The Annals of Statistics
2020, Vol. 48, No. 1, 43–63
https://doi.org/10.1214/18-AOS1790

RERANDOMIZATION IN 2^K FACTORIAL EXPERIMENTS

BY XINRAN LI[1], PENG DING[2] AND DONALD B. RUBIN[3]

[1]*Department of Statistics, University of Pennsylvania, lixinran@wharton.upenn.edu*
[2]*Department of Statistics, University of California, Berkeley, pengdingpku@berkeley.edu*
[3]*Department of Statistical Science, Temple University, donald.rubin@temple.edu*

With many pretreatment covariates and treatment factors, the classical factorial experiment often fails to balance covariates across multiple factorial effects simultaneously. Therefore, it is intuitive to restrict the randomization of the treatment factors to satisfy certain covariate balance criteria, possibly conforming to the tiers of factorial effects and covariates based on their relative importances. This is rerandomization in factorial experiments. We study the asymptotic properties of this experimental design under the randomization inference framework without imposing any distributional or modeling assumptions of the covariates and outcomes. We derive the joint asymptotic sampling distribution of the usual estimators of the factorial effects, and show that it is symmetric, unimodal and more "concentrated" at the true factorial effects under rerandomization than under the classical factorial experiment. We quantify this advantage of rerandomization using the notions of "central convex unimodality" and "peakedness" of the joint asymptotic sampling distribution. We also construct conservative large-sample confidence sets for the factorial effects.

1. Introduction. Factorial experiments, initially proposed by Fisher (1935) and Yates (1937), have been widely used in the agricultural science (see textbooks by Cochran and Cox (1950), Hinkelmann and Kempthorne (2007), Kempthorne (1952), Cox and Reid (2000)) and engineering (see textbooks by Box, Hunter and Hunter (2005), Wu and Hamada (2011)). Recently, factorial experiments also become popular in social sciences (e.g., Angrist, Lang and Oreopoulos (2009), Branson, Dasgupta and Rubin (2016), Dasgupta, Pillai and Rubin (2015)). The completely randomized factorial experiment (CRFE) balances covariates under different treatment combinations on average. However, with more pretreatment covariates and treatment factors, we have higher chance to observe unbalanced covariates with respect to multiple factorial effects. Many researchers have recognized this issue in different experimental designs (e.g., Bruhn and McKenzie (2009), Fisher (1926), Hansen and Bowers (2008), Student (1938)). To avoid this, we can force a treatment allocation to have covariate balance, which is sometimes called rerandomization (e.g., Cox (1982, 2009), Morgan and Rubin (2012)), restricted or constrained randomization (e.g., Bailey (1983), Grundy and Healy (1950), Yates (1948), Youden (1972)).

Extending Morgan and Rubin (2012)'s proposal for treatment-control experiments, Branson, Dasgupta and Rubin (2016) proposed to use rerandomization in factorial experiments to improve covariate balance, and studied finite sample properties of this design under the assumptions of equal sample sizes of all treatment combinations, Gaussianity of covariate and outcome means, and additive factorial effects. Without requiring any of these assumptions, we propose more general covariate balance criteria for rerandomization in 2^K factorial experiments, extend their theory with an asymptotic analysis of the sampling distributions of

Received September 2017; revised September 2018.
MSC2010 subject classifications. Primary 62K15, 62K10; secondary 62K05.
Key words and phrases. Covariate balance, tiers of covariates, tiers of factorial effects.

the usual factorial effect estimators and provide large-sample confidence sets for the average factorial effects.

Rerandomization in factorial experiments have two salient features that differ from rerandomization in treatment-control experiments. First, the factorial effects can have different levels of importance a priori. Many factorial experimental design principles hinge on the belief that main effects are often more important than two-way interactions, and two-way interactions are often more important than higher-order interactions (e.g., Bose (1947), Finney (1943), Wu (2015)). Consequently, we need to impose different stringencies for balancing covariates with respect to factorial effects of different importance. Second, covariates may also vary in importance based on prior knowledge about their associations with the outcome. We establish a general theory that can accommodate rerandomization with tiers of both factorial effects and covariates.

Second, in treatment-control experiments, we are often interested in a single treatment effect. In factorial experiments, however, multiple factorial effects are simultaneously of interest, motivating the asymptotic theory about the joint sampling distribution of the usual factorial effect estimators. In particular, for the joint sampling distribution, we use "central convex unimodality" (Dharmadhikari and Jogdeo (1976), Kanter (1977)) to describe its unimodal property, and "peakedness" (Sherman (1955)) to quantify the intuition that it is more "concentrated" at the true factorial effects under rerandomization than the CRFE. These two mathematical notions for multivariate distributions extend unimodality and narrower quantile ranges for univariate distributions (Li, Ding and Rubin (2018)), and they are also crucial for constructing large-sample confidence sets for factorial effects.

In sum, our asymptotic analysis further demonstrates the benefits of rerandomization in factorial experiments compared to the classical CRFE (Branson, Dasgupta and Rubin (2016)). The proposed large-sample confidence sets for factorial effects will facilitate the practical use of rerandomization in factorial experiments and the associated repeated sampling inference.

The paper proceeds as follows. Section 2 introduces the notation. Section 3 discusses sampling properties and linear projections under the CRFE. Section 4 studies rerandomization using the Mahalanobis distance criterion. Section 5 studies rerandomization with tiers of factorial effects. Section 6 contains an application to an education dataset. Section 7 concludes with possible extensions. The online Supplementary Material (Li, Ding and Rubin (2019)) contains all technical details.

2. Notation for a 2^K factorial experiment.

2.1. *Potential outcomes and causal estimands.* Consider a factorial experiment with n units and K treatment factors, where each factor has two levels, -1 and $+1$. In total there are $Q = 2^K$ treatment combinations, and for each treatment combination $1 \le q \le Q$, let $\boldsymbol{\iota}(q) = (\iota_1(q), \iota_2(q), \dots, \iota_K(q)) \in \{-1, +1\}^K$ be the levels of the K factors. We use potential outcomes to define causal effects in factorial experiments (Dasgupta, Pillai and Rubin (2015), Splawa-Neyman (1923), Branson, Dasgupta and Rubin (2016)). For unit i, let $Y_i(q)$ be the potential outcome under treatment combination q, and $\boldsymbol{Y}_i = (Y_i(1), Y_i(2), \dots, Y_i(Q))$ be the Q dimensional row vector of all potential outcomes. Let $\bar{Y}(q) = \sum_{i=1}^n Y_i(q)/n$ be the average potential outcome under treatment combination q, and $\bar{\boldsymbol{Y}} = (\bar{Y}(1), \bar{Y}(2), \dots, \bar{Y}(Q))$ be the Q dimensional row vector of all average potential outcomes. Dasgupta, Pillai and Rubin (2015) characterized each factorial effect by a Q dimensional column vector with half of its elements being -1 and the other half being $+1$. For example, the average main effect of factor k is

$$\tau_k = \frac{2}{Q}\sum_{q=1}^{Q} 1\{\iota_k(q) = 1\}\bar{Y}(q) - \frac{2}{Q}\sum_{q=1}^{Q} 1\{\iota_k(q) = -1\}\bar{Y}(q)$$

$$= \frac{1}{2^{K-1}} \sum_{q=1}^{Q} \iota_k(q)\bar{Y}(q) = \frac{1}{2^{K-1}} \bar{\boldsymbol{Y}} \boldsymbol{g}_k \quad (1 \le k \le K),$$

where $\boldsymbol{g}_k = (g_{k1}, \dots, g_{kQ})' = (\iota_k(1), \iota_k(2), \dots, \iota_k(Q))'$ is called the *generating vector* for the main effect of factor k. For an interaction effect among several factors, the $\boldsymbol{g}$-vector is an elementwise multiplication of the $\boldsymbol{g}$-vectors for the main effects of the corresponding factors. There are in total $F = 2^K - 1 = Q - 1$ factorial effects. Let $\boldsymbol{g}_f = (g_{f1}, \dots, g_{fQ})' \in \{-1, +1\}^Q$ be the generating vector for the fth factorial effect $(1 \le f \le F)$. For unit i, $\tau_{if} = 2^{-(K-1)} \boldsymbol{Y}_i \boldsymbol{g}_f$ is the fth individual factorial effect, and $\boldsymbol{\tau}_i = (\tau_{i1}, \dots, \tau_{iF})'$ is the F dimensional column vector of all individual factorial effects. Let $\tau_f = 2^{-(K-1)} \bar{\boldsymbol{Y}} \boldsymbol{g}_f$ be the fth average factorial effect, and $\boldsymbol{\tau} = (\tau_1, \dots, \tau_F)'$ be the F dimensional column vector of all average factorial effects. The definitions of the factorial effects imply $\boldsymbol{\tau}_i = 2^{-(K-1)} \sum_{q=1}^{Q} \boldsymbol{b}_q Y_i(q)$ and $\boldsymbol{\tau} = 2^{-(K-1)} \sum_{q=1}^{Q} \boldsymbol{b}_q \bar{Y}(q)$, with coefficient vectors

$$\boldsymbol{b}_1 = \begin{pmatrix} g_{11} \\ g_{21} \\ \vdots \\ g_{F1} \end{pmatrix}, \qquad \boldsymbol{b}_2 = \begin{pmatrix} g_{12} \\ g_{22} \\ \vdots \\ g_{F2} \end{pmatrix}, \qquad \dots, \qquad \boldsymbol{b}_Q = \begin{pmatrix} g_{1Q} \\ g_{2Q} \\ \vdots \\ g_{FQ} \end{pmatrix}. \tag{2.1}$$

Intuitively, the kth main effect compares the average potential outcomes when factor k is at $+1$ and -1 levels, and the interaction effect among two factors compares the average potential outcomes when both factors are at the same level and different levels. We can view a higher order interaction as the difference between two conditional lower order interactions. For example, the interaction among factors 1–3 equals the difference between the interactions of factors 1 and 2 given factor 3 at $+1$ and -1 levels. See Dasgupta, Pillai and Rubin (2015) for more details. Below we use an example to illustrate the definitions.

EXAMPLE 1. We consider factorial experiments with $K = 3$ factors, and use $(1, 2, 3)$ to denote these three factors. Table 1 shows the definitions of the $\boldsymbol{g}_f$'s and the $\boldsymbol{b}_q$'s. Specifically, the first three columns $(\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3)$ represent the levels of the three factors in all treatment combinations, and they generate the main effects of factors $(1, 2, 3)$. The remaining columns $(\boldsymbol{g}_4, \dots, \boldsymbol{g}_7)$ are the elementwise multiplications of subsets of $(\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3)$ that generate the interaction effects. The coefficient vector $\boldsymbol{b}_q$ consists of all the elements in the qth row of Table 1.

TABLE 1
$\boldsymbol{g}_f$'s and $\boldsymbol{b}_q$'s for 2^3 factorial experiments

1	2	3	12	13	23	123	
−1	−1	−1	+1	+1	+1	−1	$\boldsymbol{b}_1'$
−1	−1	+1	+1	−1	−1	+1	$\boldsymbol{b}_2'$
−1	+1	−1	−1	+1	−1	+1	$\boldsymbol{b}_3'$
−1	+1	+1	−1	−1	+1	−1	$\boldsymbol{b}_4'$
+1	−1	−1	−1	−1	+1	+1	$\boldsymbol{b}_5'$
+1	−1	+1	−1	+1	−1	−1	$\boldsymbol{b}_6'$
+1	+1	−1	+1	−1	−1	−1	$\boldsymbol{b}_7'$
+1	+1	+1	+1	+1	+1	+1	$\boldsymbol{b}_8'$
$\boldsymbol{g}_1$	$\boldsymbol{g}_2$	$\boldsymbol{g}_3$	$\boldsymbol{g}_4$	$\boldsymbol{g}_5$	$\boldsymbol{g}_6$	$\boldsymbol{g}_7$	

2.2. *Treatment assignment, covariate imbalance and rerandomization.* For each unit i, $\boldsymbol{x}_i$ represents the L dimensional column vector of pretreatment covariates. For instance, in the education example in Section 6, college freshmen receive different academic services and incentives after entering the university, and their pretreatment covariates include high school grade point average, gender, age and etc. Let Z_i be the treatment assignment, where $Z_i = q$ if unit i receives treatment combination q. Let n_q be the number of units under treatment combination q, and $\boldsymbol{Z} = (Z_1, \dots, Z_n)$ be the treatment assignment vector for all units. In the CRFE, the probability that $\boldsymbol{Z}$ takes a particular value $z = (z_1, \dots, z_n)$ is $n_1! \cdots n_Q!/n!$, where $\sum_{i=1}^n 1\{z_i = q\} = n_q$ for all q. Let $\bar{\boldsymbol{x}} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i$ be the finite population covariate mean vector; for $1 \le q \le Q$, let $\hat{\bar{\boldsymbol{x}}}(q) = n_q^{-1}\sum_{i:z_i=q} \boldsymbol{x}_i$ be the covariate mean vector for units that receive treatment combination q. For $1 \le f \le F$, the L dimensional difference-in-means vector of covariates with respect to the fth factorial effect is

$$\hat{\boldsymbol{\tau}}_{\boldsymbol{x},f} = \frac{2}{Q}\sum_{q=1}^{Q} g_{fq}\hat{\bar{\boldsymbol{x}}}(q) = \frac{1}{2^{K-1}}\sum_{q:g_{fq}=1}\hat{\bar{\boldsymbol{x}}}(q) - \frac{1}{2^{K-1}}\sum_{q:g_{fq}=-1}\hat{\bar{\boldsymbol{x}}}(q). \tag{2.2}$$

Let $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}} = (\hat{\boldsymbol{\tau}}'_{\boldsymbol{x},1}, \dots, \hat{\boldsymbol{\tau}}'_{\boldsymbol{x},F})'$ be the LF dimensional column vector of the difference-in-means of covariates with respect to all factorial effects. Although $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ has mean zero under the CRFE, for a realized value of $\boldsymbol{Z}$, covariate distributions are often imbalanced among different treatment combinations. For example, we consider a CRFE with $K = 2$ factors, $L = 4$ uncorrelated covariates and equal treatment group sizes $n_q = n/Q$. In this case, with asymptotic probability $1 - (1 - 5\%)^{4(2^2-1)} \approx 46.0\%$, at least one of the difference-in-means in (2.2) with respect to a covariate and a factorial effect standardized by its standard deviation is larger than 1.96, the 0.975-quantile of $\mathcal{N}(0, 1)$. This holds due to the asymptotic Gaussianity of $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ with zero mean and diagonal covariance matrix, implied by Proposition 1 discussed shortly.

Rerandomization is a design to prevent undesirable treatment allocations. When covariate imbalance occurs for a realized randomization under a certain criterion, we discard this unlucky realization and rerandomize the treatment assignment until this criterion is satisfied. Generally, rerandomization proceeds as follows (Morgan and Rubin (2012)): first, we collect covariate data and specify a covariate balance criterion; second, we continue randomizing the units into different treatment groups until the balance criterion is satisfied; third, we conduct the physical experiment using the accepted randomization. A major goal of this paper is to discuss the statistical analysis of the data from a rerandomized factorial experiment.

There are three additional issues on covariates. First, covariates are attributes of the units that are fixed before the experiment. Second, the covariates can be general (discrete or continuous). We can use binary indicators to represent discrete covariates. Third, the covariates can include transformations of the basic covariates and their interactions. This enables us to balance the marginal and joint distributions of the basic covariates. See Baldi Antognini and Zagoraiou (2011) and Li, Ding and Rubin (2018) for a related discussion in the treatment-control experiment.

2.3. *Additional notation.* To facilitate the discussion, for a positive semidefinite matrix $\boldsymbol{A} \in \mathbb{R}^{m\times m}$ with rank p_0, and a positive integer $p \ge p_0$, we use $\boldsymbol{A}_p^{1/2} \in \mathbb{R}^{m\times p}$ to denote a matrix such that $\boldsymbol{A}_p^{1/2}(\boldsymbol{A}_p^{1/2})' = \boldsymbol{A}$. Specifically, if $\boldsymbol{A} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^2\boldsymbol{\Gamma}'$ is the eigen-decomposition of $\boldsymbol{A}$ where $\boldsymbol{\Gamma} \in \mathbb{R}^{m\times p_0}$, $\boldsymbol{\Gamma}'\boldsymbol{\Gamma} = \boldsymbol{I}_{p_0}$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_{p_0})$, then we can choose $\boldsymbol{A}_p^{1/2} = (\boldsymbol{\Gamma}\boldsymbol{\Lambda}, \boldsymbol{0}_{m\times(p-p_0)})$. The choice of $\boldsymbol{A}_p^{1/2}$ is generally not unique. In the special case with $p = m$, we use $\boldsymbol{A}^{1/2}$ to denote the unique positive-semidefinite matrix satisfying the definition of $\boldsymbol{A}_m^{1/2}$. We use $\otimes$ for the Kronecker product of two matrices, and $\circ$ for elementwise multiplications of vectors. We say a matrix $\boldsymbol{M}_1$ is smaller than or equal to $\boldsymbol{M}_2$ and write as $\boldsymbol{M}_1 \le \boldsymbol{M}_2$, if $\boldsymbol{M}_2 - \boldsymbol{M}_1$ is positive semidefinite. We say a random vector $\boldsymbol{\phi}$ (or its distribution) is symmetric, if $\boldsymbol{\phi}$ and $-\boldsymbol{\phi}$ have the same distribution. We say a random vector is spherically symmetric, if its distribution is invariant under orthogonal transformations. In the asymptotic analysis, we use $\dot\sim$ for two sequences of random vectors converging weakly to the same distribution, after scaling by $\sqrt{n}$.

3. 2^K completely randomized factorial experiments. The sampling distributions of factorial effect estimators under rerandomization are the same as their conditional distributions under the CRFE given that the treatment assignment vector satisfies the balance criterion. Therefore, we first study the joint sampling distribution of the difference-in-means of the outcomes and covariates under the CRFE. It depends on the finite population variances and covariances: $S_{qq} = (n-1)^{-1}\sum_{i=1}^n \{Y_i(q) - \bar{Y}(q)\}^2$ and $S_{qk} = (n-1)^{-1}\sum_{i=1}^n \{Y_i(q) - \bar{Y}(q)\}\{Y_i(k) - \bar{Y}(k)\}$ for potential outcomes, $\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}} = (n-1)^{-1}\sum_{i=1}^n (\boldsymbol{\tau}_i - \boldsymbol{\tau})(\boldsymbol{\tau}_i - \boldsymbol{\tau})'$ for factorial effects, $\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{x}} = (n-1)^{-1}\sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$ for covariates and $\boldsymbol{S}_{q,\boldsymbol{x}} = \boldsymbol{S}_{\boldsymbol{x},q}' = (n-1)^{-1}\sum_{i=1}^n \{Y_i(q) - \bar{Y}(q)\}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$ for potential outcomes and covariates. The covariance $\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{x}}$ is known without any uncertainty. However, other variances or covariances (e.g., S_{qk}, $\boldsymbol{S}_{\tau\tau}$ and $\boldsymbol{S}_{q,\boldsymbol{x}}$) involve potential outcomes or individual factorial effects and are thus unknown.

3.1. *Asymptotic sampling distribution under the* CRFE. Let $Y_i^{\text{obs}} = \sum_{q=1}^Q 1\{Z_i = q\}Y_i(q)$ be the observed outcome of unit i, and $\hat{\bar{Y}}(q) = n_q^{-1}\sum_{i:Z_i=q} Y_i^{\text{obs}}$ be the average observed outcome under treatment combination q. For $1 \le f \le F$, the difference-in-means estimator for the fth average factorial effect is

$$\hat{\tau}_f = \frac{2}{Q}\sum_{q=1}^Q g_{fq}\hat{\bar{Y}}(q) = \frac{1}{2^{K-1}}\sum_{q:g_{fq}=1}\hat{\bar{Y}}(q) - \frac{1}{2^{K-1}}\sum_{q:g_{fq}=-1}\hat{\bar{Y}}(q).$$

Let $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \ldots, \hat{\tau}_F)'$ be the F dimensional column vector consisting of all factorial effect estimators.

In the finite population inference, the covariates and potential outcomes are all fixed, and the only random component is the treatment vector $\boldsymbol{Z}$. In the asymptotic analysis, we further embed the finite population into a sequence with increasing sizes, and introduce the following regularity conditions.

CONDITION 1. *As $n \to \infty$, the sequence of finite populations satisfies that for each $1 \le q \ne k \le Q$:*

(i) *the proportion of units under treatment combination q, n_q/n, has a positive limit,*
(ii) *the finite population variance and covariances S_{qq}, S_{qk}, $\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{x}}$ and $\boldsymbol{S}_{q,\boldsymbol{x}}$ have limiting values, and $\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{x}}$ and its limit are nondegenerate,*
(iii) $\max_{1\le i\le n}|Y_i(q) - \bar{Y}(q)|^2/n \to 0$ *and* $\max_{1\le i\le n}\|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|_2^2/n \to 0$.

PROPOSITION 1. *Under the* CRFE, $(\hat{\boldsymbol{\tau}}' - \boldsymbol{\tau}', \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}')'$ *has mean zero and sampling covariance matrix*

$$\boldsymbol{V} \equiv 2^{-2(K-1)}\sum_{q=1}^Q n_q^{-1}\begin{pmatrix} \boldsymbol{b}_q\boldsymbol{b}_q' S_{qq} & (\boldsymbol{b}_q\boldsymbol{b}_q') \otimes \boldsymbol{S}_{q,\boldsymbol{x}} \\ (\boldsymbol{b}_q\boldsymbol{b}_q') \otimes \boldsymbol{S}_{\boldsymbol{x},q} & (\boldsymbol{b}_q\boldsymbol{b}_q') \otimes \boldsymbol{S}_{\boldsymbol{x}\boldsymbol{x}} \end{pmatrix} - n^{-1}\begin{pmatrix} \boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$$
$$\equiv \begin{pmatrix} \boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}} & \boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{x}} \\ \boldsymbol{V}_{\boldsymbol{x}\boldsymbol{\tau}} & \boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}} \end{pmatrix}.$$

Under the CRFE *and Condition* 1, $(\hat{\boldsymbol{\tau}}' - \boldsymbol{\tau}', \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}')' \;\dot\sim\; \mathcal{N}(\boldsymbol{0}, \boldsymbol{V})$.

Proposition 1 follows from a finite population central limit theorem (Li and Ding (2017), Theorems 3 and 5), with the proof in Appendix A2 of the Supplementary Material (Li, Ding and Rubin (2019)). Proposition 1 immediately gives the sampling properties of any single factorial effect estimator. Let $S_{\tau_f\tau_f}$ be the fth diagonal element of $\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}}$, and $V_{\tau_f\tau_f} = 2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1} S_{qq} - n^{-1} S_{\tau_f\tau_f}$ be the fth diagonal element of $\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}$. Then $\hat{\tau}_f$ is unbiased for τ_f with sampling variance $V_{\tau_f\tau_f}$, and $\hat{\tau}_f - \tau_f \dot{\sim} \mathcal{N}(0, V_{\tau_f\tau_f})$. Moreover, $\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}}$ cannot be unbiasedly estimated from the observed data, and it equals $\boldsymbol{0}$ under the *additivity* defined below. Under the additivity, the individual treatment effect does not depend on covariates, that is, there is no treatment-covariate interaction.

DEFINITION 1. The factorial effects are additive if and only if the individual factorial effect $\boldsymbol{\tau}_i$ is a constant vector for all units, or, equivalently, $\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}} = \boldsymbol{0}$.

Under the CRFE, the observed sample variance $s_{qq} = (n_q - 1)^{-1}\sum_{i:Z_i=q}\{Y_i^{\text{obs}} - \hat{\bar{Y}}(q)\}^2$ is unbiased for S_{qq}, because the units receiving treatment combination q are from a simple random sample of size n_q. Similar to Splawa-Neyman (1923), we can conservatively estimate $\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}$ by $2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1}\boldsymbol{b}_q\boldsymbol{b}_q' s_{qq}$, and then construct Wald-type confidence sets for $\boldsymbol{\tau}$. Both the sampling covariance estimator and confidence sets are asymptotically conservative unless the additivity in Definition 1 holds. It is then straightforward to construct confidence sets for any linear transformations of $\boldsymbol{\tau}$.

3.2. *Linear projections.* First, we decompose the potential outcomes. Let $Y_i^{\parallel}(q) = \bar{Y}(q) + \boldsymbol{S}_{q,\boldsymbol{x}}\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{x}}^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})$ be the finite population linear projection of the $Y_i(q)$'s on the $\boldsymbol{x}_i$'s, and $Y_i^{\perp}(q) = Y_i(q) - Y_i^{\parallel}(q)$ be the corresponding residual. The finite population linear projection of $\boldsymbol{\tau}_i$ on $\boldsymbol{x}_i$ is then $\boldsymbol{\tau}_i^{\parallel} = 2^{-(K-1)}\sum_{q=1}^{Q}\boldsymbol{b}_q Y_i^{\parallel}(q)$, and the corresponding residual is $\boldsymbol{\tau}_i^{\perp} = 2^{-(K-1)}\sum_{q=1}^{Q}\boldsymbol{b}_q Y_i^{\perp}(q)$. Let $S_{qq}^{\parallel}$, $S_{qq}^{\perp}$, $\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel}$ and $\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\perp}$ be the finite population variances and covariances of $Y^{\parallel}(q)$, $Y^{\perp}(q)$, $\boldsymbol{\tau}^{\parallel}$ and $\boldsymbol{\tau}^{\perp}$, respectively. Define

$$\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel} = 2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1}\boldsymbol{b}_q\boldsymbol{b}_q' \cdot S_{qq}^{\parallel} - n^{-1}\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel},$$

$$\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\perp} = 2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1}\boldsymbol{b}_q\boldsymbol{b}_q' \cdot S_{qq}^{\perp} - n^{-1}\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\perp}$$

as analogues of the sampling covariance $\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}$ in Proposition 1, with the potential outcomes $Y_i(q)$'s replaced by the linear projections $Y_i^{\parallel}(q)$'s and the residuals $Y_i^{\perp}(q)$'s, respectively. We have $\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}} = \boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel} + \boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\perp}$.

Second, we decompose the factorial effect estimator $\hat{\boldsymbol{\tau}}$.

THEOREM 1. *Under the* CRFE, *the linear projection of* $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ *on* $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ *is* $\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{x}}\boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$, *the corresponding residual is* $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} - \boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{x}}\boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ *and they have sampling covariances*:

$$\operatorname{Cov}(\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{x}}\boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel}, \qquad \operatorname{Cov}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} - \boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{x}}\boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\perp},$$
$$\operatorname{Cov}(\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{x}}\boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}, \hat{\boldsymbol{\tau}} - \boldsymbol{\tau} - \boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{x}}\boldsymbol{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1}\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \boldsymbol{0}.$$

Theorem 1 follows from Proposition 1 and some matrix calculations, with the proof in Appendix A2 of the Supplementary Material (Li, Ding and Rubin (2019)). Let $V_{\tau_f\tau_f}^{\parallel}$ and $S_{\tau_f\tau_f}^{\parallel}$ be the fth diagonal elements of $\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel}$ and $\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel}$, respectively. The multiple correlation in the following corollary will play an important role in the asymptotic sampling distribution of $\hat{\tau}_f$ under rerandomization. We summarize its equivalent forms below.

COROLLARY 1. *Under the* CRFE, *the sampling squared multiple correlation between $\hat{\tau}_f$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ has the following equivalent forms*:

$$R_f^2 = \mathrm{Corr}^2(\hat{\tau}_f, \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}) = \frac{V_{\tau_f\tau_f}^{\parallel}}{V_{\tau_f\tau_f}} = \frac{2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1} S_{qq}^{\parallel} - n^{-1} S_{\tau_f\tau_f}^{\parallel}}{2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1} S_{qq} - n^{-1} S_{\tau_f\tau_f}}.$$

It reduces to $R_f^2 = S_{11}^{\parallel}/S_{11}$, the finite population squared multiple correlation between $Y(1)$ and $\boldsymbol{x}$ under the additivity in Definition 1.

The proof of Corollary 1 is in Appendix A2 of the Supplementary Material (Li, Ding and Rubin (2019)).

4. Rerandomization using the Mahalanobis distance. As shown in Section 3.1, although $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ has mean $\mathbf{0}$, its realized value can be very different from $\mathbf{0}$ for a particular treatment allocation. Rerandomization can avoid this drawback. In the design stage, we can force balance of the covariate means by ensuring $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ to be "small."

4.1. *Mahalanobis distance criterion.* A measure of the magnitude of $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ is the Mahalanobis distance $M \equiv \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}' \boldsymbol{V}_{\boldsymbol{xx}}^{-1} \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$. We further let a be a positive constant predetermined in the design stage. Using M as the balance criterion, we accept a treatment assignment vector $\boldsymbol{Z}$ from the CRFE if and only if $M \le a$. Below we use ReFM to denote 2^K rerandomized factorial experiments using M as the criterion, and $\mathcal{M}$ to denote the event that the treatment vector $\boldsymbol{Z}$ satisfies this criterion. From Proposition 1, M is asymptotically χ_{LF}^2, and therefore the asymptotic acceptance probability is $p_a = P(\chi_{LF}^2 \le a)$ under ReFM. In practice, we usually choose a small threshold a, or equivalently a small p_a, for example, $p_a = 0.001$. However, we do not advocate choosing p_a to be too small, because an extremely small p_a may lead to too few configurations of treatment allocations in ReFM.

4.2. *Asymptotic sampling distribution of $\hat{\boldsymbol{\tau}}$ under* ReFM. Rerandomization in the design stage accepts only the treatment assignments resulting in covariate balance, which consequently changes the sampling distribution of $\hat{\boldsymbol{\tau}}$. Understanding the asymptotic sampling distribution of $\hat{\boldsymbol{\tau}}$ is crucial for conducting the classical repeated sampling inference of $\boldsymbol{\tau}$. Intuitively, $\hat{\boldsymbol{\tau}}$ has two parts: one part is orthogonal to $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ and thus unaffected by ReFM, and the other part is the linear projection onto $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ and thus affected by ReFM. Let $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_F)$ be an F dimensional standard Gaussian random vector, and $\boldsymbol{\zeta}_{LF,a} \sim \boldsymbol{D} \mid \boldsymbol{D}'\boldsymbol{D} \le a$ be an LF dimensional truncated Gaussian random vector, where $\boldsymbol{D} = (D_1, \dots, D_{LF})' \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{LF})$. In addition, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta}_{LF,a}$ are independent. The following theorem shows the asymptotic sampling distribution of $\hat{\boldsymbol{\tau}}$.

THEOREM 2. *Under* ReFM *and Condition* 1,

$$\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} \mid \mathcal{M} \mathrel{\dot\sim} (\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\perp})^{1/2} \boldsymbol{\varepsilon} + (\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\parallel})_{LF}^{1/2} \boldsymbol{\zeta}_{LF,a}. \tag{4.1}$$

Theorem 2 holds because the sampling distribution of $\hat{\boldsymbol{\tau}}$ under rerandomization is the same as the conditional distribution of $\hat{\boldsymbol{\tau}}$ given $M \le a$. Its proof is in Appendix A3 of the Supplementary Material (Li, Ding and Rubin (2019)). We emphasize that, although the matrix $(\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\parallel})_{LF}^{1/2}$ may not be unique, the asymptotic sampling distribution (4.1) is. Therefore, the asymptotic sampling distribution of $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ under ReFM depends only on L, F, a, $\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\perp}$ and $\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\parallel}$. Theorem 2 immediately implies the asymptotic sampling distribution of a single factorial effect estimator. Let $\varepsilon_0 \sim \mathcal{N}(0, 1)$, $\eta_{LF,a} \sim D_1 \mid \boldsymbol{D}'\boldsymbol{D} \le a$ be the first coordinate of $\boldsymbol{\zeta}_{LF,a}$, and ε and $\eta_{LF,a}$ be independent.

COROLLARY 2. *Under* ReFM *and Condition* 1, *for* $1 \le f \le F$,

$$\hat{\tau}_f - \tau_f \mid \mathcal{M} \dot{\sim} \sqrt{V_{\tau_f \tau_f}} \Big(\sqrt{1 - R_f^2} \cdot \varepsilon_0 + \sqrt{R_f^2} \cdot \eta_{LF,a} \Big). \tag{4.2}$$

The proof of Corollary 2 is in Appendix A3 of the Supplementary Material (Li, Ding and Rubin (2019)). The marginal asymptotic sampling distribution (4.2) under ReFM has the same form as that under rerandomized treatment-control experiments using the Mahalanobis distance (Li, Ding and Rubin (2018)).

4.3. *Review of the central convex unimodality.* In this subsection, we review a generalization of unimodality to multivariate distributions and apply it to study the asymptotic sampling distribution (4.1). This property will be important for constructing conservative large-sample confidence sets later.

Although the definition of symmetric unimodality for univariate distribution is simple and intuitive, it is nontrivial to generalize it to multivariate distribution. Here we adopt the *central convex unimodality* proposed by Dharmadhikari and Jogdeo (1976) based on the results of Sherman (1955), which is also equivalent to the symmetric unimodality in Kanter (1977). For a set $\mathcal{B}$ of distributions on $\mathbb{R}^m$, we say that $\mathcal{B}$ is *closed convex* if it satisfies two conditions: (i) for any distributions $\nu_1, \nu_2 \in \mathcal{B}$ and for any $\lambda \in (0, 1)$, the distribution $(1-\lambda)\nu_1 + \lambda\nu_2$ is in $\mathcal{B}$, and (ii) a distribution ν is in $\mathcal{B}$ if there exists a sequence of distributions in $\mathcal{B}$ converging weakly to ν. For any set $\mathcal{C}$ of distributions, let the *closed convex hull* of $\mathcal{C}$ be the smallest closed convex set containing $\mathcal{C}$. A compact convex set in Euclidean space $\mathbb{R}^m$ is called a *convex body* if it has a nonempty interior. A set $\mathcal{K} \subset \mathbb{R}^m$ is *symmetric* if $\mathcal{K} = \{-\boldsymbol{a} : \boldsymbol{a} \in \mathcal{K}\}$. Below we introduce the definition.

DEFINITION 2. A distribution on $\mathbb{R}^m$ is *central convex unimodal* if it is in the closed convex hull of $\mathcal{U}$, where $\mathcal{U}$ is the set of all uniform distributions on symmetric convex bodies in $\mathbb{R}^m$.

The class of central convex unimodal distributions is closed under convolution, marginality, product measure and weak convergence (Kanter (1977)). A sufficient condition for the central convex unimodality is having a log-concave probability density function (Kanter (1977), Dharmadhikari and Joag-Dev (1988)). The following proposition states the central convex unimodality of the asymptotic sampling distribution of $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ under ReFM.

PROPOSITION 2. *The standard Gaussian random vector* $\boldsymbol{\varepsilon}$, *the truncated Gaussian random vector* $\boldsymbol{\zeta}_{LF,a}$ *and the asymptotic sampling distribution* (4.1) *are all central convex unimodal.*

Proposition 2 follows from the log-concavity of the densities of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta}_{LF,a}$ and the closedness of the class of central convex unimodal distributions under linear transformation and convolution. Its proof is in Appendix A3 of the Supplementary Material (Li, Ding and Rubin (2019)).

4.4. *Representation for the asymptotic sampling distribution of* $\hat{\boldsymbol{\tau}}$. In this subsection, we further represent (4.1) using well-known distributions to gain more insights. Let $\chi^2_{LF,a} \sim \chi^2_{LF} \mid \chi^2_{LF} \le a$ be a truncated χ^2 random variable, $\boldsymbol{S}$ be an LF dimensional random vector whose coordinates are independent random signs with probability $1/2$ of being ± 1 and $\boldsymbol{\beta}$ be an LF dimensional Dirichlet random vector with parameters $(1/2, \dots, 1/2)$. Let $\sqrt{\boldsymbol{\beta}}$ be the elementwise square root of the vector $\boldsymbol{\beta}$, and $v_{LF,a} = P(\chi^2_{LF+2} \le a)/P(\chi^2_{LF} \le a) \le 1$.

PROPOSITION 3. *$\boldsymbol{\zeta}_{LF,a}$ is spherically symmetric with covariance $v_{LF,a}\boldsymbol{I}_{LF}$. It follows $\boldsymbol{\zeta}_{LF,a} \sim \chi_{LF,a} \cdot \boldsymbol{S} \circ \sqrt{\boldsymbol{\beta}}$, where $(\chi_{LF,a}, \boldsymbol{S}, \boldsymbol{\beta})$ are jointly independent.*

Proposition 3 follows from the spherical symmetry of the standard multivariate Gaussian random vector, with the proof in Appendix A3 of the Supplementary Material (Li, Ding and Rubin (2019)). Proposition 3 allows for easy simulations of the asymptotic sampling distribution (4.1), which is useful for the repeated sampling inference discussed shortly. For simplicity, in the remaining paper, we assume that $\boldsymbol{V}_{\boldsymbol{\tau\tau}}$ is invertible whenever we mention its inverse; otherwise we can focus on a lower dimensional linear transformation of $\hat{\boldsymbol{\tau}}$ (Li, Ding and Rubin (2019)). Let $\boldsymbol{R} = \boldsymbol{V}_{\boldsymbol{\tau\tau}}^{-1/2}\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\parallel}\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{-1/2}$ be the matrix measuring the relative sampling covariance of $\hat{\boldsymbol{\tau}}$ explained by $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$, and $\boldsymbol{R} = \boldsymbol{\Gamma}\boldsymbol{\Pi}^2\boldsymbol{\Gamma}'$ be its eigen-decomposition, where $\boldsymbol{\Gamma} \in \mathbb{R}^{F\times F}$ is an orthogonal matrix and $\boldsymbol{\Pi}^2 = \operatorname{diag}(\pi_1^2, \dots, \pi_F^2) \in \mathbb{R}^{F\times F}$ is a diagonal matrix with nonnegative elements. The eigenvalues $(\pi_1^2, \dots, \pi_F^2)$ are the canonical correlations between the sampling distributions of $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ under the CRFE, which measure the association between the potential outcomes and covariates. Under the additivity in Definition 1, $\pi_1^2 = \cdots = \pi_F^2 = S_{11}^{\parallel}/S_{11}$. The following corollary gives an equivalent form of (4.1) highlighting the dependence on the canonical correlations $(\pi_1^2, \dots, \pi_F^2)$.

COROLLARY 3. *Under* ReFM *and Condition* 1, (4.1) *is equivalent to*

$$\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} \mid \mathcal{M} \dot\sim \boldsymbol{V}_{\boldsymbol{\tau\tau}}^{1/2}\boldsymbol{\Gamma}\{(\boldsymbol{I}_F - \boldsymbol{\Pi}^2)^{1/2}\boldsymbol{\varepsilon} + (\boldsymbol{\Pi}, \boldsymbol{0}_{F\times(L-1)F})\boldsymbol{\zeta}_{LF,a}\}. \tag{4.3}$$

The proof of Corollary 3 is in Appendix A3 of the Supplementary Material (Li, Ding and Rubin (2019)). The second term in (4.3), affected by rerandomization, depends on the canonical correlations $(\pi_1^2, \dots, \pi_F^2)$ and the asymptotic acceptance probability p_a of ReFM. Below we use a numerical example to illustrate such dependence.

EXAMPLE 2. We consider the case with $L = 1$, $K = 2$ and $F = 3$, and focus on the standardized distribution $(\boldsymbol{I}_3 - \boldsymbol{\Pi}^2)^{1/2}\boldsymbol{\varepsilon} + \boldsymbol{\Pi}\boldsymbol{\zeta}_{3,a}$, which depends on $\boldsymbol{\Pi}^2 = \operatorname{diag}(\pi_1^2, \pi_2^2, \pi_3^2)$ and $p_a = P(\chi^2_3 \le a)$. First, we fix $(\pi_2^2, \pi_3^2, p_a) = (0.5, 0.5, 0.001)$. Figure 1(a) shows the density of the first two coordinates of $\boldsymbol{\zeta}_{3,a}$ for different π_1^2. As π_1^2 increases, the density becomes more concentrated around zero, showing that the stronger the association is between the potential outcomes and covariates, the more precise the factorial effect estimators are.

Second, we fix $(\pi_1^2, \pi_2^2, \pi_3^2) = (0.5, 0.5, 0.5)$. Figure 1(b) shows the density of the first two coordinates of $\boldsymbol{\zeta}_{3,a}$ for different p_a. As the asymptotic acceptance probability p_a decreases, the density becomes more concentrated around zero, confirming the intuition that a smaller asymptotic acceptance probability gives us more precise factorial effect estimators. Note that the first $\boldsymbol{\varepsilon}$ component in the asymptotic sampling distribution (4.3) does not depend on p_a and is usually nonzero. For example, when $\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\perp}$ is positive definite, $\boldsymbol{I}_F - \boldsymbol{R} = \boldsymbol{V}_{\boldsymbol{\tau\tau}}^{-1/2}\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\perp}\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{-1/2}$ is positive definite, as well as the coefficient of $\boldsymbol{\varepsilon}$ in (4.3). Therefore, the gain of ReFM by decreasing p_a usually becomes smaller as p_a decreases.

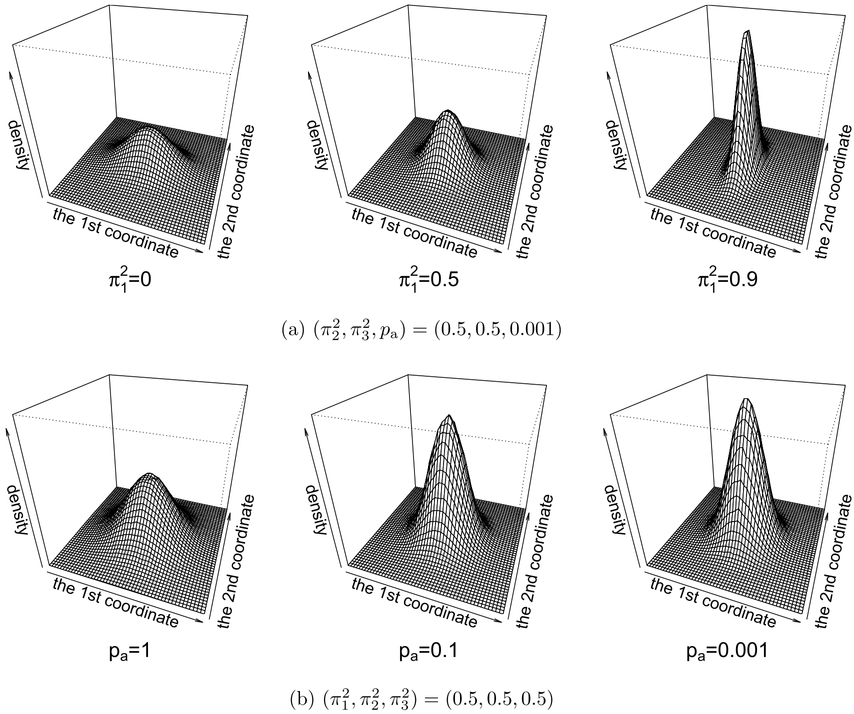

FIG. 1. *Joint density of the first two coordinates of* $(\boldsymbol{I}_3 - \boldsymbol{\Pi}^2)^{1/2}\boldsymbol{\varepsilon} + \boldsymbol{\Pi}\boldsymbol{\zeta}_{3,a}$.

4.5. *Asymptotic unbiasedness, sampling covariance and peakedness.* In this subsection, we further study the asymptotic properties of $\hat{\boldsymbol{\tau}}$ under ReFM. First, the factorial effects estimator $\hat{\boldsymbol{\tau}}$ is consistent for $\boldsymbol{\tau}$. Because covariates are potential outcomes unaffected by the treatment, the difference-in-means of any observed or unobserved covariate with respect to any factorial effect has asymptotic mean zero.

Second, we compare the asymptotic sampling covariance matrices of $\hat{\boldsymbol{\tau}}$ under ReFM and the CRFE, which also gives the reduction in asymptotic sampling covariances of difference-in-means of covariates as a special case.

THEOREM 3. *Under Condition 1, the asymptotic sampling covariance matrix of* $\hat{\boldsymbol{\tau}}$ *under* ReFM *is smaller than or equal to that under the* CRFE, *and the reduction in asymptotic sampling covariance is* $(1 - v_{LF,a})n\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel}$. *Specifically, the percentage reduction in asymptotic sampling variance* (PRIASV) *of* $\hat{\tau}_f$ *is* $(1 - v_{LF,a})R_f^2$.

Theorem 3 follows from Theorem 2 and Proposition 3, with the proof in Appendix A4 of the Supplementary Material (Li, Ding and Rubin (2019)). Rigorously, the reductions in Theorem 3 should be $(1 - v_{LF,a})\lim_{n\to\infty}(n\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel})$ and $(1 - v_{LF,a})\lim_{n\to\infty} R_f^2$. However, for descriptive simplicity, we omit the limit signs. From Theorem 3, the larger the squared multiple correlation R_f^2 is, the more PRIASV of the factorial effect estimator is through ReFM. When a is close to zero, or equivalently the asymptotic acceptance probability p_a is small, the asymptotic sampling variance of $\hat{\tau}_f$ reduces to $V_{\tau_f\tau_f}(1 - R_f^2)$, which is identical to the asymptotic sampling variance of the regression adjusted estimator under the CRFE discussed in Lu (2016).

Third, we compare the peakedness of the asymptotic sampling distributions of $\hat{\boldsymbol{\tau}}$ under ReFM and the CRFE, because of its close connection to the volumes of confidence sets for $\boldsymbol{\tau}$. Birnbaum (1948), Bickel and Lehmann (1976) and Shaked (1985) proposed some measures of dispersion for univariate distributions. Sherman (1955) and Giovagnoli and Wynn (1995) generalized them to multivariate distributions. Marshall, Olkin and Arnold (2009) discussed some related properties. Here we use the definition in Sherman (1955).

DEFINITION 3. For two symmetric random vectors $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ in $\mathbb{R}^m$, we say that $\boldsymbol{\phi}$ is more peaked than $\boldsymbol{\psi}$ and write as $\boldsymbol{\phi} \succ \boldsymbol{\psi}$, if $P(\boldsymbol{\phi} \in \mathcal{K}) \geq P(\boldsymbol{\psi} \in \mathcal{K})$ for every symmetric convex set $\mathcal{K} \subset \mathbb{R}^m$.

From Definition 3, intuitively, the more peaked a random vector is, the more "concentrated" around zero it is. Therefore, when comparing two experimental designs, the one with more peaked sampling distribution of the causal estimator gives more precise estimate for the true causal effect. That is, peakedness measures the efficiencies of the designs.

As a basic fact, the ordering of peakedness directly implies the ordering of the covariance matrices.

PROPOSITION 4. *For two symmetric random vectors $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ in $\mathbb{R}^m$ with finite second moments, if $\boldsymbol{\phi} \succ \boldsymbol{\psi}$, then* $\mathrm{Cov}(\boldsymbol{\phi}) \leq \mathrm{Cov}(\boldsymbol{\psi})$.

Proposition 4 follows from some algebra, with the proof in Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)). For two Gaussian vectors $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$, $\mathrm{Cov}(\boldsymbol{\phi}) \leq \mathrm{Cov}(\boldsymbol{\psi})$ also implies $\boldsymbol{\phi} \succ \boldsymbol{\psi}$. The reverse of Proposition 4 does not hold for general random vectors. For example, we compare a standard Gaussian random variable ε_0 and a truncated Gaussian random variable $\xi_0 \sim \varepsilon_0 \mid 0.5 \leq \varepsilon_0^2 \leq 1$. Both random variables are symmetric around zero and $\mathrm{Var}(\xi_0) < 1 = \mathrm{Var}(\varepsilon_0)$. However, ξ_0 is not more peaked than ε_0, because $P(|\xi_0| \leq 0.5) = 0 < P(|\varepsilon_0| \leq 0.5)$.

The following theorem shows that the difference-in-means estimator is more "concentrated" under ReFM than under the CRFE.

THEOREM 4. *Under Condition 1, the asymptotic sampling distribution of $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ under* ReFM *is more peaked than that under the* CRFE.

Theorem 4 holds because the truncated Gaussian random vector $\boldsymbol{\zeta}_{LF,a}$ is more peaked than the standard Gaussian random vector. Its proof is in Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)). First, Theorem 4, coupled with Proposition 4, implies the asymptotic sampling covariance of $\hat{\boldsymbol{\tau}}$ is smaller under ReFM than under the CRFE. Second, Theorem 4 shows that asymptotically, $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ has larger probability to be in any symmetric convex set under ReFM than under the CRFE. For a positive definite matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{p\times p}$ and $c \geq 0$, let $\mathcal{O}(\boldsymbol{\Lambda}, c) \equiv \{\boldsymbol{\mu} : \boldsymbol{\mu}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu} \leq c\}$. The following theorem implies that, for the special class of symmetric convex sets, $\{\mathcal{O}(\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}, c) : c \geq 0\}$, the asymptotic probability that $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ lies in $\mathcal{O}(\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}, c)$ is nondecreasing in the canonical correlation π_k^2's.

THEOREM 5. *Under* ReFM, *assume Condition 1. Let $c_{1-\alpha}$ be the solution of* $\lim_{n\to\infty} P\{\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} \in \mathcal{O}(\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}, c_{1-\alpha}) \mid \mathcal{M}\} = 1 - \alpha$ *for any* $\alpha \in (0, 1)$. *It depends only on* (L, K, a) *and the canonical correlation π_k^2's, and is nonincreasing in these canonical correlations for fixed* (L, K, a).

Theorem 5 is a multivariate extension of Theorem 2 of Li, Ding and Rubin (2018), with the proof in Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)). The set $\mathcal{O}(\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}, c_{1-\alpha})$ in Theorem 5 is a $1-\alpha$ asymptotic quantile region of $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ under ReFM. From Theorem 5, with larger canonical correlation π_k^2's, ReFM leads to more percentage reduction in volume of the $1-\alpha$ asymptotic quantile region $\mathcal{O}(\boldsymbol{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}, c_{1-\alpha})$ of $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$.

Moreover, we can establish similar conclusions as Theorems 4 and 5 for any linear transformation of $\hat{\boldsymbol{\tau}}$. This follows from two facts: (i) the peakedness relationship is invariant under linear transformations (Dharmadhikari and Joag-Dev (1988), Lemma 7.2), that is, for any $\boldsymbol{C} \in \mathbb{R}^{p\times m}$, if $\boldsymbol{\phi} \succ \boldsymbol{\psi}$, then $\boldsymbol{C}\boldsymbol{\phi} \succ \boldsymbol{C}\boldsymbol{\psi}$; (ii) the asymptotic sampling distribution of any linear transformation of $\hat{\boldsymbol{\tau}}$ has the same form as $\hat{\boldsymbol{\tau}}$, that is, a linear combination of a standard Gaussian random vector and a truncated Gaussian random vector. For conciseness, we relegate the discussion to Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)), and consider only a single factorial effect estimator in the main text. In this case, the comparison between peakedness of two univariate asymptotic sampling distributions under ReMF and the CRFE reduces to the comparison of the lengths of quantile ranges (Li, Ding and Rubin (2018)).

COROLLARY 4. *Under Condition 1, for any* $1 \le f \le F$ *and* $\alpha \in (0,1)$, *the threshold* $c_{1-\alpha}$ *for the* $1-\alpha$ *asymptotic symmetric quantile range* $[-c_{1-\alpha}V_{\tau_f\tau_f}^{1/2}, c_{1-\alpha}V_{\tau_f\tau_f}^{1/2}]$ *of* $\hat{\tau}_f - \tau_f$ *under* ReFM *is smaller than or equal to that under the* CRFE, *and is nonincreasing in* R_f^2.

The proof of Corollary 4 is in Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)). From Corollary 4, with larger squared multiple correlation R_f^2, ReFM leads to more percentage reductions in lengths of the asymptotic quantile ranges of $\hat{\tau}_f - \tau_f$.

4.6. *Conservative covariance estimator and confidence sets under* ReFM. The asymptotic sampling distribution (4.1) of $\hat{\boldsymbol{\tau}}$ under ReFM depends on $\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\perp}$ and $(\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\parallel})_{LF}^{1/2} = \boldsymbol{V}_{\boldsymbol{\tau x}}\boldsymbol{V}_{\boldsymbol{xx}}^{-1/2}$, which further depend on $S_{qq}^{\perp}$, $\boldsymbol{S}_{\boldsymbol{\tau\tau}}^{\perp}$ and $\boldsymbol{S}_{q,\boldsymbol{x}}\boldsymbol{S}_{\boldsymbol{xx}}^{-1/2}$. Under treatment combination q, define s_{qq} as the sample variance of observed outcomes, $\boldsymbol{s}_{q,\boldsymbol{x}}$ as the sample covariance between observed outcomes and covariates, $\boldsymbol{s}_{\boldsymbol{xx}}(q)$ as the sample covariance of covariates and $s_{qq}^{\perp} = s_{qq} - \boldsymbol{s}_{q,\boldsymbol{x}}\boldsymbol{s}_{\boldsymbol{xx}}^{-1}(q)\boldsymbol{s}_{\boldsymbol{x},q}$ as the sample variance of the residuals from the linear projection of observed outcomes on covariates. We estimate $\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\perp}$ by

$$\hat{\boldsymbol{V}}_{\boldsymbol{\tau\tau}}^{\perp} = 2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1} s_{qq}^{\perp}\boldsymbol{b}_q\boldsymbol{b}_q', \tag{4.4}$$

$\boldsymbol{V}_{\boldsymbol{\tau x}}$ by $\hat{\boldsymbol{V}}_{\boldsymbol{\tau x}} = 2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1}(\boldsymbol{b}_q\boldsymbol{b}_q') \otimes \{\boldsymbol{s}_{q,\boldsymbol{x}}\boldsymbol{s}_{\boldsymbol{xx}}^{-1/2}(q)\boldsymbol{S}_{\boldsymbol{xx}}^{1/2}\}$ and $(\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\parallel})_{LF}^{1/2}$ by $\hat{\boldsymbol{V}}_{\boldsymbol{\tau x}} \times \boldsymbol{V}_{\boldsymbol{xx}}^{-1/2}$. We can then obtain a covariance estimator and construct confidence sets for $\boldsymbol{\tau}$ or its linear transformations. When the threshold a is small, $\boldsymbol{\zeta}_{LF,a}$ is close to zero, and the distribution (4.1) of $\hat{\boldsymbol{\tau}}$ is close to the Gaussian distribution with mean $\boldsymbol{\tau}$ and covariance matrix $\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\perp}$. Therefore, for a parameter of interest $\boldsymbol{C}\boldsymbol{\tau}$, we recommend confidence sets of the form $\boldsymbol{C}\hat{\boldsymbol{\tau}} + \mathcal{O}(\boldsymbol{C}\hat{\boldsymbol{V}}_{\boldsymbol{\tau\tau}}^{\perp}\boldsymbol{C}', c)$. We choose the threshold c based on simulation from the estimated asymptotic sampling distribution, and let $\hat{c}_{1-\alpha}$ be the $1-\alpha$ quantile of $(\boldsymbol{C}\boldsymbol{\phi})'(\boldsymbol{C}\hat{\boldsymbol{V}}_{\boldsymbol{\tau\tau}}^{\perp}\boldsymbol{C}')^{-1}(\boldsymbol{C}\boldsymbol{\phi})$ with $\boldsymbol{\phi}$ following the estimated asymptotic sampling distribution of $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$.

THEOREM 6. *Under* ReFM *and Condition 1, consider inferring* $\boldsymbol{C}\boldsymbol{\tau}$, *where* $\boldsymbol{C}$ *has full row rank. The probability limit of the covariance estimator for* $\boldsymbol{C}\hat{\boldsymbol{\tau}}$, $\boldsymbol{C}\hat{\boldsymbol{V}}_{\boldsymbol{\tau\tau}}^{\perp}\boldsymbol{C}' + v_{LF,a}\boldsymbol{C}\hat{\boldsymbol{V}}_{\boldsymbol{\tau x}}\boldsymbol{V}_{\boldsymbol{xx}}^{-1}\hat{\boldsymbol{V}}_{\boldsymbol{x\tau}}\boldsymbol{C}'$, *is larger than or equal to the sampling covariance, and the* $1-\alpha$ *confidence set*, $\boldsymbol{C}\hat{\boldsymbol{\tau}} + \mathcal{O}(\boldsymbol{C}\hat{\boldsymbol{V}}_{\boldsymbol{\tau\tau}}^{\perp}\boldsymbol{C}', \hat{c}_{1-\alpha})$, *has asymptotic coverage rate* $\ge 1-\alpha$, *with equality holding if* $\boldsymbol{S}_{\boldsymbol{\tau\tau}}^{\perp} \to 0$ *as* $n \to \infty$.

Theorem 6 holds because the ordering of peakedness still holds by adding an independent central convex unimodal random vector. Its proof is in Appendix A6 of the Supplementary Material (Li, Ding and Rubin (2019)). The above confidence sets will be similar to the ones based on regression adjustment if the threshold a is small. Theoretically, we can extend Theorem 6 to general symmetric convex confidence sets, and we relegate this discussion to Appendix A6 in the Supplementary Material (Li, Ding and Rubin (2019)).

5. Rerandomization with tiers of factorial effects. From Corollary 1, under the additivity in Definition 1, the squared multiple correlations between $\hat{\tau}_f$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ are the same for all f: $R_1^2 = \cdots = R_F^2 = S_{11}^{\parallel}/S_{11}$. From Section 4.5, under the additivity in Definition 1, the improvement of the fth factorial effect estimator $\hat{\tau}_f$ under ReFM compared to the CRFE is asymptotically the same for all f. However, in practice, we are sometimes more interested in some factorial effects than others. For example, the main effects are often more important than higher-order interactions. Therefore, we need a balance criterion resulting in more precise estimators for the more important factorial effects.

5.1. *Tiers of factorial effects criterion.* Let $\mathcal{F} = \{1, 2, \dots, F\}$ be the set of all factorial effects. We partition $\mathcal{F}$ into H tiers $(\mathcal{F}_1, \dots, \mathcal{F}_H)$ with decreasing importance, where the $\mathcal{F}_h$'s are disjoint and $\mathcal{F} = \bigcup_{h=1}^{H} \mathcal{F}_h$. The cardinality $F_h \equiv |\mathcal{F}_h|$ represents the number of factorial effects in tier h. For example, we can partition $\mathcal{F}$ into three tiers: $\mathcal{F}_1$ contains the K main effects, $\mathcal{F}_2$ contains the $\binom{K}{2}$ interaction effects between two factors and $\mathcal{F}_3$ contains the remaining factorial effects with higher-order interactions.

Define $\gamma_{fk}^2 = \mathrm{Corr}^2(\hat{\tau}_f, \hat{\boldsymbol{\tau}}_{\boldsymbol{x},k})$. When the fth factorial effect is more important, we would like to put more restriction on the difference-in-means vector $\hat{\boldsymbol{\tau}}_{\boldsymbol{x},k}$ with larger squared multiple correlation γ_{fk}^2. Although general results for the relative magnitudes of the γ_{fk}^2's appear too complicated, below we give a proposition under the additivity, which serves as a guideline for the choice of the balance criterion.

PROPOSITION 5. *Under the* CRFE, *assume the additivity in Definition* 1. *The squared multiple correlations satisfy* $\max_{1\le k\le F} \gamma_{fk}^2 = \gamma_{ff}^2 = R_f^2 = S_{11}^{\parallel}/S_{11}$ *for* $1 \le f \le F$. *The squared multiple partial correlation between* $\hat{\tau}_f$ *and* $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ *given* $\hat{\boldsymbol{\tau}}_{\boldsymbol{x},f}$ *is zero, that is, the residuals from the linear projections of* $\hat{\tau}_f$ *and* $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ *on* $\hat{\boldsymbol{\tau}}_{\boldsymbol{x},f}$ *are uncorrelated. If further* $n_1 = \cdots = n_Q = n/Q$, *then* $\gamma_{fk}^2 = 0$ *for* $k \neq f$.

Proposition 5 follows from some algebra, with the proof in Appendix A2 of the Supplementary Material (Li, Ding and Rubin (2019)). From Proposition 5, with the additivity and under the CRFE, $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ explains $\hat{\tau}_f$ in the linear projection only through $\hat{\boldsymbol{\tau}}_{\boldsymbol{x},f}$. Therefore, it is desirable to impose more restriction on the difference-in-means of covariates with respect to more important factorial effects under rerandomization.

5.2. *Orthogonalization with tiers of factorial effects.* For $1 \le h \le H$, let $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}[\mathcal{F}_h]$ be the subvector of $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$, consisting of the difference-in-means of covariates $\hat{\boldsymbol{\tau}}_{\boldsymbol{x},f}$ with respect to factorial effect $f \in \mathcal{F}_h$. From Section 5.1, the smaller the h is, the more restriction we want to impose on $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}[\mathcal{F}_h]$. However, due to the correlations among the $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}[\mathcal{F}_h]$'s, restrictions on one also restrict others. For example, balancing $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}[\mathcal{F}_1]$ partially balances $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}[\mathcal{F}_2]$. Therefore, instead of unnecessarily balancing for all factorial effects in tier h, we balance only the part that is orthogonal to the factorial effects in previous tiers.

Let $\tilde{\boldsymbol{B}} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} \boldsymbol{b}_q \boldsymbol{b}_q'$. From Proposition 1, the sampling covariance of $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ under the CRFE, $\boldsymbol{V}_{\boldsymbol{xx}} = \tilde{\boldsymbol{B}} \otimes \boldsymbol{S}_{\boldsymbol{xx}}$, contains two components: $\tilde{\boldsymbol{B}}$ determined by the coefficient vector $\boldsymbol{b}_q$'s and $\boldsymbol{S}_{\boldsymbol{xx}}$ determined by the covariates. Below we introduce a blockwise

Gram–Schmidt orthogonalization of the coefficient vector $\boldsymbol{b}_q$'s, taking into account the tiers of factorial effects. Let $\mathcal{F}_{\overline{h}} = \bigcup_{l=1}^{h} \mathcal{F}_l$ be the factorial effects in the first h tiers. We use $\boldsymbol{b}_q[\mathcal{F}_h]$ and $\boldsymbol{b}_q[\mathcal{F}_{\overline{h}}]$ to denote the subvectors of $\boldsymbol{b}_q$ with indices in $\mathcal{F}_h$ and $\mathcal{F}_{\overline{h}}$, and $\tilde{\boldsymbol{B}}[\mathcal{F}_h, \mathcal{F}_{\overline{h}}]$ and $\tilde{\boldsymbol{B}}[\mathcal{F}_{\overline{h}}, \mathcal{F}_{\overline{h}}]$ to denote the submatrices of $\tilde{\boldsymbol{B}}$ with indices in $\mathcal{F}_h \times \mathcal{F}_{\overline{h}}$ and $\mathcal{F}_{\overline{h}} \times \mathcal{F}_{\overline{h}}$. For each $1 \le q \le Q$, we define the orthogonalized coefficient vector $\boldsymbol{c}_q = (\boldsymbol{c}_q'[1], \dots, \boldsymbol{c}_q'[H])'$ as $\boldsymbol{c}_q[1] = \boldsymbol{b}_q[\mathcal{F}_1]$, and for $2 \le h \le H$,

$$\boldsymbol{c}_q[h] = \boldsymbol{b}_q[\mathcal{F}_h] - \tilde{\boldsymbol{B}}[\mathcal{F}_h, \mathcal{F}_{\overline{h-1}}]\{\tilde{\boldsymbol{B}}[\mathcal{F}_{\overline{h-1}}, \mathcal{F}_{\overline{h-1}}]\}^{-1}\boldsymbol{b}_q[\mathcal{F}_{\overline{h-1}}]. \tag{5.1}$$

The difference-in-means vector of covariates with respect to orthogonalized coefficient vectors is

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{x}} \equiv \begin{pmatrix} \hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[1] \\ \vdots \\ \hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[H] \end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^{Q} \begin{pmatrix} \boldsymbol{c}_q[1] \\ \vdots \\ \boldsymbol{c}_q[H] \end{pmatrix} \otimes \hat{\bar{\boldsymbol{x}}}(q). \tag{5.2}$$

By construction, $\tilde{\boldsymbol{C}} \equiv 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} \boldsymbol{c}_q \boldsymbol{c}_q'$ is block diagonal, and thus the sampling covariance of $\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}$ under the CRFE, $\tilde{\boldsymbol{C}} \otimes \boldsymbol{S}_{\boldsymbol{xx}}$, is also block diagonal. The following proposition summarizes these results.

PROPOSITION 6. *Under the* CRFE, $(\hat{\boldsymbol{\tau}}' - \boldsymbol{\tau}', \hat{\boldsymbol{\theta}}_{\boldsymbol{x}}')'$ *has mean zero and sampling covariance*:

$$\operatorname{Cov}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}, \hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h]) \equiv \boldsymbol{W}_{\boldsymbol{\tau x}}[h] = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} (\boldsymbol{b}_q \boldsymbol{c}_q'[h]) \otimes \boldsymbol{S}_{q,\boldsymbol{x}},$$

$$\operatorname{Cov}(\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h]) \equiv \boldsymbol{W}_{\boldsymbol{xx}}[h] = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} (\boldsymbol{c}_q[h] \boldsymbol{c}_q'[h]) \otimes \boldsymbol{S}_{\boldsymbol{xx}},$$

and $\operatorname{Cov}(\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h], \hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[\tilde{h}]) = \boldsymbol{0}$ *if* $h \neq \tilde{h}$.

Proposition 6 follows from some algebra, with the proof in Appendix A2 of the Supplementary Material (Li, Ding and Rubin (2019)). From Proposition 6, $(\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[1], \dots, \hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[H])$ are mutually uncorrelated under the CRFE, and thus are essentially from a blockwise Gram–Schmidt orthogonalization of $(\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}[\mathcal{F}_1], \dots, \hat{\boldsymbol{\tau}}_{\boldsymbol{x}}[\mathcal{F}_H])$. We define the Mahalanobis distance in tier h as

$$M_h = \hat{\boldsymbol{\theta}}_{\boldsymbol{x}}'[h] (\boldsymbol{W}_{\boldsymbol{xx}}[h])^{-1} \hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h] \quad (1 \le h \le H). \tag{5.3}$$

Let $(a_1, \dots, a_H)$ be H positive constants predetermined in the design stage. Under rerandomization with tiers of factorial effects, denoted by ReFMT_F, a randomization is acceptable if and only if $M_h \le a_h$ for all $1 \le h \le H$. Below we use $\mathcal{T}_\text{F}$ to denote the event that the treatment vector $\boldsymbol{Z}$ satisfies this criterion. From the finite population central limit theorem, asymptotically, M_h is $\chi^2_{LF_h}$, and $(M_1, \dots, M_H)$ are jointly independent. Therefore, the asymptotic acceptance probability under ReFMT_F is $p_a = \prod_{h=1}^{H} P(\chi^2_{LF_h} \le a_h)$. We usually choose small a_h's. The magnitude of a_h's depend on the relative importance of the factorial effects in all tiers. See Morgan and Rubin (2015) for a related discussion.

With equal treatment group sizes, M_h has a simpler form.

PROPOSITION 7. *When* $n_1 = \cdots = n_Q = n/Q$, *the coefficient* $\boldsymbol{c}_q[h]$ *in* (5.1) *reduces to* $\boldsymbol{b}_q[\mathcal{F}_h]$, *the difference-in-means of covariates* $\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h]$ *in* (5.2) *reduces to* $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}[\mathcal{F}_h]$ *and the Mahalanobis distance* M_h *in* (5.3) *reduces to* $M_h = n/4 \cdot \sum_{f \in \mathcal{F}_h} \hat{\boldsymbol{\tau}}_{\boldsymbol{x},f}' \boldsymbol{S}_{\boldsymbol{xx}}^{-1} \hat{\boldsymbol{\tau}}_{\boldsymbol{x},f}$.

Proposition 7 follows from some algebra, with the proof in Appendix A2 of the Supplementary Material (Li, Ding and Rubin (2019)). In Proposition 7, if further each tier contains exactly one factorial effect, ReFMT_{F} reduces to the rerandomization scheme discussed in Branson, Dasgupta and Rubin (2016).

5.3. *Asymptotic sampling distribution of* $\hat{\boldsymbol{\tau}}$. In this subsection, we study the asymptotic sampling distribution of $\hat{\boldsymbol{\tau}}$ under ReFMT_{F}. Let $\boldsymbol{W}^{\parallel}_{\boldsymbol{\tau}\boldsymbol{\tau}}[h] = \boldsymbol{W}_{\boldsymbol{\tau}\boldsymbol{x}}[h](\boldsymbol{W}_{\boldsymbol{x}\boldsymbol{x}}[h])^{-1}\boldsymbol{W}_{\boldsymbol{x}\boldsymbol{\tau}}[h]$ be the sampling covariance matrix of $\hat{\boldsymbol{\tau}}$ explained by $\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h]$ in the linear projection under the CRFE. Extending earlier notation, let $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_F)$, and $\boldsymbol{\zeta}_{LF_h,a_h} \sim \boldsymbol{D}_h \mid \boldsymbol{D}_h'\boldsymbol{D}_h \le a_h$ be a truncated Gaussian vector with LF_h dimensions, where $\boldsymbol{D}_h = (D_{h1}, \dots, D_{h,LF_h})' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{LF_h})$. In addition, $(\boldsymbol{\varepsilon}, \boldsymbol{\zeta}_{LF_1,a_1}, \dots, \boldsymbol{\zeta}_{LF_H,a_H})$ are jointly independent.

THEOREM 7. *Under* ReFMT_{F} *and Condition* 1,

$$\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} \mid \mathcal{T}_{\text{F}} \dot\sim (\boldsymbol{V}^{\perp}_{\boldsymbol{\tau}\boldsymbol{\tau}})^{1/2}\boldsymbol{\varepsilon} + \sum_{h=1}^{H} (\boldsymbol{W}^{\parallel}_{\boldsymbol{\tau}\boldsymbol{\tau}}[h])^{1/2}_{LF_h}\boldsymbol{\zeta}_{LF_h,a_h}. \tag{5.4}$$

The proof of Theorem 7, similar to that of Theorem 2, is in Appendix A3 of the Supplementary Material (Li, Ding and Rubin (2019)).

Let $W^{\parallel}_{\tau_f\tau_f}[h]$ be the fth diagonal element of $\boldsymbol{W}^{\parallel}_{\boldsymbol{\tau}\boldsymbol{\tau}}[h]$. The squared multiple correlation between $\hat{\tau}_f$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h]$ under the CRFE is then $\rho_f^2[h] = W^{\parallel}_{\tau_f\tau_f}[h]/V_{\tau_f\tau_f}$. When treatment group sizes are equal, $\rho_f^2[h]$ reduces to $\rho_f^2[h] = \sum_{k:k\in\mathcal{F}_h} \gamma_{fk}^2$ for all f; if further the additivity holds, $\rho_f^2[h]$ reduces to $S^{\parallel}_{11}/S_{11}$ if $f \in \mathcal{F}_h$, and zero otherwise. Because the $\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h]$'s are from a blockwise Gram–Schmidt orthogonalization of $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$, the squared multiple correlation between $\hat{\tau}_f$ and $\hat{\boldsymbol{\tau}}_{\boldsymbol{x}}$ can be decomposed as $R_f^2 = \sum_{h=1}^{H} \rho_f^2[h]$. The following corollary shows the marginal asymptotic sampling distribution of a single factorial effect estimator. Let $\varepsilon_0 \sim \mathcal{N}(0, 1)$, $\eta_{LF_h,a_h} \sim D_{h1} \mid \boldsymbol{D}_h'\boldsymbol{D}_h \le a_h$ be the first coordinate of $\boldsymbol{\zeta}_{LF_h,a_h}$, and $(\varepsilon_0, \eta_{LF_1,a_1}, \dots, \eta_{LF_H,a_H})$ be jointly independent.

COROLLARY 5. *Under* ReFMT_{F} *and Condition* 1, *for* $1 \le f \le F$,

$$\hat{\tau}_f - \tau_f \mid \mathcal{T}_{\text{F}} \dot\sim \sqrt{V_{\tau_f\tau_f}}\left(\sqrt{1 - R_f^2} \cdot \varepsilon_0 + \sum_{h=1}^{H} \sqrt{\rho_f^2[h]} \cdot \eta_{LF_h,a_h}\right). \tag{5.5}$$

The proof of Corollary 5 is in Appendix A3 of the Supplementary Material (Li, Ding and Rubin (2019)).

5.4. *Asymptotic unbiasedness*, *sampling covariance and peakedness*. Based on the asymptotic distributions in Section 5.3, we study the asymptotic properties of the factorial effect estimators. First, $(\boldsymbol{\varepsilon}, \boldsymbol{\zeta}_{LF_1,a_1}, \dots, \boldsymbol{\zeta}_{LF_H,a_H})$ are all central convex unimodal from Proposition 2, and thus the asymptotic sampling distribution (5.4) of $\hat{\boldsymbol{\tau}}$ under ReFMT_{F} is also central convex unimodal. The symmetry of the asymptotic sampling distributions ensures that the factorial effect estimator $\hat{\boldsymbol{\tau}}$ is consistent for $\boldsymbol{\tau}$ under ReFMT_{F}, which implies that the difference-in-means of any observed or unobserved covariate with respect to any factorial effect has asymptotic mean zero.

Second, we compare the asymptotic sampling covariance matrices of $\hat{\boldsymbol{\tau}}$ under ReFMT_{F} and the CRFE. For each $1 \le h \le H$, let $v_{LF_h,a_h} = P(\chi^2_{LF_h+2} \le a_h)/P(\chi^2_{LF_h} \le a_h) \le 1$.

THEOREM 8. *Under Condition 1, $\hat{\boldsymbol{\tau}}$ has smaller asymptotic sampling covariance under* $\mathrm{ReFMT_F}$ *than that under the* CRFE, *and the reduction in asymptotic sampling covariance is $n\sum_{h=1}^{H}(1-v_{LF_h,a_h})\boldsymbol{W}^{\parallel}_{\boldsymbol{\tau\tau}}[h]$. Specifically, for each $1 \le f \le F$, the* PRIASV *of $\hat{\tau}_f$ is $\sum_{h=1}^{H}(1-v_{LF_h,a_h})\rho_f^2[h]$.*

Theorem 8 follows from Theorem 7 and Proposition 3, with the proof in Appendix A4 of the Supplementary Material (Li, Ding and Rubin (2019)). When the threshold a_h's are close to zero, the asymptotic sampling variance of $\hat{\tau}_f$ reduces to $V_{\tau_f\tau_f}(1-R_f^2)$, which is identical to the asymptotic sampling variance of the regression adjusted estimator under the CRFE (Lu (2016)).

Third, we compare the peakedness of asymptotic sampling distributions of $\hat{\boldsymbol{\tau}}$ under $\mathrm{ReFMT_F}$ and the CRFE.

THEOREM 9. *Under Condition 1, the asymptotic sampling distribution of $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ under* $\mathrm{ReFMT_F}$ *is more peaked than that under the* CRFE.

The proof of Theorem 9, similar to that of Theorem 4, is in Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)). We then consider the specific symmetric convex set $\mathcal{O}(\boldsymbol{V}_{\boldsymbol{\tau\tau}}, c)$. Unfortunately, considering joint quantile region for $\boldsymbol{\tau}$ is technically challenging in general, and we consider the case where the following condition holds.

CONDITION 2. *There exists an orthogonal matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{F\times F}$ such that*

$$\boldsymbol{\Gamma}'\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{-1/2}\boldsymbol{W}^{\parallel}_{\boldsymbol{\tau\tau}}[h]\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{-1/2}\boldsymbol{\Gamma} = \mathrm{diag}(\omega_{h1}^2, \dots, \omega_{hF}^2) \quad (1 \le h \le H),$$

where $(\omega_{h1}^2, \dots, \omega_{hF}^2)$ are the canonical correlations between $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h]$ under the CRFE.

Condition 2 holds automatically when $H = 1$. Moreover, the additivity in Definition 1 implies Condition 2 for general $H \ge 1$. The following proposition states this result. By construction, $\boldsymbol{b}_q = \boldsymbol{\Psi}\boldsymbol{c}_q$, where $\boldsymbol{\Psi} \in \mathbb{R}^{F\times F}$ is the common linear transformation matrix for all $1 \le q \le Q$. Recall that $\tilde{\boldsymbol{B}} = 2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1}\boldsymbol{b}_q\boldsymbol{b}_q'$, and $\tilde{\boldsymbol{C}} = 2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1}\boldsymbol{c}_q\boldsymbol{c}_q'$.

PROPOSITION 8. *Under the additivity in Definition 1, Condition 2 holds with orthogonal matrix $\boldsymbol{\Gamma} = \tilde{\boldsymbol{B}}^{-1/2}\boldsymbol{\Psi}\tilde{\boldsymbol{C}}^{1/2}$, and the canonical correlations between $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{x}}[h]$ have exactly F_h nonzero elements, which are all equal to $S_{11}^{\parallel}/S_{11}$.*

Proposition 8 follows from some algebra, with the proof in Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)).

THEOREM 10. *Under* $\mathrm{ReFMT_F}$, *assume that Conditions 1 and 2 hold. Let $c_{1-\alpha}$ be the solution of $\lim_{n\to\infty} P\{\hat{\boldsymbol{\tau}} - \boldsymbol{\tau} \in \mathcal{O}(\boldsymbol{V}_{\boldsymbol{\tau\tau}}, c_{1-\alpha}) \mid \mathcal{T}_{\mathrm{F}}\} = 1-\alpha$. It depends only on L, F_h's, a_h's and $(\omega_{h1}^2, \dots, \omega_{hF}^2)$'s, and is nonincreasing in ω_{hf}^2 for $1 \le h \le H$ and $1 \le f \le F$.*

The proof of Theorem 10, similar to that of Theorem 5, is in Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)).

Because the peakedness relationship is invariant under linear transformations, and any linear transformation of $\hat{\boldsymbol{\tau}}$ has an asymptotic sampling distribution of the same form as $\hat{\boldsymbol{\tau}}$, we can establish similar conclusions as Theorems 9 and 10 for any linear transformations of $\hat{\boldsymbol{\tau}}$. We relegate the details to Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)), and consider only the asymptotic sampling distribution of a single factorial effect estimator below.

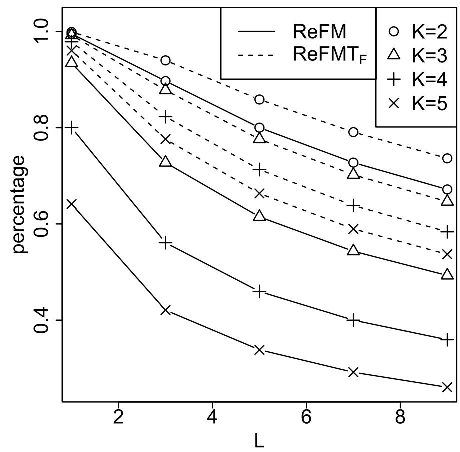

FIG. 2. PRIASV *of main effect estimators divided by* R_f^2.

COROLLARY 6. *Under Condition 1, for any* $1 \le f \le F$ *and* $\alpha \in (0,1)$, *the threshold* $c_{1-\alpha}$ *for* $1-\alpha$ *asymptotic symmetric quantile range* $[-c_{1-\alpha}V_{\tau_f\tau_f}^{1/2}, c_{1-\alpha}V_{\tau_f\tau_f}^{1/2}]$ *of* $\hat{\tau}_f - \tau_f$ *under* ReFMT$_\text{F}$ *is smaller than or equal to that under the* CRFE, *and is nonincreasing in* $\rho_f^2[h]$ *for* $1 \le h \le H$.

The proof of Corollary 6 is in Appendix A5 of the Supplementary Material (Li, Ding and Rubin (2019)). From Corollary 6, with larger squared multiple correlation $\rho_f^2[h]$, ReFMT$_\text{F}$ yields more percentage reductions of quantile ranges.

The example below shows the advantage of ReFMT$_\text{F}$ over ReFM.

EXAMPLE 3. We consider experiments with K factors and L dimensional covariates. Assume the additivity in Definition 1, which implies that R_f^2 is the same for all factorial effects f. Suppose that we are more interested in the K main effects than the interaction effects. We divide the F effects into 2 tiers, where tier 1 contains the $F_1 = K$ main effects and tier 2 contains the remaining $F_2 = 2^K - 1 - K$ interaction effects. From Proposition 5, we can derive $\rho_k^2[1] = R_f^2$ and $\rho_k^2[2] = 0$ for the main effect $1 \le k \le K$. We compare two rerandomization schemes with the same asymptotic acceptance probability: ReFM with $p_a = 0.001$ and ReFMT$_\text{F}$ with thresholds (a_1, a_2) satisfying $P(\chi^2_{LF_1} \le a_1) = 0.002$ and $P(\chi^2_{LF_2} \le a_2) = 0.5$. Figure 2 shows the PRIASV, divided by R_f^2, of the main effect estimators for both rerandomization schemes. It shows that the advantage of ReFMT$_\text{F}$ increases as the numbers of factors and covariates increase.

5.5. *Conservative covariance estimator and confidence sets under* ReFMT$_\text{F}$. We estimate $\boldsymbol{V}_{\boldsymbol{\tau\tau}}^{\perp}$ by $\hat{\boldsymbol{V}}_{\boldsymbol{\tau\tau}}^{\perp}$ in (4.4), $\boldsymbol{W}_{\boldsymbol{\tau x}}[h]$ by $\hat{\boldsymbol{W}}_{\boldsymbol{\tau x}}[h] = 2^{-2(K-1)}\sum_{q=1}^{Q} n_q^{-1}(\boldsymbol{b}_q \boldsymbol{c}_q'[h]) \otimes \{\boldsymbol{s}_{q,\boldsymbol{x}} \boldsymbol{s}_{\boldsymbol{xx}}^{-1/2}(q)\boldsymbol{S}_{\boldsymbol{xx}}^{1/2}\}$, and $(\boldsymbol{W}_{\boldsymbol{\tau\tau}}^{\parallel}[h])_{LF_h}^{1/2}$ by $\hat{\boldsymbol{W}}_{\boldsymbol{\tau x}}[h](\boldsymbol{W}_{\boldsymbol{xx}}[h])^{-1/2}$. We can then obtain a covariance estimator and construct confidence sets for $\boldsymbol{\tau}$ or its linear transformations. Similar to ReFM, for a parameter of interest $\boldsymbol{C\tau}$, we recommend confidence sets of the form $\boldsymbol{C}\hat{\boldsymbol{\tau}} + \mathcal{O}(\boldsymbol{C}\hat{\boldsymbol{V}}_{\boldsymbol{\tau\tau}}^{\perp}\boldsymbol{C}', c)$, where we choose the threshold c by simulating random draws from the estimated asymptotic sampling distribution. Let $\hat{c}_{1-\alpha}$ be the $1-\alpha$ quantile of $(\boldsymbol{C\phi})'(\boldsymbol{C}\hat{\boldsymbol{V}}_{\boldsymbol{\tau\tau}}^{\perp}\boldsymbol{C}')^{-1}(\boldsymbol{C\phi})$ with $\boldsymbol{\phi}$ following the estimated asymptotic sampling distribution of $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ under ReFMT$_\text{F}$.

THEOREM 11. *Under* ReFMT$_F$ *and Condition* 1, *consider inferring* $\boldsymbol{C}\boldsymbol{\tau}$, *where* $\boldsymbol{C}$ *has full row rank. The probability limit of the covariance estimator,* $\boldsymbol{C}\hat{\boldsymbol{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\perp}\boldsymbol{C}' + \sum_{h=1}^{H} v_{LF_h,a_h}\boldsymbol{C} \times \hat{\boldsymbol{W}}_{\boldsymbol{\tau}\boldsymbol{x}}[h](\boldsymbol{W}_{\boldsymbol{x}\boldsymbol{x}}[h])^{-1}\hat{\boldsymbol{W}}_{\boldsymbol{x}\boldsymbol{\tau}}[h]\boldsymbol{C}'$, *for* $\boldsymbol{C}\hat{\boldsymbol{\tau}}$ *is larger than or equal to the actual sampling covariance, and the* $1-\alpha$ *confidence set,* $\boldsymbol{C}\hat{\boldsymbol{\tau}} + \mathcal{O}(\boldsymbol{C}\hat{\boldsymbol{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\perp}\boldsymbol{C}', \hat{c}_{1-\alpha})$, *has asymptotic coverage rate* $\geq 1-\alpha$, *with equality holding if* $\boldsymbol{S}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\perp} \to 0$ *as* $n \to \infty$.

The proof of Theorem 11, similar to that of Theorem 6, is in Appendix A6 of the Supplementary Material (Li, Ding and Rubin (2019)). The above confidence sets will be similar to the ones based on regression adjustment if the threshold a_h's are small (Lu (2016)). Moreover, we can also extend Theorem 11 to general symmetric convex confidence sets (Li, Ding and Rubin (2019)).

6. An education example. We illustrate the theory of rerandomization using a dataset from the Student Achievement and Retention Project (Angrist, Lang and Oreopoulos (2009)), a 2^2 CRFE at one of the satellite campuses of a large Canadian university. One treatment factor is the Student Support Program (SSP), which provides students some services for study. The other treatment factor is the Student Fellowship Program (SFP), which awards students scholarships for achieving a target first year grade point average (GPA). There were 1,006 students in the control group receiving neither SSP nor SFP (i.e., $(-1,-1)$), 250 students offered only SFP (i.e., $(-1,+1)$), 250 students offered only SSP (i.e., $(+1,-1)$) and 150 students offered both SSP and SFP (i.e., $(+1,+1)$). We include $L=5$ pretreatment covariates: high school GPA, gender, age, indicators for whether the student was living at home and whether the student rarely put off studying for tests, and exclude students with missing covariate values. This results in treatment groups of sizes $(856, 216, 208, 118)$ for treatment combinations $(-1,-1)$, $(-1,+1)$, $(+1,-1)$ and $(+1,+1)$, respectively.

To demonstrate the advantage of rerandomization, we compare the CRFE and ReFMT$_F$ in terms of the sampling distributions of the factorial effects estimator. However, the sampling distributions depend on all the potential outcomes including the missing ones. To make the simulation more realistic, we impute all of the missing potential outcomes based on simple model fitting. Specifically, we fit a linear regression of the observed GPA on the levels of two treatment factors, all covariates and the interactions between these covariates, and then impute all the missing potential outcomes based on the fitted model. We further truncate all the potential outcomes to $[0,4]$ to mimic the values of GPA. Note that the generating models for the missing potential outcomes are not linear in the covariates. For the simulated data set, the sampling squared multiple correlations between factorial effect estimators and the difference-in-means of covariates are $(R_1^2, R_2^2, R_3^2) = (0.247, 0.244, 0.245)$.

We divide the three factorial effects into two tiers, where tier 1 contains $F_1 = 2$ main effects, and tier 2 contains $F_2 = 1$ interaction effect, and choose thresholds (a_1, a_2) such that $P(\chi^2_{LF_1} \leq a_1) = 0.002$ and $P(\chi^2_{LF_2} \leq a_2) = 0.5$. Table 2 shows the empirical and theoretical percentage reductions in the sampling variances and the lengths of 95% symmetric quantile ranges for the three factorial effect estimators under ReFMT$_F$, compared to the CRFE. From Table 2, the asymptotic approximations work fairly well, and ReFMT$_F$ improves the precision of the two average main effects estimators more than that of the average interaction effect estimator.

We then consider confidence sets for the two average main effects (τ_1, τ_2) under both designs. The empirical coverage probabilities of 95% confidence sets discussed in Sections 3.1 and 5.5 under the CRFE and ReFMT$_F$ are, respectively, 96.4% and 96.5%, showing that both confidence sets are slightly conservative. Moreover, the percentage reduction in the average volume of 95% confidence sets under ReFMT$_F$ compared to the CRFE is 20.5%, and

TABLE 2
Comparison of the factorial effect estimators between the CRFE *and* ReFMT$_{\mathrm{F}}$. *The second and third columns show the percentage reductions in variances, and the fourth and fifth columns show the percentage reductions in the lengths of* 95% *quantile ranges*

	Reduction in variance		Reduction in quantile range	
Factorial effect	Empirical	Theoretical	Empirical	Theoretical
Main effect of SSP	20.2%	21.2%	10.7%	11.2%
Main effect of SFP	20.4%	20.9%	10.8%	11.1%
Interaction effect	14.4%	14.9%	7.7%	7.8%

the corresponding percentage increase in sample size needed for the CRFE to obtain 95% confidence set of the same average volume as ReFMT$_{\mathrm{F}}$ is about 25.8%.

To end this section, we investigate the dependence of the PRIASVs on the choices of thresholds (a_1, a_2). Let $p_{ah} \equiv P(\chi^2_{LF_h} \le a_h)$ be the asymptotic acceptance probability for tier h ($h = 1, 2$). Fixing the overall asymptotic acceptance probability $p_a \equiv p_{a1} p_{a2}$ at 0.001, Figure 3 shows the PRIASVs of all factorial effect estimators as functions of p_{a1}. We can see that (1) more stringent restrictions on the first tier of factorial effects (i.e., the two main effects) lead to larger PRIASVs of the corresponding estimators, but (2) the PRIASV of the estimator of the second tier of factorial effect (i.e., the interaction effect) is a nonmonotone function of p_{a1}. Therefore, in practice, we are facing a trade-off, which depends on the a priori relative importance of the factorial effects.

7. Extension. When covariates have varying importance for the potential outcomes, we can further consider a balance criterion using tiers of covariates, that is, rerandomized factorial experiments with tiers of both covariates and factorial effects. We discuss this balance criterion and demonstrate its advantage in Appendix A1 of the Supplementary Material (Li, Ding and Rubin (2019)).

Acknowledgments. We thank the Associate Editor, two reviewers and Dr. Zach Branson for helpful comments.

The second author was supported by the Institute for Education Science (IES Grant R305D150040) and the National Science Foundation Grant DMS-713152).

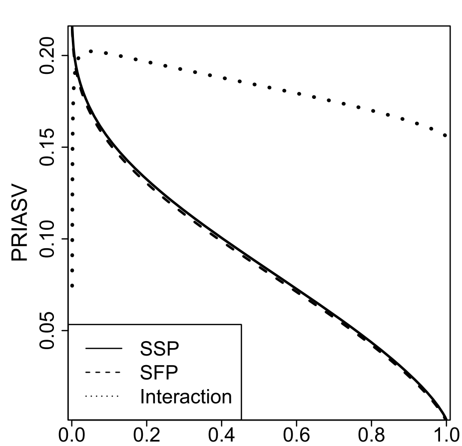

FIG. 3. PRIASVs *of all factorial effects as* p_{a1}, *with* $p_a = p_{a1} p_{a2}$ *fixed at* 0.001.

The third author was supported by the National Institute of Allergy and Infectious Diseases/NIH (R01AI102710), National Science Foundation (IIS-1409177), Office of Naval Research (N00014-17-1-2131) and a Google Faculty Fellowship.

SUPPLEMENTARY MATERIAL

Supplement to "Rerandomization in 2^K factorial experiments" (DOI: 10.1214/18-AOS1790SUPP; .pdf). We study the theoretical properties of 2^K rerandomized factorial experiments with tiers of both covariates and factorial effects, and prove all the theorems, corollaries and propositions.

REFERENCES

ANGRIST, J., LANG, D. and OREOPOULOS, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *Am. Econ. J. Appl. Econ.* **1** 136–163.

BAILEY, R. A. (1983). Restricted randomization. *Biometrika* **70** 183–198. MR0742988 https://doi.org/10.1093/biomet/70.1.183

BALDI ANTOGNINI, A. and ZAGORAIOU, M. (2011). The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika* **98** 519–535. MR2836404 https://doi.org/10.1093/biomet/asr021

BICKEL, P. J. and LEHMANN, E. L. (1976). Descriptive statistics for nonparametric models. III. Dispersion. *Ann. Statist.* **4** 1139–1158. MR0474620

BIRNBAUM, Z. W. (1948). On random variables with comparable peakedness. *Ann. Math. Stat.* **19** 76–81. MR0024099 https://doi.org/10.1214/aoms/1177730293

BOSE, R. C. (1947). Mathematical theory of the symmetrical factorial design. *Sankhyā* **8** 107–166. MR0026781

BOX, G. E. P., HUNTER, J. S. and HUNTER, W. G. (2005). *Statistics for Experimenters*: *Design*, *Innovation*, *and Discovery*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR2140250

BRANSON, Z., DASGUPTA, T. and RUBIN, D. B. (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. *Ann. Appl. Stat.* **10** 1958–1976. MR3592044 https://doi.org/10.1214/16-AOAS959

BRUHN, M. and MCKENZIE, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *Am. Econ. J. Appl. Econ.* **1** 200–232.

COCHRAN, W. G. and COX, G. M. (1950). *Experimental Designs*. Wiley, New York. MR0034558

COX, D. R. (1982). Randomization and concomitant variables in the design of experiments. In *Statistics and Probability*: *Essays in Honor of C. R. Rao* (P. R. K. G. Kallianpur and J. K. Ghosh, eds.) 197–202. North-Holland, Amsterdam. MR0659470

COX, D. R. (2009). Randomization in the design of experiments. *Int. Stat. Rev.* **77** 415–429.

COX, D. R. and REID, N. (2000). *The Theory of the Design of Experiments*. Chapman and Hall/CRC, New York.

DASGUPTA, T., PILLAI, N. S. and RUBIN, D. B. (2015). Causal inference from 2^K factorial designs by using potential outcomes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 727–753. MR3382595 https://doi.org/10.1111/rssb.12085

DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality*, *Convexity*, *and Applications*. *Probability and Mathematical Statistics*. Academic Press, Boston, MA. MR0954608

DHARMADHIKARI, S. W. and JOGDEO, K. (1976). Multivariate unimodality. *Ann. Statist.* **4** 607–613. MR0415719

FINNEY, D. J. (1943). The fractional replication of factorial arrangements. *Ann. Hum. Genet.* **12** 291–301.

FISHER, R. A. (1926). The arrangement of field experiments. *J. Minist. Agric. G.B.* **33** 503–513.

FISHER, R. A. (1935). *The Design of Experiments*, 1st ed. Oliver and Boyd, Edinburgh.

GIOVAGNOLI, A. and WYNN, H. P. (1995). Multivariate dispersion orderings. *Statist. Probab. Lett.* **22** 325–332. MR1333191 https://doi.org/10.1016/0167-7152(94)00084-L

GRUNDY, P. M. and HEALY, M. J. R. (1950). Restricted randomization and quasi-Latin squares. *J. Roy. Statist. Soc. Ser. B* **12** 286–291.

HANSEN, B. B. and BOWERS, J. (2008). Covariate balance in simple, stratified and clustered comparative studies. *Statist. Sci.* **23** 219–236. MR2516821 https://doi.org/10.1214/08-STS254

HINKELMANN, K. and KEMPTHORNE, O. (2007). *Design and Analysis of Experiments*, *Vol.* 1, *Introduction to Experimental Design*, 2nd ed. Wiley, New York.

KANTER, M. (1977). Unimodality and dominance for symmetric random vectors. *Trans. Amer. Math. Soc.* **229** 65–85. MR0445580 https://doi.org/10.2307/1998500

KEMPTHORNE, O. (1952). *The Design and Analysis of Experiments*. Wiley, New York. MR0045368

LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Amer. Statist. Assoc.* **112** 1759–1769. MR3750897 https://doi.org/10.1080/01621459.2017.1295865

LI, X., DING, P. and RUBIN, D. B. (2018). Asymptotic theory of rerandomization in treatment-control experiments. *Proc. Natl. Acad. Sci. USA* **115** 9157–9162. MR3856113 https://doi.org/10.1073/pnas.1808191115

LI, X., DING, P. and RUBIN, D. B. (2020). Supplement to "Rerandomization in 2^K factorial experiments." https://doi.org/10.1214/18-AOS1790SUPP.

LU, J. (2016). Covariate adjustment in randomization-based causal inference for 2^K factorial designs. *Statist. Probab. Lett.* **119** 11–20. MR3555264 https://doi.org/10.1016/j.spl.2016.07.010

MARSHALL, A. W., OLKIN, I. and ARNOLD, B. C. (2009). *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. Springer, New York.

MORGAN, K. L. and RUBIN, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Ann. Statist.* **40** 1263–1282. MR2985950 https://doi.org/10.1214/12-AOS1008

MORGAN, K. L. and RUBIN, D. B. (2015). Rerandomization to balance tiers of covariates. *J. Amer. Statist. Assoc.* **110** 1412–1421. MR3449036 https://doi.org/10.1080/01621459.2015.1079528

SHAKED, M. (1985). Ordering distributions by dispersion. In *Encyclopedia of Statistical Sciences, Vol.* 6 (N. L. Johnson and S. Kotz, eds.) 485–490. Wiley, New York.

SHERMAN, S. (1955). A theorem on convex sets with applications. *Ann. Math. Stat.* **26** 763–767. MR0074845 https://doi.org/10.1214/aoms/1177728435

SPLAWA-NEYMAN, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles (with discussion). Section 9. *Statist. Sci.* **5** 465–472. Translated from the Polish and edited by D. M. Dąbrowska and T. P. Speed. MR1092986

STUDENT (1938). Comparison between balanced and random arrangements of field plots. *Biometrika* **29** 363–378.

WU, C. F. J. (2015). Post-Fisherian experimentation: From physical to virtual. *J. Amer. Statist. Assoc.* **110** 612–620. MR3367251 https://doi.org/10.1080/01621459.2014.914441

WU, C. F. J. and HAMADA, M. S. (2011). *Experiments: Planning, Analysis, and Optimization*. Wiley, Hoboken, NJ.

YATES, F. (1937). The design and analysis of factorial experiments. Technical Communication No. 35, Imperial Bureau of Soil Sciences, Harpenden.

YATES, F. (1948). Comment to F. J. Anscombe. *J. R. Stat. Soc., A* **111** 204–205.

YOUDEN, W. J. (1972). Randomization and experimentation. *Technometrics* **14** 13–22.