\mathbf{T}_r = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p)$.

proposition provides the theoretical support for Algorithm 1.

Proposition 2. *Let* $\mathbf{T}^*$ *and* $\mathbf{T}_r$ *be the optimal solution of Problem* (4.11) *and the output of Algorithm* 1*, respectively. Then, with probability* 1, $\mathbf{T}_r \in \Theta$ *and* $E(\text{tr}(\mathbf{T}_r^T \Sigma(\mathbf{R}_*)\mathbf{T}_r)) \geq (2/\pi)\text{tr}(\mathbf{T}^{*T}\Sigma(\mathbf{R}_*)\mathbf{T}^*)$.

Proposition 2 shows that based on the optimal solution of Problem (4.12), the output $\mathbf{T}_r$ of Algorithm 1 provides a feasible solution for Problem (4.11) with probability 1. In addition, the expected efficiency of $\mathbf{T}_r$ relative

to the optimal solution of Problem (4.11) is at least $2/\pi$ (this is an efficiency lower bound independent of the scale of the original problem!). Bhat et al. (2020) derived a similar conclusion in the context of A/B testing, so the result of Proposition 2 can be regarded as a generalization of the conclusion in Bhat et al. (2020) to the case of p-dimensional treatments.

Next, we investigate the approximation capability of the output $\mathbf{T}_r$ of Algorithm 1 with respect to the ϕ_s-optimal design $\mathbf{T}_s^*$ of Problem (4.10). For simplicity, we define $\mathbf{M}_r = \mathbf{T}_r^T \Sigma(\mathbf{R}_*)\mathbf{T}_r$ and $\mathbf{M}_s^* = \mathbf{T}_s^{*T} \Sigma(\mathbf{R}_*)\mathbf{T}_s^*$. Subsequently, we establish the following lower bound for any ϕ_s-efficiency of $\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_r)$ relative to $\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_s^*)$.

Theorem 5. *For any $0 \le s \le \infty$, the expected relative ϕ_s-efficiency*

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

where $\kappa_s(\mathbf{M}_s^*) = \kappa(\mathbf{M}_s^*)$ if $s < 1$; $\kappa_s(\mathbf{M}_s^*) = \min\{p^{1-1/s}, \kappa(\mathbf{M}_s^*)\}$ if $s \ge 1$, and $e_s \ge 1$ is a constant independent of $\mathbf{T}_r$.

Note that the GLS estimator $\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_s^*)$ is computationally infeasible unless the sample size n and dimension p are very small. Theorem 5 ensures that $\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_r)$ serves as a practical alternative, offering a theoretically guaranteed approximation, that is, the expected ϕ_s-efficiency of $\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_r)$ relative to $\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_s^*)$ is bounded below by a positive constant. This means

that the $\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_r)$'s precision is guaranteed to not degrade arbitrarily compared to the ideal estimator, even in high-dimensional settings. The conditional numbers $\kappa(\mathbf{M}_r)$ and $\kappa(\mathbf{M}_s^*)$ have a negative impact on the lower bound of the efficiency, because as these conditional numbers increase, the GLS estimators $\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_r)$ and $\hat{\boldsymbol{\alpha}}(\mathbf{R}_*, \mathbf{T}_s^*)$ become increasingly unstable.

It is worth noting that the random allocation Algorithm 1 does not depend on a specific ϕ_s optimality criterion. For a given s, the stability and effectiveness of Algorithm 1 can be further improved by integrating the outputs of the M runs of Algorithm 1. Specifically, let

$$\mathbf{T}_{s,M} = \arg \min_{\mathbf{T}_{r,i}, i \in [M]} \operatorname{tr} \left((\mathbf{T}_{r,i}^T \Sigma(\mathbf{R}_*) \mathbf{T}_{r,i})^{-s} / p \right)^{1/s},$$

where $\mathbf{T}_{r,i}$ is the i-th output of Algorithm 1. We call $(\mathbf{T}_{s,M}, \mathbf{R}_*)$ the semidefinite relaxation based matching (SDRM). The total time complexity of the SDRM scheme is $O(n^2 \log n/\epsilon^2 + n^3 + n^2 pM)$, where the first, second, and third terms come from solving the semidefinite programming problem in (4.12), computing the Cholesky decomposition of $\mathbf{S}^*$, and drawing pM independent n-dimensional Gaussian samples, respectively. By using the stochastic proximal point algorithm (Vono et al., 2022) to generate Gaussian distribution samples, the time complexity can be reduced to $O(n^2 \log n/\epsilon^2 + n^2 pM)$. In addition, the time for solving the semidefinite programming problem in (4.12) can be further reduced by using the block

coordinate descent algorithm (Waldspurger et al., 2015).

5. Simulation

In this section, we assume the following relationship between the responses and the covariates:

$$\mathbf{Y} \sim N(\mathbf{T}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}, \mathbf{R}), \tag{5.13}$$

where the treatment effect $\boldsymbol{\alpha} = (2, 3 \times \mathbf{1}_{(p-1)\times 1})^T$, the fixed effect $\boldsymbol{\beta} = -\mathbf{1}_{q\times 1}$, and $N(\boldsymbol{\mu}, \Sigma)$ represents the multivariate normal distribution with location parameter $\boldsymbol{\mu}$ and scale matrix Σ. We set $p = 2$, $q = m = 13$, $k = 3$, $\mathbf{Z} = \mathbf{U}$, and consider the following covariates distributions: $u_{1i} \equiv 1$, $\mathbf{u}_{2i} \overset{i.i.d.}{\sim} \mathrm{MN}(2; 0.9, 0.1)$, $\mathbf{u}_{3i} \overset{i.i.d.}{\sim} \mathrm{MN}(10; 0.1 \times \mathbf{1}^T_{10\times 1})$, $i \in [n]$, where $\mathrm{MN}(n; p_1, \ldots, p_n)$ is a multinomial distribution.

We define $\Sigma_t = \mathbf{U}_t\mathbf{U}_t^T/\mathrm{tr}(\mathbf{U}_t\mathbf{U}_t^{\mathbf{T}})$, where $\mathbf{U}_t = (\mathbf{u}_{t1}, \cdots, \mathbf{u}_{tn})^T$, for $t \in [k]$, $\Sigma_0 = (\rho^{|i-j|})$, if Σ_0 follows a first-order autoregressive (AR) structure, and $\Sigma_0 = (\mathbf{M}^{-1} - \rho\mathbf{W})^{-1}$, if Σ_0 follows a conditional autoregressive (CAR) structure, where $\mathbf{M}$ is a diagonal matrix with diagonal elements $m_{ii} = (1 + \sum_{j\in[n]} w_{ij})^{-1}$, $\mathbf{W}$ is a random adjacency matrix, satisfying $w_{ii} = 0$, $P(w_{ij} = 1) = P(w_{ij} = 0) = 0.5$. In this simulation, we set $\rho \sim U[0.2, 0.8]$

in Σ_0. Let

$$\Omega = \{\mathbf{R} = \sum_{t=0}^{k} \sigma_t^2 \Sigma_t \mid \sigma_0^2 \sim U[10^{-3}, 1], \sigma_1^2, \cdots, \sigma_k^2 \overset{i.i.d.}{\sim} U[0,1]\},$$

and

$$\overline{\Omega} = \{\mathbf{R} + \mathbf{K} \mid \mathbf{R} \in \Omega, \mathbf{K} = v\mathbf{V}\mathbf{V}^T/\mathrm{tr}(\mathbf{V}\mathbf{V}^{\mathbf{T}}), V_{ij} \overset{i.i.d.}{\sim} U[0,1], v \sim U[0,1]\},$$

represent parametric and nonparametric covariance structures, respectively. The true covariance structure $\mathbf{R}$ in (5.13) is randomly sampled from the following cases,

R1. Ω, where Σ_0 follows the first-order AR structure;

R2. $\overline{\Omega}$, where Σ_0 follows the first-order AR structure;

R3. Ω, where Σ_0 follows the CAR structure;

R4. $\overline{\Omega}$, where Σ_0 follows the CAR structure.

In this section, we compare the performance of the GLS estimator $\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T}_0)$ based on different experimental schemes listed in Table 1. In the baseline scheme (BI), each treatment is assigned to each individual with equal probability, and the working covariance matrix $\mathbf{R}_0 = \mathbf{I}_n$. According to Theorem 3, the minimax covariance structure $\mathbf{R}_* = \Sigma_0 + (1/2)\sum_{t=1}^{k} \Sigma_t$. As suggested by Zhang and Kang (2022), we set $\rho = 0.5$ in Σ_0 of $\mathbf{R}_*$. The

schemes of combining the column-balanced design and column-orthogonal design with the minimax covariance structure ($\mathbf{R}_0 = \mathbf{R}_*$) are abbreviated as BM and OM, respectively. The random sampling and SDR matching methods based on the minimax covariance structure ($\mathbf{R}_0 = \mathbf{R}_*$) are abbreviated as RSM and SDRM, respectively. Thus, the true covariance structure $\mathbf{R}$ is misspecified by $\mathbf{R}_0$ in the BI, BM, OM, RSM, and SDRM schemes. We denote the SDR matching method based on the true covariance structure ($\mathbf{R}_0 = \mathbf{R}$) as ORACLE, which corresponds to the case of the covariance structure is correctly specified. In this simulation, we set $s = 1$, and use $M = 10^4$ independent samplings for the RSM, SDRM, and ORACLE schemes.

For each experimental scheme $(\mathbf{R}_0, \mathbf{T}_0)$, we calculate the following three metrics:

(i). Regret: $\|\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R_0}, \mathbf{T}_0)) - \mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T}))\| / \|\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T}))\|$;

(ii). Relative ϕ_0-efficiency: $\det(\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T}_0))) / \det(\mathrm{cov}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})))$;

(iii). Mean Squared Error-efficiency (MSE-efficiency): $\mathrm{MSE}(\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})) / \mathrm{MSE}(\hat{\boldsymbol{\alpha}}(\mathbf{R}_0, \mathbf{T}_0))$,

where $\hat{\boldsymbol{\alpha}}(\mathbf{R}, \mathbf{T})$ is the GLS estimator based on the ORACLE scheme, and $\det(\cdot)$ is the determinant operator. The summary table of various 128-run schemes evaluated across $L = 100$ randomly generated true covariance

Table 1: Designs and covariance structures of experimental schemes.

Method	Design	Covariance Structure
BI	Column-balanced	$\mathbf{R}_0 = \mathbf{I}_n$
BM	Column-balanced	$\mathbf{R}_0 = \mathbf{R}_*$
OM	Column-orthogonal	$\mathbf{R}_0 = \mathbf{R}_*$
RSM	Random Sampling Matching	$\mathbf{R}_0 = \mathbf{R}_*$
SDRM	SDR Matching	$\mathbf{R}_0 = \mathbf{R}_*$
ORACLE	SDR Matching	$\mathbf{R}_0 = \mathbf{R}$

structures is displayed in Table 2. From these results, we can draw the following conclusions: when the random error follows an AR structure (R1 and R2), the performance of the BM scheme is significantly better than that of the BI scheme. The OM scheme further reduces the uncertainty in the estimation of the treatment effect by combining the column-orthogonal design, which is consistent with the conclusion in Theorem 3. In addition, the RSM scheme utilizes the structure of the ϕ_s-optimal design, and its performance is better than that of the OM scheme. Across all true covariance structures, the SDRM scheme consistently exhibits the best performance among the five experimental schemes.

Table 2: Summary table across various 128-run experimental schemes.

	BI	BM	OM	RSM	SDRM
			Average regret (standard deviation)		
R1	2.75(1.88)	1.11(0.30)	0.80(0.19)	0.65(0.15)	0.03(0.02)
R2	3.20(5.07)	2.66(6.98)	1.80(4.40)	2.00(6.02)	1.25(5.15)
R3	0.21(0.08)	0.21(0.08)	0.15(0.05)	0.07(0.03)	0.03(0.02)
R4	1.77(10.73)	1.78(10.83)	1.58(9.37)	1.45(9.61)	1.32(7.74)
			Average ϕ_0-efficiency (standard deviation)		
R1	0.16(0.12)	0.28(0.07)	0.45(0.10)	0.42(0.06)	1.00(0.01)
R2	0.15(0.11)	0.20(0.10)	0.31(0.14)	0.28(0.13)	0.60(0.28)
R3	0.80(0.05)	0.80(0.05)	0.82(0.04)	0.92(0.02)	1.00(0.00)
R4	0.53(0.21)	0.53(0.21)	0.55(0.21)	0.61(0.23)	0.66(0.25)
			Average MSE-efficiency (standard deviation)		
R1	0.51(0.06)	0.61(0.07)	0.71(0.09)	0.72(0.07)	1.00(0.06)
R2	0.50(0.06)	0.56(0.06)	0.67(0.08)	0.65(0.07)	0.91(0.10)
R3	0.88(0.08)	0.88(0.08)	0.89(0.07)	0.96(0.08)	1.00(0.07)
R4	0.80(0.07)	0.80(0.07)	0.80(0.06)	0.86(0.07)	0.90(0.07)

It is worth noting that when the true covariance structure belongs to the parametric class Ω (R1 and R3), the performance of the SDRM scheme is almost the same as that of the ORACLE. This is consistent with the conclusion in Theorem 4. When the random error follows a CAR structure (R3 and R4), the performance of the BM scheme is almost the same as that of the BI scheme, and the performance of other experimental schemes is similar to that when the error follows an AR structure. The difference between the BM and BI schemes under different random error assumptions is due to the robustness of the minimax covariance structure against the misspecification of the true covariance structure (Theorem 4), and the AR structure is essentially different from the independent and identically distributed error structure, that is, $\Sigma_0 \notin [\Omega_{\mathbf{I}_n}]$. Figure S3 in the Supplementary Materials shows that the average Frobenius norm of $\mathbf{I}_n - (\mathbf{M} - \rho\mathbf{W})^{-1}$ for different values of ρ and n is less than 10^{-3}. Therefore, the CAR structure is close to the independent and identically distributed structure, that is, the performance of the BM scheme is close to that of the BI scheme. It is worth mentioning that under the general covariance structure, the conclusion should be similar to that under the AR structure. Furthermore, as demonstrated in Table S1 in the Supplementary Materials, when covariates distributions exhibit greater complexity, the advantages of the proposed

approaches become more pronounced.

6. Case Study

This case study is based on the Amazon Mechanical Turk dataset from the UCI Machine Learning Repository. After removing missing values, the dataset contains 9843 records and 23 attributes, including 21 attributes describing driving scenarios such as destination, current time, weather conditions, and whether there are passengers, as well as two treatment variables: coupon type and coupon validity period. Coupon types include bars, takeaway food restaurants, coffee houses, cheap restaurants, and expensive restaurants. The coupon validity periods include 2 hours and 24 hours. For more information on this dataset, refer to Wang et al. (2017).

Our main goal is to assess the impact of coupon distribution on the consumption intention of drivers. We encode the q-level qualitative variables into q binary dummy variables. Therefore, there are 64 coupon distribution strategies composed of 6 two-level treatment variables. For example, a feasible strategy is to distribute bar and cafe coupons with a validity period of 2 hours. Note that in addition to coupon distribution strategies, different driving scenarios also affect drivers' consumption intentions. For example, if the coupon's usage location is in the same direction as the driver's desti-

nation, they are more likely to consume. Therefore, we use the LMM (5.13) to model the relationship between drivers' consumption intentions, coupon distribution strategies, and driving scenarios.

Different experimental schemes in Table 1 are compared based on a synthetic dataset, where drivers' covariates are randomly sampled from the Amazon Mechanical Turk dataset, and responses are generated by the LMM (5.13) considering fixed effects for the first 5 covariates, with the true covariance structure belonging to Ω. To ensure the reliability of comparison results, we require that the sample and population have the same number of categories for qualitative variables to maintain similarity, which is easily satisfied through repeated sampling. Three covariates describing the distance between the driver's current location and the coupon location are treated as three-dimensional covariates corresponding to the same random effect, resulting in a total of $k = 19$ groups of covariates. The true values of treatment effects and fixed effects are set as $\boldsymbol{\alpha} = (2, 3, 3, 3, 3, 3)^T$ and $\boldsymbol{\beta} = -\mathbf{1}_{5\times 1}$, respectively.

We compare the average MSE (AMSE) of estimating treatment effects for different experimental schemes under 100 randomly generated covariance structures from Ω. Under each true covariance structure, the MSE is calculated based on 100 independent response datasets.

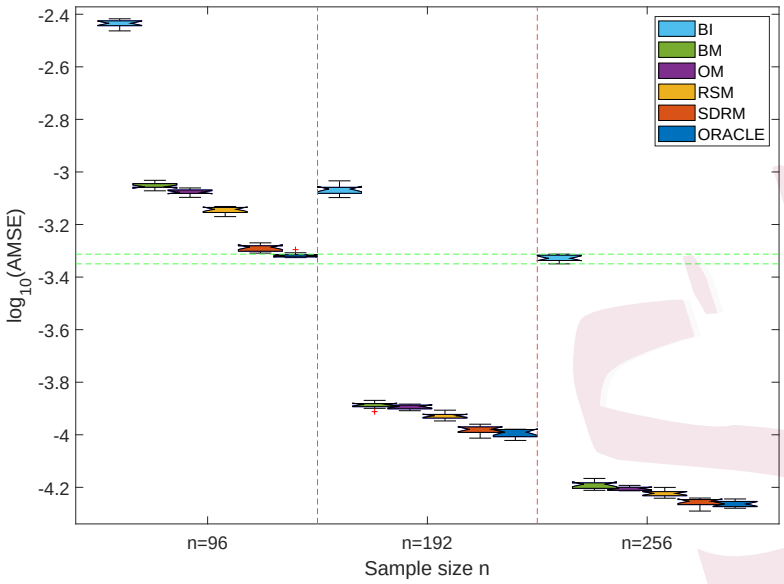


Figure 2: AMSEs of various experimental schemes in the case study when random errors is the CAR structure.

Figure 2 shows the AMSEs of various schemes in 100 repetitions when random errors follow the CAR structure. We can draw the following conclusions: at each sample size, the experimental schemes based on the minimax covariance structure (BM, OM, RSM, and SDRM) achieve smaller AMSEs than the BI scheme. The OM scheme further reduces the estimation error of the BI scheme by minimizing the upper bound of the regret. On this basis, our last two schemes (RSM and SDRM) further improve estimation accuracy by optimally matching the design with the minimax covariance

structure. Among these five experimental designs, the SDRM scheme performs the best. Notably, the 96-run SDRM scheme performs almost as well as the 256-run BI scheme, significantly demonstrating the advantages of the SDRM scheme under the LMM. Additional simulation results for other network structures are summarized in Figures S4 and S5 in the Supplementary Materials. From these results, we conclude that under the AR structure, the boxplots of various experimental schemes are wider because all subjects are connected. Other conclusions are consistent with those under the CAR structure. These results indicate that our experimental schemes can be applied to controlled experiments with complex covariates and network structures.

7. Discussion

This paper focuses on the impact of covariates and network structures of subjects on treatment effect estimation, aiming to improve estimation accuracy through optimizing experimental designs. First, the study employs a linear mixed effects model framework to effectively address the uncertainty of covariates and network dependencies. Based on this, minimax robust schemes and optimal matching schemes are proposed. Simulation experiments and real data analysis show that the SDRM scheme proposed in this

paper exhibits significant advantages under the LMM.

Future researches include two important directions: First, exploring optimal designs for heterogeneous treatment effects, such as analyzing the differential impacts of treatments across different age and income groups. Second, conducting in-depth research on modeling and analyzing interference effects of treatment allocation on related subjects, such as scenarios where online game duration is influenced by both activity attractiveness and social relationships. Such problems require integrating network interference models with optimization design methods, and relevant research can build on the theoretical framework of Chen et al. (2023) to provide new approaches for causal inference in complex social systems.

Supplementary Materials

The Supplementary Materials include two applications of the proposed robust experimental schemes: A/B testing and sequential experiments, supplementary simulation results, and proofs for all the theoretical results.

Acknowledgements

The authors thank the Editor, the Associate Editor, and the two reviewers for their valuable comments and suggestions. This work was supported by

the National Natural Science Foundation of China (12131001), the Fundamental Research Funds for Central Universities, LPMC, and KLMDASR.

References

Atkinson, A., A. Donev, and R. Tobias (2007). *Optimum Experimental Designs, with SAS*, Volume 34. OUP Oxford.

Bai, Y., J. Liu, and M. Tabord-Meehan (2024). Inference for matched tuples and fully blocked factorial designs. *Quantitative Economics 15*(2), 279–330.

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

Branson, Z., T. Dasgupta, and D. B. Rubin (2016). Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education High School Study. *The Annals of Applied Statistics 10*(4), 1958 – 1976.

Chen, Q., B. Li, L. Deng, and Y. Wang (2023). Optimized covariance design for ab test on social network under interference. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pp. 37448–37471.

Haizler, T. and D. M. Steinberg (2021). Factorial designs for online experiments. *Technometrics 63*(1), 1–12.

Harville, D. (1976). Extension of the gauss-markov theorem to include the estimation of random effects. *The Annals of Statistics 4*(2), 384–395.

Jiang, J., D. Legrand, R. Severn, and R. Miikkulainen (2020). A comparison of the taguchi method and evolutionary optimization in multivariate testing. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–6. IEEE.

Kohavi, R., R. Longbotham, D. Sommerfield, and R. M. Henne (2009). Controlled experiments on the web: survey and practical guide. *Data Mining and Knowledge Discovery 18*, 140–181.

Kohavi, R., D. Tang, and Y. Xu (2020). *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press.

Larsen, N., J. Stallrich, S. Sengupta, A. Deng, R. Kohavi, and N. T. Stevens (2024). Statistical challenges in online controlled experiments: A review of a/b testing methodology. *The American Statistician 78*(2), 135–149.

Li, X. and P. Ding (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(1), 241–268.

Liu, H., J. Ren, and Y. Yang (2024). Randomization-based joint central limit theorem and efficient covariate adjustment in randomized block 2^k factorial experiments. *Journal of the American Statistical Association 119*(545), 136–150.

Nesterov, Y. (1998). Semidefinite relaxation and nonconvex quadratic optimization. *Optimization Methods and Software 9*(1-3), 141–160.

Nesterov, Y. (2007). Smoothing technique and its applications in semidefinite optimization. *Mathematical Programming 110*(2), 245–259.

REFERENCES

Pashley, N. E. and M.-A. C. Bind (2023). Causal inference for multiple treatments using fractional factorial designs. *Canadian Journal of Statistics 51*(2), 444–468.

Pokhilko, V., Q. Zhang, L. Kang, et al. (2019). D-optimal design for network A/B testing. *Journal of Statistical Theory and Practice 13*(4), 1–23.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association 100*(469), 322–331.

Sadeghi, S., P. Chien, and N. Arora (2020). Sliced designs for multi-platform online experiments. *Technometrics 62*(3), 387–402.

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

Verbeke, G., G. Molenberghs, and G. Verbeke (1997). *Linear Mixed Models for Longitudinal Data*. Springer.

Vono, M., N. Dobigeon, and P. Chainais (2022). High-dimensional gaussian sampling: a review and a unifying approach based on a stochastic proximal point algorithm. *SIAM Review 64*(1), 3–56.

Waldspurger, I., A. d'Aspremont, and S. Mallat (2015). Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming 149*, 47–81.

Wang, T., C. Rudin, F. Doshi-Velez, Y. Liu, E. Klampfl, and P. MacNeille (2017). A bayesian framework for learning rule sets for interpretable classification. *The Journal of Machine*

Learning Research 18(1), 2357–2393.

Wiens, D. P. (2015). Robustness of Design. In A. M. Dean, M. Morris, J. Stufken, and D. Bingham (Eds.), *Handbook of Design and Analysis of Experiments*, pp. 719–753. CRC Press Taylor & Francis Group.

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.

Zhao, A. and P. Ding (2022). Regression-based causal inference with factorial experiments: estimands, model specifications and design-based properties. *Biometrika 109*(3), 799–815.

Zhao, A. and P. Ding (2023). Covariate adjustment in multiarmed, possibly factorial experiments. *Journal of the Royal Statistical Society Series B: Statistical Methodology 85*(1), 1–23.

NITFID, School of Statistics and Data Science, Nankai University, Tianjin 300071, China

E-mail: shxu@mail.nankai.edu.cn

NITFID, School of Statistics and Data Science, Nankai University, Tianjin 300071, China

E-mail: ydzhou@nankai.edu.cn