# Covariate balancing for high-dimensional samples in controlled experiments

Xi Luo[a], Penggao Yan[b*], Ran Yan[c*], Shuaian Wang[a]

[a] *Department of Logistics and Maritime Studies, Faculty of Business, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR*

[b] *Department of Aeronautical and Aviation Engineering, Faculty of Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR*

[c] *School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore*

* *Corresponding authors, E-mails: peng-gao.yan@connect.polyu.hk; ran.yan@ntu.edu.sg*

## Abstract

In controlled experiments, achieving covariate balancing across all groups is crucial as it ensures that the estimated treatment effects are not confounded by the effects of covariates. This study proposes a mixed-integer nonlinear programming model to address the covariate balancing problem. Specifically, we introduce a new covariate imbalance measure, which is the maximum discrepancy in both the first and second central moments between any two groups. The second central moment can effectively capture the correlation of covariates in a physical sense, which is crucial for partitioning high-dimensional samples. A mixed-integer nonlinear programming model is constructed to minimize the proposed measure to obtain the optimal partitioning results. The nonlinear model is then linearized to accelerate the optimization process. We conduct computational experiments based on simulated datasets, including one-dimensional, two-dimensional, and three-dimensional Gaussian distributed samples, and a real clinic trial dataset. Compared to the conventional discrepancy-based method, our method achieves a 54.81% and a 40.6% reduction in the maximum discrepancy of partitioning results in the two-dimensional simulated Gaussian samples and the real clinic trial dataset, respectively. These results demonstrate the superiority of the proposed model in partitioning high-dimensional samples with correlated covariates compared with the conventional discrepancy-based method.

## Keywords

Experiment design; Covariate balance; Controlled experiment; Partitioning problem; High-dimensional samples

## 1. Introduction

Controlled experiments are essential in scientific research to identify cause-and-effect relationships (Pocock et al., 2013; Wei et al., 2024). In these experiments, researchers manipulate one or more independent variables while keeping other factors constant to observe the effects on the dependent variable(s). The dependent variables are the responses that are measured to assess the impact of the independent variables, which can be designed by the researchers. Researchers need to partition experimental units into different groups, one of which is usually chosen as the control group in the controlled experiment. This group acts as the baseline, while the other groups serve as the experimental groups, each with a slightly different treatment (Ferrier and Valdmanis, 2004; Mergoni and De Witte, 2022). To ensure the validity of the experimental findings, researchers need to achieve covariate balancing across all groups to ensure that the estimated treatment effects are not confounded by the effects of covariates (Tam et al., 2018; Kwon et al., 2019; Ben-Michael et al., 2021).

In practice, researchers often use a completely randomized design that partitions experimental units randomly into different groups (Rosenberger, 2015). When the number of experimental units is sufficiently large, randomization tends to achieve covariate balancing among groups (Moulton, 2004; Deaton and Cartwright, 2018; Li et al., 2021). However, experimenters usually conduct an experiment only once, or a few times at most, each using a single instance of the complete randomization. This may lead to differences between the groups in each iteration, possibly failing to achieve the necessary balance of covariates (Turner et al., 2020; Li et al., 2021; Harshaw et al., 2024). In particular, when the number of experimental units is small or moderate, the covariate imbalance between groups might be more acute (Bertsimas et al., 2015; Li et al., 2021). This imbalance can either mask or amplify the

true effect of the treatment, thus compromising the accuracy and reliability of the experiment.

Recently, operations research approaches have been proposed to achieve covariates balancing among groups by optimizing a proper balance measure (Nikolaev et al., 2013; Bertsimas et al., 2015, 2019; Kwon et al., 2019; Bhat et al., 2020; Josey et al., 2021). Among these methods, Bertsimas et al. (2015)' s approach has gained popularity and has a wide impact on the covariate balancing problem (Vazquez and Wong, 2024; Harshaw et al., 2024). Specifically, Bertsimas et al. (2015) proposed to minimize the maximum discrepancy in both the first and second raw moments between any two groups with the standardized samples. A mixed-integer linear programming model is proposed in their work. However, an underlying assumption has been made in these approaches that the pre-standardization process can ensure a zero sample mean in each group, potentially hindering the partitioning performance in situations with multiple covariates.

In this study, we propose to improve the method of Bertsimas et al. (2015). We chose the method of Bertsimas et al. (2015) over other optimization methods because it is model-free and provides theoretical foundations for other methods, such as Bertsimas et al. (2019). Our proposed modification on Bertsimas et al. (2015) can also be easily extended to other methods. Specifically, we propose a mixed-integer nonlinear programming model that aims to minimize the maximum discrepancy in both the first and second center moments between any two groups. Note that we use the second central moment[1] (Rice, 2007) instead of the second raw moment in calculating the second moment of group samples, which relaxes the assumption that each group has a zero mean. In addition, the second central moment reflects the correlation of covariates in a physical sense, which is crucial for partitioning high-

[1] $\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$ is the $k$th central moment, in which $\mu$ refers to the mean value.

dimensional samples where the correlation of covariates is not a trivial problem. Furthermore, we linearize the original model to improve computational efficiency.

In simulated experiments with normally distributed data, we illustrate that the proposed model can achieve better partitioning performance than the optimization model proposed by Bertsimas et al. (2015). Furthermore, we compare the partitioning performance of the two methods using a real clinical trial with 18 subjects and three continuous covariates. Experimental results show that the strength of the proposed model is further enhanced when the number of covariates increases and the correlation between them is stronger. The contributions of this study are twofold:

First, we introduce a novel mixed-integer linear programming model to minimize the maximum discrepancy in both the first and second central moments between any two groups. Unlike Bertsimas et al. (2015), we use the second central moment instead of the second raw moment to calculate the discrepancy in the second moment between groups. The specific advantage of our model lies in the fact that the off-diagonal elements of the second central moment can reflect the correlation of covariates in a physical sense. This makes the calculation of discrepancy between groups more accurate.

Then, we experimentally demonstrate the effect of covariate correlations on partitioning high-dimensional simulated samples and real clinic trial samples. By comparing the partitioning performance of the proposed method and the method of Bertsimas et al. (2015) on 2-D and 3-D Gaussian samples with different correlation coefficient parameters, we observe that the proposed method exhibits lower group discrepancy in the partitioning results. Furthermore, the lower discrepancy in the partitioning results derived by the proposed method is observed when the objective

function pays more attention to the discrepancy in the second moment. As the primary difference between the two methods is that only the proposed method implicitly considers the covariate correlations in calculating the second moment discrepancy, our results suggest that covariate correlations could be a critical factor in balanced partitioning for high-dimensional samples. In addition, the assignment of larger weight to the second moment discrepancy in the objective function also makes the model of Bertsimas et al., (2015) more inaccurate in calculating discrepancy between groups, resulting in higher group discrepancy in the partitioning results obtained by the method of Bertsimas et al., (2015).

## 2. Literature review

To mitigate the impact of covariate imbalance on experimental results, either post-hoc balancing or a priori balancing methods can be used (Kallus, 2018). Post-hoc balancing is designed for observational data. It involves using a completely randomized design to assign treatments to experimental units before data is collected, followed by the use of adjustment methods during the data analysis stage. Common methods include post-stratification (McHugh and Matts, 1983; Miratrix et al., 2013), analysis of covariance (Wang et al., 2019), and propensity score matching (Rosenbaum and Rubin, 1983; Kane et al., 2020; Arbona et al., 2023). In contrast, a priori balancing involves considering the balance of covariates during the experimental design stage by carefully planning the design to ensure a similar distribution of covariates among the groups. This approach aims to prevent potential biases rather than correcting them during the data analysis stage (Li et al., 2021). Classic methods include randomized block design (Fisher, 1936), pairwise-matched allocation (Greevy et al., 2004), and rerandomization (Rubin, 2012, 2015).

Recently, researchers have proposed operations research approaches to balance the covariates among groups by optimizing a proper balance measure (Nikolaev et al., 2013; Bertsimas et al., 2015, 2019; Kwon et al., 2019; Bhat et al., 2020; Josey et al., 2021). Nikolaev et al. (2013) proposed the balance optimization subset selection (BOSS) method by minimizing the difference between the bin-represented distributions of two groups. However, this bin-based method results in a biased estimate of the distribution and does not guarantee an optimal solution (Kwon et al., 2019). To solve this problem, Tam (2018) proposed to use the actual covariate values instead of binned values for distribution calculations in the BOSS framework. Instead of directly minimizing the difference between distributions, Bertsimas et al. (2015) proposed to minimize the maximum discrepancy in both the first and second raw moments between any two groups with the standardized samples. A mixed-integer linear programming model is proposed in their work. Bertsimas et al. (2019) further improved this model by considering the uncertainty in covariates through robust optimization. Vazquez and Wong (2024) verified the superiority of Bertsimas et al. (2019) in balancing covariate distributions by comparing it with other optimization methods, including Pocock and Simon (1975), Nishi and Takaichi (2004), and Ma and Hu (2013).

Bertsimas et al. (2015) minimized the maximum discrepancy in both the first and second raw moments[2] (Papoulis, 1984) between any two groups with the standardized samples. If the sample mean is zero in each group, the second raw moment represents the sample covariance within the group. However, Bertsimas et al. (2015) conducted the standardization process based on all samples, only ensuring that the mean of all samples is zero. The mean of each group may not be zero. Therefore, the

[2] $\frac{1}{n}\sum_{i=1}^{n} x_i^k$ is the $k$th raw moment, where $x_1, \dots, x_n$ are independent and identically distributed realization (samples) from a random variable $X$.

second raw moments cannot characterize the sample covariance and thus cannot represent the correlations among covariates in each group. As the experimental section shows, such a zero-sample-mean assumption can considerably reduce the computational load but with the compromise of partitioning performance. Partitioning performance deteriorates further as the dimension of the samples and the correlation between the covariates increase. However, in many applications, the experimental units have multiple covariates that can influence the response measurement along with the treatment (Festing and Altman, 2002; Schneeweiss et al., 2009; Ning et al., 2020; Hochbaum et al., 2022). Therefore, a countermeasure is needed to relax such an assumption when dealing with problems with multiple covariates, which is the focus of this paper.

## 3. Optimization approach

Given pre-treatment values of samples $\mathbf{w}_i,\ i = 1, \dots, n$, each sample is an $r$-dimensional vector. We use $w_{is},\ s = 1, \dots, r$ to denote the $s$th covariate of sample $i$, and then $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{ir}]$. We assume that these samples are the realizations of a random variable $\mathbf{w}$, which is subject to a certain distribution. Our proposal is to create $m$ groups each containing $k$ samples (assume that $n = mk$ for simplification) such that the discrepancy in the first moment and $\rho$ times the discrepancy in the second moment are minimized between any two groups. The parameter $\rho$ controls the trade-off between the discrepancy in the first and second moments and is chosen by the researcher.

### *3.1 Data normalization*

We first normalize all samples to have a zero mean and identity variance. The original mean and covariance of the samples are given by $\widehat{\boldsymbol{\mu}} = \frac{1}{k}\sum_{i=1}^{n} \mathbf{w}_i$ and $\widehat{\boldsymbol{\Sigma}} = (\sum_{i=1}^{n}(\mathbf{w}_i - \widehat{\boldsymbol{\mu}})(\mathbf{w}_i - \widehat{\boldsymbol{\mu}})^T)/n$. The normalization is achieved by the following transformation (Kessy et al., 2018):

$$\mathbf{w}_i' = \boldsymbol{\Gamma}(\mathbf{w}_i - \hat{\boldsymbol{\mu}}), i = 1,2,\ldots,n, \tag{1}$$

where $\mathbf{w}_i'$ is the $i$th normalized sample and $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^{\mathrm{T}}$. $\mathbf{U}$ and $\boldsymbol{\Lambda}$ are obtained from the principal component analysis, i.e.,

$$\hat{\boldsymbol{\Sigma}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}. \tag{2}$$

Specifically, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r]$ is the eigenmatrix, where $\mathbf{u}_j$ $(j = 1,2,\ldots,r)$ is the unit eigenvector of $\hat{\boldsymbol{\Sigma}}$. $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$, where $\lambda_j$ $(j = 1,2,\ldots,r)$ is the eigenvalue with respect to the eigne-vector $\mathbf{u}_j$. By assuming that $\hat{\boldsymbol{\Sigma}}$ is a positive definite matrix, all eigenvalues of $\hat{\boldsymbol{\Sigma}}$ are positive real numbers and we have $\boldsymbol{\Lambda}^{-1/2} = \mathrm{diag}\left(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \ldots, \frac{1}{\sqrt{\lambda_r}}\right)$. The procedures for proving that normalized samples have a zero mean and identity covariance are shown in Appendix A. In the following sections, the establishment of the optimization model is based on the normalized data. We then introduce the binary decision variable $x_{ip}$, where $x_{ip} = 1$ means that the $i$th normalized sample $\mathbf{w}_i'$ is assigned to group $p$, and $x_{ip} = 0$ otherwise.

### *3.2 Discrepancy in the first and second moments*

Given two groups $p$ and $q$, the discrepancy in the first moment between groups $p$ and $q$ in the $s$th covariate is represented as follows:

$$\frac{1}{k}\sum_{i=1}^{n}(w_{is}' x_{ip} - w_{is}' x_{iq}), \forall s = 1, \ldots, r. \tag{3}$$

To derive the discrepancy in the second moment between groups $p$ and $q$, we first need to calculate the covariance matrix of the two groups. The procedures for calculating the covariance matrix of group $p$ are as follows:

(i) Calculate the mean of group $p$ by

$$\boldsymbol{\mu}_p = \left[\frac{1}{k}\sum_{i=1}^{n} w_{i1}' x_{ip}, \frac{1}{k}\sum_{i=1}^{n} w_{i2}' x_{ip}, \ldots, \frac{1}{k}\sum_{i=1}^{n} w_{ir}' x_{ip}\right] = [\mu_{p1}, \mu_{p2}, \ldots, \mu_{pr}]. \tag{4}$$

(ii) Calculate the covariance matrix of group $p$. All samples in group $p$ can be represented as

$$\mathbf{W}_p = \begin{pmatrix} \mathbf{w}'_{p:1} \\ \mathbf{w}'_{p:2} \\ \vdots \\ \mathbf{w}'_{p:k} \end{pmatrix} = \begin{pmatrix} w'_{p:11} & w'_{p:12} & \cdots & w'_{p:1r} \\ w'_{p:21} & w'_{p:22} & \cdots & w'_{p:2r} \\ \vdots & \vdots & \ddots & \vdots \\ w'_{p:k1} & w'_{p:k2} & \cdots & w'_{p:kr} \end{pmatrix}, \tag{5}$$

where $\mathbf{w}'_{p:i}$ represents the $i$th normalized sample in group $p$ and $w'_{p:is}$ denotes the $s$th covariate of the $i$th normalized sample in the group $p$. Next, we subtract the mean vector $\boldsymbol{\mu}_p$ from each sample in the group $p$ to obtain $\mathbf{W}'_p$, as shown in the following formula:

$$\mathbf{W}'_p = \begin{pmatrix} w'_{p:11} - \mu_{p1} & w'_{p:12} - \mu_{p2} & \cdots & w'_{p:1r} - \mu_{pr} \\ w'_{p:21} - \mu_{p1} & w'_{p:22} - \mu_{p2} & \cdots & w'_{p:2r} - \mu_{pr} \\ \vdots & \vdots & \ddots & \vdots \\ w'_{p:k1} - \mu_{p1} & w'_{p:k2} - \mu_{p2} & \cdots & w'_{p:kr} - \mu_{pr} \end{pmatrix}. \tag{6}$$

The covariance matrix of group $p$ can be represented as

$$\begin{aligned} \boldsymbol{\Sigma}'_p &= \frac{1}{k}\left(\mathbf{W}'_p\right)^{\mathrm{T}} \mathbf{W}'_p \\ &= \frac{1}{k}\begin{pmatrix} w'_{p:11} - \mu_{p1} & \cdots & w'_{p:k1} - \mu_{p1} \\ \vdots & \ddots & \vdots \\ w'_{p:1r} - \mu_{pr} & \cdots & w'_{p:kr} - \mu_{pr} \end{pmatrix}\begin{pmatrix} w'_{p:11} - \mu_{p1} & \cdots & w'_{p:1r} - \mu_{pr} \\ \vdots & \ddots & \vdots \\ w'_{p:k1} - \mu_{p1} & \cdots & w'_{p:kr} - \mu_{pr} \end{pmatrix} \\ &= \frac{1}{k}\begin{pmatrix} \sum_{i=1}^{k}\left(w'_{p:i1} - \mu_{p1}\right)^2 & \cdots & \sum_{i=1}^{k}\left(w'_{p:i1} - \mu_{p1}\right)\left(w'_{p:ir} - \mu_{pr}\right) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{k}\left(w'_{p:ir} - \mu_{pr}\right)\left(w'_{p:i1} - \mu_{p1}\right) & \cdots & \sum_{i=1}^{k}\left(w'_{p:ir} - \mu_{pr}\right)^2 \end{pmatrix} \end{aligned} \tag{7}$$

Next, we introduce $x_{ip}$ to reformulate Eq. (7), which can be represented as

$$\boldsymbol{\Sigma}'_p = \frac{1}{k}\begin{pmatrix} \sum_{i=1}^{n}\left(w'_{i1} - \mu_{p1}\right)^2 x_{ip} & \cdots & \sum_{i=1}^{n}\left(w'_{i1} - \mu_{p1}\right)\left(w'_{ir} - \mu_{pr}\right)x_{ip} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n}\left(w'_{ir} - \mu_{pr}\right)\left(w'_{i1} - \mu_{p1}\right)x_{ip} & \cdots & \sum_{i=1}^{n}\left(w'_{ir} - \mu_{pr}\right)^2 x_{ip} \end{pmatrix}. \tag{8}$$

Similarly, we can obtain the covariance matrix of group $q$, as shown in Eq. (9):

$$\boldsymbol{\Sigma}'_q = \frac{1}{k}\begin{pmatrix} \sum_{i=1}^{n}\left(w'_{i1} - \mu_{q1}\right)^2 x_{iq} & \cdots & \sum_{i=1}^{n}\left(w'_{i1} - \mu_{q1}\right)\left(w'_{ir} - \mu_{qr}\right)x_{iq} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n}\left(w'_{ir} - \mu_{qr}\right)\left(w'_{i1} - \mu_{q1}\right)x_{iq} & \cdots & \sum_{i=1}^{n}\left(w'_{ir} - \mu_{qr}\right)^2 x_{iq} \end{pmatrix}. \tag{9}$$

(iii) Perform element-wise subtraction on $\mathbf{\Sigma}'_p$ and $\mathbf{\Sigma}'_q$. The element of $\mathbf{\Sigma}'_p - \mathbf{\Sigma}'_q$ at location $(s, s')$ can be represented as

$$\frac{1}{k}\sum_{i=1}^{n}\left[\left(w'_{is} - \mu_{ps}\right)\left(w'_{is'} - \mu_{ps'}\right)x_{ip} - \left(w'_{is} - \mu_{qs}\right)\left(w'_{is'} - \mu_{qs'}\right)x_{iq}\right], \tag{10}$$

which is the discrepancy in the second moment between group $p$ and group $q$ at location $(s, s')$.

### *3.3 Mixed-integer linear programming for partitioning problem*

After deriving the discrepancy in the first and second moments between groups $p$ and $q$, the partitioning problem involving multi-dimensional samples can be formulated as a mixed-integer nonlinear programming model using $m(1 + 2n - m)/2$ binary decision variables and $m(m - 1)r(r + 3)/4$ continuous decision variables, as follows:

[M1]

$$\min_{x_{ip}} \max_{p \neq q} \left(\sum_{s=1}^{r} M_{pqs} + \rho \sum_{s=1}^{r} V_{pqss} + 2\rho \sum_{s=1}^{r-1}\sum_{s'=s+1}^{r} V_{pqss'}\right), \tag{11}$$

s.t.

$$M_{pqs} \geq \left|\frac{1}{k}\sum_{i=1}^{n} w'_{is}\left(x_{ip} - x_{iq}\right)\right| \ \forall p = 1, \dots, m-1, q = p+1, \dots, m, s = 1, \dots, r, \tag{12}$$

$$V_{pqss'} \geq \left|\frac{1}{k}\sum_{i=1}^{n}\left[\left(w'_{is} - \mu_{ps}\right)\left(w'_{is'} - \mu_{ps'}\right)x_{ip} - \left(w'_{is} - \mu_{qs}\right)\left(w'_{is'} - \mu_{qs'}\right)x_{iq}\right]\right| \forall p = 1, \dots, m-1, q = p+1, \dots, m, s = 1, \dots, r, s' = s, \dots, r, \tag{13}$$

$$\sum_{i=1}^{n} x_{ip} = k \ \forall p = 1, \dots, m, \tag{14}$$

$$\sum_{p=1}^{m} x_{ip} = 1 \ \forall i = 1, \dots, n, \tag{15}$$

$$x_{ip} = 0 \ \forall p = 2, \dots, m, \forall i = 1, \dots, p-1, \tag{16}$$

$$x_{ip} \in \{0,1\} \ \forall i = 1, \dots, n, \forall p = 1, \dots, m. \tag{17}$$

The objective function (11) minimizes the maximum discrepancy in the first and second moments between any two groups. The first term is the sum of the discrepancy in the first moment over all covariates between groups $p$ and $q$, the second term is the sum of the diagonal elements of $\mathbf{\Sigma}'_p - \mathbf{\Sigma}'_q$,

and the third term is the sum of the off-diagonal elements of $\mathbf{\Sigma}'_p - \mathbf{\Sigma}'_q$. The combination of the second and third terms is the sum of the discrepancy in the second moment over all covariates between groups $p$ and $q$. Constraints (12) require that the discrepancy in the first moment between groups $p$ and $q$ in the objective function be the largest among all group pairs, while constraints (13) require the same for the discrepancy in the second moment. Constraints (14) ensure that each group contains $k$ samples. Constraints (15) guarantee that every sample can only belong to one group. Constraints (16) reduce the redundancy in the branch-and-bound tree due to permutation symmetry, as explained by Bertsimas et al. (2015). Moreover, Eq. (12) can be written as

$$M_{pqs} \geq \frac{1}{k}\sum_{i=1}^{n} w'_{is}\left(x_{ip} - x_{iq}\right), \forall p = 1, \dots, m-1, q = p+1, \dots, m, s = 1, \dots, r \tag{18}$$

$$M_{pqs} \geq \frac{1}{k}\sum_{i=1}^{n} w'_{is}\left(x_{iq} - x_{ip}\right), \forall p = 1, \dots, m-1, q = p+1, \dots, m, s = 1, \dots, r, \tag{19}$$

and Eq. (13) can be written as

$$\begin{aligned} V_{pqss'} \geq\ & \frac{1}{k}\sum_{i=1}^{n}\left[w'_{is}w'_{is'} + \mu_{ps}\mu_{ps'} - \mu_{ps}w'_{is'} - \mu_{ps'}w'_{is}\right]x_{ip} \\ & -\frac{1}{k}\sum_{i=1}^{n}\left[w'_{is}w'_{is'} + \mu_{qs}\mu_{qs'} - \mu_{qs}w'_{is'} - \mu_{qs'}w'_{is}\right]x_{iq}, \\ & \forall p = 1, \dots, m, q = p+1, \dots, m, s = 1, \dots, r, s' = s, \dots, r, \end{aligned} \tag{20}$$

$$\begin{aligned} V_{pqss'} \geq\ & \frac{1}{k}\sum_{i=1}^{n}\left[w'_{is}w'_{is'} + \mu_{qs}\mu_{qs'} - \mu_{qs}w'_{is'} - \mu_{qs'}w'_{is}\right]x_{iq} \\ & -\frac{1}{k}\sum_{i=1}^{n}\left[w'_{is}w'_{is'} + \mu_{ps}\mu_{ps'} - \mu_{ps}w'_{is'} - \mu_{ps'}w'_{is}\right]x_{ip}, \\ & \forall p = 1, \dots, m, q = p+1, \dots, m, s = 1, \dots, r, s' = s, \dots, r. \end{aligned} \tag{21}$$

With some rearrangement, we can obtain the final expression as

$$\begin{aligned} V_{pqss'} \geq\ & \frac{1}{k}\sum_{i=1}^{n} w'_{is}w'_{is'}\left(x_{ip} - x_{iq}\right) \\ & -\left(\left[\frac{1}{k}\sum_{i=1}^{n} w'_{is'}x_{ip}\right]\left[\frac{1}{k}\sum_{i=1}^{n} w'_{is}x_{ip}\right] - \left[\frac{1}{k}\sum_{i=1}^{n} w'_{is'}x_{iq}\right]\left[\frac{1}{k}\sum_{i=1}^{n} w'_{is}x_{iq}\right]\right), \\ & \forall p = 1, \dots, m, q = p+1, \dots, m, s = 1, \dots, r, s' = s, \dots, r, \end{aligned} \tag{22}$$

$$\begin{aligned} V_{pqss'} \geq\ & \frac{1}{k}\sum_{i=1}^{n} w'_{is}w'_{is'}\left(x_{iq} - x_{ip}\right) \\ & -\left(\left[\frac{1}{k}\sum_{i=1}^{n} w'_{is'}x_{iq}\right]\left[\frac{1}{k}\sum_{i=1}^{n} w'_{is}x_{iq}\right] - \left[\frac{1}{k}\sum_{i=1}^{n} w'_{is'}x_{ip}\right]\left[\frac{1}{k}\sum_{i=1}^{n} w'_{is}x_{ip}\right]\right), \\ & \forall p = 1, \dots, m, q = p+1, \dots, m, s = 1, \dots, r, s' = s, \dots, r. \end{aligned} \tag{23}$$

[M1] is a mixed-integer nonlinear programming model with nonlinear terms shown in Eqs. (22) and

(23). To linearize constraints (22) and (23), we introduce a binary decision variable $y_{ijp}$ to represent $x_{ip} \times x_{jp}$. Constraints (22) and (23) are presented as follows:

$$V_{pqss'} \geq \frac{1}{k}\sum_{i=1}^{n} w'_{is}w'_{is'}\left(x_{ip} - x_{iq}\right) - \frac{1}{k^2}\left[\left(\sum_{i=1}^{n} w'_{is}w'_{is'}x_{ip} + \sum_{i=1}^{n} w'_{is'}\sum_{j=1,j\neq i}^{n} w_{js}y_{ijp}\right)\right.$$
$$\left.-\left(\sum_{i=1}^{n} w'_{is}w'_{is'}x_{iq} + \sum_{i=1}^{n} w'_{is'}\sum_{j=1,j\neq i}^{n} w_{js}y_{ijq}\right)\right],$$
$$\forall p = 1,\dots,m, q = p+1,\dots,m, s = 1,\dots,r, s' = s,\dots,r, \quad (24)$$

$$V_{pqss'} \geq \frac{1}{k}\sum_{i=1}^{n} w'_{is}w'_{is'}\left(x_{iq} - x_{ip}\right) - \frac{1}{k^2}\left[\left(\sum_{i=1}^{n} w'_{is}w'_{is'}x_{iq} + \sum_{i=1}^{n} w'_{is'}\sum_{j=1,j\neq i}^{n} w_{js}y_{ijq}\right)\right.$$
$$\left.-\left(\sum_{i=1}^{n} w'_{is}w'_{is'}x_{ip} + \sum_{i=1}^{n} w'_{is'}\sum_{j=1,j\neq i}^{n} w_{js}y_{ijp}\right)\right],$$
$$\forall p = 1,\dots,m, q = p+1,\dots,m, s = 1,\dots,r, s' = s,\dots,r, \quad (25)$$

$$y_{ijp} \leq x_{ip}, \forall i = 1,\dots,n, \forall j = 1,\dots,n, j \neq i, \forall p = 1,\dots,m, \quad (26)$$

$$y_{ijp} \leq x_{jp}, \forall i = 1,\dots,n, \forall j = 1,\dots,n, j \neq i, \forall p = 1,\dots,m, \quad (27)$$

$$y_{ijp} \geq x_{ip} + x_{jp} - 1, \forall i = 1,\dots,n, \forall j = 1,\dots,n, j \neq i, \forall p = 1,\dots,m, \quad (28)$$

$$y_{ijp} \geq 0, \forall i = 1,\dots,n, \forall j = 1,\dots,n, j \neq i, \forall p = 1,\dots,m. \quad (29)$$

Therefore, the linearized model, denoted by [M1-linearized], has the same objective function as model [M1], subject to constraints (14), (15), (16), (17), (18), (19), (24), (25), (26), (27), (28), and (29). It should be noted that by assuming that the mean values of all groups are zero, constraints (13) in model [M1] can degrade to that proposed by Bertsimas et al. (2015), as shown below:

$$V_{pqss'} \geq \left|\frac{1}{k}\sum_{i=1}^{n}\left[\left(w'_{is} - 0\right)\left(w'_{is'} - 0\right)x_{ip} - \left(w'_{is} - 0\right)\left(w'_{is'} - 0\right)x_{iq}\right]\right|, \forall p =$$
$$1,\dots,m, q = p+1,\dots,m, s = 1,\dots,r, s' = s,\dots,r, \quad (30)$$

Then, Eq. (30) can be rewritten as follows:

$$V_{pqss'} \geq \frac{1}{k}\sum_{i=1}^{n}\left[w'_{is}w'_{is'}x_{ip} - w'_{is}w'_{is'}x_{iq}\right], \forall p =$$
$$1,\dots,m, q = p+1,\dots,m, s = 1,\dots,r, s' = s,\dots,r, \quad (31)$$

$$V_{pqss'} \geq \frac{1}{k}\sum_{i=1}^{n} w'_{is}w'_{is'}\left(x_{ip} - x_{iq}\right), \forall p =$$
$$1,\dots,m, q = p+1,\dots,m, s = 1,\dots,r, s' = s,\dots,r. \quad (32)$$

Therefore, based on this assumption, model [M1] can be converted to the optimization model proposed

by Bertsimas et al. (2015), which is denoted by model [M0] in this study. Clearly, the only difference between [M1] and [M0] lies in their second moment discrepancy constraint, as shown in Eqs. (13) and (30), respectively. Although simple, such a difference has a significant impact on the final partitioning results, which are shown in the next section. From the perspective of elementwise second moment discrepancy demonstrated in Eq. (10), model [M1] tries to minimize the difference of covariance for each pair of covariates between any two groups. As covariance is directly related to the correlation coefficients between covariates, the optimal partitioning results of [M1] would be that the correlation of the covariates of each group reaches maximum similarity. In contrast, [M0] tries to minimize the difference of the second raw moment between any two groups, which cannot reflect the correlation in a physical sense. Such a difference between [M1] and [M0] is essential for partitioning high-dimensional samples where the correlation between covariates is not trivial. The extensive experiments in Section 4 support this argument. The notation used in this study is summarized in Table 1.

Table 1. Notation

| Parameters | |
|---|---|
| $\mathbf{w}_i$ | The $i$th sample with $r$ dimensions |
| $\mathbf{w}_i'$ | The $i$th normalized sample with $r$ dimensions |
| $w_{is}$ | The $s$th covariate of the $i$th sample |
| $m$ | The number of groups |
| $k$ | The number of samples in each group |
| $n$ | The number of samples, which can be calculated by $n = mk$ |
| $\rho$ | The trade-off between the discrepancy in the first and second moments |
| $\hat{\boldsymbol{\mu}}$ | The mean of all samples |
| $\hat{\boldsymbol{\Sigma}}$ | The covariance matrix of all samples |
| $\mathbf{u}_j$ | The $j$th unit eigenvector of covariance matrix $\hat{\boldsymbol{\Sigma}}$ |
| $\lambda_j$ | The $j$th eigenvalue with respect to the $j$th unit eigenvector $\mathbf{u}_j$ |
| $\mathbf{U}$ | Eigenmatrix $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r]$ |
| $\boldsymbol{\Lambda}$ | Diagonal matrix with the diagonal element being the eigenvalue, i.e., $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$ |
| $p$ | Group $p$ |
| $q$ | Group $q$ |
| $\boldsymbol{\mu}_p$ | The mean of samples in group $p$ |
| $\mu_{ps}$ | The $s$th element in the mean of samples in group $p$ |
| $\mu_{qs}$ | The $s$th element in the mean of samples in group $q$ |
| $\mathbf{W}_p$ | The $i$th row of the matrix represents the $i$th normalized sample in group $p$. |
| $\mathbf{W}_p'$ | The $i$th row of the matrix represents the $i$th normalized sample in group $p$ minus the mean of |

| Parameters | |
|---|---|
| | samples in group $p$. |
| $\mathbf{w}'_{p:i}$ | The $i$th normalized sample in group $p$ |
| $w'_{p:is}$ | The $s$th covariate of the $i$th normalized sample in the group $p$ |
| $\mathbf{\Sigma}'_p$ | Covariance matrix of group $p$ |
| $\mathbf{\Sigma}'_q$ | Covariance matrix of group $q$ |
| Decision variables | |
| $x_{ip}$ | Binary decision variable that equals 1 if the $i$th normalized sample $\mathbf{w}'_i$ is assigned to group $p$ and 0 otherwise |
| $y_{ijp}$ | Binary decision variable that equals to 1 if the $i$th and the $j$th normalized samples are allocated to the same group $p$ |
| $M_{pqs}$ | Continuous decision variable, which represents the discrepancy in the first moment over the $s$th covariate between group $p$ and group $q$ |
| $V_{pqss}$ | Continuous decision variable, which represents the element of $\mathbf{\Sigma}'_p - \mathbf{\Sigma}'_q$ at location $(s, s)$ |
| $V_{pqss'}$ | Continuous decision variable, which represents the element of $\mathbf{\Sigma}'_p - \mathbf{\Sigma}'_q$ at location $(s, s')$ |

## 4. Computational experiments

In this experiment, we aim to evenly divide $n$ samples into $m$ groups using both the proposed model and the model proposed by Bertsimas et al. (2015). Three cases, namely the partitioning problem with one-dimensional (1-D) samples, the partitioning problem with two-dimensional (2-D) samples, and the partitioning problem with three-dimensional (3-D) samples, are investigated. In each case, parameter $\rho$ is adjusted to control the trade-off between the discrepancy in the first and second moments, as defined in Eq. (11), with values of $\rho$ set to 0.1, 1.0, and 5.0. We generate the sample data from multivariate normal distributions using the MATLAB $mvnrnd$ function and control the shape of the distribution by adjusting its mean $\boldsymbol{\mu}$ and covariance $\mathbf{\Sigma}$. The settings of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are listed in Table 2. We note that the parameter $c$ controls the covariance of 2-D and 3-D samples. As the diagonal elements of the covariance matrices are set to 1, $c$ can be regarded as the correlation coefficients of the covariates. We implement the proposed models, i.e., models [M0] and [M1-linearized], in Gurobi v11.0.1. All experiments are conducted using a desktop computer (Intel Core i7-12700H CPU, 2.30 GHz).

Table 2. Settings of mean and covariance for generating 1-D, 2-D and 3-D samples.

| Case | Mean $\boldsymbol{\mu}$ | Covariance $\boldsymbol{\Sigma}$ | Correlation coefficients $c$ |
|---|---|---|---|
| 1-D sample | 0.2 | 0.6 | / |
| 2-D sample | $[0.2, 0.3]^{\mathrm{T}}$ | $\begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}$ | {0.0,0.1, … ,0.9} |
| 3-D sample | $[0.2, 0.3, 0.4]^{\mathrm{T}}$ | $\begin{pmatrix} 1 & c & c \\ c & 1 & c \\ c & c & 1 \end{pmatrix}$ | {0.0,0.1, … ,0.9} |

### *4.1 Case of partitioning 20 samples*

Figure 1 shows the experimental results of partitioning $n = 20$ samples into $m = 4$ groups under the three cases. In the case of partitioning 1-D samples, when $\rho$ is set to 1.0 and 5.0, the maximum discrepancy is reduced by 2.67% and 17.30% respectively using model [M1-linearized] compared with model [M0]. In such cases, the objective function pays more attention to the discrepancy in the second moment and reduces the importance of the discrepancy in the first moment. As the characterization of the discrepancy in the second moment using model [M1-linearized] is more precise than that using model [M0], it is anticipated that model [M1-linearized] performs better on the partitioning results. However, such an improvement is reduced by the decrease in $\rho$. As can be seen, the partition performance of the three methods is almost the same when $\rho = 0.1$.

The experimental results of partitioning 2-D samples in the case of $\rho = 0.1$, $\rho = 1.0$, and $\rho = 5.0$ are plotted in Figure 1(d), Figure 1(e), and Figure 1(f), respectively. When $\rho = 1.0$ for different settings of correlation coefficients, the maximum discrepancy of the partitioning results obtained using model [M1-linearized] is consistently smaller than that obtained using model [M0]. In particular, the reduction of the maximum discrepancy using model [M1-linearized] is more than 50% when $c$ is 0.9. When $\rho$ is set to 5.0, the strength of model [M1-linearized] increases further. Similar to the 1-D case, the partitioning performance of the three methods is almost the same when $\rho = 0.1$. However, when

$c$ is set to 0.5, 0.8, and 0.9, model [M1-linearized] shows a slight improvement compared with model [M0], with a maximum discrepancy of 0.63%, 1.69%, and 1.85% respectively, smaller than that of model [M0]. A possible explanation is that the increased sample dimension increases the discrepancy between any two groups. Similar findings are observed when partitioning 3-D samples. The statistical results for the maximum discrepancy obtained using each method are listed in Tables B.1, B.2, and B.3 in Appendix B.

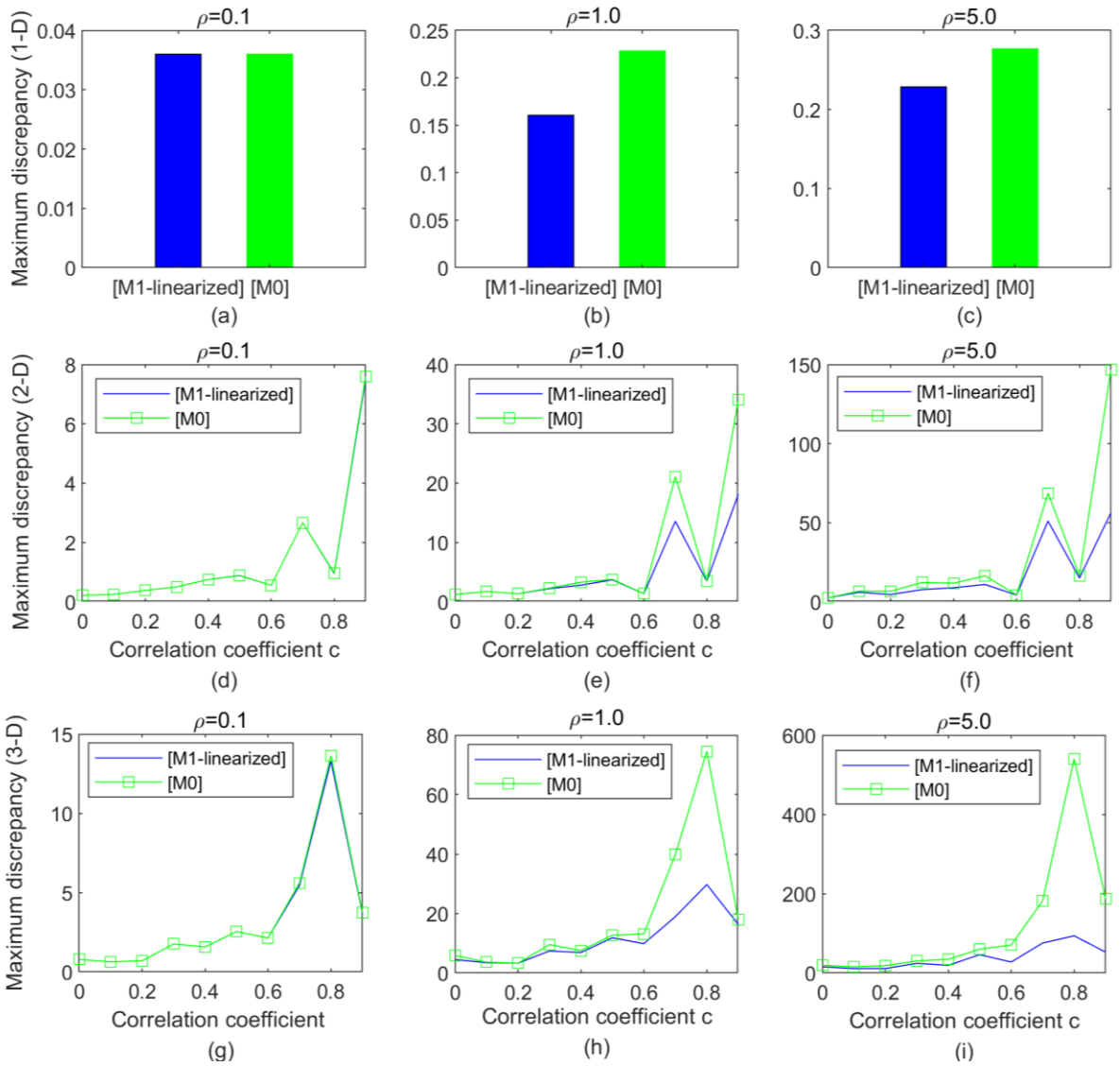


Figure 1. Maximum discrepancy of the partition results obtained from models [M0] and [M1-linearized] when 20 1-D, 2-D and 3-D samples are evenly divided into four groups.

### *4.2 Case of partitioning 40 samples*

In this section, we further examine the impact of the number of samples on partitioning performance. The experiment involves evenly dividing 40 1-D, 2-D, and 3-D samples into four groups. During the experiment, we find that the proposed model [M1-linearized] cannot obtain the optimal solution within a reasonable time limit, while model [M0] can obtain the optimal solution within one hour. Therefore, we limit the computation time of model [M1-linearized] to one hour.

Figure 2 shows the experimental results of partitioning $n = 40$ samples into $m = 4$ groups under three cases. In the case of partitioning 1-D samples, as shown in Figures 2(a) and 2(b), when $\rho$ is set to 0.1 and 1.0, the partitioning performance of the two methods is the same. However, when $\rho$ is set to 5.0, as shown in Figure 2(c), the maximum discrepancy of partitioning results derived from model [M1-linearized] is reduced by 34.50%. This result indicates that when the number of samples increases, the maximum discrepancy of the partitioning results obtained by model [M1-linearized] is very close to, or even smaller than that obtained by model [M0], even though model [M1-linearized] is unable to obtain an optimal solution within a reasonable time limit. This conclusion is strongly supported by the case of partitioning 2-D samples. Specifically, as shown in Figures 2(d) and 2(e), the curves of maximum discrepancy obtained by model [M0] are lower than those obtained by model [M1-linearized] when $\rho$ is set to 0.1 and 1.0. Nevertheless, the gap between the two curves becomes smaller as $\rho$ increases. When $\rho = 5.0$ (as shown in Figure 2(f)), the results become more promising in that the partitioning results obtained by model [M1-linearized] are better than those obtained by model [M0] for some settings of correlation coefficients. Specifically, the maximum discrepancy of the partitioning results obtained using model [M1-linearized] is 7.52%, 54.81%, 5.08% and 43.82%

smaller than that obtained using model [M0] when $c$ is 0.1, 0.3, 0.4 and 0.6, respectively. Similar results are observed in the case of partitioning 3-D samples, and the strength of model [M1-linearized] is further enhanced. This is reflected in the fact that for some correlation coefficient settings at $\rho = 1.0$ (as shown in Figure 2(h)), the partitioning results obtained using model [M1-linearized] are even better than those obtained using model [M0]. In addition, in the case of $\rho = 5.0$ (as shown in Figure 2(i)), the maximum discrepancy of the grouping results obtained using model [M1-linearized] is smaller than that obtained using model [M0] for all settings of $c$ (except for $c = 0.1$). This result indicates that as the sample dimension and the weight of the discrepancy in the second moment increase, model [M1-linearized] can obtain better grouping results than model [M0] even if the computation time of model [M1-linearized] is limited to one hour.

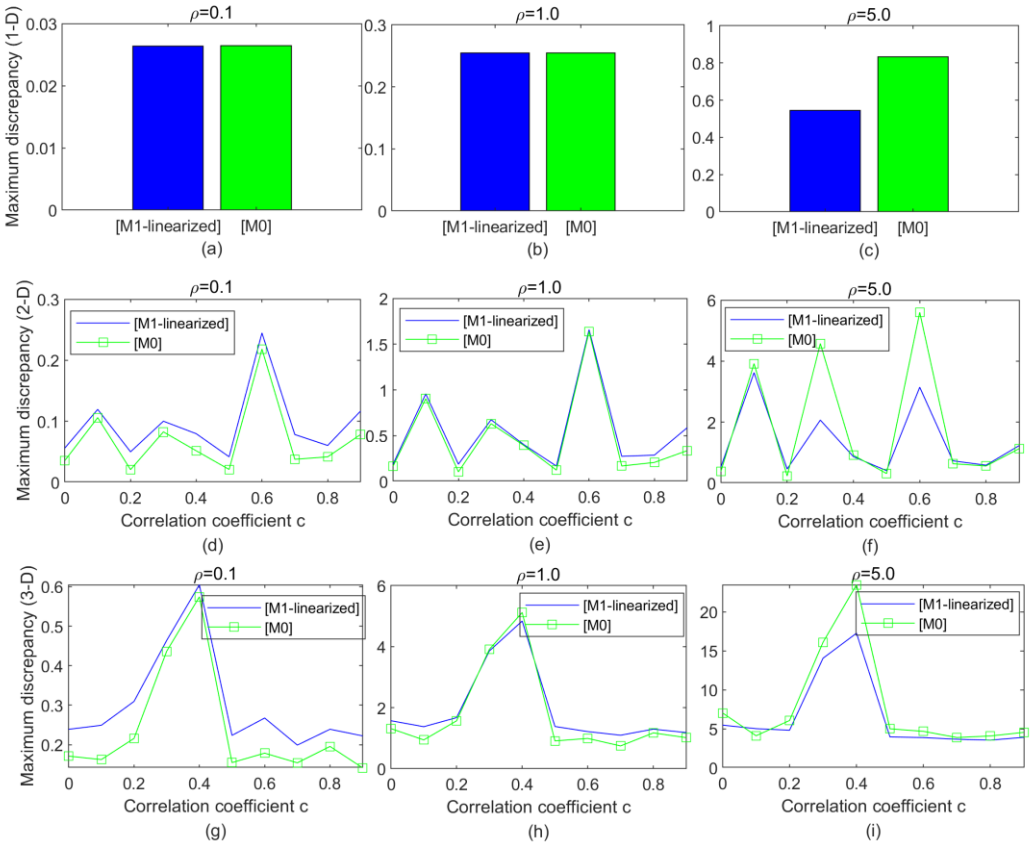


Figure 2. Maximum discrepancy of the partition results obtained from models [M0] and [M1-linearized] when 40 1-D, 2-D and 3-D samples are evenly divided into four groups.

### *4.3 Case of pembrolizumab clinical trial*

In a clinical trial that investigates the safety of pembrolizumab in combination with sequential or concomitant body radiotherapy in metastatic bladder cancer, Sundar et al. (2019) recruited 18 patients with covariates including age, sex, hemoglobin concentration, PD-L1 modified proportion score, smoking status, etc. We compare the ability of models [M0] and [M1-linearized] to balance the continuous covariates, including age, the PD-L1 modified proportion score, and hemoglobin concentration, when the patients are equally divided into two and three groups. The maximum discrepancies of partitioning 18 patients into two groups in the case of $\rho = 0.1$, $\rho = 1.0$, and $\rho = 5.0$ are plotted in Figure 3(a). It can be observed that model [M1-linearized] achieves the same performance as model [M0] regardless of the value of trade-off parameter $\rho$. In the case of three groups, as shown in Figure 3(b), model [M1-linearization] still has the same performance as the model [M0] when $\rho$ is set to 0.1 and 1.0. However, the maximum discrepancy of the partitioning results obtained by model [M1-linearized] is 40.6% smaller than that obtained by model [M0] when $\rho = 5.0$. Figure 4 shows the correlation coefficients among the three covariates in the pembrolizumab clinic trial. A considerable correlation between PD-L1 modified proportion score and hemoglobin concentration is observed. Since the proposed model [M1-linearized] considers the impacts of correlations in designing the objective function, it is reasonable for the model [M1-linearized] to achieve better partitioning results on the pembrolizumab clinic trial. The findings in this real-data experiment are consistent with those in the simulated experiments.

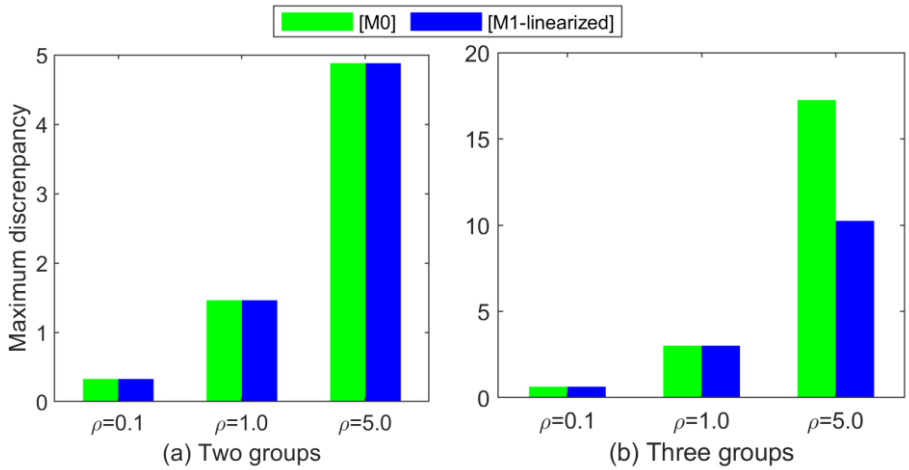


Figure 3. Maximum discrepancy of the partition results obtained from models [M0] and [M1-linearized] in the pembrolizumab clinic trial.

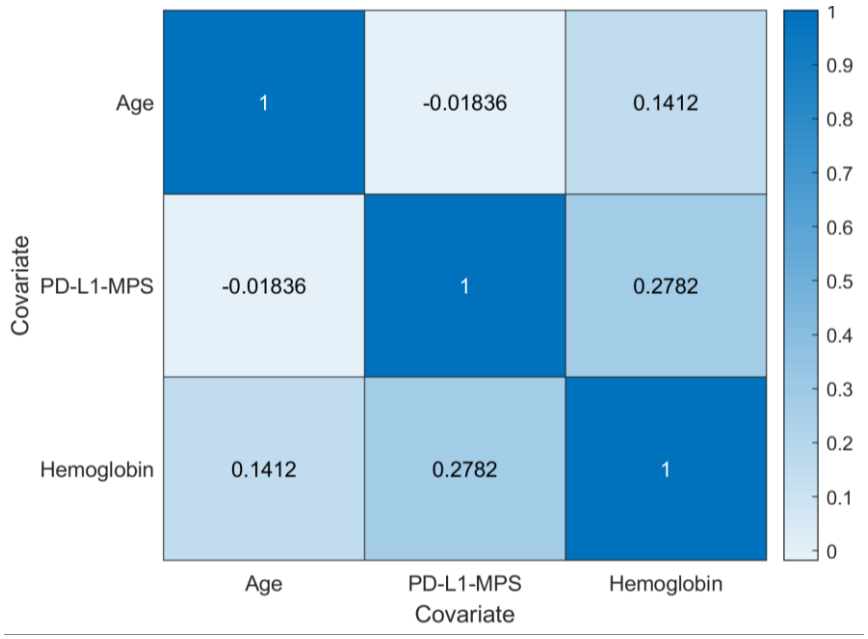


Figure 4. The correlation coefficients among covariates in the pembrolizumab clinic trial.

## 5. Discussion

### *5.1 Implications on high-dimensional samples in partitioning*

The experimental results demonstrate the strength of the proposed method in partitioning high-dimensional samples with strong correlations when compared with model [M0] (Bertsimas et al., 2015). The primary reason is that we implicitly consider the correlation between covariates when calculating

the second moment discrepancy between groups. Specifically, the second moment discrepancy is calculated based on the second central moment, as shown in Eqs. (22) and (23). The off-diagonal elements of the second central moment reflect the correlation of covariates in a physical sense, making the calculation of discrepancy between groups more accurate. This finding is beneficial for enhancing the covariate balance between groups in controlled experiments, thus further strengthening the credibility of inferred causal relationships. For example, in a clinical trial for a new antidiabetic drug, the experimental units are patients participating in the trial. The covariates of the experimental units include age, gender, weight, and other medical information, where age and weight typically exhibit correlation. In such cases, our optimization model can better consider these correlations, providing more balanced grouping results, thereby excluding factors other than drugs as much as possible when analyzing the causal relationship between antidiabetic drug use and diabetes remission. Therefore, we recommend considering the impact of correlations when partitioning samples with multiple covariates.

### *5.2 Implications on cross-validation technique*

Cross-validation is a technique used to evaluate and select machine learning models (Schaffer, 1993; Nti et al., 2021), which involves a process of equally dividing the dataset into $K$ subsets (often referred to as folds). The machine learning model is trained on the $K-1$ folds and evaluated on the remaining one fold. This training process is repeated $K$ times, with each fold being used for validation. Finally, the average of the $K$ evaluation results is taken as the performance metric of the machine learning model. Random partitioning is commonly adopted to partition data into $K$ folds. When the sample size is large, random partitioning ensures that the different subsets have similar distributions, allowing the model to be trained on representative data from the $K-1$ folds and appropriately

evaluated on a validation set from the remaining fold. However, when the sample size is moderate or small, the distributions among different subsets may differ, which can lead to biased or unreliable performance estimates. The proposed mixed-integer linear programming model that aims to achieve covariate balancing can be naturally used in the cross-validation process. More importantly, data used for training machine learning models typically have high-dimensional features. Complex relationships and correlations may exist among the features. Since our model has advantages in effectively capturing the correlation of covariates, it is promising to deploy our model to partition high-dimensional samples in the cross-validation process.

### *5.3 Practical consideration on computational efficiency*

In the proposed model [M1], there exist non-linear terms as shown in the constraints (22) and (23), which increase the computational complexity of model [M1] compared with model [M0]. To address this, we linearize model [M1] and obtain model [M1-linearized], which significantly accelerates the optimization process. However, this linearization process introduces additional decision variables. In the case of partitioning $n$ samples with $r$ covariates into $m$ groups, [M1-linearized] has $n^2m - \frac{m(m-1)}{2}$ binary decision variables, while model [M0] only has $\frac{m(2n-m+1)}{2}$ binary decision variables. Therefore, as the sample size $n$ increases, the number of decision variables in [M1-linearized] increases quadratically, while the number of decision variables in model [M0] increases linearly. Consequently, for the same partitioning tasks, it is faster to solve model [M0] than to solve model [M1-linearized]. This difference becomes more pronounced as $n$ increases, making it challenging to obtain the optimal solution of model [M1-linearized] within a reasonable time limit. Since the increase in the computational load mainly comes from the calculation of covariate

correlations, it is recommended to adopt model [M0] when the correlations among covariates are not significant and the computational resources are limited. However, when the covariates show significant correlations, model [M1-linearized] is recommended to produce more balanced partitioning results.

## 6. Conclusion

In this study, we propose a mixed-integer linear programming model for partitioning experimental units in controlled experiments. We propose a novel imbalance measure, i.e., the maximum discrepancy in both the first and second central moments between any two groups. A mixed-integer nonlinear programming model ([M1]) is then established intuitively based on the proposed imbalance measure. By adopting a simple linearization technique, we linearize the non-linear model to obtain the finalized mixed-integer linear programming model ([M1-linearized]) to accelerate the optimization process.

To evaluate the effectiveness of the proposed model, we generate samples from 1-D, 2-D, and 3-D Gaussian distributions. In terms of generating the 2-D and 3-D samples, the correlation coefficient parameter $c$ takes discrete values within the range [0.1, 0.9] at intervals of 0.1 to adjust the correlation between covariates. The setup of these generated data simulates the scenarios encountered in controlled experiments, where experimental units have multiple covariates with varying degrees of correlation between them. In the case of partitioning 20 samples, the proposed model [M1-linearized] achieves better partitioning results than the [M0] model, where a reduction of 54.81% in the maximum discrepancy of the partitioning results is observed when $c$ is set to 0.3 and the weighting parameter $\rho$ is set to 5.0. In the case of partitioning 40 samples, our results show that as the sample dimension and the weight of the discrepancy in the second central moment increase, model [M1-linearized] can

obtain better grouping results than model [M0].

Furthermore, we compare the partitioning performance of models [M0] and [M1-linearized] on the pembrolizumab clinical trial. Computational results show that the maximum discrepancy of the partitioning results obtained by model [M1-linearized] is 40.6% smaller than that obtained by model [M0] when $\rho = 5.0$. These results on the simulated and real datasets indicate that our optimization model can better account for correlations between covariates, providing more balanced grouping results. Therefore, we recommend considering the impact of correlations when partitioning samples with multiple covariates.

One limitation of this study is the considerable computation load of the proposed method. As discussed in Section 5.3, the main computation burden comes from the calculation of covariance matrix, especially the off-diagonal elements of in the covariance matrix. To address this issue, we might calculate the covariance matrix in an approximate way. For example, if the correlation of some covariate pairs is smaller than a predefined threshold, we can set the corresponding off-diagonal element in the covariance matrix to zero. However, the choice of the threshold becomes a crucial problem in balancing the computational efficiency and the partitioning performance, which will be addressed in our future work. Another promising method is to employ the evolution algorithm by exploiting its flexibility and robustness in solving complex optimization problems, as demonstrated in Tam (2018). However, the adoption of the evolution algorithm possibly requires certain modification on the objective function, which should be investigated further in the future research.

**Data Availability Statement**

Computer code, simulated Gaussian distributed dataset, and pembrolizumab clinic trial data are

published in Github. Readers can access these files via this link: https://github.com/Luoxi-scholar/CovBa.git.

**Disclosure of Interest**

The authors report there are no competing interests to declare.

**Funding**

This work was supported by the Start-Up Grant from Nanyang Technological University, Singapore and the National Natural Science Foundation of China [Grant Nos. 72071173, 72371221].

**Appendix A. Proof of zero mean and identity covariance of normalized samples**

To prove that the normalized samples have zero mean and identity covariance, we rewrite the transform in Eq. (1) as follows,

$$\mathbf{w}' = \boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}}), \tag{33}$$

where $\mathbf{w}'$ is the random variable representing the normalized samples, i.e., $\mathbf{w}_1', \mathbf{w}_2', \dots, \mathbf{w}_n'$ are realizations of the random variable $\mathbf{w}'$. The mean of $\mathbf{w}'$ can be written as:

$$\mathbb{E}(\mathbf{w}') = \mathbb{E}[\boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}})] = \boldsymbol{\Gamma}(\mathbb{E}[\mathbf{w}] - \widehat{\boldsymbol{\mu}}). \tag{34}$$

Here we use the mean to calculate $\mathbb{E}[\mathbf{w}]$, i.e., $\mathbb{E}[\mathbf{w}] = \widehat{\boldsymbol{\mu}}$. Therefore,

$$\mathbb{E}(\mathbf{w}') = 0. \tag{35}$$

The covariance matrix of $\mathbf{w}'$ can be written as:

$$\begin{aligned}
\mathrm{var}(\mathbf{w}') &= \mathrm{var}[\boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}})] \\
&= \mathbb{E}\left[\left[\boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}}) - \mathbb{E}\big(\boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}})\big)\right]\left[\boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}}) - \mathbb{E}\big(\boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}})\big)\right]^T\right] \\
&= \mathbb{E}\Big[[\boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}})][\boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}})]^{\mathrm{T}}\Big] \\
&= \mathbb{E}[\boldsymbol{\Gamma}(\mathbf{w} - \widehat{\boldsymbol{\mu}})(\mathbf{w} - \widehat{\boldsymbol{\mu}})^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}]
\end{aligned}$$

$$= \boldsymbol{\Gamma}\mathbb{E}[(\mathbf{w}-\widehat{\boldsymbol{\mu}})(\mathbf{w}-\widehat{\boldsymbol{\mu}})^{\mathrm{T}}]\boldsymbol{\Gamma}^{\mathrm{T}}$$

$$= \boldsymbol{\Gamma} var[(\mathbf{w}-\widehat{\boldsymbol{\mu}})]\boldsymbol{\Gamma}^{\mathrm{T}}$$

$$= \boldsymbol{\Gamma} var[\mathbf{w}]\boldsymbol{\Gamma}^{\mathrm{T}} \tag{36}$$

Here we use the covariance to calculate $\mathrm{var}[\mathbf{w}]$, i.e., $\mathrm{var}[\mathbf{w}] = \widehat{\boldsymbol{\Sigma}}$. Therefore,

$$\mathrm{var}(\mathbf{w}') = \boldsymbol{\Gamma}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Gamma}^{\mathrm{T}} = \boldsymbol{\Lambda}^{-1/2}\mathbf{U}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}\mathbf{U}\boldsymbol{\Lambda}^{-1/2}. \tag{37}$$

By substituting Eq. (2) into Eq. (37),

$$\mathrm{var}(\mathbf{w}') = \boldsymbol{\Lambda}^{-1/2}(\mathbf{U}^{T}\mathbf{U})\boldsymbol{\Lambda}(\mathbf{U}^{\mathrm{T}}\mathbf{U})\boldsymbol{\Lambda}^{-1/2}. \tag{38}$$

Since $\mathbf{U}$ is the eigenmatrix that contains eigenvectors, which means each column vector are orthogonal to each other,

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}. \tag{39}$$

Therefore,

$$\mathrm{var}(\mathbf{w}') = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2} = \mathbf{I}. \tag{40}$$

As can be seen, the normalized samples by Eq. (1) will have zero mean and identity covariance.

## Appendix B. Partitioning results

The maximum discrepancy of the partition results obtained by models [M1-linearized] and [M0] is denoted by $D_{[\mathrm{M1}]}$ and $D_{[\mathrm{M0}]}$, respectively.

Table B.1. Maximum discrepancy of the partition results obtained from the proposed model and Bertsimas's model when 20 1-D samples are evenly divided into four groups

| Trade-off parameter $\rho$ | $D_{[\mathrm{M1}]}$ | $D_{[\mathrm{M0}]}$ | $\frac{(D_{[\mathrm{M1}]} - D_{[\mathrm{M0}]})}{D_{[\mathrm{M0}]}} \times 100\%$ |
|---|---|---|---|
| $\rho = 0.1$ | 0.0360 | 0.0360 | 0.00% |
| $\rho = 1.0$ | 0.1603 | 0.1647 | **-2.67%** |
| $\rho = 5.0$ | 0.2285 | 0.2763 | **-17.30%** |

Table B.2. Maximum discrepancy of the partition results obtained from the proposed model and Bertsimas's model when 20 2-D samples are evenly divided into four groups

| Trade-off parameter $\rho$ | $c$ | $D_{[\mathrm{M1}]}$ | $D_{[\mathrm{M0}]}$ | $\frac{(D_{[\mathrm{M1}]} - D_{[\mathrm{M0}]})}{D_{[\mathrm{M0}]}} \times 100\%$ |
|---|---|---|---|---|
| $\rho = 0.1$ | 0.0 | 0.2199 | 0.2199 | 0.00% |
| | 0.1 | 0.2496 | 0.2496 | 0.00% |
| | 0.2 | 0.3842 | 0.3842 | 0.00% |
| | 0.3 | 0.5052 | 0.5052 | 0.00% |
| | 0.4 | 0.7510 | 0.7510 | 0.00% |
| | 0.5 | 0.8843 | 0.8899 | **-0.63%** |
| | 0.6 | 0.5540 | 0.5540 | 0.00% |
| | 0.7 | 2.6699 | 2.6699 | 0.00% |
| | 0.8 | 0.9566 | 0.9731 | **-1.69%** |
| | 0.9 | 7.4556 | 7.5965 | **-1.85%** |
| $\rho = 1.0$ | 0.0 | 1.2234 | 1.2315 | **-0.66%** |
| | 0.1 | 1.7140 | 1.7439 | **-1.72%** |
| | 0.2 | 1.3631 | 1.3631 | 0.00% |
| | 0.3 | 2.1467 | 2.3134 | **-7.20%** |
| | 0.4 | 2.7066 | 3.2992 | **-17.96%** |
| | 0.5 | 3.4274 | 3.7890 | **-9.54%** |
| | 0.6 | 1.3662 | 1.3757 | **-0.69%** |
| | 0.7 | 10.9443 | 21.0933 | **-48.11%** |
| | 0.8 | 3.1190 | 3.4804 | **-10.38%** |
| | 0.9 | 16.5729 | 34.0665 | **-51.35%** |
| $\rho = 5.0$ | 0.0 | 2.4562 | 2.4714 | **-0.62%** |
| | 0.1 | 4.6504 | 6.7594 | **-31.20%** |
| | 0.2 | 3.4843 | 6.6977 | **-47.98%** |
| | 0.3 | 6.3043 | 12.2300 | **-48.45%** |
| | 0.4 | 7.5106 | 11.7787 | **-36.24%** |
| | 0.5 | 6.5115 | 16.4101 | **-60.32%** |
| | 0.6 | 3.2689 | 4.3830 | **-25.42%** |
| | 0.7 | 33.6952 | 68.6417 | **-50.91%** |
| | 0.8 | 8.6957 | 16.5129 | **-47.34%** |
| | 0.9 | 47.5715 | 146.9449 | **-67.63%** |

Table B.3. Discrepancy of the partition results obtained from the proposed model and Bertsimas's model when 20 3-D samples are evenly divided into four groups

| Trade-off parameter $\rho$ | $c$ | $D_{[\mathrm{M1}]}$ | $D_{[\mathrm{M0}]}$ | $\frac{(D_{[\mathrm{M1}]} - D_{[\mathrm{M0}]})}{D_{[\mathrm{M0}]}} \times 100\%$ |
|---|---|---|---|---|
| $\rho = 0.1$ | 0.0 | 0.7838 | 0.7838 | 0.00% |
| | 0.1 | 0.6318 | 0.6318 | 0.00% |
| | 0.2 | 0.6957 | 0.6957 | 0.00% |
| | 0.3 | 1.7593 | 1.7593 | 0.00% |
| | 0.4 | 1.5735 | 1.5735 | 0.00% |
| | 0.5 | 2.5409 | 2.5409 | 0.00% |
| | 0.6 | 2.1364 | 2.1364 | 0.00% |
| | 0.7 | 5.4530 | 5.5987 | **-2.60%** |
| | 0.8 | 13.3067 | 13.6182 | **-2.29%** |
| | 0.9 | 3.6982 | 3.7431 | **-1.20%** |
| $\rho = 1.0$ | 0.0 | 4.4935 | 5.8691 | **-23.44%** |
| | 0.1 | 3.4261 | 3.6042 | **-4.94%** |
| | 0.2 | 3.2745 | 3.2745 | 0.00% |
| | 0.3 | 7.3757 | 9.5362 | **-22.66%** |
| | 0.4 | 6.8707 | 7.5209 | **-8.65%** |
| | 0.5 | 11.8188 | 12.6581 | **-6.63%** |
| | 0.6 | 9.8433 | 13.1225 | **-24.99%** |
| | 0.7 | 18.9389 | 39.8604 | **-52.49%** |
| | 0.8 | 29.7225 | 74.5422 | **-60.13%** |
| | 0.9 | 16.2893 | 17.9768 | **-9.39%** |
| $\rho = 5.0$ | 0.0 | 15.4513 | 19.3897 | **-20.31%** |
| | 0.1 | 10.9357 | 15.3796 | **-28.89%** |
| | 0.2 | 11.0186 | 18.3034 | **-39.80%** |
| | 0.3 | 24.1568 | 30.2377 | **-20.11%** |
| | 0.4 | 19.1977 | 34.3673 | **-44.14%** |
| | 0.5 | 45.9795 | 59.9309 | **-23.28%** |
| | 0.6 | 27.4676 | 70.2984 | **-60.93%** |
| | 0.7 | 75.4534 | 181.9854 | **-58.54%** |
| | 0.8 | 93.5671 | 540.1609 | **-82.68%** |
| | 0.9 | 52.4544 | 187.0272 | **-71.95%** |

Table B.4. Discrepancy of the partition results obtained from the proposed model and Bertsimas's model when 40 1-D samples are evenly divided into four groups

| Trade-off parameter $\rho$ | $D_{[\mathrm{M1}]}$ | $D_{[\mathrm{M0}]}$ | $\frac{(D_{[\mathrm{M2}]} - D_{[\mathrm{M0}]})}{D_{[\mathrm{M0}]}} \times 100\%$ |
|---|---|---|---|
| $\rho = 0.1$ | 0.0264* | 0.0264 | 0.00% |
| $\rho = 1.0$ | 0.2546* | 0.2546 | 0.00% |
| $\rho = 5.0$ | 0.5460* | 0.8336 | **-34.50%** |

Note: * indicates the solution obtained by limiting the solving time of the model to one hour.

Table B.5. Discrepancy of the partition results obtained from the proposed model and Bertsimas's model when 40 2-D samples are evenly divided into four groups

| Trade-off parameter $\rho$ | $c$ | $D_{[\mathrm{M1}]}$ | $D_{[\mathrm{M0}]}$ | $\frac{(D_{[\mathrm{M1}]} - D_{[\mathrm{M0}]})}{D_{[\mathrm{M0}]}} \times 100\%$ |
|---|---|---|---|---|
| $\rho = 0.1$ | 0.0 | 0.0562 | 0.0357 | 57.51% |
| | 0.1 | 0.1195 | 0.1060 | 12.77% |
| | 0.2 | 0.0500 | 0.0208 | 140.18% |
| | 0.3 | 0.1001 | 0.0826 | 21.18% |
| | 0.4 | 0.0796 | 0.0516 | 54.17% |
| | 0.5 | 0.0418 | 0.0210 | 98.69% |
| | 0.6 | 0.2447 | 0.2183 | 12.09% |
| | 0.7 | 0.0783 | 0.0377 | 107.55% |
| | 0.8 | 0.0604 | 0.0417 | 44.72% |
| | 0.9 | 0.1167 | 0.0783 | 48.99% |
| $\rho = 1.0$ | 0.0 | 1.5718* | 1.3139 | 13.37% |
| | 0.1 | 1.3765* | 0.9429 | 5.78% |
| | 0.2 | 1.6730* | 1.5680 | 80.13% |
| | 0.3 | 3.8587* | 3.9119 | 6.84% |
| | 0.4 | 4.8393* | 5.1187 | 2.26% |
| | 0.5 | 1.3830* | 0.9093 | 35.88% |
| | 0.6 | 1.2159* | 0.9925 | 1.30% |
| | 0.7 | 1.0970* | 0.7457 | 62.67% |
| | 0.8 | 1.2965* | 1.1731 | 38.34% |
| | 0.9 | 1.1837* | 1.0173 | 74.27% |
| $\rho = 5.0$ | 0.0 | 15.4513* | 19.3897 | 37.92% |
| | 0.1 | 10.9357* | 15.3796 | **-7.52%** |
| | 0.2 | 11.0186* | 18.3034 | 95.91% |
| | 0.3 | 24.1568* | 30.2377 | **-54.81%** |
| | 0.4 | 19.1977* | 34.3673 | **-5.08%** |
| | 0.5 | 45.9795* | 59.9309 | 30.21% |
| | 0.6 | 27.4676* | 70.2984 | **-43.82%** |
| | 0.7 | 75.4534* | 181.9854 | 14.73% |
| | 0.8 | 93.5671* | 540.1609 | 4.77% |
| | 0.9 | 52.4544* | 187.0272 | 8.79% |

Table B.6. Discrepancy of the partition results obtained from the proposed model and Bertsimas's model when 40 3-D samples are evenly divided into four groups

| Trade-off parameter $\rho$ | $c$ | $D_{[\mathrm{M1}]}$ | $D_{[\mathrm{M0}]}$ | $\frac{(D_{[\mathrm{M1}]} - D_{[\mathrm{M0}]})}{D_{[\mathrm{M0}]}} \times 100\%$ |
|---|---|---|---|---|
| $\rho = 0.1$ | 0.0 | 0.2389* | 0.1710 | 39.71% |
| | 0.1 | 0.2492* | 0.1623 | 53.48% |
| | 0.2 | 0.3095* | 0.2159 | 43.37% |
| | 0.3 | 0.4640* | 0.4366 | 6.26% |
| | 0.4 | 0.6049* | 0.5742 | 5.34% |
| | 0.5 | 0.2236* | 0.1547 | 44.53% |
| | 0.6 | 0.2675* | 0.1786 | 49.79% |
| | 0.7 | 0.1989* | 0.1541 | 29.11% |
| | 0.8 | 0.2391* | 0.1956 | 22.25% |
| | 0.9 | 0.2223* | 0.1414 | 57.21% |
| $\rho = 1.0$ | 0.0 | 1.5718* | 1.3139 | 19.63% |
| | 0.1 | 1.3765* | 0.9429 | 45.98% |
| | 0.2 | 1.6730* | 1.5680 | 6.70% |
| | 0.3 | 3.8587* | 3.9119 | **-1.36%** |
| | 0.4 | 4.8393* | 5.1187 | **-5.46%** |
| | 0.5 | 1.3830* | 0.9093 | 52.10% |
| | 0.6 | 1.2159* | 0.9925 | 22.51% |
| | 0.7 | 1.0970* | 0.7457 | 47.11% |
| | 0.8 | 1.2965* | 1.1731 | 10.52% |
| | 0.9 | 1.1837* | 1.0173 | 16.36% |
| $\rho = 5.0$ | 0.0 | 15.4513* | 19.3897 | **-22.16%** |
| | 0.1 | 10.9357* | 15.3796 | 22.02% |
| | 0.2 | 11.0186* | 18.3034 | **-20.56%** |
| | 0.3 | 24.1568* | 30.2377 | **-12.71%** |
| | 0.4 | 19.1977* | 34.3673 | **-26.46%** |
| | 0.5 | 45.9795* | 59.9309 | **-20.77%** |
| | 0.6 | 27.4676* | 70.2984 | **-17.11%** |
| | 0.7 | 75.4534* | 181.9854 | **-5.71%** |
| | 0.8 | 93.5671* | 540.1609 | **-13.17%** |
| | 0.9 | 52.4544* | 187.0272 | **-13.93%** |

**References**

Arbona, A., López-Estrada, S., Prior, D., Rialp, J., 2023. How much do companies know what contributes to education? Journal of the Operational Research Society 74(12), 1–13.

Ben-Michael, E., Feller, A., Hirshberg, D., Zubizarreta, J., 2021. The balancing act in causal inference. arXiv preprint arXiv:2110.14831.

Bertsimas, D., Johnson, M., Kallus, N., 2015. The power of optimization over randomization in designing experiments involving small samples. Operations Research 63(4), 868–876.

Bertsimas, D., Korolko, N., Weinstein, A. M., 2019. Covariate-adaptive optimization in online clinical trials. Operations Research 67(4), 1150–1161.

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

Deaton, A., Cartwright, N., 2018. Understanding and misunderstanding randomized controlled trials. Social Science & Medicine 210, 2–21.

Festing, M., Altman, D., 2002. Guidelines for the design and statistical analysis of experiments using laboratory animals. ILAR Journal 43(4), 244–258.

Fisher, R., 1936. Design of experiments. British Medical Journal 1(3923), 554.

Ferrier, G. D., Valdmanis, V. G., 2004. Do mergers improve hospital productivity? Journal of the Operational Research Society 55(10), 1071–1080.

Greevy, R., Lu, B., Silber, J., Rosenbaum, P., 2004. Optimal multivariate matching before randomization. Biostatistics 5(2), 263–275.

Harshaw, C., Sävje, F., Spielman, D. A., Zhang, P., 2024. Balancing covariates in randomized

experiments with the Gram–Schmidt walk design. Journal of the American Statistical Association, 1–13.

Hochbaum, D.S., Rao, X., Sauppe, J., 2022. Network flow methods for the minimum covariate imbalance problem. European Journal of Operational Research 300(3), 827–836.

Josey, K. P., Juarez-Colunga, E., Yang, F., Ghosh, D., 2021. A framework for covariate balance using Bregman distances. Scandinavian Journal of Statistics 48(3), 790–816.

Kallus, N., 2018. Optimal a priori balance in the design of controlled experiments. Journal of the Royal Statistical Society Series B: Statistical Methodology 80(1), 85–112.

Kane, L., Fang, T., Galetta, M., Goyal, D., Nicholson, K., Kepler, C., Vaccaro, A., Schroeder, G., 2020. Propensity score matching: a statistical method. Clinical Spine Surgery 33(3), 120–122.

Kessy, A., Lewin, A., Strimmer, K., 2018. Optimal whitening and decorrelation. The American Statistician 72(4), 309–314.

Kwon, H. Y., Sauppe, J. J., Jacobson, S. H., 2019. Bias in balance optimization subset selection: Exploration through examples. Journal of the Operational Research Society 70(1), 67–80.

Li, Y., Kang, L., Huang, X., 2021. Covariate balancing based on kernel density estimates for controlled experiments. Statistical Theory and Related Fields 5(2), 102–113.

Ma, Z., Hu, F., 2013. Balancing continuous covariates based on kernel densities. Contemporary Clinical Trials 34(2), 262–269.

McHugh, R., Matts, J., 1983. Post-stratification in the randomized clinical trial. Biometrics 39(1), 217–225.

Mergoni, A., De Witte, K., 2022. Estimating the causal impact of an intervention on efficiency in a

dynamic setting. Journal of the Operational Research Society 73(10), 2275–2293.

Miratrix, L., Sekhon, J., Yu, B., 2013. Adjusting treatment effect estimates by post-stratification in randomized experiments. Journal of the Royal Statistical Society Series B: Statistical Methodology 75(2), 369–396.

Morgan, K., Rubin, D., 2012. Rerandomization to improve covariate balance in experiments. The Annals of Statistics 40(2), 1236–1282.

Morgan, K., Rubin, D., 2015. Rerandomization to balance tiers of covariates. Journal of the American Statistical Association 110(512), 1412–1421.

Moulton, L., 2004. Covariate-based constrained randomization of group-randomized trials. Clinical Trials 1(3), 297–305.

Nti, I. K., Nyarko-Boateng, O., Aning, J., 2021. Performance of machine learning algorithms with different K values in K-fold cross-validation. International Journal of Information Technology and Computer Science 13(6), 61–71.

Nikolaev, A. G., Jacobson, S. H., Cho, W. K. T., Sauppe, J. J., Sewell, E. C., 2013. Balance optimization subset selection (BOSS): An alternative approach for causal inference with observational data. Operations Research 61(2), 398–412.

Ning, Y., Sida, P., Imai, K., 2020. Robust estimation of causal effects via a high-dimensional covariate balancing propensity score. Biometrika 107(3), 533–554.

Nishi, T., Takaichi, A., 2004. An extended minimization method to assure similar means of continuous prognostic variables between treatment groups. Japanese Journal of Biometrics 24(2), 43–55.

Papoulis, A., 1984. Probability, Random Variables and Stochastic Processes (2nd ed.). McGraw-Hill,

New York, the USA.

Pocock, S. J., 2013. Clinical trials: A Practical Approach. John Wiley & Sons.

Pocock, S. J., Simon, R., 1975. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 30(1), 103–115.

Rice, J., 2007. Mathematical Statistics and Data Analysis. Duxbury Press, Pacific Grove.

Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70(1), 41–55.

Rosenberger, W. F., Lachin, J. M., 2015. Randomization in Clinical Trials: theory and practice. John Wiley & Sons.

Sundahl, N., Vandekerkhove, G., Decaestecker, K., Meireson, A., De Visschere, P., Fonteyne, V., & Ost, P. (2019). Randomized phase 1 trial of pembrolizumab with sequential versus concomitant stereotactic body radiotherapy in metastatic urothelial carcinoma. European urology, 75(5), 707-711. https://doi.org/10.1016/j.eururo.2019.01.009

Schaffer, C., 1993. Selecting a classification method by cross-validation. Machine learning 13, 135–143.

Schneeweiss, S., Rassen, J., Glynn, R., Avorn, J., Mogun, H., Brookhart, M., 2009. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 20(4), 512–522.

Tam Cho, W. K., 2018. An evolutionary algorithm for subset selection in causal inference models. Journal of the Operational Research Society 69(4), 630–644.

Turner, P., Meka, R., Rigollet, P., 2020. Balancing Gaussian vectors in high dimension. In Proceedings

of Thirty Third Conference on Learning Theory, 3455–3486.

Vazquez, A. R., Wong, W. K., 2024. Mathematical programming tools for randomization purposes in small two-arm clinical trials: A case study with real data. Pharmaceutical Statistics. 1–19.

Wang, B., Ogburn, E., Rosenblum, M., 2019. Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions. Biometrics 75(4), 1391–1400.

Wei, W., Ma, X., Wang, J., 2024. Fair adaptive experiments. Advances in Neural Information Processing Systems, 36.