

RERANDOMIZATION IN 2^K FACTORIAL EXPERIMENTS

BY XINRAN LI¹, PENG DING² AND DONALD B. RUBIN³

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 $1 \leq q \leq Q$, let $\iota(q) = (\iota_1(q), \iota_2(q), \dots, \iota_K(q)) \in \{-1, +1\}^K$ 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 $\mathbf{Y}_i = (Y_i(1), Y_i(2), \dots, Y_i(Q))$ 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 $\bar{\mathbf{Y}} = (\bar{Y}(1), \bar{Y}(2), \dots, \bar{Y}(Q))$ 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

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

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

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

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

Intuitively, the k th 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 $\mathbf{g}_f$'s and the $\mathbf{b}_q$'s. Specifically, the first three columns ($\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3$) 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 ($\mathbf{g}_4, \dots, \mathbf{g}_7$) are the elementwise multiplications of subsets of $(\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3)$ that generate the interaction effects. The coefficient vector $\mathbf{b}_q$ consists of all the elements in the q th row of Table 1.

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

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

2.2. Treatment assignment, covariate imbalance and rerandomization. For each unit i , $\mathbf{x}_i$ 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 $\mathbf{Z} = (Z_1, \dots, Z_n)$ be the treatment assignment vector for all units. In the CRFE, the probability that $\mathbf{Z}$ takes a particular value $\mathbf{z} = (z_1, \dots, z_n)$ is $n_1! \cdots n_Q! / n!$, where $\sum_{i=1}^n 1\{z_i = q\} = n_q$ for all q . Let $\bar{\mathbf{x}} = n^{-1} \sum_{i=1}^n \mathbf{x}_i$ be the finite population covariate mean vector; for $1 \leq q \leq Q$, let $\hat{\mathbf{x}}(q) = n_q^{-1} \sum_{i: z_i = q} \mathbf{x}_i$ be the covariate mean vector for units that receive treatment combination q . For $1 \leq f \leq F$, the L dimensional difference-in-means vector of covariates with respect to the f th factorial effect is

$$(2.2) \quad \hat{\mathbf{t}}_{\mathbf{x},f} = \frac{2}{Q} \sum_{q=1}^Q g_{fq} \hat{\mathbf{x}}(q) = \frac{1}{2^{K-1}} \sum_{q: g_{fq}=1} \hat{\mathbf{x}}(q) - \frac{1}{2^{K-1}} \sum_{q: g_{fq}=-1} \hat{\mathbf{x}}(q).$$

Let $\hat{\mathbf{t}}_{\mathbf{x}} = (\hat{\mathbf{t}}'_{\mathbf{x},1}, \dots, \hat{\mathbf{t}}'_{\mathbf{x},F})'$ be the LF dimensional column vector of the difference-in-means of covariates with respect to all factorial effects. Although $\hat{\mathbf{t}}_{\mathbf{x}}$ has mean zero under the CRFE, for a realized value of $\mathbf{Z}$, 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 $\hat{\mathbf{t}}_{\mathbf{x}}$ 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 $\mathbf{A} \in \mathbb{R}^{m \times m}$ with rank p_0 , and a positive integer $p \geq p_0$, we use $\mathbf{A}_p^{1/2} \in \mathbb{R}^{m \times p}$ to denote a matrix such that $\mathbf{A}_p^{1/2} (\mathbf{A}_p^{1/2})' = \mathbf{A}$. Specifically, if $\mathbf{A} = \mathbf{\Gamma} \mathbf{\Lambda}^2 \mathbf{\Gamma}'$ is the eigen-decomposition of $\mathbf{A}$ where $\mathbf{\Gamma} \in \mathbb{R}^{m \times p_0}$, $\mathbf{\Gamma}' \mathbf{\Gamma} = \mathbf{I}_{p_0}$ and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_{p_0})$, then we can choose $\mathbf{A}_p^{1/2} = (\mathbf{\Gamma} \mathbf{\Lambda}, \mathbf{0}_{m \times (p-p_0)})$. The choice of $\mathbf{A}_p^{1/2}$ is generally not unique. In the special case with $p = m$, we use $\mathbf{A}^{1/2}$ to denote the unique positive-semidefinite matrix satisfying the definition of $\mathbf{A}_m^{1/2}$. We use $\otimes$ for the Kronecker product of two matrices, and $\circ$ for elementwise multiplications of vectors. We say a matrix $\mathbf{M}_1$ is smaller than or equal to $\mathbf{M}_2$ and write as $\mathbf{M}_1 \leq \mathbf{M}_2$,

if $\mathbf{M}_2 - \mathbf{M}_1$ 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 $\sim$ 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, $\mathbf{S}_{\tau\tau} = (n-1)^{-1} \sum_{i=1}^n (\boldsymbol{\tau}_i - \boldsymbol{\tau})(\boldsymbol{\tau}_i - \boldsymbol{\tau})'$ for factorial effects, $\mathbf{S}_{xx} = (n-1)^{-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$ for covariates and $\mathbf{S}_{q,x} = \mathbf{S}'_{x,q} = (n-1)^{-1} \sum_{i=1}^n \{Y_i(q) - \bar{Y}(q)\} (\mathbf{x}_i - \bar{\mathbf{x}})'$ for potential outcomes and covariates. The covariance $\mathbf{S}_{xx}$ is known without any uncertainty. However, other variances or covariances (e.g., S_{qk} , $\mathbf{S}_{\tau\tau}$ and $\mathbf{S}_{q,x}$) 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 $\hat{Y}(q) = n_q^{-1} \sum_{i: Z_i=q} Y_i^{\text{obs}}$ be the average observed outcome under treatment combination q . For $1 \leq f \leq F$, the difference-in-means estimator for the f th average factorial effect is

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

Let $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \dots, \hat{\tau}_F)'$ 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 $\mathbf{Z}$. 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 $n \rightarrow \infty$, the sequence of finite populations satisfies that for each $1 \leq q \neq k \leq Q$:

- (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} , $\mathbf{S}_{xx}$ and $\mathbf{S}_{q,x}$ have limiting values, and $\mathbf{S}_{xx}$ and its limit are nondegenerate,
- (iii) $\max_{1 \leq i \leq n} |Y_i(q) - \bar{Y}(q)|^2/n \rightarrow 0$ and $\max_{1 \leq i \leq n} \|\mathbf{x}_i - \bar{\mathbf{x}}\|_2^2/n \rightarrow 0$.

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

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

Under the CRFE and Condition 1, $(\hat{\boldsymbol{\tau}}' - \boldsymbol{\tau}', \hat{\boldsymbol{\tau}}'_x)' \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$.

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 f th diagonal element of $\mathbf{S}_{\tau\tau}$, 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 f th diagonal element of $\mathbf{V}_{\tau\tau}$. Then $\hat{\tau}_f$ is unbiased for τ_f with sampling variance $V_{\tau_f \tau_f}$, and $\hat{\tau}_f - \tau_f \sim \mathcal{N}(0, V_{\tau_f \tau_f})$. Moreover, $\mathbf{S}_{\tau\tau}$ cannot be unbiasedly estimated from the observed data, and it equals $\mathbf{0}$ 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 τ_i is a constant vector for all units, or, equivalently, $\mathbf{S}_{\tau\tau} = \mathbf{0}$.

Under the CRFE, the observed sample variance $s_{qq} = (n_q - 1)^{-1} \sum_{i: Z_i=q} \{Y_i^{\text{obs}} - \hat{Y}(q)\}^2$ 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 $\mathbf{V}_{\tau\tau}$ by $2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q' s_{qq}$, and then construct Wald-type confidence sets for τ . 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 τ .

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

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

as analogues of the sampling covariance $\mathbf{V}_{\tau\tau}$ 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 $\mathbf{V}_{\tau\tau} = \mathbf{V}_{\tau\tau}^{\parallel} + \mathbf{V}_{\tau\tau}^{\perp}$.

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

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

$$\begin{aligned} \text{Cov}(\mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1} \hat{\tau}_x) &= \mathbf{V}_{\tau\tau}^{\parallel}, & \text{Cov}(\hat{\tau} - \tau - \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1} \hat{\tau}_x) &= \mathbf{V}_{\tau\tau}^{\perp}, \\ \text{Cov}(\mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1} \hat{\tau}_x, \hat{\tau} - \tau - \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1} \hat{\tau}_x) &= \mathbf{0}. \end{aligned}$$

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 f th diagonal elements of $\mathbf{V}_{\tau\tau}^{\parallel}$ and $\mathbf{S}_{\tau\tau}^{\parallel}$, 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 $\hat{\tau}_x$ has the following equivalent forms:*

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

It reduces to $R_f^2 = S_{11}^{\parallel}/S_{11}$, the finite population squared multiple correlation between $Y(1)$ and 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 $\hat{\tau}_x$ 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 $\hat{\tau}_x$ to be “small.”

4.1. Mahalanobis distance criterion. A measure of the magnitude of $\hat{\tau}_x$ is the Mahalanobis distance $M \equiv \hat{\tau}_x' V_{xx}^{-1} \hat{\tau}_x$. 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 $\mathbf{Z}$ from the CRFE if and only if $M \leq a$. 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 $\mathbf{Z}$ satisfies this criterion. From Proposition 1, M is asymptotically χ_{LF}^2 , and therefore the asymptotic acceptance probability is $p_a = P(\chi_{LF}^2 \leq a)$ 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 $\hat{\tau}$ under ReFM. Rerandomization in the design stage accepts only the treatment assignments resulting in covariate balance, which consequently changes the sampling distribution of $\hat{\tau}$. Understanding the asymptotic sampling distribution of $\hat{\tau}$ is crucial for conducting the classical repeated sampling inference of τ . Intuitively, $\hat{\tau}$ has two parts: one part is orthogonal to $\hat{\tau}_x$ and thus unaffected by ReFM, and the other part is the linear projection onto $\hat{\tau}_x$ and thus affected by ReFM. Let $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_F)$ be an F dimensional standard Gaussian random vector, and $\boldsymbol{\zeta}_{LF,a} \sim \mathbf{D} \mid \mathbf{D}'\mathbf{D} \leq a$ be an LF dimensional truncated Gaussian random vector, where $\mathbf{D} = (D_1, \dots, D_{LF})' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{LF})$. In addition, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\zeta}_{LF,a}$ are independent. The following theorem shows the asymptotic sampling distribution of $\hat{\tau}$.

THEOREM 2. *Under ReFM and Condition 1,*

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

Theorem 2 holds because the sampling distribution of $\hat{\tau}$ under rerandomization is the same as the conditional distribution of $\hat{\tau}$ given $M \leq a$. Its proof is in Appendix A3 of the Supplementary Material (Li, Ding and Rubin (2019)). We emphasize that, although the matrix $(\mathbf{V}_{\tau\tau}^{\parallel})_{LF}^{1/2}$ may not be unique, the asymptotic sampling distribution (4.1) is. Therefore,

the asymptotic sampling distribution of $\hat{\tau} - \tau$ under ReFM depends only on L , F , a , $V_{\tau\tau}^\perp$ and $V_{\tau\tau}^\parallel$. Theorem 2 immediately implies the asymptotic sampling distribution of a single factorial effect estimator. Let $\varepsilon_0 \sim \mathcal{N}(0, 1)$, $\eta_{LF,a} \sim D_1 \mid \mathbf{D}'\mathbf{D} \leq a$ be the first coordinate of $\zeta_{LF,a}$, and ε and $\eta_{LF,a}$ be independent.

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

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

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 $\mathcal{K} = \{-\mathbf{a} : \mathbf{a} \in \mathcal{K}\}$. 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 $\hat{\tau} - \tau$ under ReFM.

PROPOSITION 2. *The standard Gaussian random vector $\boldsymbol{\varepsilon}$, the truncated Gaussian random vector $\zeta_{LF,a}$ 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 $\zeta_{LF,a}$ 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 $\hat{\tau}$.* In this subsection, we further represent (4.1) using well-known distributions to gain more insights. Let $\chi_{LF,a}^2 \sim \chi_{LF}^2 \mid \chi_{LF}^2 \leq a$ be a truncated χ^2 random variable, $\mathbf{S}$ 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 $v_{LF,a} = P(\chi_{LF+2}^2 \leq a) / P(\chi_{LF}^2 \leq a) \leq 1$.

PROPOSITION 3. $\boldsymbol{\zeta}_{LF,a}$ is spherically symmetric with covariance $v_{LF,a} \mathbf{I}_{LF}$. It follows $\boldsymbol{\zeta}_{LF,a} \sim \chi_{LF,a} \cdot \mathbf{S} \circ \sqrt{\boldsymbol{\beta}}$, where $(\chi_{LF,a}, \mathbf{S}, \boldsymbol{\beta})$ 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 $\mathbf{V}_{\tau\tau}$ is invertible whenever we mention its inverse; otherwise we can focus on a lower dimensional linear transformation of $\hat{\tau}$ (Li, Ding and Rubin (2019)). Let $\mathbf{R} = \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau\tau}^{\parallel} \mathbf{V}_{\tau\tau}^{-1/2}$ be the matrix measuring the relative sampling covariance of $\hat{\tau}$ explained by $\hat{\tau}_x$, and $\mathbf{R} = \mathbf{\Gamma} \mathbf{\Pi}^2 \mathbf{\Gamma}'$ be its eigen-decomposition, where $\mathbf{\Gamma} \in \mathbb{R}^{F \times F}$ is an orthogonal matrix and $\mathbf{\Pi}^2 = \text{diag}(\pi_1^2, \dots, \pi_F^2) \in \mathbb{R}^{F \times F}$ 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 $\hat{\tau}$ and $\hat{\tau}_x$ under the CRFE, which measure the association between the potential outcomes and covariates. Under the additivity in Definition 1, $\pi_1^2 = \dots = \pi_F^2 = S_{11}^{\parallel} / S_{11}$. 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

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

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 $(\mathbf{I}_3 - \mathbf{\Pi}^2)^{1/2} \boldsymbol{\varepsilon} + \mathbf{\Pi} \boldsymbol{\zeta}_{3,a}$, which depends on $\mathbf{\Pi}^2 = \text{diag}(\pi_1^2, \pi_2^2, \pi_3^2)$ and $p_a = P(\chi_3^2 \leq a)$. 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 $\mathbf{V}_{\tau\tau}^{\perp}$ is positive definite, $\mathbf{I}_F - \mathbf{R} = \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau\tau}^{\perp} \mathbf{V}_{\tau\tau}^{-1/2}$ 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.

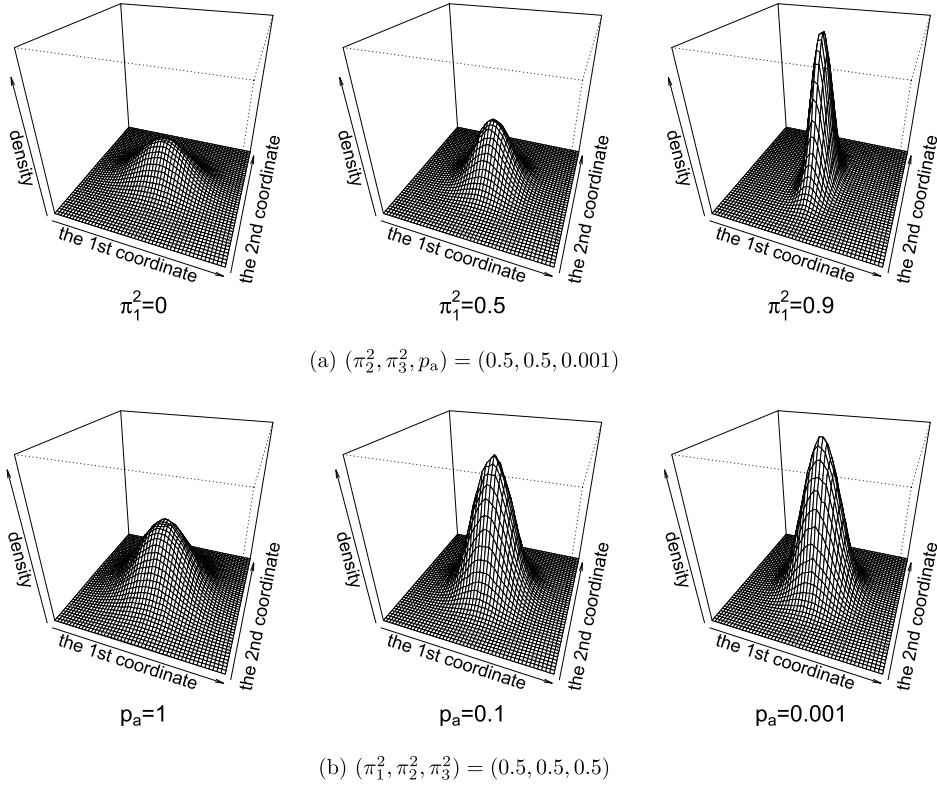


FIG. 1. Joint density of the first two coordinates of $(\mathbf{I}_3 - \mathbf{\Pi}^2)^{1/2}\boldsymbol{\varepsilon} + \mathbf{\Pi}\boldsymbol{\zeta}_{3,a}$.

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 $(1 - v_{LF,a})n\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel}$. Specifically, the percentage reduction in asymptotic sampling variance (PRIASV) of $\hat{\boldsymbol{\tau}}_f$ 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 $(1 - v_{LF,a})\lim_{n \rightarrow \infty}(n\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^{\parallel})$ and $(1 - v_{LF,a})\lim_{n \rightarrow \infty}R_f^2$. 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 $\hat{\boldsymbol{\tau}}_f$ reduces to $V_{\boldsymbol{\tau}_f\boldsymbol{\tau}_f}(1 - R_f^2)$, 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 ϕ and ψ in $\mathbb{R}^m$, we say that ϕ is more peaked than ψ and write as $\phi \succ \psi$, if $P(\phi \in \mathcal{K}) \geq P(\psi \in \mathcal{K})$ 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 ϕ and ψ in $\mathbb{R}^m$ with finite second moments, if $\phi \succ \psi$, then $\text{Cov}(\phi) \leq \text{Cov}(\psi)$.

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 ϕ and ψ , $\text{Cov}(\phi) \leq \text{Cov}(\psi)$ also implies $\phi \succ \psi$. 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 $\text{Var}(\xi_0) < 1 = \text{Var}(\varepsilon_0)$. 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 $\hat{\tau} - \tau$ under ReFM is more peaked than that under the CRFE.

Theorem 4 holds because the truncated Gaussian random vector $\xi_{LF,a}$ 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 $\hat{\tau}$ is smaller under ReFM than under the CRFE. Second, Theorem 4 shows that asymptotically, $\hat{\tau} - \tau$ has larger probability to be in any symmetric convex set under ReFM than under the CRFE. For a positive definite matrix $\Lambda \in \mathbb{R}^{p \times p}$ and $c \geq 0$, let $\mathcal{O}(\Lambda, c) \equiv \{\mu : \mu' \Lambda^{-1} \mu \leq c\}$. The following theorem implies that, for the special class of symmetric convex sets, $\{\mathcal{O}(V_{\tau\tau}, c) : c \geq 0\}$, the asymptotic probability that $\hat{\tau} - \tau$ lies in $\mathcal{O}(V_{\tau\tau}, c)$ 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 $\lim_{n \rightarrow \infty} P\{\hat{\tau} - \tau \in \mathcal{O}(V_{\tau\tau}, c_{1-\alpha}) \mid \mathcal{M}\} = 1 - \alpha$ 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 $\mathcal{O}(V_{\tau\tau}, c_{1-\alpha})$ in Theorem 5 is a $1 - \alpha$ asymptotic quantile region of $\hat{\tau} - \tau$ 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 $\mathcal{O}(V_{\tau\tau}, c_{1-\alpha})$ of $\hat{\tau} - \tau$.

Moreover, we can establish similar conclusions as Theorems 4 and 5 for any linear transformation of $\hat{\tau}$. 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 $C \in \mathbb{R}^{p \times m}$, if $\phi \succ \psi$, then $C\phi \succ C\psi$; (ii) the asymptotic sampling distribution of any linear transformation of $\hat{\tau}$ has the same form as $\hat{\tau}$, 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 $1 \leq f \leq F$ 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 $\hat{\tau}$ under ReFM depends on $V_{\tau\tau}^\perp$ and $(V_{\tau\tau}^\parallel)_{LF}^{1/2} = V_{\tau x} V_{xx}^{-1/2}$, which further depend on $S_{qq}^\perp$, $S_{\tau\tau}^\perp$ and $S_{q,x} S_{xx}^{-1/2}$. Under treatment combination q , define s_{qq} as the sample variance of observed outcomes, $s_{q,x}$ as the sample covariance between observed outcomes and covariates, $s_{xx}(q)$ as the sample covariance of covariates and $s_{qq}^\perp = s_{qq} - s_{q,x} s_{xx}^{-1}(q) s_{x,q}$ as the sample variance of the residuals from the linear projection of observed outcomes on covariates. We estimate $V_{\tau\tau}^\perp$ by

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

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

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

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 $\hat{\tau}_x$ are the same for all f : $R_1^2 = \dots = R_F^2 = S_{11}^{\parallel}/S_{11}$. From Section 4.5, under the additivity in Definition 1, the improvement of the f th 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 $\gamma_{fk}^2 = \text{Corr}^2(\hat{\tau}_f, \hat{\tau}_{x,k})$. When the f th factorial effect is more important, we would like to put more restriction on the difference-in-means vector $\hat{\tau}_{x,k}$ 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 $\max_{1 \leq k \leq F} \gamma_{fk}^2 = \gamma_{ff}^2 = R_f^2 = S_{11}^{\parallel}/S_{11}$ for $1 \leq f \leq F$. The squared multiple partial correlation between $\hat{\tau}_f$ and $\hat{\tau}_x$ given $\hat{\tau}_{x,f}$ is zero, that is, the residuals from the linear projections of $\hat{\tau}_f$ and $\hat{\tau}_x$ on $\hat{\tau}_{x,f}$ are uncorrelated. If further $n_1 = \dots = n_Q = n/Q$, 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, $\hat{\tau}_x$ explains $\hat{\tau}_f$ in the linear projection only through $\hat{\tau}_{x,f}$. 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 $1 \leq h \leq H$, let $\hat{\tau}_x[\mathcal{F}_h]$ be the subvector of $\hat{\tau}_x$, consisting of the difference-in-means of covariates $\hat{\tau}_{x,f}$ 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 $\hat{\tau}_x[\mathcal{F}_h]$. However, due to the correlations among the $\hat{\tau}_x[\mathcal{F}_h]$'s, restrictions on one also restrict others. For example, balancing $\hat{\tau}_x[\mathcal{F}_1]$ partially balances $\hat{\tau}_x[\mathcal{F}_2]$. 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 $\tilde{\mathbf{B}} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q'$. From Proposition 1, the sampling covariance of $\hat{\tau}_x$ under the CRFE, $\mathbf{V}_{xx} = \tilde{\mathbf{B}} \otimes \mathbf{S}_{xx}$, contains two components: $\tilde{\mathbf{B}}$ determined by the coefficient vector $\mathbf{b}_q$'s and $\mathbf{S}_{xx}$ determined by the covariates. Below we introduce a blockwise

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

$$(5.1) \quad \mathbf{c}_q[h] = \mathbf{b}_q[\mathcal{F}_h] - \tilde{\mathbf{B}}[\mathcal{F}_h, \mathcal{F}_{\bar{h}-1}] \{ \tilde{\mathbf{B}}[\mathcal{F}_{\bar{h}-1}, \mathcal{F}_{\bar{h}-1}] \}^{-1} \mathbf{b}_q[\mathcal{F}_{\bar{h}-1}].$$

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

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

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

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

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

and $\text{Cov}(\hat{\boldsymbol{\theta}}_{\mathbf{x}}[h], \hat{\boldsymbol{\theta}}_{\mathbf{x}}[\tilde{h}]) = \mathbf{0}$ 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, $(\hat{\boldsymbol{\theta}}_{\mathbf{x}}[1], \dots, \hat{\boldsymbol{\theta}}_{\mathbf{x}}[H])$ are mutually uncorrelated under the CRFE, and thus are essentially from a blockwise Gram–Schmidt orthogonalization of $(\hat{\boldsymbol{\tau}}_{\mathbf{x}}[\mathcal{F}_1], \dots, \hat{\boldsymbol{\tau}}_{\mathbf{x}}[\mathcal{F}_H])$. We define the Mahalanobis distance in tier h as

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

Let $(a_1, \dots, a_H)$ be H positive constants predetermined in the design stage. Under rerandomization with tiers of factorial effects, denoted by $\text{ReFMT}_{\mathbf{F}}$, a randomization is acceptable if and only if $M_h \leq a_h$ for all $1 \leq h \leq H$. Below we use $\mathcal{T}_{\mathbf{F}}$ to denote the event that the treatment vector $\mathbf{Z}$ satisfies this criterion. From the finite population central limit theorem, asymptotically, M_h is $\chi^2_{L_{F_h}}$, and $(M_1, \dots, M_H)$ are jointly independent. Therefore, the asymptotic acceptance probability under $\text{ReFMT}_{\mathbf{F}}$ is $p_a = \prod_{h=1}^H P(\chi^2_{L_{F_h}} \leq a_h)$. 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 $n_1 = \dots = n_Q = n/Q$, the coefficient $\mathbf{c}_q[h]$ in (5.1) reduces to $\mathbf{b}_q[\mathcal{F}_h]$, the difference-in-means of covariates $\hat{\boldsymbol{\theta}}_{\mathbf{x}}[h]$ in (5.2) reduces to $\hat{\boldsymbol{\tau}}_{\mathbf{x}}[\mathcal{F}_h]$ and the Mahalanobis distance M_h in (5.3) reduces to $M_h = n/4 \cdot \sum_{f \in \mathcal{F}_h} \hat{\boldsymbol{\tau}}'_{\mathbf{x},f} \mathbf{S}_{\mathbf{x}\mathbf{x}}^{-1} \hat{\boldsymbol{\tau}}_{\mathbf{x},f}$.*

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

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

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 $W_{\tau_f \tau_f}^{\parallel}[h]$ be the f th diagonal element of $\mathbf{W}_{\tau\tau}^{\parallel}[h]$. The squared multiple correlation between $\hat{\tau}_f$ and $\hat{\theta}_x[h]$ under the CRFE is then $\rho_f^2[h] = W_{\tau_f \tau_f}^{\parallel}[h] / V_{\tau_f \tau_f}$. 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 $S_{11}^{\parallel} / S_{11}$ if $f \in \mathcal{F}_h$, and zero otherwise. Because the $\hat{\theta}_x[h]$'s are from a blockwise Gram–Schmidt orthogonalization of $\hat{\tau}_x$, the squared multiple correlation between $\hat{\tau}_f$ and $\hat{\tau}_x$ 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 $\varepsilon_0 \sim \mathcal{N}(0, 1)$, $\eta_{LF_h, a_h} \sim D_{h1} \mid \mathbf{D}_h' \mathbf{D}_h \leq a_h$ 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 $1 \leq f \leq F$,*

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

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 $\hat{\tau}$ under ReFMT_F is also central convex unimodal. The symmetry of the asymptotic sampling distributions ensures that the factorial effect estimator $\hat{\tau}$ is consistent for τ 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 $\hat{\tau}$ under ReFMT_F and the CRFE. For each $1 \leq h \leq H$, let $v_{LF_h, a_h} = P(\chi_{LF_h+2}^2 \leq a_h) / P(\chi_{LF_h}^2 \leq a_h) \leq 1$.

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

THEOREM 9. *Under Condition 1, the asymptotic sampling distribution of $\hat{\tau} - \tau$ under 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 $\mathcal{O}(\mathbf{V}_{\tau\tau}, c)$. Unfortunately, considering joint quantile region for τ is technically challenging in general, and we consider the case where the following condition holds.

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

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

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

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

PROPOSITION 8. *Under the additivity in Definition 1, Condition 2 holds with orthogonal matrix $\mathbf{\Gamma} = \tilde{\mathbf{B}}^{-1/2} \mathbf{\Psi} \tilde{\mathbf{C}}^{1/2}$, and the canonical correlations between $\hat{\tau}$ and $\hat{\theta}_x[h]$ 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 ReFMT_F , assume that Conditions 1 and 2 hold. Let $c_{1-\alpha}$ be the solution of $\lim_{n \rightarrow \infty} P\{\hat{\tau} - \tau \in \mathcal{O}(\mathbf{V}_{\tau\tau}, c_{1-\alpha}) \mid \mathcal{F}_F\} = 1 - \alpha$. 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 $1 \leq h \leq H$ and $1 \leq f \leq F$.*

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 $\hat{\tau}$ has an asymptotic sampling distribution of the same form as $\hat{\tau}$, we can establish similar conclusions as Theorems 9 and 10 for any linear transformations of $\hat{\tau}$. 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.

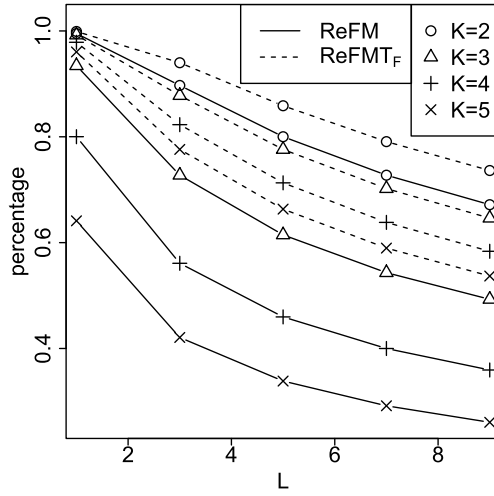


FIG. 2. PRIASV of main effect estimators divided by R_f^2 .

COROLLARY 6. Under Condition 1, for any $1 \leq f \leq F$ 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_F is smaller than or equal to that under the CRFE, and is nonincreasing in $\rho_f^2[h]$ for $1 \leq h \leq H$.

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_F yields more percentage reductions of quantile ranges.

The example below shows the advantage of ReFMT_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 $1 \leq k \leq K$. We compare two rerandomization schemes with the same asymptotic acceptance probability: ReFM with $p_a = 0.001$ and ReFMT_F with thresholds (a_1, a_2) satisfying $P(\chi_{L F_1}^2 \leq a_1) = 0.002$ and $P(\chi_{L F_2}^2 \leq a_2) = 0.5$. 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_F increases as the numbers of factors and covariates increase.

5.5. Conservative covariance estimator and confidence sets under ReFMT_F . We estimate $V_{\tau\tau}^\perp$ by $\hat{V}_{\tau\tau}^\perp$ in (4.4), $W_{\tau x}[h]$ by $\hat{W}_{\tau x}[h] = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{b}_q \mathbf{c}'_q[h]) \otimes \{\mathbf{s}_{q,x} \mathbf{s}_{xx}^{-1/2}(q) \mathbf{S}_{xx}^{1/2}\}$, and $(W_{\tau\tau}^\parallel[h])_{L F_h}^{1/2}$ by $\hat{W}_{\tau x}[h] (W_{xx}[h])^{-1/2}$. We can then obtain a covariance estimator and construct confidence sets for τ or its linear transformations. Similar to ReFM , for a parameter of interest $C\tau$, we recommend confidence sets of the form $C\hat{\tau} + \mathcal{O}(C\hat{V}_{\tau\tau}^\perp C', c)$, 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 $(C\phi)'(C\hat{V}_{\tau\tau}^\perp C')^{-1}(C\phi)$ with ϕ following the estimated asymptotic sampling distribution of $\hat{\tau} - \tau$ under ReFMT_F .

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

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 $P(\chi_{LF_1}^2 \leq a_1) = 0.002$ and $P(\chi_{LF_2}^2 \leq a_2) = 0.5$. 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_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

Factorial effect	Reduction in variance		Reduction in quantile range	
	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_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 $p_{ah} \equiv P(\chi_{L_{F_h}}^2 \leq a_h)$ 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).

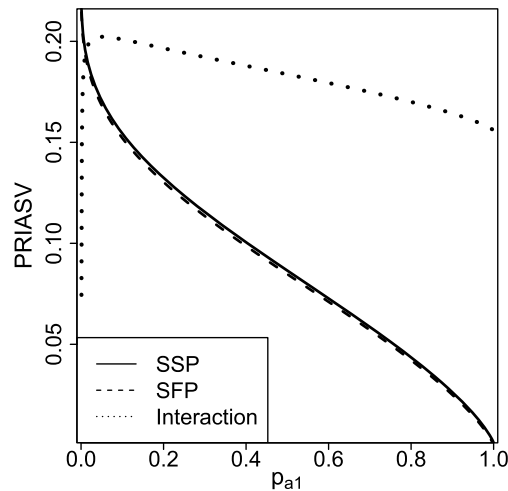


FIG. 3. PRIASVs of all factorial effects as p_{a1} , with $p_a = p_{a1}p_{a2}$ fixed at 0.001.

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](https://doi.org/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](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](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](https://doi.org/10.1214/aoms/1177730293)
- BIRNBAUM, Z. W. (1948). On random variables with comparable peakedness. *Ann. Math. Stat.* **19** 76–81. [MR0024099 https://doi.org/10.1214/aoms/1177730293](https://doi.org/10.1214/aoms/1177730293)
- BOSE, R. C. (1947). Mathematical theory of the symmetrical factorial design. *Sankhyā* **8** 107–166. [MR0026781](https://doi.org/10.1214/aoms/1177730293)
- 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](https://doi.org/10.1214/aoms/1177730293)
- 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](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](https://doi.org/10.1214/aoms/1177730293)
- 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](https://doi.org/10.1214/aoms/1177730293)
- 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](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](https://doi.org/10.1214/aoms/1177730293)
- DHARMADHIKARI, S. W. and JOGDEO, K. (1976). Multivariate unimodality. *Ann. Statist.* **4** 607–613. [MR0415719](https://doi.org/10.1214/aoms/1177730293)
- 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](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](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](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.
- YODEN, W. J. (1972). Randomization and experimentation. *Technometrics* **14** 13–22.

SUPPLEMENTARY MATERIAL TO “RERANDOMIZATION IN 2^K FACTORIAL EXPERIMENTS”

BY XINRAN LI^{*}, PENG DING[†] AND DONALD B. RUBIN[‡]

University of Pennsylvania^{}, University of California, Berkeley[†], Tsinghua
University and Temple University[‡]*

Appendix [A1](#) studies the theoretical properties of 2^K rerandomized factorial experiments with tiers of both covariates and factorial effects.

Appendix [A2](#) shows the sampling covariances, asymptotic sampling distributions, and squared multiple correlations under the CRFE. It includes the proofs of Theorem [1](#), Corollary [1](#), and Propositions [1](#), [5–7](#) and [A1](#).

Appendix [A3](#) proves the asymptotic sampling distributions of $\hat{\tau}$ under ReFM, ReFMT_F and ReFMT_{CF}. It includes the proofs of Theorems [2](#), [7](#) and [A1](#), Corollaries [2](#), [3](#), [5](#) and [A1](#), and Propositions [2](#) and [3](#).

Appendix [A4](#) compares the asymptotic sampling covariances of $\hat{\tau}$ under rerandomizations and the CRFE. It includes the proofs of Theorems [3](#), [8](#) and [A2](#).

Appendix [A5](#) compares the peakedness of the asymptotic sampling distributions under rerandomizations and the CRFE. It includes the proofs of Theorems [4](#), [5](#), [9](#) and [10](#), Corollaries [4](#) and [6](#), and Propositions [4](#), [8](#) and [A2](#).

Appendix [A6](#) proves the asymptotic conservativeness of covariance estimators and symmetric convex confidence sets under rerandomizations. It includes the proof of Theorem [6](#).

A1. Tiers of both covariates and factorial effects.

A1.1. *Tiers and orthogonalized covariates.* When covariates have varying importance for the potential outcomes, [Morgan and Rubin \[2015\]](#) proposed rerandomization with tiers of covariates in treatment-control experiments. It is important to consider the tiers of covariates with tiers of factorial effects. We first partition the covariates into T tiers with decreasing importance, and use $\mathbf{x}_i[t]$ to denote the L_t dimensional covariates in tier t . Let $\mathbf{x}_i[\bar{t}] = (\mathbf{x}_i[1], \dots, \mathbf{x}_i[t])$ be the covariates in the first t tiers, $\mathbf{S}_{\mathbf{x}[\bar{t}]\mathbf{x}[\bar{t}]}$ be the finite population covariance of the covariates in the first t tiers, and $\mathbf{S}_{\mathbf{x}[t],\mathbf{x}[\overline{t-1}]}$ be the finite population covariance between $\mathbf{x}[t]$ and $\mathbf{x}[\overline{t-1}]$. We then apply a block-wise Gram–Schmidt orthogonalization to the $\mathbf{x}_i[t]$ ’s: $\mathbf{e}_i[1] = \mathbf{x}_i[1]$,

and

$$\mathbf{e}_i[t] = \mathbf{x}_i[t] - \mathbf{S}_{\mathbf{x}[t], \mathbf{x}[t-1]} \mathbf{S}_{\mathbf{x}[t-1], \mathbf{x}[t-1]}^{-1} \mathbf{x}_i[t-1], \quad (2 \leq t \leq T)$$

where $\mathbf{e}_i[t]$ is the residual from the linear projection of the covariates $\mathbf{x}_i[t]$ in tier t onto the space spanned by the covariates in previous tiers. Li et al. [2018] call $\mathbf{e}_i[t]$ the orthogonalized covariates in tier t .

A1.2. Tiers of covariates and factorial effects criterion. Using the notation in Section 5, we partition the F factorial effects into H tiers with decreasing importance, i.e., $\mathcal{F} = \bigcup_{h=1}^H \mathcal{F}_h$, and the corresponding block-wise Gram-Schmidt orthogonalization of $(\mathbf{b}_q[\mathcal{F}_1], \dots, \mathbf{b}_q[\mathcal{F}_H])$ is $(\mathbf{c}_q[1], \dots, \mathbf{c}_q[H])$ defined in (5.1). Let $\hat{\mathbf{e}}_{[t]}(q) = n_q^{-1} \sum_{i: Z_i=q} \mathbf{e}_i[t]$ be the mean of the orthogonalized covariates in tier t under treatment combination q . The difference-in-means of orthogonalized covariates in tier t with coefficients $\mathbf{c}_q$'s is

$$(A1) \quad \hat{\boldsymbol{\theta}}_{\mathbf{e}[t]} = \begin{pmatrix} \hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[1] \\ \vdots \\ \hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[H] \end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^Q \begin{pmatrix} \mathbf{c}_q[1] \\ \vdots \\ \mathbf{c}_q[H] \end{pmatrix} \otimes \hat{\mathbf{e}}_{[t]}(q), \quad (1 \leq t \leq T).$$

Let $\mathbf{S}_{\mathbf{e}[t]\mathbf{e}[t]}$ be the finite population covariance of the orthogonalized covariates in tier t , and $\mathbf{S}_{q,\mathbf{e}[t]}$ be the finite population covariance between the $Y_i(q)$'s and $\mathbf{e}_i[t]$'s.

PROPOSITION A1. *Under the CRFE, $(\hat{\boldsymbol{\tau}}' - \boldsymbol{\tau}', \hat{\boldsymbol{\theta}}'_{\mathbf{e}[1]}, \dots, \hat{\boldsymbol{\theta}}'_{\mathbf{e}[T]})'$ has mean zero and sampling covariance:*

$$\begin{aligned} \text{Cov} \left(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}, \hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[h] \right) &= \mathbf{W}_{\boldsymbol{\tau}\mathbf{e}[t]}[h] = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{b}_q \mathbf{c}'_q[h]) \otimes \mathbf{S}_{q,\mathbf{e}[t]}, \\ \text{Cov} \left(\hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[h] \right) &= \mathbf{W}_{\mathbf{e}[t]\mathbf{e}[t]}[h] = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{c}_q[h] \mathbf{c}'_q[h]) \otimes \mathbf{S}_{\mathbf{e}[t]\mathbf{e}[t]}, \end{aligned}$$

and $\text{Cov}(\hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[h], \hat{\boldsymbol{\theta}}_{\mathbf{e}[\tilde{t}]}[\tilde{h}]) = \mathbf{0}$ if $t \neq \tilde{t}$ or $h \neq \tilde{h}$, i.e., $\hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[h]$'s are mutually uncorrelated.

From Proposition A1, we define the Mahalanobis distance for orthogonalized covariates in tier t with respect to factorial effects in tier h as

$$M_{t,h} = \hat{\boldsymbol{\theta}}'_{\mathbf{e}[t]}[h] (\mathbf{W}_{\mathbf{e}[t]\mathbf{e}[t]}[h])^{-1} \hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[h], \quad (1 \leq h \leq H, 1 \leq t \leq T).$$

TABLE A1
Mahalanobis distances: covariates and factorial effects in different tiers

		Factorial effects			
		Tier 1	Tier 2	...	Tier H
Covariates	Tier 1	$M_{1,1}$	$M_{1,2}$	...	$M_{1,H}$
	Tier 2	$M_{2,1}$	$M_{2,2}$	...	$M_{2,H}$
	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
	Tier T	$M_{T,1}$	$M_{T,2}$	...	$M_{T,H}$

Table A1 displays the Mahalanobis distances for all tiers of covariates and factorial effects. As discussed in Section 5, we should balance more the orthogonalized covariates with respect to more important factorial effects, e.g., putting decreasing restrictions on $M_{t,1}, M_{t,2}, \dots, M_{t,H}$ for a given t . With the pre-assumed importance of covariates, we should also balance more the covariates in more important tiers, e.g., putting decreasing restrictions on $M_{1,h}, M_{2,h}, \dots, M_{T,h}$ for a given h . Generally, the restrictions on the Mahalanobis distances in Table A1 should decrease from left to right and from top to bottom, i.e., if $h \leq \tilde{h}$ and $t \leq \tilde{t}$, we should put more restriction on $M_{t,h}$ than that on $M_{\tilde{t},\tilde{h}}$. This implies a partial order of importance on the set $\mathcal{S} = \{(t, h) : 1 \leq t \leq T, 1 \leq h \leq H\}$ of all combinations of tiers of covariates and factorial effects. In practice, we can divide the set $\mathcal{S}$ into J tiers $(\mathcal{S}_1, \dots, \mathcal{S}_J)$ with decreasing importance, which are coherent with the partial order on $\mathcal{S}$.

EXAMPLE A1. A choice of the $\mathcal{S}_j$'s is the triangular tiers, where $J = \min\{T, H\}$, and

$$\begin{aligned}\mathcal{S}_j &= \{(t, h) : h + t = j + 1, 1 \leq h \leq H, 1 \leq t \leq T\}, \quad (1 \leq j \leq J - 1) \\ \mathcal{S}_J &= \{(t, h) : h + t > J, 1 \leq h \leq H, 1 \leq t \leq T\}.\end{aligned}$$

Let $(a_1, \dots, a_J)$ be J positive constants predetermined in the design stage. Under rerandomized factorial experiments with tiers of covariates and factorial effects, denoted by ReFMT_{CF} , we accept only those treatment assignments with $\sum_{(t,h) \in \mathcal{S}_j} M_{t,h} \leq a_j$, for all $1 \leq j \leq J$. Below we use $\mathcal{T}_{\text{CF}}$ to denote the event that the treatment vector $\mathbf{Z}$ is accepted under ReFMT_{CF} . From the finite population central limit theorem, asymptotically, $M_{t,h}$ is $\chi_{L_t F_h}^2$, and the $M_{t,h}$'s are jointly independent. For each $1 \leq j \leq J$, let $\lambda_j = \sum_{(t,h) \in \mathcal{S}_j} L_t F_h$. The asymptotic acceptance probability under ReFMT_{CF} is then $p_a = \prod_{j=1}^J P(\chi_{\lambda_j}^2 \leq a_j)$.

By the same logic as Proposition 7, with equal treatment groups sizes, $M_{t,h}$ reduces to $M_{t,h} = n/4 \cdot \sum_{f \in \mathcal{F}_h} \hat{\tau}'_{e[t],f} \mathbf{S}_{e[t]e[t]}^{-1} \hat{\tau}_{e[t],f}$, where $\hat{\tau}_{e[t],f} =$

$2^{-(K-1)} \sum_{q=1}^Q g_{fq} \hat{\mathbf{e}}_{[t]}(q)$ is the difference-in-means of orthogonalized covariate $\mathbf{e}[t]$ with respect to the f th factorial effect.

A1.3. Asymptotic sampling distribution of $\hat{\boldsymbol{\tau}}$. For each tier $\mathcal{S}_j$, let $\mathbf{U}_{\boldsymbol{\tau}\mathbf{e}}[j] \in \mathbb{R}^{F \times \lambda_j}$ be a matrix consisting of the columns of matrices $\{\mathbf{W}_{\boldsymbol{\tau}\mathbf{e}[t]}[h]\}_{(t,h) \in \mathcal{S}_j}$, $\mathbf{U}_{\mathbf{e}\mathbf{e}}[j] \in \mathbb{R}^{\lambda_j \times \lambda_j}$ be a block diagonal matrix with $\{\mathbf{W}_{\mathbf{e}[t]\mathbf{e}[t]}[h]\}_{(t,h) \in \mathcal{S}_j}$ as the diagonal components, and $\mathbf{U}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel[j] = \mathbf{U}_{\boldsymbol{\tau}\mathbf{e}}[j](\mathbf{U}_{\mathbf{e}\mathbf{e}}[j])^{-1}\mathbf{U}_{\mathbf{e}\boldsymbol{\tau}}[j]$ be the sampling covariance matrix of $\hat{\boldsymbol{\tau}}$ explained by $\{\hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[h]\}_{(t,h) \in \mathcal{S}_j}$ in the linear projection under the CRFE. Let $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_F)$, and $\boldsymbol{\zeta}_{\lambda_j, a_j} \sim \mathbf{D}_j \mid \mathbf{D}_j' \mathbf{D}_j \leq a_j$, where $\mathbf{D}_j = (D_{j1}, \dots, D_{j, \lambda_j})' \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{\lambda_j})$. In addition, $(\boldsymbol{\varepsilon}, \boldsymbol{\zeta}_{\lambda_1, a_1}, \dots, \boldsymbol{\zeta}_{\lambda_J, a_J})$ are jointly independent.

THEOREM A1. *Under ReFMT_{CF} and Condition 1,*

$$(A2) \quad \hat{\boldsymbol{\tau}} - \boldsymbol{\tau} \mid \mathcal{T}_{CF} \quad \dot{\sim} \quad \left(\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp\right)^{1/2} \boldsymbol{\varepsilon} + \sum_{j=1}^J \left(\mathbf{U}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel[j]\right)_{\lambda_j}^{1/2} \boldsymbol{\zeta}_{\lambda_j, a_j}.$$

Let $U_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel[j]$ be the f th diagonal element of $\mathbf{U}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel[j]$, and $\beta_f^2[j] = U_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel[j]/V_{\boldsymbol{\tau}\boldsymbol{\tau}}$ be the proportion of variance of $\hat{\tau}_f$ explained by $\{\hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[h]\}_{(t,h) \in \mathcal{S}_j}$ in the linear projection. Because the $\{\hat{\boldsymbol{\theta}}_{\mathbf{e}[t]}[h]\}_{(t,h) \in \mathcal{S}_j}$'s are essentially from a block-wise Gram-Schmidt orthogonalization of $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$, we have $\sum_{j=1}^J \beta_f^2[j] = R_f^2$. Let $\varepsilon_0 \sim \mathcal{N}(0, 1)$, $\eta_{\lambda_j, a_j} \sim D_{j1} \mid \mathbf{D}_j' \mathbf{D}_j \leq a_j$ be the first coordinate of $\boldsymbol{\zeta}_{\lambda_j, a_j}$, and $(\varepsilon_0, \eta_{\lambda_1, a_1}, \dots, \eta_{\lambda_J, a_J})$ be jointly independent. Theorem A1 immediately implies the following asymptotic sampling distribution of a single factorial effect estimator.

COROLLARY A1. *Under ReFMT_{CF} and Condition 1, for $1 \leq f \leq F$,*

$$(A3) \quad \hat{\tau}_f - \tau_f \mid \mathcal{T}_{CF} \quad \dot{\sim} \quad \sqrt{V_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp} \left(\sqrt{1 - R_f^2} \cdot \varepsilon_0 + \sum_{j=1}^J \sqrt{\beta_f^2[j]} \cdot \eta_{\lambda_j, a_j} \right).$$

A1.4. Asymptotic unbiasedness, sampling covariance and peakedness. First, the asymptotic sampling distribution (A2) is central convex unimodal. Therefore, the factorial effect estimator $\hat{\boldsymbol{\tau}}$ is asymptotically unbiased for $\boldsymbol{\tau}$ under ReFMT_{CF} , and the difference-in-means of covariates with respect to any factorial effect has mean zero asymptotically.

Second, we consider the asymptotic sampling covariance of $\hat{\boldsymbol{\tau}}$ under ReFMT_{CF} .

THEOREM A2. *Under Condition 1, $\hat{\boldsymbol{\tau}}$ has smaller asymptotic sampling covariance matrix under ReFMT_{CF} than that under the CRFE, and the*

reduction in asymptotic sampling covariance is $n \sum_{j=1}^J (1 - v_{\lambda_j, a_j}) \mathbf{U}_{\tau\tau}^\parallel[j]$. For each $1 \leq f \leq F$, the PRIASV of $\hat{\tau}_f$ is $\sum_{j=1}^J (1 - v_{\lambda_j, a_j}) \beta_f^2[j]$.

Third, we consider the peakedness of (A2).

THEOREM A3. *Under Condition 1, the asymptotic sampling distribution of $\hat{\tau} - \tau$ under ReFMT_{CF} is more peaked than that under the CRFE.*

We then consider specific symmetric convex sets of form $\mathcal{O}(\mathbf{V}_{\tau\tau}, c)$. Similar to Section 5.4, due to some technical difficulties, we consider only the case under the following condition.

CONDITION A1. *There exists an orthogonal matrix $\mathbf{\Gamma} \in \mathbb{R}^{F \times F}$ such that*

$$\mathbf{\Gamma}' \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{U}_{\tau\tau}^\parallel[j] \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{\Gamma} = \text{diag}(\kappa_{j1}^2, \dots, \kappa_{jF}^2), \quad (1 \leq j \leq J)$$

where $(\kappa_{j1}^2, \dots, \kappa_{jF}^2)$ are the canonical correlations between $\hat{\tau}$ and $\{\hat{\theta}_{e[t]}[h]\}_{(t,h) \in \mathcal{S}_j}$ under the CRFE.

Condition A1 holds automatically when $J = 1$. Moreover, the following proposition shows that the additivity is a sufficient condition for Condition A1. Recall that Ψ is the linear transformation from $\mathbf{c}_q$ to $\mathbf{b}_q$, $\tilde{\mathbf{B}} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q'$, and $\tilde{\mathbf{C}} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{c}_q \mathbf{c}_q'$.

PROPOSITION A2. *Under the additivity in Definition 1, Condition A1 holds with orthogonal matrix $\mathbf{\Gamma} = \tilde{\mathbf{B}}^{-1/2} \Psi \tilde{\mathbf{C}}^{1/2}$.*

THEOREM A4. *Under ReFMT_{CF} , assume that Conditions 1 and A1 hold. Let $c_{1-\alpha}$ be the solution of $\lim_{n \rightarrow \infty} P\{\hat{\tau} - \tau \in \mathcal{O}(\mathbf{V}_{\tau\tau}, c_{1-\alpha}) \mid \mathcal{T}_{\text{CF}}\} = 1 - \alpha$. It depends only on K , λ_j 's, a_j 's, and $(\kappa_{j1}^2, \dots, \kappa_{jF}^2)$'s, and is nonincreasing in κ_{jf}^2 for $1 \leq j \leq J$ and $1 \leq f \leq F$.*

Because the peakedness relationship is invariant under linear transformations, and any linear transformation of $\hat{\tau}$ has asymptotic sampling distribution of the same form as $\hat{\tau}$, we can establish similar conclusions as Theorems A3 and A4 for any linear transformations of $\hat{\tau}$. We relegate the details to Appendix A5, and consider only the marginal asymptotic sampling distribution of a single factorial effect estimator here.

COROLLARY A2. *Under Condition 1, for any $1 \leq f \leq F$ 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_{CF} is smaller than or equal to that under the CRFE, and is nonincreasing in $\beta_f^2(j)$, for any $1 \leq j \leq J$.*

EXAMPLE A2. We consider again the setting in Example 3 in the main text, and further assume that the finite population partial covariance between potential outcome $Y(1)$ and other covariates given the first covariate is zero, i.e., only the first covariate is important. We divide the covariates into two tiers, where tier 1 contains only the first covariate and tier 2 contains the remaining covariates, and then construct triangular tiers in Example A1, where $\mathcal{S}_1$ consists of the combination of main effects and first covariate with $\lambda_1 = K$, and $\mathcal{S}_2$ consists of the remaining combinations of factorial effects and covariates with $\lambda_2 = (2^K - 1)L - K$. We choose thresholds (a_1, a_2) such that $P(\chi_{\lambda_1}^2 \leq a_1) = 0.002$ and $P(\chi_{\lambda_2}^2 \leq a_2) = 0.5$. Then for the main effect $1 \leq k \leq K$, $\beta_k^2[1] = R_k^2$ and $\beta_k^2[2] = 0$. Figure A1 shows the PRIASV, divided by R_f^2 , of the main effect estimators for ReFMT_F and ReFMT_{CF} . It shows that the advantage of further using tiers of covariates increases as numbers of factors and covariates increase. $\square$

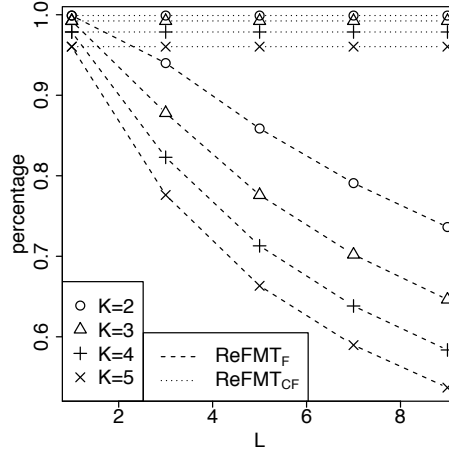


Fig A1: PRIASV of main effect estimators under ReFMT_F and ReFMT_{CF} , divided by R_f^2

A1.5. Conservative covariance estimator and confidence set under ReFMT_{CF} .

We define $\mathbf{s}_{q,e[t]}$ and $\mathbf{s}_{e[t]e[t]}(q)$ as the sample covariances of the observed outcomes and orthogonalized covariates under treatment combination q . We estimate $\mathbf{V}_{\tau\tau}^\perp$ by $\hat{\mathbf{V}}_{\tau\tau}^\perp$ in (4.4), $\mathbf{W}_{\tau e[t]}[h]$ by $\hat{\mathbf{W}}_{\tau e[t]}[h] = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{b}_q \mathbf{c}'_q[h]) \otimes \{ \mathbf{s}_{q,e[t]} \mathbf{s}_{e[t]e[t]}^{-1/2}(q) \mathbf{S}_{e[t]e[t]}^{1/2} \}$, $\mathbf{U}_{\tau e[j]}$ by $\hat{\mathbf{U}}_{\tau e[j]}$ consisting of the corresponding $\hat{\mathbf{W}}_{\tau e[t]}[h]$'s, and $(\mathbf{U}_{\tau\tau}^\parallel[j])_{\lambda_j}^{1/2} = \mathbf{U}_{\tau e[j]}(\mathbf{U}_{ee[j]})^{-1/2}$ by $\hat{\mathbf{U}}_{\tau e[j]}(\mathbf{U}_{ee[j]})^{-1/2}$. We can then obtain a covariance estimator and construct confidence sets

for τ or its lower dimensional linear transformation. Similar to ReFM, for a parameter of interest $C\tau$, we recommend to use confidence sets of the form $C\hat{\tau} + \mathcal{O}(C\hat{V}_{\tau\tau}^\perp C', c)$, and choose the threshold c by simulating the estimated asymptotic distribution. Let $\hat{c}_{1-\alpha}$ be the $1 - \alpha$ quantile of $(C\phi)'(C\hat{V}_{\tau\tau}^\perp C')^{-1}(C\phi)$ with ϕ following the estimated asymptotic sampling distribution of $\hat{\tau} - \tau$ under ReFMT_{CF} .

THEOREM A5. *Under ReFMT_{CF} and Condition 1, consider inferring $C\tau$, where C has full row rank. The probability limit of covariance estimator for $C\hat{\tau}$, $C\hat{V}_{\tau\tau}^\perp C' + \sum_{j=1}^J v_{\lambda_j, a_j} C\hat{U}_{\tau e}[j](U_{ee}[j])^{-1}\hat{U}_{e\tau}[j]C'$, is larger than or equal to the actual sampling covariance, and the $1 - \alpha$ confidence set, $C\hat{\tau} + \mathcal{O}(C\hat{V}_{\tau\tau}^\perp C', \hat{c}_{1-\alpha})$, has asymptotic coverage rate $\geq 1 - \alpha$, with equality holding if $S_{\tau\tau}^\perp \rightarrow 0$ as $n \rightarrow \infty$.*

The above confidence sets is similar to the ones based on regression adjustment if the threshold a_j 's are small [Lu, 2016]. Moreover, we will extend Theorem A5 to general symmetric convex confidence sets in Appendix A6.

A2. Sampling properties under the CRFE.

A2.1. Lemmas for matrices. For any positive integer m , we use $\mathbf{1}_m$ to denote an m dimensional column vector with all elements one, $\mathbf{I}_m$ to denote an $m \times m$ identity matrix, and $\mathbf{J}_m$ to denote an $m \times m$ matrix with all elements one.

LEMMA A1. *Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$ be four matrices.*

- (1) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$.
- (2) If $\mathbf{A}$ and $\mathbf{B}$ have the same dimension, then $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}$, and $\mathbf{C} \otimes (\mathbf{A} + \mathbf{B}) = \mathbf{C} \otimes \mathbf{A} + \mathbf{C} \otimes \mathbf{B}$.
- (3) If one can form the matrix products $\mathbf{AC}$ and $\mathbf{BD}$, then $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$.
- (4) If $\mathbf{A}$ and $\mathbf{B}$ are invertible, then $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$.

Recall that $\mathbf{g}_f = (g_{f1}, \dots, g_{fQ})'$ is the generating vector for the f th factorial effect, $\mathbf{b}_q = (g_{1q}, \dots, g_{Fq})'$ is the coefficient vector for the treatment combination q , and $\tilde{\mathbf{B}} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q'$.

LEMMA A2. *For any $1 \leq q, k \leq Q$,*

$$(A1) \quad 2^{-2(K-1)} n_q^{-1} n_k^{-1} \mathbf{b}_q' \tilde{\mathbf{B}}^{-1} \mathbf{b}_k = n_q^{-1} \times 1\{q = k\} - n^{-1}.$$

PROOF OF LEMMA A2. Let $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_Q) = (\mathbf{g}_1, \dots, \mathbf{g}_F)' \in \mathbb{R}^{F \times Q}$, and $\mathbf{N} = \text{diag}(n_1, \dots, n_Q) \in \mathbb{R}^{Q \times Q}$. Then $\tilde{\mathbf{B}} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q' = 2^{-2(K-1)} \mathbf{B} \mathbf{N}^{-1} \mathbf{B}'$, and $2^{-2(K-1)} n_q^{-1} n_k^{-1} \mathbf{b}_q' \tilde{\mathbf{B}}^{-1} \mathbf{b}_k$ is the (q, k) th element of matrix $\mathbf{C} = \mathbf{N}^{-1} \mathbf{B}' (\mathbf{B} \mathbf{N}^{-1} \mathbf{B}')^{-1} \mathbf{B} \mathbf{N}^{-1}$.

First, we show that there exists a constant c such that $\mathbf{C} = \mathbf{N}^{-1} + c \mathbf{J}_Q$. By definition, $\mathbf{B} \mathbf{C} = \mathbf{B} \mathbf{N}^{-1} \mathbf{B}' (\mathbf{B} \mathbf{N}^{-1} \mathbf{B}')^{-1} \mathbf{B} \mathbf{N}^{-1} = \mathbf{B} \mathbf{N}^{-1}$. Thus,

$$(A2) \quad \mathbf{0} = \mathbf{B}(\mathbf{C} - \mathbf{N}^{-1}) = (\mathbf{g}_1, \dots, \mathbf{g}_F)' (\mathbf{C} - \mathbf{N}^{-1}).$$

By the properties of generating vectors, $(\mathbf{1}_Q, \mathbf{g}_1, \dots, \mathbf{g}_F)$ constitute an orthogonal basis of $\mathbb{R}^Q$. Equation (A2) implies that each column of $\mathbf{C} - \mathbf{N}^{-1}$ is orthogonal to $\mathbf{g}_1, \dots, \mathbf{g}_F$, and thus has to be $c \mathbf{1}_{Q \times 1}$ for some constant c . Because $\mathbf{C} - \mathbf{N}^{-1}$ is a symmetric matrix, $\mathbf{C} - \mathbf{N}^{-1}$ must be $c \mathbf{J}_Q$ for some constant c .

Second, we show that $c = -n^{-1}$ and $\mathbf{C} = \mathbf{N}^{-1} - n^{-1} \mathbf{J}_Q$. On the one hand, $\mathbf{C} = \mathbf{N}^{-1} + c \mathbf{J}_Q$ implies

$$\text{tr}(\mathbf{C} \mathbf{N}) = \text{tr}(\mathbf{I}) + c \cdot \text{tr}(\mathbf{J}_Q \mathbf{N}) = Q + c \sum_{q=1}^Q n_q = Q + cn;$$

on the other hand, the definition of $\mathbf{C}$ implies

$$\begin{aligned} \text{tr}(\mathbf{C} \mathbf{N}) &= \text{tr} \left\{ \mathbf{N}^{-1} \mathbf{B}' (\mathbf{B} \mathbf{N}^{-1} \mathbf{B}')^{-1} \mathbf{B} \right\} = \text{tr} \left\{ \mathbf{B} \mathbf{N}^{-1} \mathbf{B}' (\mathbf{B} \mathbf{N}^{-1} \mathbf{B}')^{-1} \right\} \\ &= F = Q - 1. \end{aligned}$$

Therefore, $c = -n^{-1}$, $\mathbf{C} = \mathbf{N}^{-1} - n^{-1} \mathbf{J}_Q$, and Lemma A2 holds. $\square$

LEMMA A3. *If two matrices $\mathbf{A}$ and $\mathbf{B}$ in $\mathbb{R}^{p \times m}$ satisfy $\mathbf{A} \mathbf{A}' = \mathbf{B} \mathbf{B}'$, then there exists an orthogonal matrix $\mathbf{\Gamma} \in \mathbb{R}^{m \times m}$ such that $\mathbf{A} = \mathbf{B} \mathbf{\Gamma}$.*

PROOF OF LEMMA A3. First, we consider the case with $p = m$, i.e., $\mathbf{A}$ and $\mathbf{B}$ are square matrices. From the polar decomposition, $\mathbf{A} = (\mathbf{A} \mathbf{A}')^{1/2} \mathbf{\Gamma}_1$ and $\mathbf{B} = (\mathbf{B} \mathbf{B}')^{1/2} \mathbf{\Gamma}_2$, where $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are orthogonal matrices. Therefore,

$$\mathbf{B} \mathbf{\Gamma}_2' \mathbf{\Gamma}_1 = (\mathbf{B} \mathbf{B}')^{1/2} \mathbf{\Gamma}_2 \mathbf{\Gamma}_2' \mathbf{\Gamma}_1 = (\mathbf{A} \mathbf{A}')^{1/2} \mathbf{\Gamma}_1 = \mathbf{A},$$

where $\mathbf{\Gamma}_2' \mathbf{\Gamma}_1$ is an orthogonal matrix. Thus, Lemma A3 holds when $p = m$.

Second, we consider the case with $p < m$. Define two square matrices:

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{A} \\ \mathbf{0}_{(m-p) \times m} \end{pmatrix} \in \mathbb{R}^{m \times m}, \quad \mathbf{B}_1 = \begin{pmatrix} \mathbf{B} \\ \mathbf{0}_{(m-p) \times m} \end{pmatrix} \in \mathbb{R}^{m \times m}.$$

We can verify $\mathbf{A}_1 \mathbf{A}_1' = \mathbf{B}_1 \mathbf{B}_1'$. From the first case for square matrices, there exists an orthogonal matrix $\mathbf{\Gamma} \in \mathbb{R}^{m \times m}$ such that $\mathbf{A}_1 = \mathbf{B}_1 \mathbf{\Gamma}$, which immediately implies $\mathbf{A} = \mathbf{B} \mathbf{\Gamma}$. Thus, Lemma A3 holds when $p < m$.

Third, we consider the case with $p > m$. Let $\mathbf{W}_0 \in \mathbb{R}^{p \times p_0}$ be a matrix whose columns are an orthonormal basis of $\{\mathbf{x} : \mathbf{A}'\mathbf{x} = 0\} \subset \mathbb{R}^p$, and $\mathbf{W}_1 \in \mathbb{R}^{p \times (p-p_0)}$ be the matrix such that $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_0) \in \mathbb{R}^{p \times p}$ is an orthogonal matrix. We can verify $\mathbf{W}_0' \mathbf{A} = \mathbf{0}_{p_0 \times m}$. Because

$$\mathbf{W}_0' \mathbf{B} (\mathbf{W}_0' \mathbf{B})' = \mathbf{W}_0' \mathbf{B} \mathbf{B}' \mathbf{W}_0 = \mathbf{W}_0' \mathbf{A} \mathbf{A}' \mathbf{W}_0 = \mathbf{0}_{p_0 \times p_0},$$

we have $\mathbf{W}_0' \mathbf{B} = \mathbf{0}_{p_0 \times m}$. Because $\{\mathbf{x} : \mathbf{A}'\mathbf{x} = 0\}$ is of dimension at least $p - m$, we have $p_0 \geq p - m$ and therefore $p - p_0 \leq m$. Because the two matrices $\mathbf{W}_1' \mathbf{A}$ and $\mathbf{W}_1' \mathbf{B}$ in $\mathbb{R}^{(p-p_0) \times m}$ satisfy $\mathbf{W}_1' \mathbf{A} \mathbf{A}' \mathbf{W}_1 = \mathbf{W}_1' \mathbf{B} \mathbf{B}' \mathbf{W}_1$, from the second case, there exists an orthogonal matrix $\mathbf{\Gamma} \in \mathbb{R}^{m \times m}$ such that $\mathbf{W}_1' \mathbf{A} = \mathbf{W}_1' \mathbf{B} \mathbf{\Gamma}$. Thus,

$$\mathbf{W}' \mathbf{B} \mathbf{\Gamma} = \begin{pmatrix} \mathbf{W}_1' \\ \mathbf{W}_0' \end{pmatrix} \mathbf{B} \mathbf{\Gamma} = \begin{pmatrix} \mathbf{W}_1' \mathbf{B} \mathbf{\Gamma} \\ \mathbf{W}_0' \mathbf{B} \mathbf{\Gamma} \end{pmatrix} = \begin{pmatrix} \mathbf{W}_1' \mathbf{A} \\ \mathbf{0}_{p_0 \times m} \end{pmatrix} = \begin{pmatrix} \mathbf{W}_1' \mathbf{A} \\ \mathbf{W}_0' \mathbf{A} \end{pmatrix} = \mathbf{W}' \mathbf{A}.$$

Because $\mathbf{W}$ is an orthogonal matrix, we have $\mathbf{B} \mathbf{\Gamma} = \mathbf{A}$. Thus, Lemma A3 holds when $p > m$. $\square$

A2.2. Covariances between $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\tau}}_{\mathbf{x}}$.

PROOF OF PROPOSITION 1. We first write the vector $(\hat{\boldsymbol{\tau}}', \hat{\boldsymbol{\tau}}_{\mathbf{x}}')'$ as a linear combination of average observed outcomes and covariates:

(A3)

$$\begin{pmatrix} \hat{\boldsymbol{\tau}} \\ \hat{\boldsymbol{\tau}}_{\mathbf{x}} \end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^Q \begin{pmatrix} \mathbf{b}_q \hat{Y}(q) \\ g_{1q} \hat{\mathbf{x}}(q) \\ \vdots \\ g_{Fq} \hat{\mathbf{x}}(q) \end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^Q \begin{pmatrix} \mathbf{b}_q & \mathbf{0}_{F \times L} \\ \mathbf{0}_{FL \times 1} & \mathbf{b}_q \otimes \mathbf{I}_L \end{pmatrix} \begin{pmatrix} \hat{Y}(q) \\ \hat{\mathbf{x}}(q) \end{pmatrix}.$$

We can view covariates as “outcomes” unaffected by the treatment, and view $(Y_i(q), \mathbf{x}_i')'$ as the potential outcome vector under treatment combination q . Using Li and Ding [2017, Theorem 3], $(\hat{\boldsymbol{\tau}}', \hat{\boldsymbol{\tau}}_{\mathbf{x}}')'$ has sampling mean

$$2^{-(K-1)} \sum_{q=1}^Q \begin{pmatrix} \mathbf{b}_q & \mathbf{0}_{F \times L} \\ \mathbf{0}_{FL \times 1} & \mathbf{b}_q \otimes \mathbf{I}_L \end{pmatrix} \begin{pmatrix} \bar{Y}(q) \\ \bar{\mathbf{x}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\tau} \\ \mathbf{0}_{FL \times 1} \end{pmatrix},$$

and sampling covariance matrix

$$2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \begin{pmatrix} \mathbf{b}_q & \mathbf{0}_{F \times L} \\ \mathbf{0}_{FL \times 1} & \mathbf{b}_q \otimes \mathbf{I}_L \end{pmatrix} \begin{pmatrix} S_{qq} & \mathbf{S}_{q,\mathbf{x}} \\ \mathbf{S}_{\mathbf{x},q} & \mathbf{S}_{\mathbf{x}\mathbf{x}} \end{pmatrix} \begin{pmatrix} \mathbf{b}_q & \mathbf{0}_{F \times L} \\ \mathbf{0}_{FL \times 1} & \mathbf{b}_q \otimes \mathbf{I}_L \end{pmatrix}'$$

$$\begin{aligned}
& -n^{-1} \begin{pmatrix} \mathbf{S}_{\tau\tau} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\
& = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \begin{pmatrix} \mathbf{b}_q \mathbf{b}_q' S_{qq} & \mathbf{b}_q \mathbf{S}_{q,x} \cdot (\mathbf{b}_q \otimes \mathbf{I}_L)' \\ (\mathbf{b}_q \otimes \mathbf{I}_L) \cdot \mathbf{S}_{x,q} \mathbf{b}_q' & (\mathbf{b}_q \otimes \mathbf{I}_L) \mathbf{S}_{xx} (\mathbf{b}_q \otimes \mathbf{I}_L)' \end{pmatrix} \\
& -n^{-1} \begin{pmatrix} \mathbf{S}_{\tau\tau} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.
\end{aligned}$$

From Lemma A1,

$$\begin{aligned}
\mathbf{b}_q \mathbf{S}_{q,x} \cdot (\mathbf{b}_q \otimes \mathbf{I}_L)' &= (\mathbf{b}_q \otimes \mathbf{S}_{q,x}) (\mathbf{b}_q' \otimes \mathbf{I}_L) = (\mathbf{b}_q \mathbf{b}_q') \otimes \mathbf{S}_{q,x}, \\
(\mathbf{b}_q \otimes \mathbf{I}_L) \mathbf{S}_{xx} (\mathbf{b}_q \otimes \mathbf{I}_L)' &= (\mathbf{b}_q \otimes \mathbf{I}_L) (1 \otimes \mathbf{S}_{xx}) (\mathbf{b}_q' \otimes \mathbf{I}_L) = (\mathbf{b}_q \mathbf{b}_q') \otimes \mathbf{S}_{xx}.
\end{aligned}$$

Therefore, we deduce the mean and sampling covariance of $(\hat{\boldsymbol{\tau}}', \hat{\boldsymbol{\tau}}_x')$ in the form of Proposition 1. The asymptotic Gaussianity of $(\hat{\boldsymbol{\tau}}', \hat{\boldsymbol{\tau}}_x)'$ follows directly from Condition 1 and Li and Ding [2017, Theorem 5]. $\square$

PROOF OF PROPOSITION 6. We first rewrite

$$\begin{pmatrix} \hat{\boldsymbol{\tau}} - \boldsymbol{\tau} \\ \hat{\boldsymbol{\theta}}_x \end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^Q \begin{pmatrix} \mathbf{b}_q \hat{Y}(q) - \mathbf{b}_q \bar{Y}(q) \\ \mathbf{c}_q \otimes \hat{\mathbf{x}}(q) \end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^Q \begin{pmatrix} \mathbf{b}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{c}_q \otimes \mathbf{I}_L \end{pmatrix} \begin{pmatrix} \hat{Y}(q) - \bar{Y}(q) \\ \hat{\mathbf{x}}(q) \end{pmatrix}.$$

From Li and Ding [2017, Theorem 3], $(\hat{\boldsymbol{\tau}}' - \boldsymbol{\tau}', \hat{\boldsymbol{\theta}}_x')'$ has mean zero and sampling covariance

$$\begin{aligned}
& 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \begin{pmatrix} \mathbf{b}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{c}_q \otimes \mathbf{I}_L \end{pmatrix} \begin{pmatrix} S_{qq} & \mathbf{S}_{q,x} \\ \mathbf{S}_{x,q} & \mathbf{S}_{xx} \end{pmatrix} \begin{pmatrix} \mathbf{b}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{c}_q \otimes \mathbf{I}_L \end{pmatrix}' - n^{-1} \begin{pmatrix} \mathbf{S}_{\tau\tau} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\
& = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \begin{pmatrix} \mathbf{b}_q \mathbf{b}_q' S_{qq} & (\mathbf{b}_q \mathbf{c}_q') \otimes \mathbf{S}_{q,x} \\ (\mathbf{c}_q \mathbf{b}_q') \otimes \mathbf{S}_{x,q} & (\mathbf{c}_q \mathbf{c}_q') \otimes \mathbf{S}_{xx} \end{pmatrix} - n^{-1} \begin{pmatrix} \mathbf{S}_{\tau\tau} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.
\end{aligned}$$

From the Gram–Schmidt orthogonalization of the $\mathbf{c}_q$'s,

$$\sum_{q=1}^Q n_q^{-1} \mathbf{c}_q \mathbf{c}_q' = \sum_{q=1}^Q n_q^{-1} \begin{pmatrix} \mathbf{c}_q[1] \mathbf{c}_q'[1] & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{c}_q[2] \mathbf{c}_q'[2] & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{c}_q[H] \mathbf{c}_q'[H] \end{pmatrix}.$$

Therefore, Proposition 6 holds. $\square$

PROOF OF PROPOSITION A1. We first rewrite

$$\begin{aligned} \begin{pmatrix} \hat{\tau} - \tau \\ \hat{\theta}_{e[1]} \\ \vdots \\ \hat{\theta}_{e[T]} \end{pmatrix} &= 2^{-(K-1)} \sum_{q=1}^Q \begin{pmatrix} \mathbf{b}_q \hat{Y}(q) - \mathbf{b}_q \bar{Y}(q) \\ \mathbf{c}_q \otimes \hat{\mathbf{e}}_{[1]}(q) \\ \vdots \\ \mathbf{c}_q \otimes \hat{\mathbf{e}}_{[T]}(q) \end{pmatrix} \\ &= 2^{-(K-1)} \sum_{q=1}^Q \begin{pmatrix} \mathbf{b}_q & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{c}_q \otimes \mathbf{I}_{L_1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{c}_q \otimes \mathbf{I}_{L_T} \end{pmatrix} \begin{pmatrix} \hat{Y}(q) - \bar{Y}(q) \\ \hat{\mathbf{e}}_{[1]}(q) \\ \vdots \\ \hat{\mathbf{e}}_{[T]}(q) \end{pmatrix}. \end{aligned}$$

From Li and Ding [2017, Theorem 3], $(\hat{\tau}' - \tau', \hat{\theta}'_{e[1]}, \dots, \hat{\theta}'_{e[T]})'$ has mean zero and sampling covariance

$$\begin{aligned} &2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \left\{ \begin{pmatrix} \mathbf{b}_q & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{c}_q \otimes \mathbf{I}_{L_1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{c}_q \otimes \mathbf{I}_{L_T} \end{pmatrix} \right. \\ &\quad \cdot \begin{pmatrix} S_{qq} & \mathbf{S}_{q,e[1]} & \dots & \mathbf{S}_{q,e[T]} \\ \mathbf{S}_{e[1],q} & \mathbf{S}_{e[1]e[1]} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{e[T],q} & \mathbf{0} & \dots & \mathbf{S}_{e[T]e[T]} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{b}_q & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{c}_q \otimes \mathbf{I}_{L_1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{c}_q \otimes \mathbf{I}_{L_T} \end{pmatrix}' \Big\} \\ &\quad - n^{-1} \begin{pmatrix} \mathbf{S}_{\tau\tau} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix} \\ &= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \begin{pmatrix} \mathbf{b}_q \mathbf{b}_q' S_{qq} & (\mathbf{b}_q \mathbf{c}_q') \otimes \mathbf{S}_{q,e[1]} & \dots & (\mathbf{b}_q \mathbf{c}_q') \otimes \mathbf{S}_{q,e[T]} \\ (\mathbf{c}_q \mathbf{b}_q') \otimes \mathbf{S}_{e[1],q} & (\mathbf{c}_q \mathbf{c}_q') \otimes \mathbf{S}_{e[1]e[1]} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{c}_q \mathbf{b}_q') \otimes \mathbf{S}_{e[T],q} & \mathbf{0} & \dots & (\mathbf{c}_q \mathbf{c}_q') \otimes \mathbf{S}_{e[T]e[T]} \end{pmatrix} \\ &\quad - n^{-1} \begin{pmatrix} \mathbf{S}_{\tau\tau} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix}. \end{aligned}$$

Therefore, Proposition A1 holds. $\square$

A2.3. Linear projections and squared multiple correlations.

PROOF OF THEOREM 1. From Lemma A1, the sampling covariance matrix of $\hat{\tau}$ explained by $\hat{\tau}_x$ in the linear projection satisfies

$$\begin{aligned}
& \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1} \mathbf{V}_{x\tau} \\
&= 2^{-4(K-1)} \left\{ \sum_{q=1}^Q n_q^{-1} (\mathbf{b}_q \mathbf{b}_q') \otimes \mathbf{S}_{q,x} \right\} \left(\tilde{\mathbf{B}} \otimes \mathbf{S}_{xx} \right)^{-1} \left\{ \sum_{k=1}^Q n_k^{-1} (\mathbf{b}_k \mathbf{b}_k') \otimes \mathbf{S}_{x,k} \right\} \\
&= 2^{-4(K-1)} \sum_{q=1}^Q \sum_{k=1}^Q \left\{ \sum_{q=1}^Q n_q^{-1} (\mathbf{b}_q \mathbf{b}_q') \otimes \mathbf{S}_{q,x} \right\} \left(\tilde{\mathbf{B}}^{-1} \otimes \mathbf{S}_{xx}^{-1} \right) \left\{ \sum_{k=1}^Q n_k^{-1} (\mathbf{b}_k \mathbf{b}_k') \otimes \mathbf{S}_{x,k} \right\} \\
&= 2^{-4(K-1)} \sum_{q=1}^Q \sum_{k=1}^Q \left\{ n_q^{-1} (\mathbf{b}_q \mathbf{b}_q') \cdot \tilde{\mathbf{B}}^{-1} \cdot n_k^{-1} (\mathbf{b}_k \mathbf{b}_k') \right\} \otimes (\mathbf{S}_{q,x} \mathbf{S}_{xx}^{-1} \mathbf{S}_{x,k}) \\
&= 2^{-2(K-1)} \sum_{q=1}^Q \sum_{k=1}^Q \left\{ \mathbf{b}_q \left(2^{-2(K-1)} n_q^{-1} n_k^{-1} \mathbf{b}_q' \tilde{\mathbf{B}}^{-1} \mathbf{b}_k \right) \mathbf{b}_k' \right\} \cdot (\mathbf{S}_{q,x} \mathbf{S}_{xx}^{-1} \mathbf{S}_{x,k}),
\end{aligned}$$

where in the last equality the Kronecker product reduces to the matrix product because $\mathbf{S}_{q,x} \mathbf{S}_{xx}^{-1} \mathbf{S}_{x,k}$ is a scalar. Using Lemma A2, we have

$$\begin{aligned}
& \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1} \mathbf{V}_{x\tau} \\
&= 2^{-2(K-1)} \left[\sum_{q=1}^Q (n_q^{-1} \mathbf{b}_q \mathbf{b}_q') \cdot (\mathbf{S}_{q,x} \mathbf{S}_{xx}^{-1} \mathbf{S}_{x,q}) - \sum_{q=1}^Q \sum_{k=1}^Q n^{-1} \mathbf{b}_q \mathbf{b}_k' \cdot (\mathbf{S}_{q,x} \mathbf{S}_{xx}^{-1} \mathbf{S}_{x,k}) \right] \\
&= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q' \cdot S_{qq}^{\parallel} - n^{-1} \left(2^{-(K-1)} \sum_{q=1}^Q \mathbf{b}_q \mathbf{S}_{q,x} \right) \mathbf{S}_{xx}^{-1} \left(2^{-(K-1)} \sum_{k=1}^Q \mathbf{S}_{x,k} \mathbf{b}_k' \right) \\
&= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q' \cdot S_{qq}^{\parallel} - n^{-1} \mathbf{S}_{\tau,x} \mathbf{S}_{xx}^{-1} \mathbf{S}_{x,\tau} \\
&= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q' \cdot S_{qq}^{\parallel} - n^{-1} \mathbf{S}_{\tau\tau}^{\parallel} \equiv \mathbf{V}_{\tau\tau}^{\parallel}.
\end{aligned}$$

Therefore, the sampling covariance of the residual from the linear projection of $\hat{\tau}$ on $\hat{\tau}_x$ satisfies

$$\begin{aligned}
& \mathbf{V}_{\tau\tau} - \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1} \mathbf{V}_{x\tau} \\
&= \mathbf{V}_{\tau\tau} - \mathbf{V}_{\tau\tau}^{\parallel} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q' \cdot (S_{qq} - S_{qq}^{\parallel}) - n^{-1} (\mathbf{S}_{\tau\tau} - \mathbf{S}_{\tau\tau}^{\parallel})
\end{aligned}$$

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

Theorem 1 holds. $\square$

PROOF OF COROLLARY 1. From Proposition 1 and Theorem 1, under the CRFE, the variance of $\hat{\tau}_f$ is $V_{\tau_f\tau_f}$, and the variance of $\hat{\tau}_f$ explained by $\hat{\tau}_{\mathbf{x}}$ in the linear projection is $V_{\tau_f\tau_f}^\parallel$. Therefore, the squared multiple correlation between $\hat{\tau}_f$ and $\hat{\tau}_{\mathbf{x}}$ is

$$R^2(f) = \frac{V_{\tau_f\tau_f}^\parallel}{V_{\tau_f\tau_f}} = \frac{2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} S_{qq}^\parallel - n^{-1} S_{\tau_f\tau_f}^\parallel}{2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} S_{qq} - n^{-1} S_{\tau_f\tau_f}}.$$

Under the additivity, $S_{11} = \dots = S_{QQ}$, $S_{11}^\parallel = \dots = S_{QQ}^\parallel$, and $S_{\tau_f\tau_f} = S_{\tau_f\tau_f}^\parallel = 0$, which further imply $R_f^2 = S_{11}^\parallel / S_{11}$. $\square$

PROOF OF PROPOSITION 5. Under the additivity, from Proposition 1, the sampling variance and covariances are $\text{Var}(\hat{\tau}_f) = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} S_{11}$, $\text{Cov}(\hat{\tau}_{\mathbf{x},k}) = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{S}_{\mathbf{x}\mathbf{x}}$, and $\text{Cov}(\hat{\tau}_f, \hat{\tau}_{\mathbf{x},k}) = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} g_{fq} g_{kq} \mathbf{S}_{1,\mathbf{x}}$. Let $\mathbf{g}_0 = \mathbf{1}_Q$. By the property of 2^K factorial experiments, for any $1 \leq f, k \leq F$, there exists $0 \leq m \leq F$ such that $\mathbf{g}_f \circ \mathbf{g}_k = \mathbf{g}_m$, recalling that $\circ$ denotes element-wise multiplication. Define $f \star k$ as the index such that $\mathbf{g}_f \circ \mathbf{g}_k = \mathbf{g}_{f \star k}$. Let $w_q = n_q^{-1} / \sum_{k=1}^Q n_k^{-1}$ be the weight inversely proportional to the number of units under treatment combination q . We can simplify the variance of $\hat{\tau}_f$ explained by $\hat{\tau}_{\mathbf{x},k}$ under the CRFE as

$$\begin{aligned} & \text{Cov}(\hat{\tau}_f, \hat{\tau}_{\mathbf{x},k}) \text{Cov}^{-1}(\hat{\tau}_{\mathbf{x},k}) \text{Cov}(\hat{\tau}_{\mathbf{x},k}, \hat{\tau}_f) \\ &= 2^{-2(K-1)} \left(\sum_{q=1}^Q n_q^{-1} g_{f \star k, q} \mathbf{S}_{1,\mathbf{x}} \right) \left(\sum_{q=1}^Q n_q^{-1} \mathbf{S}_{\mathbf{x}\mathbf{x}} \right)^{-1} \left(\sum_{q=1}^Q n_q^{-1} g_{f \star k, q} \mathbf{S}_{\mathbf{x},1} \right) \\ &= 2^{-2(K-1)} \left(\sum_{q=1}^Q n_q^{-1} \right) \left(\sum_{q=1}^Q w_q g_{f \star k, q} \right)^2 S_{11}^\parallel. \end{aligned}$$

Therefore, the squared multiple correlation between $\hat{\tau}_f$ and $\hat{\tau}_{\mathbf{x},k}$ satisfies

$$\begin{aligned} \gamma_{fk}^2 &= \frac{\text{Cov}(\hat{\tau}_f, \hat{\tau}_{\mathbf{x},k}) \text{Cov}^{-1}(\hat{\tau}_{\mathbf{x},k}) \text{Cov}(\hat{\tau}_{\mathbf{x},k}, \hat{\tau}_f)}{\text{Var}(\hat{\tau}_f)} \\ &= \frac{2^{-2(K-1)} \left(\sum_{q=1}^Q n_q^{-1} \right) \left(\sum_{q=1}^Q w_q g_{f \star k, q} \right)^2 S_{11}^\parallel}{2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} S_{11}} \end{aligned}$$

$$\begin{aligned}
&= \left(\sum_{q=1}^Q w_q g_{f \star k, q} \right)^2 S_{11}^{\parallel} / S_{11} = \left(\sum_{q: g_{f \star k, q}=1} w_q - \sum_{q: g_{f \star k, q}=-1} w_q \right)^2 S_{11}^{\parallel} / S_{11} \\
&\leq \left(\sum_{q=1}^Q w_q \right)^2 S_{11}^{\parallel} / S_{11} = S_{11}^{\parallel} / S_{11},
\end{aligned}$$

where the equality holds if $f \star k = 0$, i.e., $k = f$. Using Corollary 1, we have $\gamma_{fk}^2 \leq \gamma_{ff}^2 = R_f^2 = S_{11}^{\parallel} / S_{11}$. Moreover, because $\gamma_{ff}^2 = R_f^2$, under the CRFE, the variance of $\hat{\tau}_f$ explained by $\hat{\tau}_{\mathbf{x}}$ is the same as that explained by $\hat{\tau}_{\mathbf{x}, f}$. Therefore, the squared multiple partial correlation between $\hat{\tau}_f$ and $\hat{\tau}_{\mathbf{x}}$ given $\hat{\tau}_{\mathbf{x}, f}$ is zero. If $n_1 = \dots = n_Q = n/Q$, then $w_1 = \dots = w_Q = Q^{-1}$, and thus $\sum_{q=1}^Q w_q g_{f \star k, q} = Q^{-1} \sum_{q=1}^Q g_{f \star k, q} = 0$ if $k \neq f$. Therefore, $\gamma_{fk}^2 = 0$ for $k \neq f$. $\square$

PROOF OF PROPOSITION 7. With equal treatment group sizes, from the definition of coefficient vector $\mathbf{b}_q$, we have

$$\begin{aligned}
\tilde{\mathbf{B}} &= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q' = (4/Q^2)(n/Q)^{-1} \cdot \sum_{q=1}^Q \mathbf{b}_q \mathbf{b}_q' \\
&= 4/(nQ) \cdot (\mathbf{b}_1, \dots, \mathbf{b}_Q) \cdot (\mathbf{b}_1, \dots, \mathbf{b}_Q)' \\
&= 4/(nQ) \cdot (\mathbf{g}_1, \dots, \mathbf{g}_F)' \cdot (\mathbf{g}_1, \dots, \mathbf{g}_F).
\end{aligned}$$

By the property of 2^K factorial experiments, these generating vectors satisfy that, for $1 \leq f \neq k \leq F$, $\mathbf{g}'_f \mathbf{g}_k = 0$ and $\mathbf{g}'_f \mathbf{g}_f = Q$. Thus, $\tilde{\mathbf{B}}$ further reduces to $\tilde{\mathbf{B}} = 4/(nQ) \cdot Q \mathbf{I}_F = (4/n) \mathbf{I}_F$, which is a diagonal matrix. Therefore, in (5.1), $\tilde{\mathbf{B}}[\mathcal{F}_h, \mathcal{F}_{h-1}] = \mathbf{0}$, and $\mathbf{c}_q[h]$ reduces to $\mathbf{b}_q[\mathcal{F}_h]$. Consequently, $\hat{\boldsymbol{\theta}}_{\mathbf{x}}[h]$ in (5.2) reduces to $\hat{\tau}_{\mathbf{x}}[\mathcal{F}_h]$. Because

$$\begin{aligned}
\mathbf{W}_{\mathbf{x}\mathbf{x}}[h] &= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{c}_q[h] \mathbf{c}_q'[h]) \otimes \mathbf{S}_{\mathbf{x}\mathbf{x}} \\
&= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{b}_q[\mathcal{F}_h] \mathbf{b}_q'[\mathcal{F}_h]) \otimes \mathbf{S}_{\mathbf{x}\mathbf{x}} = \tilde{\mathbf{B}}[\mathcal{F}_h, \mathcal{F}_h] \otimes \mathbf{S}_{\mathbf{x}\mathbf{x}} \\
&= 4/n \cdot \mathbf{I}_{F_h} \otimes \mathbf{S}_{\mathbf{x}\mathbf{x}},
\end{aligned}$$

we can simplify M_h as

$$M_h = \hat{\boldsymbol{\theta}}_{\mathbf{x}}'[h] (\mathbf{W}_{\mathbf{x}\mathbf{x}}[h])^{-1} \hat{\boldsymbol{\theta}}_{\mathbf{x}}[h] = \hat{\tau}_{\mathbf{x}}'[\mathcal{F}_h] (4/n \cdot \mathbf{I}_{F_h} \otimes \mathbf{S}_{\mathbf{x}\mathbf{x}})^{-1} \hat{\tau}_{\mathbf{x}}[\mathcal{F}_h]$$

$$= n/4 \cdot \sum_{f \in \mathcal{F}_h} \hat{\tau}'_{x,f} \mathbf{S}_{xx}^{-1} \hat{\tau}_{x,f}.$$

Therefore, Proposition 7 holds. $\square$

A3. Asymptotic sampling distributions of $\hat{\tau}$.

A3.1. *Lemmas for central convex unimodality.* We use $\mathcal{B}_m(r)$ to denote the ball in $\mathbb{R}^m$ with center zero and radius r .

LEMMA A4. *The class of central convex unimodal distributions is closed under convolution, marginality, product measure, and weak convergence.*

PROOF OF LEMMA A4. See Kanter [1977], Dharmadhikari and Joag-Dev [1988] and Dai [1989]. $\square$

The following two lemmas extends Lemma A4 to all linear transformations. Although this extension is straightforward, we give a proof below for completeness.

LEMMA A5. *If $\psi \in \mathbb{R}^m$ is central convex unimodal, then for any $p \geq 1$, $(\psi', \mathbf{0}'_{p \times 1})' \in \mathbb{R}^{m+p}$ is central convex unimodal.*

PROOF OF LEMMA A5. For any random vector ϕ uniformly distributed on a symmetric convex body $\mathcal{K} \subset \mathbb{R}^m$, we define $\phi_r \in \mathbb{R}^{m+p}$ as the random vector uniformly distributed on $\mathcal{K} \times \mathcal{B}_p(r) \subset \mathbb{R}^{m+p}$. Because $\mathcal{K} \times \mathcal{B}_p(r)$ is a symmetric convex set, ϕ_r is central convex unimodal. As r goes to zero, ϕ_r converges weakly to $(\phi', \mathbf{0}'_{p \times 1})'$. Therefore, $(\phi', \mathbf{0}'_{p \times 1})'$ is central convex unimodal. By taking mixtures and weak limits, we deduce Lemma A5. $\square$

LEMMA A6. *If $\psi \in \mathbb{R}^m$ is central convex unimodal, then for any matrix $\mathbf{C} \in \mathbb{R}^{p \times m}$, $\mathbf{C}\psi \in \mathbb{R}^p$ is central convex unimodal.*

PROOF OF LEMMA A6. First, we consider the case where $\mathbf{C}$ is an $m \times m$ invertible matrix. For any random vector ϕ uniformly distributed on a symmetric convex body $\mathcal{K} \subset \mathbb{R}^m$, $\mathbf{C}\phi$ is uniformly distributed on $\mathbf{C}\mathcal{K} \equiv \{\mathbf{C}\mathbf{x} : \mathbf{x} \in \mathcal{K}\}$. By taking mixtures and weak limits, we can know that Lemma A6 holds when $\mathbf{C}$ is an $m \times m$ invertible matrix.

Second, we consider the case where $\mathbf{C}$ has full row rank, i.e., $\text{rank}(\mathbf{C}) = p \leq m$. In this case, there exists an $(m-p) \times m$ matrix $\tilde{\mathbf{C}}$ such that $(\mathbf{C}', \tilde{\mathbf{C}}')$ is an $m \times m$ invertible matrix. From the first case, we know for any central

convex unimodal random vector $\psi \in \mathbb{R}^m$, $(\mathbf{C}', \bar{\mathbf{C}}')'\psi$ is also central convex unimodal. From Lemma A4, its subvector $\mathbf{C}\psi$ is also central convex unimodal.

Third, we consider a general matrix $\mathbf{C}$ of rank p_1 . Let $\mathbf{W}$ be a matrix permuting the rows of $\mathbf{C}$ such that the first p_1 rows of $\mathbf{WC}$ is linearly independent, and the remaining $p_2 = p - p_1$ rows of $\mathbf{WC}$ are all linear combinations of the first p_1 rows, i.e., $(\mathbf{WC})' = (\mathbf{C}'_1, \mathbf{C}'_2)$, where $\mathbf{C}_1 \in \mathbb{R}^{p_1 \times m}$, $\mathbf{C}_2 \in \mathbb{R}^{p_2 \times m}$, $\text{rank}(\mathbf{C}_1) = p_1$, and $\mathbf{C}_2 = \mathbf{\Gamma}\mathbf{C}_1$. Thus,

$$\begin{aligned} \mathbf{C}\psi &= \mathbf{W}^{-1}(\mathbf{WC})\psi = \mathbf{W}^{-1} \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{pmatrix} \psi = \mathbf{W}^{-1} \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{\Gamma}\mathbf{C}_1 \end{pmatrix} \psi = \mathbf{W}^{-1} \begin{pmatrix} \mathbf{C}_1\psi \\ \mathbf{\Gamma}\mathbf{C}_1\psi \end{pmatrix} \\ &= \mathbf{W}^{-1} \begin{pmatrix} \mathbf{I}_{p_1} & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{\Gamma} & \mathbf{I}_{p_2} \end{pmatrix} \begin{pmatrix} \mathbf{C}_1\psi \\ \mathbf{0}_{p_2 \times 1} \end{pmatrix} \equiv \tilde{\mathbf{\Gamma}} \begin{pmatrix} \mathbf{C}_1\psi \\ \mathbf{0}_{p_2 \times 1} \end{pmatrix}. \end{aligned}$$

From the second case, $\mathbf{C}_1\psi$ is central convex unimodal. From Lemma A5, $((\mathbf{C}_1\psi)', \mathbf{0}'_{p_2 \times 1})$ is also central convex unimodal. Because the $p \times p$ matrix $\tilde{\mathbf{\Gamma}}$ is invertible, from the first case, we know $\mathbf{C}\psi$ is central convex unimodal. $\square$

LEMMA A7. *Let ψ_1 and ψ_2 be two independent m dimensional random vectors. If both ψ_1 and ψ_2 are central convex unimodal, then $\psi_1 + \psi_2$ is central convex unimodal.*

PROOF OF LEMMA A7. See Dharmadhikari and Joag-Dev [1988, Theorem 2.20]. $\square$

LEMMA A8. *If a random vector in $\mathbb{R}^m$ has a log-concave density, then it is central convex unimodal.*

PROOF OF LEMMA A8. See Kanter [1977, Lemma 3.1] and Dharmadhikari and Joag-Dev [1988, Theorem 2.15]. $\square$

A3.2. Properties of the truncated Gaussian random vectors.

PROOF OF PROPOSITION 2. For any m , the densities of ε and $\zeta_{m,a}$ are

$$f(\mathbf{x}) = (2\pi)^{-m/2} \exp(-\mathbf{x}'\mathbf{x}/2), \quad g(\mathbf{x}) = \frac{1\{\mathbf{x}'\mathbf{x} \leq a\}}{P(\chi_m^2 \leq a)} (2\pi)^{-m/2} \exp(-\mathbf{x}'\mathbf{x}/2),$$

and the log-densities of ε and $\zeta_{m,a}$ are

$$\log f(\mathbf{x}) = -(m/2) \log(2\pi) - \mathbf{x}'\mathbf{x}/2,$$

$$\log g(\mathbf{x}) = \begin{cases} -\log \{P(\chi_m^2 \leq a)\} - (m/2) \log(2\pi) - \mathbf{x}'\mathbf{x}/2, & \text{if } \mathbf{x}'\mathbf{x} \leq a, \\ -\infty, & \text{otherwise.} \end{cases}$$

It is straightforward to show that both $\log f(\cdot)$ and $\log g(\cdot)$ are concave functions. From Lemma A8, both ε and $\zeta_{m,a}$ are central convex unimodal. Because both ε and $\zeta_{LF,a}$ are central convex unimodal, using Lemma A6, both $(\mathbf{V}_{\tau\tau}^\perp)^{1/2}\varepsilon$ and $(\mathbf{V}_{\tau\tau}^\parallel)^{1/2}_{LF}\zeta_{LF,a}$ are central convex unimodal. From Lemma A7, we have Proposition 2. $\square$

PROOF OF PROPOSITION 3. First, we prove that $\zeta_{LF,a}$ is spherically symmetric. Let $\mathbf{D} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{LF})$. For any orthogonal matrix $\mathbf{\Gamma}$, $\mathbf{\Gamma}\mathbf{D} \sim \mathbf{D}$, and thus,

$$\zeta_{LF,a} \sim \mathbf{D} \mid \mathbf{D}'\mathbf{D} \leq a \sim \mathbf{\Gamma}\mathbf{D} \mid (\mathbf{\Gamma}\mathbf{D})'\mathbf{\Gamma}\mathbf{D} \leq a \sim \mathbf{\Gamma}\mathbf{D} \mid \mathbf{D}'\mathbf{D} \leq a \sim \mathbf{\Gamma}\zeta_{LF,a}.$$

Second, from Morgan and Rubin [2012, Theorem 3.1], $\text{Cov}(\zeta_{LF,a}) = v_{LF,a}\mathbf{I}_{LF}$. Third, we show the representation for $\zeta_{LF,a}$. By the spherical symmetry of the standard Gaussian random vector $\mathbf{D}$, $(\mathbf{D}'\mathbf{D})^{-1/2}\mathbf{D} \sim \mathbf{U} = (U_1, \dots, U_{LF})'$ is uniformly distributed on the LF dimensional unit sphere, $\mathbf{D}'\mathbf{D}$ follows χ_{LF}^2 , and they are independent. Therefore,

$$\mathbf{D} \mid \mathbf{D}'\mathbf{D} \leq a \sim (\mathbf{D}'\mathbf{D})^{1/2} \cdot (\mathbf{D}'\mathbf{D})^{-1/2}\mathbf{D} \mid \mathbf{D}'\mathbf{D} \leq a \sim \chi_{LF,a}\mathbf{U}.$$

Let $\text{sign}(U_i)$ be the sign of U_i . Given $(|U_1|, \dots, |U_{LF}|)$, the sign vector $(\text{sign}(U_1), \dots, \text{sign}(U_{LF}))$ has the same probability to be any value in $\{-1, 1\}^m$. Thus, $(\text{sign}(U_1), \dots, \text{sign}(U_{LF})) \sim \mathbf{S}$, and it is independent of $(|U_1|, \dots, |U_{LF}|)$. Because $\mathbf{U} \sim (\mathbf{D}'\mathbf{D})^{-1/2}\mathbf{D}$, and $(D_1^2, \dots, D_{LF}^2)$ are independent and identically Gamma distributed with shape parameter 1/2 and rate parameter 1/2, we have

$$(U_1^2, \dots, U_{LF}^2)' = (\mathbf{D}'\mathbf{D})^{-1} \cdot (D_1^2, \dots, D_{LF}^2)' \sim \text{Dirichlet}(1/2, \dots, 1/2) \sim \beta.$$

Therefore, $\mathbf{U} = (\text{sign}(U_1), \dots, \text{sign}(U_{LF}))' \circ (|U_1|, \dots, |U_{LF}|)' \sim \mathbf{S} \circ \sqrt{\beta}$, which further implies $\zeta_{LF,a} \sim \chi_{LF,a} \cdot \mathbf{S} \circ \sqrt{\beta}$. $\square$

The following lemma helps to simplify the asymptotic sampling distributions of $\hat{\tau}$ under the CRFE, ReFMT_F and ReFMT_{CF}.

LEMMA A9. *Let $\zeta_{m,a} \sim \mathbf{D} \mid \mathbf{D}'\mathbf{D} \leq a$ be an m dimensional truncated Gaussian random vector, where $\mathbf{D} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$. If two matrices $\mathbf{A}$ and $\mathbf{B}$ in $\mathbb{R}^{p \times m}$ satisfy $\mathbf{A}\mathbf{A}' = \mathbf{B}\mathbf{B}'$, then $\mathbf{A}\zeta_{m,a} \sim \mathbf{B}\zeta_{m,a}$.*

PROOF OF LEMMA A9. From Lemma A3, there exists an orthogonal matrix $\mathbf{\Gamma} \in \mathbb{R}^{m \times m}$ such that $\mathbf{A} = \mathbf{B}\mathbf{\Gamma}$. From Proposition 3, by the spherical symmetry of $\zeta_{m,a}$, $\mathbf{A}\zeta_{m,a} = \mathbf{B}\mathbf{\Gamma}\zeta_{m,a} \sim \mathbf{B}\zeta_{m,a}$. $\square$

A3.3. Asymptotic sampling distributions of $\hat{\tau}$ under ReFM. To prove Theorem 2, we need the following three lemmas. We further introduce the following regularity condition for a general covariate balance criteria depending only on $\hat{\tau}_x$ and V_{xx} .

CONDITION A2. Let $\mathbf{B} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$. For any $\mathbf{\Lambda} > 0$, the 0-1 function $\kappa(\sqrt{n}\hat{\tau}_x, nV_{xx}) \equiv \kappa(\hat{\tau}_x, V_{xx})$ satisfies

- (a) κ is almost surely continuous,
- (b) $\text{Var}\{\mathbf{B} \mid \kappa(\mathbf{B}, \mathbf{\Lambda}) = 1\}$, as a function of $\mathbf{\Lambda}$, is continuous,
- (c) $P\{\kappa(\mathbf{B}, \mathbf{\Lambda}) = 1\} > 0$,
- (d) $\kappa(\boldsymbol{\mu}, \mathbf{\Lambda}) = \kappa(-\boldsymbol{\mu}, \mathbf{\Lambda})$, for all $\boldsymbol{\mu}$.

We can verify that the balance criteria for ReFM, ReFMT_F and ReFMT_{CF} depend only on $\hat{\tau}_x$ and V_{xx} and satisfy Condition A2.

LEMMA A10. Let $(\sqrt{n}\tilde{\tau}', \sqrt{n}\tilde{\tau}'_x)$ be a random vector following $\mathcal{N}(0, n\mathbf{V})$. Then as $n \rightarrow \infty$, under Conditions 1 and A2, the two conditional distributions, $\sqrt{n}(\hat{\tau}' - \tau', \hat{\tau}'_x) \mid \kappa(\hat{\tau}_x, V_{xx}) = 1$ and $\sqrt{n}(\tilde{\tau}', \tilde{\tau}'_x) \mid \kappa(\tilde{\tau}_x, V_{xx}) = 1$, converge weakly to the same distribution, i.e.,

$$\left(\begin{array}{c} \hat{\tau} - \tau \\ \hat{\tau}_x \end{array} \right) \Bigg| \kappa(\hat{\tau}_x, V_{xx}) = 1 \quad \sim \quad \left(\begin{array}{c} \tilde{\tau} \\ \tilde{\tau}_x \end{array} \right) \Bigg| \kappa(\tilde{\tau}_x, V_{xx}) = 1.$$

PROOF OF LEMMA A10. The proof of this lemma is almost identical to Li et al. [2018, Proposition A1 and Corollary A1]. We omit it. $\square$

PROOF OF THEOREM 2. From Lemma A10, under Condition 1, $\hat{\tau} - \tau \mid \mathcal{M} \sim \tilde{\tau} \mid \tilde{\tau}'_x V_{xx}^{-1} \tilde{\tau}_x \leq a$, where $(\tilde{\tau}', \tilde{\tau}'_x) \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$. The linear projection of $\tilde{\tau}$ on $\tilde{\tau}_x$ is $V_{\tau x} V_{xx}^{-1} \tilde{\tau}_x$, and the corresponding residual is $\tilde{\tau}_\varepsilon = \tilde{\tau} - V_{\tau x} V_{xx}^{-1} \tilde{\tau}_x$. From Theorem 1, $\tilde{\tau}_\varepsilon$ has variance $V_{\tau\tau}^\perp$. Let $\mathbf{D} = V_{xx}^{-1/2} \tilde{\tau}_x \sim \mathcal{N}(\mathbf{0}, I_{LF})$ be the standardization of $\tilde{\tau}_x$. We can then simplify the asymptotic sampling distribution of $\hat{\tau}$ under ReFM as:

$$\begin{aligned} \hat{\tau} - \tau \mid \mathcal{M} &\sim \tilde{\tau} \mid \tilde{\tau}'_x V_{xx}^{-1} \tilde{\tau}_x \leq a \sim \tilde{\tau}_\varepsilon + V_{\tau x} V_{xx}^{-1} \tilde{\tau}_x \mid \tilde{\tau}'_x V_{xx}^{-1} \tilde{\tau}_x \leq a \\ &\sim \tilde{\tau}_\varepsilon + V_{\tau x} V_{xx}^{-1/2} \mathbf{D} \mid \mathbf{D}' \mathbf{D} \leq a \\ &\sim \left(V_{\tau\tau}^\perp \right)^{1/2} \varepsilon + V_{\tau x} V_{xx}^{-1/2} \zeta_{LF,a}, \end{aligned}$$

where $\varepsilon \sim \mathcal{N}(\mathbf{0}, I_F)$ is independent of $\zeta_{LF,a}$. From Theorem 1, $V_{\tau\tau}^\parallel = V_{\tau x} V_{xx}^{-1} V_{x\tau} = V_{\tau x} V_{xx}^{-1/2} (V_{\tau x} V_{xx}^{-1/2})'$. Using Lemma A9, we have $V_{\tau x} V_{xx}^{-1/2} \zeta_{LF,a} \sim (V_{\tau\tau}^\parallel)^{1/2} \zeta_{LF,a}$. Therefore, Theorem 2 holds. $\square$

PROOF OF COROLLARY 2. Let $\mathbf{e}_f = (0, \dots, 0, 1, 0, \dots, 0)'$ be an F dimensional unit vector with f th element one. From Theorem 2,

$$\begin{aligned} \hat{\tau}_f - \tau_f \mid \mathcal{M} &\sim \mathbf{e}_f' \left(\mathbf{V}_{\tau\tau}^\perp \right)^{1/2} \boldsymbol{\varepsilon} + \mathbf{e}_f' (\mathbf{V}_{\tau\tau}^\parallel)_{LF}^{1/2} \boldsymbol{\zeta}_{LF,a} \\ &\sim \sqrt{\mathbf{e}_f' \mathbf{V}_{\tau\tau}^\perp \mathbf{e}_f} \cdot \varepsilon_0 + \sqrt{\mathbf{e}_f' \mathbf{V}_{\tau\tau}^\parallel \mathbf{e}_f} \cdot \mathbf{c}' \boldsymbol{\zeta}_{LF,a}, \end{aligned}$$

where $\mathbf{c}' = (\mathbf{e}_f' \mathbf{V}_{\tau\tau}^\parallel \mathbf{e}_f)^{-1/2} \mathbf{e}_f' (\mathbf{V}_{\tau\tau}^\parallel)_{LF}^{1/2}$ is a unit vector with length one. By the spherical symmetry of $\boldsymbol{\zeta}_{LF,a}$ in Proposition 3, $\mathbf{c}' \boldsymbol{\zeta}_{LF,a} \sim \eta_{LF,a}$ [Li et al., 2018, Lemma A1]. By the definition of R_f^2 in Corollary 1, we can further simplify the asymptotic sampling distribution of $\hat{\tau}_f$ as

$$\begin{aligned} \hat{\tau}_f - \tau_f \mid \mathcal{M} &\sim \sqrt{\mathbf{e}_f' \mathbf{V}_{\tau\tau} \mathbf{e}_f - \mathbf{e}_f' \mathbf{V}_{\tau\tau}^\parallel \mathbf{e}_f} \cdot \varepsilon_0 + \sqrt{\mathbf{e}_f' \mathbf{V}_{\tau\tau}^\parallel \mathbf{e}_f} \cdot \eta_{LF,a} \\ &\sim \sqrt{V_{\tau_f \tau_f} - V_{\tau_f \tau_f}^\parallel} \cdot \varepsilon_0 + \sqrt{V_{\tau_f \tau_f}^\parallel} \cdot \eta_{LF,a} \\ &\sim \sqrt{V_{\tau_f \tau_f}} \left(\sqrt{1 - V_{\tau_f \tau_f}^\parallel / V_{\tau_f \tau_f}} \cdot \varepsilon_0 + \sqrt{V_{\tau_f \tau_f}^\parallel / V_{\tau_f \tau_f}} \cdot \eta_{LF,a} \right) \\ &\sim \sqrt{V_{\tau_f \tau_f}} \left(\sqrt{1 - R_f^2} \cdot \varepsilon_0 + \sqrt{R_f^2} \cdot \eta_{LF,a} \right). \end{aligned}$$

Therefore, Corollary 2 holds. $\square$

PROOF OF COROLLARY 3. Recall that $\mathbf{R} = \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau\tau}^\parallel \mathbf{V}_{\tau\tau}^{-1/2} = \mathbf{\Gamma} \mathbf{\Pi}^2 \mathbf{\Gamma}'$ is the eigen-decomposition of $\mathbf{R}$, where $\mathbf{\Gamma} \in \mathbb{R}^{F \times F}$ is an orthogonal matrix, and $\mathbf{\Pi}^2 \in \mathbb{R}^{F \times F}$ is a diagonal matrix containing the eigenvalues of $\mathbf{R}$. Let $\mathbf{\Omega} = \mathbf{\Pi}^{-1} \mathbf{\Gamma}' \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1/2}$. Then $\mathbf{\Omega} \mathbf{\Omega}' = \mathbf{\Pi}^{-1} \mathbf{\Gamma}' \mathbf{R} \mathbf{\Gamma} \mathbf{\Pi}^{-1} = \mathbf{I}_F$, and $\mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1/2} = \mathbf{\Gamma} \mathbf{\Pi} \mathbf{\Omega}$ is the singular value decomposition of $\mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1/2}$. Note that $\mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau\tau}^\perp \mathbf{V}_{\tau\tau}^{-1/2} = \mathbf{I}_F - \mathbf{R} = \mathbf{\Gamma} (\mathbf{I}_F - \mathbf{\Pi}^2) \mathbf{\Gamma}'$. We can simplify the asymptotic distribution (4.1) as

$$\begin{aligned} \hat{\tau} - \tau \mid \mathcal{M} &\sim \left(\mathbf{V}_{\tau\tau}^\perp \right)^{1/2} \boldsymbol{\varepsilon} + \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1/2} \boldsymbol{\zeta}_{LF,a} \\ &= \mathbf{V}_{\tau\tau}^{1/2} \left\{ \mathbf{V}_{\tau\tau}^{-1/2} \left(\mathbf{V}_{\tau\tau}^\perp \right)^{1/2} \boldsymbol{\varepsilon} + \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1/2} \boldsymbol{\zeta}_{LF,a} \right\} \\ &\sim \mathbf{V}_{\tau\tau}^{1/2} \left\{ \left(\mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau\tau}^\perp \mathbf{V}_{\tau\tau}^{-1/2} \right)^{1/2} \boldsymbol{\varepsilon} + \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{V}_{\tau x} \mathbf{V}_{xx}^{-1/2} \boldsymbol{\zeta}_{LF,a} \right\} \\ &= \mathbf{V}_{\tau\tau}^{1/2} \left[\left\{ \mathbf{\Gamma} (\mathbf{I}_F - \mathbf{\Pi}^2) \mathbf{\Gamma}' \right\}^{1/2} \boldsymbol{\varepsilon} + \mathbf{\Gamma} \mathbf{\Pi} \mathbf{\Omega} \boldsymbol{\zeta}_{LF,a} \right] \\ &\sim \mathbf{V}_{\tau\tau}^{1/2} \left\{ \mathbf{\Gamma} (\mathbf{I}_F - \mathbf{\Pi}^2)^{1/2} \boldsymbol{\varepsilon} + \mathbf{\Gamma} \mathbf{\Pi} \mathbf{\Omega} \boldsymbol{\zeta}_{LF,a} \right\}. \end{aligned}$$

Because $\mathbf{\Omega} \mathbf{\Omega}' = \mathbf{I}_F$, there exists a matrix $\tilde{\mathbf{\Omega}} \in \mathbb{R}^{(L-1)F \times LF}$ such that $(\mathbf{\Omega}', \tilde{\mathbf{\Omega}}') \in \mathbb{R}^{LF \times LF}$ is orthogonal. From Proposition 3, $(\mathbf{\Omega}', \tilde{\mathbf{\Omega}}')' \boldsymbol{\zeta}_{LF,a} \sim$

$\zeta_{LF,a}$. Therefore, we can simplify the asymptotic distribution (4.1) as

$$\begin{aligned}\hat{\tau} - \tau \mid \mathcal{M} &\sim V_{\tau\tau}^{1/2} \left\{ \Gamma(I_F - \Pi^2)^{1/2} \varepsilon + \Gamma \Pi \Omega \zeta_{LF,a} \right\} \\ &\sim V_{\tau\tau}^{1/2} \Gamma \left\{ (I_F - \Pi^2)^{1/2} \varepsilon + (\Pi, 0) \begin{pmatrix} \Omega \\ \tilde{\Omega} \end{pmatrix} \zeta_{LF,a} \right\} \\ &\sim V_{\tau\tau}^{1/2} \Gamma \left\{ (I_F - \Pi^2)^{1/2} \varepsilon + (\Pi, 0) \zeta_{LF,a} \right\}.\end{aligned}$$

Therefore, Corollary 3 holds. $\square$

A3.4. Asymptotic sampling distribution of $\hat{\tau}$ under ReFMT_F.

PROOF OF THEOREM 7. Let $(\tilde{\tau}', \tilde{\theta}'_x)' = (\tilde{\tau}', \tilde{\theta}'_x[1], \dots, \tilde{\theta}'_x[H])'$ be a random vector following Gaussian distribution with mean zero and covariance matrix

$$\begin{pmatrix} V_{\tau\tau} & W_{\tau x}[1] & \dots & W_{\tau x}[H] \\ W_{x\tau}[1] & W_{xx}[1] & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ W_{x\tau}[H] & \mathbf{0} & \dots & W_{xx}[H] \end{pmatrix}.$$

From Lemma A10, under Condition 1,

$$\hat{\tau} - \tau \mid \mathcal{T}_F \sim \tilde{\tau} \mid \left\{ \tilde{\theta}'_x[h] (W_{xx}[h])^{-1} \tilde{\theta}_x[h] \leq a_h, 1 \leq h \leq H \right\},$$

where $\tilde{\theta}_x[h]$'s are mutually uncorrelated. The linear projection of $\tilde{\tau}$ on $\tilde{\theta}_x$ is $\sum_{h=1}^H W_{\tau x}[h] (W_{xx}[h])^{-1} \tilde{\theta}_x[h]$. Let $\tilde{\tau}_\varepsilon$ be the corresponding residual, which, by the identical covariance structure between $(\tilde{\tau}', \tilde{\theta}'_x)'$ and $(\hat{\tau} - \tau', \hat{\theta}'_x)'$, has the same covariance as the sampling covariance of $\hat{\tau}_\varepsilon$, the residual from the linear projection of $\hat{\tau}$ on $\hat{\theta}_x$. Because $\hat{\theta}_x$ and $\hat{\tau}_x$ are linear transformations of each other, $\hat{\tau}_\varepsilon$ is the same as the residual from the linear projection of $\hat{\tau}$ on $\hat{\tau}_x$. Thus, from Theorem 1, $\text{Cov}(\tilde{\tau}_\varepsilon) = \text{Cov}(\hat{\tau}_\varepsilon) = V_{\tau\tau}^\perp$. Let $\varepsilon = (V_{\tau\tau}^\perp)^{-1/2} \tilde{\tau}_\varepsilon \sim \mathcal{N}(0, I_F)$ be the standardization of $\tilde{\tau}_\varepsilon$, and $D_h = (W_{xx}[h])^{-1/2} \tilde{\theta}_x[h]$ be the standardization of $\tilde{\theta}_x[h]$. We have

$$\begin{aligned}&\hat{\tau} - \tau \mid \mathcal{T}_F \\ &\sim \tilde{\tau} \mid \left\{ \tilde{\theta}'_x[h] (W_{xx}[h])^{-1} \tilde{\theta}_x[h] \leq a_h, h = 1, \dots, H \right\} \\ &\sim \tilde{\tau}_\varepsilon + \sum_{h=1}^H W_{\tau x}[h] (W_{xx}[h])^{-1} \tilde{\theta}_x[h] \mid \left\{ \tilde{\theta}'_x[h] (W_{xx}[h])^{-1} \tilde{\theta}_x[h] \leq a_h, 1 \leq h \leq H \right\}\end{aligned}$$

$$\begin{aligned}
& \sim (\mathbf{V}_{\tau\tau}^\perp)^{1/2} \varepsilon + \sum_{h=1}^H \mathbf{W}_{\tau x}[h] (\mathbf{W}_{xx}[h])^{-1/2} \mathbf{D}_h \mid \{ \mathbf{D}'_h \mathbf{D}_h \leq a_h, 1 \leq h \leq H \} \\
& \sim (\mathbf{V}_{\tau\tau}^\perp)^{1/2} \varepsilon + \sum_{h=1}^H \mathbf{W}_{\tau x}[h] (\mathbf{W}_{xx}[h])^{-1/2} \boldsymbol{\zeta}_{LF_h, a_h}.
\end{aligned}$$

Because $\mathbf{W}_{\tau\tau}^\parallel[h] = \mathbf{W}_{\tau x}[h] (\mathbf{W}_{xx}[h])^{-1/2} \{ \mathbf{W}_{\tau x}[h] (\mathbf{W}_{xx}[h])^{-1/2} \}'$ by definition, Theorem 7 holds by Lemma A9. $\square$

PROOF OF COROLLARY 5. Recall that $\mathbf{e}_f$ is a unit vector with the f th coordinate being one. From Theorem 7, the asymptotic sampling distribution of $\hat{\tau}_f$ under ReFMT_F is

$$\begin{aligned}
& \hat{\tau}_f - \tau_f \mid \mathcal{T}_F \\
& \sim \mathbf{e}'_f (\mathbf{V}_{\tau\tau}^\perp)^{1/2} \varepsilon + \sum_{h=1}^H \mathbf{e}'_f (\mathbf{W}_{\tau\tau}^\parallel[h])^{1/2}_{LF_h} \boldsymbol{\zeta}_{LF_h, a_h} \\
& \sim \sqrt{\mathbf{e}'_f \mathbf{V}_{\tau\tau}^\perp \mathbf{e}_f} \cdot \varepsilon_0 + \sum_{h=1}^H \sqrt{\mathbf{e}'_f \mathbf{W}_{\tau\tau}^\parallel[h] \mathbf{e}_f} \cdot \mathbf{c}'_h \boldsymbol{\zeta}_{LF_h, a_h} \\
& \sim \sqrt{V_{\tau_f \tau_f}} \left(\sqrt{1 - V_{\tau_f \tau_f}^\parallel / V_{\tau_f \tau_f}} \cdot \varepsilon_0 + \sum_{h=1}^H \sqrt{W_{\tau_f \tau_f}^\parallel[h] / V_{\tau_f \tau_f}} \cdot \mathbf{c}'_h \boldsymbol{\zeta}_{LF_h, a_h} \right)
\end{aligned}$$

where $\mathbf{c}'_h = (\mathbf{e}'_f \mathbf{W}_{\tau\tau}^\parallel[h] \mathbf{e}_f)^{-1/2} \mathbf{e}'_f (\mathbf{W}_{\tau\tau}^\parallel[h])^{1/2}_{LF_h}$ is a unit vector with length one. By the definitions of R_f^2 and the $\rho_f^2[h]$'s, and the spherical symmetry of the $\boldsymbol{\zeta}_{LF_h, a_h}$'s from Proposition 3,

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

Therefore, Corollary 5 holds. $\square$

A3.5. Asymptotic sampling distribution of $\hat{\tau}$ under ReFMT_{CF} .

PROOF OF THEOREM A1. For each tier $\mathcal{S}_j$, let $\hat{\boldsymbol{\delta}}_e[j]$ be the concatenation of $\hat{\boldsymbol{\theta}}_{e[t]}[h]$ with $(t, h) \in \mathcal{S}_j$, and $\hat{\boldsymbol{\delta}}'_e = (\hat{\boldsymbol{\delta}}'_e[1], \dots, \hat{\boldsymbol{\delta}}'_e[J])$. From Proposition A1, under the CRFE, $(\hat{\boldsymbol{\tau}}' - \boldsymbol{\tau}', \hat{\boldsymbol{\delta}}'_e)'$ has sampling mean zero and sampling

covariance matrix

$$(A1) \quad \begin{pmatrix} \mathbf{V}_{\tau\tau} & \mathbf{U}_{\tau e}[1] & \dots & \mathbf{U}_{\tau e}[J] \\ \mathbf{U}_{e\tau}[1] & \mathbf{U}_{ee}[1] & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{U}_{e\tau}[J] & \mathbf{0} & \dots & \mathbf{U}_{ee}[J] \end{pmatrix}.$$

Let $(\tilde{\tau}', \tilde{\delta}_e')'$ be a Gaussian random vector with mean zero and covariance matrix (A1). From Lemma A10, under Condition 1,

$$\hat{\tau} - \tau \mid \mathcal{T}_{CF} \sim \tilde{\tau} \mid \left\{ \tilde{\delta}_e'[j](\mathbf{U}_{ee}[j])^{-1} \tilde{\delta}_e[j] \leq a_j, j = 1, \dots, J \right\}.$$

Using the same logic as the proof of Theorem 7, we can simplify the asymptotic sampling distribution of $\hat{\tau}$ under ReFMT_{CF} as

$$\hat{\tau} - \tau \mid \mathcal{T}_{CF} \sim \left(\mathbf{V}_{\tau\tau}^\perp \right)^{1/2} \varepsilon + \sum_{j=1}^J \mathbf{U}_{\tau e}[j] (\mathbf{U}_{ee}[j])^{-1/2} \zeta_{\lambda_j, a_j}.$$

Because by definition $\mathbf{U}_{\tau e}[j](\mathbf{U}_{ee}[j])^{-1/2} \{ \mathbf{U}_{\tau e}[j](\mathbf{U}_{ee}[j])^{-1/2} \}' = \mathbf{U}_{\tau\tau}^\parallel[j]$, from Lemma A9, Theorem A1 holds. $\square$

PROOF OF COROLLARY A1. Recall that $\mathbf{e}_f$ is a unit vector with the f th coordinate being one. From Theorem A1, the asymptotic sampling distribution of $\hat{\tau}_f$ under ReFMT_{CF} is

$$\begin{aligned} & \hat{\tau}_f - \tau_f \mid \mathcal{T}_{CF} \\ & \sim \mathbf{e}_f' \left(\mathbf{V}_{\tau\tau}^\perp \right)^{1/2} \varepsilon + \sum_{j=1}^J \mathbf{e}_f' \left(\mathbf{U}_{\tau\tau}^\parallel[j] \right)_{\lambda_j}^{1/2} \zeta_{\lambda_j, a_j} \\ & \sim \sqrt{\mathbf{e}_f' \mathbf{V}_{\tau\tau}^\perp \mathbf{e}_f} \cdot \varepsilon_0 + \sum_{j=1}^J \sqrt{\mathbf{e}_f' \mathbf{U}_{\tau\tau}^\parallel[j] \mathbf{e}_f} \cdot \mathbf{c}_j' \zeta_{\lambda_j, a_j} \\ & \sim \sqrt{V_{\tau_f \tau_f}} \left(\sqrt{1 - V_{\tau_f \tau_f}^\parallel / V_{\tau_f \tau_f}} \cdot \varepsilon_0 + \sum_{j=1}^J \sqrt{U_{\tau_f \tau_f}^\parallel[j] / V_{\tau_f \tau_f}} \cdot \mathbf{c}_j' \zeta_{\lambda_j, a_j} \right), \end{aligned}$$

where $\mathbf{c}_j' = (\mathbf{e}_f' \mathbf{U}_{\tau\tau}^\parallel[j] \mathbf{e}_f)^{-1/2} \mathbf{e}_f' (\mathbf{U}_{\tau\tau}^\parallel[j])_{\lambda_j}^{1/2}$ is a unit vector with length one. By the definitions of R_f^2 and $\beta_f^2[j]$'s, and the spherical symmetry of ζ_{λ_j, a_j} 's from Proposition 3,

$$\hat{\tau}_f - \tau_f \mid \mathcal{T}_{CF} \sim \sqrt{V_{\tau_f \tau_f}} \left(\sqrt{1 - R_f^2} \cdot \varepsilon_0 + \sum_{j=1}^J \sqrt{\beta_f^2[j]} \cdot \eta_{\lambda_j, a_j} \right).$$

Therefore, Corollary A1 holds. $\square$

A4. Reduction in asymptotic sampling covariances. We use Var_a and Cov_a to denote the variance and covariance of the asymptotic distributions of sequences of random variables and random vectors, respectively.

PROOF OF THEOREM 3. First, we calculate the reduction in the asymptotic sampling covariance of $\hat{\tau}$. From Proposition 1, $\text{Cov}_a\{\sqrt{n}(\hat{\tau} - \tau)\} = \lim_{n \rightarrow \infty} n\mathbf{V}_{\tau\tau} = \lim_{n \rightarrow \infty} n\mathbf{V}_{\tau\tau}^\perp + \lim_{n \rightarrow \infty} n\mathbf{V}_{\tau\tau}^\parallel$. For notational simplicity, we omit the limiting signs. From Theorem 2 and Proposition 3,

$$\begin{aligned} & \text{Cov}_a\{\sqrt{n}(\hat{\tau} - \tau) \mid \mathcal{M}\} \\ &= \text{Cov} \left\{ \left(n\mathbf{V}_{\tau\tau}^\perp \right)^{1/2} \varepsilon \right\} + n \left(\mathbf{V}_{\tau\tau}^\parallel \right)_{LF}^{1/2} \text{Cov}(\zeta_{LF,a}) \left\{ \left(\mathbf{V}_{\tau\tau}^\parallel \right)_{LF}^{1/2} \right\}' \\ &= n\mathbf{V}_{\tau\tau}^\perp + v_{LF,a} \cdot n \left(\mathbf{V}_{\tau\tau}^\parallel \right)_{LF}^{1/2} \left\{ \left(\mathbf{V}_{\tau\tau}^\parallel \right)_{LF}^{1/2} \right\}' \\ &= n\mathbf{V}_{\tau\tau}^\perp + v_{LF,a} \cdot n\mathbf{V}_{\tau\tau}^\parallel. \end{aligned}$$

Therefore, the reduction in asymptotic sampling covariance is $(1 - v_{LF,a})n\mathbf{V}_{\tau\tau}^\parallel$.

Second, we consider the PRIASV of $\hat{\tau}_f$. From Proposition 1 and Corollary 2, the asymptotic sampling variance of $\hat{\tau}_f$ are

$$\begin{aligned} \text{Var}_a\{\sqrt{n}(\hat{\tau}_f - \tau_f)\} &= nV_{\tau_f\tau_f}, \\ \text{Var}_a\{\sqrt{n}(\hat{\tau}_f - \tau_f) \mid \mathcal{M}\} &= nV_{\tau_f\tau_f}(1 - R_f^2 + R_f^2 \cdot v_{LF,a}) \\ &= nV_{\tau_f\tau_f}\{1 - (1 - v_{LF,a})R_f^2\}. \end{aligned}$$

Therefore, the PRIASV of $\hat{\tau}_f$ is $(1 - v_{LF,a})R_f^2$. $\square$

PROOF OF THEOREM 8. First, we calculate the reduction in the asymptotic sampling covariance of $\hat{\tau}$. From Theorem 7 and Proposition 3,

$$\begin{aligned} & \text{Cov}_a\{\sqrt{n}(\hat{\tau} - \tau) \mid \mathcal{T}_F\} \\ &= n\mathbf{V}_{\tau\tau}^\perp + n \sum_{h=1}^H \left(\mathbf{W}_{\tau\tau}^\parallel[h] \right)_{LF_h}^{1/2} \text{Cov}(\zeta_{LF_h,a_h}) \left\{ \left(\mathbf{W}_{\tau\tau}^\parallel[h] \right)_{LF_h}^{1/2} \right\}' \\ &= n\mathbf{V}_{\tau\tau}^\perp + n \sum_{h=1}^H v_{LF_h,a_h} \mathbf{W}_{\tau\tau}^\parallel[h]. \end{aligned}$$

Because $\hat{\tau}_x$ and $\hat{\theta}_x$ are linear transformations of each other, the sampling covariances of $\hat{\tau}$ explained by $\hat{\tau}_x$ and $\hat{\theta}_x$ are the same, i.e., $\mathbf{V}_{\tau\tau}^\parallel = \sum_{h=1}^H \mathbf{W}_{\tau\tau}^\parallel[h]$. Thus,

$$\text{Cov}_a\{\sqrt{n}(\hat{\tau} - \tau)\} = n\mathbf{V}_{\tau\tau}^\perp + n\mathbf{V}_{\tau\tau}^\parallel = n\mathbf{V}_{\tau\tau}^\perp + n \sum_{h=1}^H \mathbf{W}_{\tau\tau}^\parallel[h].$$

Therefore, the reduction in asymptotic sampling covariance of $\hat{\boldsymbol{\tau}}$ is

$$\text{Cov}_a\{\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})\} - \text{Cov}_a\{\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \mid \mathcal{T}_F\} = n \sum_{h=1}^H (1 - v_{LF_h, a_h}) \mathbf{W}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel[h].$$

Second, we consider the PRIASV of $\hat{\tau}_f$. From Proposition 1 and Corollary 5, the asymptotic sampling variance of $\hat{\tau}_f$ under the CRFE and ReFMT_F are

$$\begin{aligned} \text{Var}_a\{\sqrt{n}(\hat{\tau}_f - \tau_f)\} &= nV_{\tau_f\tau_f}, \\ \text{Var}_a\{\sqrt{n}(\hat{\tau}_f - \tau_f) \mid \mathcal{T}_F\} &= nV_{\tau_f\tau_f} \left(1 - R_f^2 + \sum_{h=1}^H \rho_f^2[h] v_{LF_h, a_h} \right). \end{aligned}$$

By definition of $\rho_f^2[h]$ and the fact that $\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel = \sum_{h=1}^H \mathbf{W}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel[h]$, we have $R_f^2 = \sum_{h=1}^H \rho_f^2[h]$. Therefore, the PRIASV of $\hat{\tau}_f$ under ReFMT_F is

$$\begin{aligned} & \frac{nV_{\tau_f\tau_f} - nV_{\tau_f\tau_f} \left(1 - R_f^2 + \sum_{h=1}^H \rho_f^2[h] v_{LF_h, a_h} \right)}{nV_{\tau_f\tau_f}} \\ &= R_f^2 - \sum_{h=1}^H \rho_f^2[h] v_{LF_h, a_h} = \sum_{h=1}^H (1 - v_{LF_h, a_h}) \rho_f^2[h]. \end{aligned}$$

□

PROOF OF THEOREM A2. Note that the $\{\hat{\boldsymbol{\theta}}_{e[t]}[h]\}_{(t,h) \in \mathcal{S}_j}$'s are from a block-wise Gram-Schmidt orthogonalization of $\hat{\boldsymbol{\tau}}_x$. The proof is similar to that of Theorem 8. We omit it. □

A5. Peakedness of the asymptotic sampling distributions.

A5.1. *Lemmas and propositions for peakedness.* Recall that we say a random vector $\boldsymbol{\phi} \in \mathbb{R}^m$ is more peaked than another random vector $\boldsymbol{\psi} \in \mathbb{R}^m$ 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 $\mathcal{K} \subset \mathbb{R}^m$.

LEMMA A11. *If two m dimensional symmetric random vectors $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ satisfy $\boldsymbol{\phi}_1 \succ \boldsymbol{\phi}_2$, then for any matrix $\mathbf{C} \in \mathbb{R}^{p \times m}$, $\mathbf{C}\boldsymbol{\phi}_1 \succ \mathbf{C}\boldsymbol{\phi}_2$.*

LEMMA A12. *Let $\boldsymbol{\psi}, \boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ be three independent m dimensional symmetric random vectors. If $\boldsymbol{\psi}$ is central symmetric unimodal and $\boldsymbol{\phi}_1 \succ \boldsymbol{\phi}_2$, then $\boldsymbol{\psi} + \boldsymbol{\phi}_1 \succ \boldsymbol{\psi} + \boldsymbol{\phi}_2$.*

PROOF OF LEMMAS A11 AND A12. See Dharmadhikari and Joag-Dev [1988, Lemma 7.2 and Theorem 7.5]. $\square$

The following lemma states that truncating a standard Gaussian random vector within a ball makes it more peaked. Although the result seems intuitive, the proof below is a little tedious due to some technical reasons.

LEMMA A13. *Let $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ and $\zeta_{m,a} \sim \varepsilon \mid \varepsilon' \varepsilon \leq a$. Then $\zeta_{m,a} \succ \varepsilon$.*

PROOF OF LEMMA A13. For any symmetric convex set $\mathcal{K}$, let $\|\varepsilon\|_2 = (\varepsilon' \varepsilon)^{1/2}$ be the l_2 -norm of ε , and $G(r) = P(\varepsilon \in \mathcal{K} \mid \|\varepsilon\|_2 = r)$ be the conditional probability that ε is in $\mathcal{K}$ given the length of ε . Let ϕ be a random vector uniformly distributed on m dimensional unit sphere. By the spherical symmetry of ε , we can simplify $G(r)$ as

$$G(r) = P\left(\|\varepsilon\|_2 \cdot \frac{\varepsilon}{\|\varepsilon\|_2} \in \mathcal{K} \mid \|\varepsilon\|_2 = r\right) = P(r\phi \in \mathcal{K}) = P(\phi \in r^{-1}\mathcal{K}),$$

where $r^{-1}\mathcal{K} = \{r^{-1}\mathbf{x} : \mathbf{x} \in \mathcal{K}\}$. For any $r_1 \geq r_2 \geq 0$, if $\tilde{\mathbf{x}} \in r_1^{-1}\mathcal{K}$, then $r_1\tilde{\mathbf{x}} \in \mathcal{K}$. By the symmetric convexity of $\mathcal{K}$, we then have $r_2\tilde{\mathbf{x}} = r_2/r_1 \cdot (r_1\tilde{\mathbf{x}}) \in \mathcal{K}$, i.e., $\tilde{\mathbf{x}} \in r_2^{-1}\mathcal{K}$. Thus, $r_1^{-1}\mathcal{K} \subset r_2^{-1}\mathcal{K}$, which further implies $G(r_1) \leq G(r_2)$. Therefore, $G(r)$ is a nonincreasing function of $r \in [0, \infty)$. We can represent the probabilities that $\zeta_{m,a}$ and ε belong to $\mathcal{K}$, respectively, as follows:

$$\begin{aligned} P(\zeta_{m,a} \in \mathcal{K}) &= P(\varepsilon \in \mathcal{K} \mid \varepsilon' \varepsilon \leq a) = \mathbb{E} \{P(\varepsilon \in \mathcal{K} \mid \varepsilon' \varepsilon \leq a, \|\varepsilon\|_2) \mid \varepsilon' \varepsilon \leq a\} \\ &= \mathbb{E} \{G(\|\varepsilon\|_2) \mid \varepsilon' \varepsilon \leq a\} = \mathbb{E} \{G(\chi_m) \mid \chi_m^2 \leq a\}, \end{aligned}$$

and $P(\varepsilon \in \mathcal{K}) = \mathbb{E}\{G(\chi_m)\}$. By the monotone nonincreasing property of $G(r)$, we have

$$\begin{aligned} P(\varepsilon \in \mathcal{K}) &= \mathbb{E} \{G(\chi_m)\} \\ &= P(\chi_m^2 \leq a) \mathbb{E} \{G(\chi_m) \mid \chi_m^2 \leq a\} + P(\chi_m^2 > a) \mathbb{E} \{G(\chi_m) \mid \chi_m^2 > a\} \\ &\leq P(\chi_m^2 \leq a) \mathbb{E} \{G(\chi_m) \mid \chi_m^2 \leq a\} + P(\chi_m^2 > a) G(\sqrt{a}) \\ &\leq P(\chi_m^2 \leq a) \mathbb{E} \{G(\chi_m) \mid \chi_m^2 \leq a\} + P(\chi_m^2 > a) \mathbb{E} \{G(\chi_m) \mid \chi_m^2 \leq a\} \\ &= \mathbb{E} \{G(\chi_m) \mid \chi_m^2 \leq a\} = P(\zeta_{m,a} \in \mathcal{K}). \end{aligned}$$

Therefore, Lemma A13 holds. $\square$

PROOF OF PROPOSITION 4. From Lemma A11, for any vector $\mathbf{c} \in \mathbb{R}^m$, $\mathbf{c}'\phi \succ \mathbf{c}'\psi$. Thus,

$$\text{Var}(\mathbf{c}'\phi) = \mathbb{E} \{(\mathbf{c}'\phi)^2\} = \int_0^\infty \left[1 - P \{(\mathbf{c}'\phi)^2 \leq t\}\right] dt$$

$$\begin{aligned}
&= \int_0^\infty \{1 - P(\mathbf{c}'\phi \in [-t, t])\} dt \\
&\leq \int_0^\infty \{1 - P(\mathbf{c}'\psi \in [-t, t])\} dt = \text{Var}(\mathbf{c}'\psi).
\end{aligned}$$

Because the above inequality holds for any $\mathbf{c}$, $\text{Cov}(\phi) \leq \text{Cov}(\psi)$. $\square$

A5.2. *Propositions for simultaneous diagonalization.*

PROOF OF PROPOSITION 8. Recall that Ψ is the matrix such that $\mathbf{b}_q = \Psi \mathbf{c}_q$. By the construction of the $\mathbf{c}_q$'s,

$$\tilde{C} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{c}_q \mathbf{c}_q') \equiv \text{diag} \left\{ \tilde{C}[1, 1], \dots, \tilde{C}[H, H] \right\}$$

is a block diagonal matrix, and $\tilde{B} = \Psi \tilde{C} \Psi'$. Partition $\tilde{C}$ into $\tilde{C} = (\tilde{C}[1, 1], \dots, \tilde{C}[H, H])$. From Proposition 6, under the additivity, we have

$$\begin{aligned}
\mathbf{W}_{\tau x}[h] &= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{b}_q \mathbf{c}_q'[h]) \otimes \mathbf{S}_{1,x} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\Psi \mathbf{c}_q \mathbf{c}_q'[h]) \otimes \mathbf{S}_{1,x} \\
&= \left(\Psi \tilde{C}[h] \right) \otimes \mathbf{S}_{1,x}, \\
\mathbf{W}_{xx}[h] &= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{c}_q[h] \mathbf{c}_q'[h]) \otimes \mathbf{S}_{xx} = \tilde{C}[h, h] \otimes \mathbf{S}_{xx}.
\end{aligned}$$

Thus, under the additivity, for each $1 \leq h \leq H$, $\mathbf{W}_{\tau\tau}^\parallel[h]$ reduces to

$$\begin{aligned}
\mathbf{W}_{\tau\tau}^\parallel[h] &= \mathbf{W}_{\tau x}[h] \mathbf{W}_{xx}^{-1}[h] \mathbf{W}_{x\tau}[h] \\
&= \left\{ \Psi \tilde{C}[h] \left(\tilde{C}[h, h] \right)^{-1} (\tilde{C}[h, h])' \Psi' \right\} \otimes (\mathbf{S}_{1,x} \mathbf{S}_{xx}^{-1} \mathbf{S}_{x,1}) \\
&= S_{11}^\parallel \cdot \Psi \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, \tilde{C}[h, h], \mathbf{0}, \dots, \mathbf{0}) \cdot \Psi' \\
&= S_{11}^\parallel \cdot \Psi \cdot \tilde{C}^{1/2} \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, \mathbf{I}_{F_h}, \mathbf{0}, \dots, \mathbf{0}) \cdot \tilde{C}^{1/2} \cdot \Psi'
\end{aligned}$$

Under the additivity, $\mathbf{V}_{\tau\tau} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q' \cdot S_{11} = S_{11} \cdot \tilde{B}$. Thus,

$$\begin{aligned}
&\mathbf{V}_{\tau\tau}^{-1/2} \mathbf{W}_{\tau\tau}^\parallel[h] \mathbf{V}_{\tau\tau}^{-1/2} \\
&= S_{11}^{-1/2} \tilde{B}^{-1/2} \left\{ S_{11}^\parallel \cdot \Psi \cdot \tilde{C}^{1/2} \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, \mathbf{I}_{F_h}, \mathbf{0}, \dots, \mathbf{0}) \cdot \tilde{C}^{1/2} \cdot \Psi' \right\} \cdot S_{11}^{-1/2} \tilde{B}^{-1/2} \\
&= S_{11}^\parallel / S_{11} \cdot \tilde{B}^{-1/2} \Psi \tilde{C}^{1/2} \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, \mathbf{I}_{F_h}, \mathbf{0}, \dots, \mathbf{0}) \cdot \tilde{C}^{1/2} \Psi' \tilde{B}^{-1/2}
\end{aligned}$$

$$= \mathbf{\Gamma} \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, S_{11}^{\parallel}/S_{11} \cdot \mathbf{I}_{F_h}, \mathbf{0}, \dots, \mathbf{0}) \cdot \mathbf{\Gamma}',$$

where $\mathbf{\Gamma} = \tilde{\mathbf{B}}^{-1/2} \mathbf{\Psi} \tilde{\mathbf{C}}^{1/2}$. Because

$$\mathbf{\Gamma} \mathbf{\Gamma}' = \tilde{\mathbf{B}}^{-1/2} \mathbf{\Psi} \tilde{\mathbf{C}}^{1/2} \tilde{\mathbf{C}}^{1/2} \mathbf{\Psi}' \tilde{\mathbf{B}}^{-1/2} = \tilde{\mathbf{B}}^{-1/2} (\mathbf{\Psi} \tilde{\mathbf{C}} \mathbf{\Psi}') \tilde{\mathbf{B}}^{-1/2} = \tilde{\mathbf{B}}^{-1/2} \tilde{\mathbf{B}} \tilde{\mathbf{B}}^{-1/2} = \mathbf{I}_F,$$

we have that $\mathbf{\Gamma}$ is an orthogonal matrix. Note that $\text{diag}(\mathbf{0}, \dots, \mathbf{0}, S_{11}^{\parallel}/S_{11} \cdot \mathbf{I}_{F_h}, \mathbf{0}, \dots, \mathbf{0})$ is a diagonal matrix with exactly F_h nonzero elements, which are all equal to $S_{11}^{\parallel}/S_{11}$. Therefore, Condition 2 and Proposition 8 hold. $\square$

PROOF OF PROPOSITION A2. Recall $\mathbf{\Psi}$ is the matrix such that $\mathbf{b}_q = \mathbf{\Psi} \mathbf{c}_q$. From the proof of Proposition 8, $\tilde{\mathbf{C}} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{c}_q \mathbf{c}_q')$ is a block diagonal matrix, and $\tilde{\mathbf{B}} = \mathbf{\Psi} \tilde{\mathbf{C}} \mathbf{\Psi}'$. From Proposition A1, under the additivity, $\mathbf{W}_{\tau e[t]}[h]$ and $\mathbf{W}_{e[t]e[t]}[h]$ reduce to

$$\begin{aligned} \mathbf{W}_{\tau e[t]}[h] &= 2^{-2(K-1)} \left\{ \mathbf{\Psi} \sum_{q=1}^Q n_q^{-1} (\mathbf{c}_q \mathbf{c}_q'[h]) \right\} \otimes \mathbf{S}_{1,e[t]} = (\mathbf{\Psi} \tilde{\mathbf{C}}[h]) \otimes \mathbf{S}_{1,e[t]} \\ \mathbf{W}_{e[t]e[t]}[h] &= 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} (\mathbf{c}_q[h] \mathbf{c}_q'[h]) \otimes \mathbf{S}_{e[t]e[t]} = \tilde{\mathbf{C}}[h, h] \otimes \mathbf{S}_{e[t]e[t]}, \end{aligned}$$

which then implies

$$\begin{aligned} \mathbf{W}_{e[t]e[t]}^{\parallel}[h] &\equiv \mathbf{W}_{\tau e[t]}[h] \mathbf{W}_{e[t]e[t]}^{-1}[h] \mathbf{W}_{\tau e[t]}'[h] \\ &= \left\{ \mathbf{\Psi} \tilde{\mathbf{C}}[h] (\tilde{\mathbf{C}}[h, h])^{-1} (\tilde{\mathbf{C}}[h, h])' \mathbf{\Psi}' \right\} \otimes (\mathbf{S}_{1,e[t]} \mathbf{S}_{e[t]e[t]}^{-1} \mathbf{S}_{e[t],1}) \\ &= (\mathbf{S}_{1,e[t]} \mathbf{S}_{e[t]e[t]}^{-1} \mathbf{S}_{e[t],1}) \cdot \mathbf{\Psi} \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, \tilde{\mathbf{C}}[h, h], \mathbf{0}, \dots, \mathbf{0}) \cdot \mathbf{\Psi}' \\ &= S_{11}^{\parallel}[t] \cdot \mathbf{\Psi} \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, \tilde{\mathbf{C}}[h, h], \mathbf{0}, \dots, \mathbf{0}) \cdot \mathbf{\Psi}' \\ &= S_{11}^{\parallel}[t] \cdot \mathbf{\Psi} \cdot \tilde{\mathbf{C}}^{1/2} \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, \mathbf{I}_{F_h}, \mathbf{0}, \dots, \mathbf{0}) \cdot \tilde{\mathbf{C}}^{1/2} \cdot \mathbf{\Psi}', \end{aligned}$$

where $S_{11}^{\parallel}[t] = \mathbf{S}_{1,e[t]} \mathbf{S}_{e[t]e[t]}^{-1} \mathbf{S}_{e[t],1}$ is the finite population variance of the linear projection of potential outcome $Y(1)$ on orthogonalized covariates in tier t . Because $\mathbf{V}_{\tau\tau}$ reduces to $\mathbf{V}_{\tau\tau} = S_{11} \tilde{\mathbf{B}}$ under the additivity, we have

$$\begin{aligned} &\mathbf{V}_{\tau\tau}^{-1/2} \mathbf{W}_{e[t]e[t]}^{\parallel}[h] \mathbf{V}_{\tau\tau}^{-1/2} \\ &= S_{11}^{\parallel}[t]/S_{11} \cdot \tilde{\mathbf{B}}^{-1/2} \left\{ \mathbf{\Psi} \tilde{\mathbf{C}}^{1/2} \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, \mathbf{I}_{F_h}, \mathbf{0}, \dots, \mathbf{0}) \cdot \tilde{\mathbf{C}}^{1/2} \mathbf{\Psi}' \right\} \tilde{\mathbf{B}}^{-1/2} \\ &= S_{11}^{\parallel}[t]/S_{11} \cdot \tilde{\mathbf{B}}^{-1/2} \mathbf{\Psi} \tilde{\mathbf{C}}^{1/2} \cdot \text{diag}(\mathbf{0}, \dots, \mathbf{0}, \mathbf{I}_{F_h}, \mathbf{0}, \dots, \mathbf{0}) \cdot \tilde{\mathbf{C}}^{1/2} \mathbf{\Psi}' \tilde{\mathbf{B}}^{-1/2} \end{aligned}$$

$$= \mathbf{\Gamma} \mathbf{\Omega}_{[t]}[h] \mathbf{\Gamma}',$$

where $\mathbf{\Omega}_{[t]}[h] = \text{diag}(\mathbf{0}, \dots, \mathbf{0}, S_{11}^{\parallel}[t]/S_{11} \cdot \mathbf{I}_{F_h}, \mathbf{0}, \dots, \mathbf{0})$ is a diagonal matrix, and $\mathbf{\Gamma} = \tilde{\mathbf{B}}^{-1/2} \mathbf{\Psi} \tilde{\mathbf{C}}^{1/2}$ is an orthogonal matrix. By definition, we then have

$$\begin{aligned} & \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{U}_{\tau\tau}^{\parallel}[j] \mathbf{V}_{\tau\tau}^{-1/2} \\ &= \mathbf{V}_{\tau\tau}^{-1/2} \left(\sum_{(t,h) \in \mathcal{S}_j} \mathbf{W}_{e[t]e[t]}^{\parallel}[h] \right) \mathbf{V}_{\tau\tau}^{-1/2} = \sum_{(t,h) \in \mathcal{S}_j} \mathbf{V}_{\tau\tau}^{-1/2} \mathbf{W}_{e[t]e[t]}^{\parallel}[h] \mathbf{V}_{\tau\tau}^{-1/2} \\ &= \sum_{(t,h) \in \mathcal{S}_j} \mathbf{\Gamma} \mathbf{\Omega}_{[t]}[h] \mathbf{\Gamma}' = \mathbf{\Gamma} \left(\sum_{(t,h) \in \mathcal{S}_j} \mathbf{\Omega}_{[t]}[h] \right) \mathbf{\Gamma}', \quad (1 \leq j \leq J) \end{aligned}$$

where $\sum_{(t,h) \in \mathcal{S}_j} \mathbf{\Omega}_{[t]}[h]$ is a diagonal matrix. Therefore, Condition A1 and Proposition A2 hold. $\square$

A5.3. Peakedness under ReFM.

PROOF OF THEOREM 4. Let $\mathbf{D} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{LF})$. From Lemma A13, $\zeta_{LF,a} \succ \mathbf{D}$. From Lemma A11, $(\mathbf{V}_{\tau\tau}^{\parallel})_{LF}^{1/2} \zeta_{LF,a} \succ (\mathbf{V}_{\tau\tau}^{\parallel})_{LF}^{1/2} \mathbf{D}$. From Proposition 2, ε is central convex unimodal. From Lemma A6, $(\mathbf{V}_{\tau\tau}^{\perp})^{1/2} \varepsilon$ is also central convex unimodal. Then, from Lemma A12,

$$(\mathbf{V}_{\tau\tau}^{\perp})^{1/2} \varepsilon + (\mathbf{V}_{\tau\tau}^{\parallel})_{LF}^{1/2} \zeta_{LF,a} \succ (\mathbf{V}_{\tau\tau}^{\perp})^{1/2} \varepsilon + (\mathbf{V}_{\tau\tau}^{\parallel})_{LF}^{1/2} \mathbf{D} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_{\tau\tau}).$$

Therefore, Theorem 4 holds. $\square$

To prove Theorem 5, we need the following three lemmas.

LEMMA A14. *Let $\varepsilon, \eta \sim \mathcal{N}(0, 1)$ be two independent standard Gaussian random variables. For any $1 \geq \rho \geq \tilde{\rho} \geq 0$, $a \geq 0$ and $c \geq 0$,*

$$P\left(|\sqrt{1-\rho^2} \cdot \varepsilon_0 + \rho \cdot \eta| \leq c \mid \eta^2 \leq a\right) \geq P\left(|\sqrt{1-\tilde{\rho}^2} \cdot \varepsilon_0 + \tilde{\rho} \cdot \eta| \leq c \mid \eta^2 \leq a\right).$$

PROOF OF LEMMA A14. It follows directly from Li et al. [2018, Lemma A3], and is also a special case of Das Gupta et al. [1972, Theorem 2.1]. $\square$

The following lemma extends the above lemma to the multivariate case.

LEMMA A15. *Let ε and η be two independent m dimensional standard Gaussian random vectors, $(\rho_1, \dots, \rho_m)$ be m constants in $[0, 1]$, and Δ be*

a diagonal matrix with diagonal elements $(\rho_1, \dots, \rho_m)$. For any $r \geq 0$, the probability

$$P \left\{ (\mathbf{I}_m - \mathbf{\Delta}^2)^{1/2} \boldsymbol{\varepsilon} + \mathbf{\Delta} \boldsymbol{\eta} \in \mathcal{B}_m(r) \mid \boldsymbol{\eta}' \boldsymbol{\eta} \leq a \right\}$$

is nondecreasing in $(\rho_1, \dots, \rho_m)$. Specifically, for any constants $(\rho_1, \dots, \rho_m)$ and $(\tilde{\rho}_1, \dots, \tilde{\rho}_m)$ in $[0, 1]$, if $\rho_j \geq \tilde{\rho}_j$ for $1 \leq j \leq m$, then for any $r \geq 0$,

$$\begin{aligned} & P \left\{ (\mathbf{I}_F - \mathbf{\Delta}^2)^{1/2} \boldsymbol{\varepsilon} + \mathbf{\Delta} \boldsymbol{\eta} \in \mathcal{B}_m(r) \mid \boldsymbol{\eta}' \boldsymbol{\eta} \leq a \right\} \\ (A1) \quad & \geq P \left\{ (\mathbf{I}_F - \tilde{\mathbf{\Delta}}^2)^{1/2} \boldsymbol{\varepsilon} + \tilde{\mathbf{\Delta}} \boldsymbol{\eta} \in \mathcal{B}_m(r) \mid \boldsymbol{\eta}' \boldsymbol{\eta} \leq a \right\}. \end{aligned}$$

where $\mathbf{\Delta} = \text{diag}(\rho_1, \dots, \rho_m)$ and $\tilde{\mathbf{\Delta}} = \text{diag}(\tilde{\rho}_1, \dots, \tilde{\rho}_m)$.

PROOF OF LEMMA A15. To prove Lemma A15, it suffices to prove that for any constants $(\rho_1, \dots, \rho_m)$ and $(\tilde{\rho}_1, \dots, \tilde{\rho}_m)$ in $[0, 1]$, if there exists $1 \leq k \leq m$ such that $\rho_k \geq \tilde{\rho}_k$ and $\rho_j = \tilde{\rho}_j$ for $j \neq k$, then (A1) holds for any $r \geq 0$. By symmetry, we consider only the case with $k = 1$. Let $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ independent of $\boldsymbol{\varepsilon}$, and $\boldsymbol{\varepsilon}_{-1} = (\varepsilon_2, \dots, \varepsilon_m)$ and $\boldsymbol{\eta}_{-1} = (\eta_2, \dots, \eta_m)$ be the subvectors of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\eta}$ excluding the first coordinates. Define

$$\mathcal{B}(r, \boldsymbol{\varepsilon}_{-1}, \boldsymbol{\eta}_{-1}, \rho_2, \dots, \rho_m) = \left\{ x : \begin{pmatrix} x \\ \sqrt{1 - \rho_2^2} \cdot \varepsilon_2 + \rho_2 \eta_2 \\ \vdots \\ \sqrt{1 - \rho_m^2} \cdot \varepsilon_m + \rho_m \eta_m \end{pmatrix} \in \mathcal{B}_m(r) \right\} \subset \mathbb{R},$$

which is either an empty set or a symmetric closed interval on the real line.

For any $r \geq 0$, $\boldsymbol{\varepsilon}_{-1}$ and $\boldsymbol{\eta}_{-1}$,

$$\begin{aligned} & P \left\{ (\mathbf{I}_F - \mathbf{\Delta}^2)^{1/2} \boldsymbol{\varepsilon} + \mathbf{\Delta} \boldsymbol{\eta} \in \mathcal{B}_m(r) \mid \boldsymbol{\eta}' \boldsymbol{\eta} \leq a, \boldsymbol{\varepsilon}_{-1}, \boldsymbol{\eta}_{-1} \right\} \\ &= P \left\{ \sqrt{1 - \rho_1^2} \cdot \varepsilon_1 + \rho_1 \cdot \eta_1 \in \mathcal{B}(r, \boldsymbol{\varepsilon}_{-1}, \boldsymbol{\eta}_{-1}, \rho_2, \dots, \rho_m) \mid \eta_1^2 \leq a - \boldsymbol{\eta}'_{-1} \boldsymbol{\eta}_{-1}, \boldsymbol{\varepsilon}_{-1}, \boldsymbol{\eta}_{-1} \right\} \\ &\geq P \left\{ \sqrt{1 - \tilde{\rho}_1^2} \cdot \varepsilon_1 + \tilde{\rho}_1 \cdot \eta_1 \in \mathcal{B}(r, \boldsymbol{\varepsilon}_{-1}, \boldsymbol{\eta}_{-1}, \tilde{\rho}_2, \dots, \tilde{\rho}_m) \mid \eta_1^2 \leq a - \boldsymbol{\eta}'_{-1} \boldsymbol{\eta}_{-1}, \boldsymbol{\varepsilon}_{-1}, \boldsymbol{\eta}_{-1} \right\} \\ (A2) \quad &= P \left\{ (\mathbf{I}_F - \tilde{\mathbf{\Delta}}^2)^{1/2} \boldsymbol{\varepsilon} + \tilde{\mathbf{\Delta}} \boldsymbol{\eta} \in \mathcal{B}_m(r) \mid \boldsymbol{\eta}' \boldsymbol{\eta} \leq a, \boldsymbol{\varepsilon}_{-1}, \boldsymbol{\eta}_{-1} \right\}, \end{aligned}$$

where the second last inequality follows from Lemma A14. Taking expectations of both sides of (A2), we obtain (A1). $\square$

LEMMA A16. Let $\mathbf{B}_0$ and $\mathbf{B}_1$ be two $m \times m$ positive semi-definite matrix, $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ be a standard Gaussian random vector, and $\zeta_{p,a} \sim \mathbf{D} \mid \mathbf{D}'\mathbf{D} \leq a$ be a truncated Gaussian random vector, where $p \geq m$ and $a \geq 0$. Define $\phi \sim (\mathbf{B}_0)^{1/2} \varepsilon + (\mathbf{B}_1)^{1/2}_p \zeta_{p,a}$. If $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$ is positive definite, then for any $\alpha \in (0, 1)$, the threshold $c_{1-\alpha}$ for $1 - \alpha$ quantile region $\{\mu : \mu' \mathbf{B}^{-1} \mu \leq c_{1-\alpha}\}$ of ϕ depends only on (m, p, a) and the eigenvalues of $\mathbf{B}^{-1/2} \mathbf{B}_1 \mathbf{B}^{-1/2}$, and is nonincreasing in these eigenvalues.

PROOF OF LEMMA A16. Let $\mathbf{B}^{-1/2} \mathbf{B}_1 \mathbf{B}^{-1/2} = \mathbf{\Gamma} \mathbf{\Delta}^2 \mathbf{\Gamma}'$ be the eigen-decomposition of $\mathbf{B}^{-1/2} \mathbf{B}_1 \mathbf{B}^{-1/2}$, where $\mathbf{\Gamma} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix, and $\mathbf{\Delta}^2 = \text{diag}(\rho_1^2, \dots, \rho_m^2)$ is a diagonal matrix. From Lemma A9,

$$\begin{aligned} \mathbf{\Gamma}' \mathbf{B}^{-1/2} \phi &\sim \mathbf{\Gamma}' \mathbf{B}^{-1/2} (\mathbf{B}_0)^{1/2} \varepsilon + \mathbf{\Gamma}' \mathbf{B}^{-1/2} (\mathbf{B}_1)^{1/2}_p \zeta_{p,a} \\ &\sim \left(\mathbf{\Gamma}' \mathbf{B}^{-1/2} \mathbf{B}_0 \mathbf{B}^{-1/2} \mathbf{\Gamma} \right)^{1/2} \varepsilon + \left(\mathbf{\Gamma}' \mathbf{B}^{-1/2} \mathbf{B}_1 \mathbf{B}^{-1/2} \mathbf{\Gamma} \right)^{1/2}_p \zeta_{p,a} \\ &\sim \left\{ \mathbf{\Gamma}' (\mathbf{I}_m - \mathbf{\Gamma} \mathbf{\Delta}^2 \mathbf{\Gamma}') \mathbf{\Gamma} \right\}^{1/2} \varepsilon + (\mathbf{\Gamma}' \mathbf{\Gamma} \mathbf{\Delta}^2 \mathbf{\Gamma}')^{1/2}_p \zeta_{p,a} \\ &\sim (\mathbf{I}_m - \mathbf{\Delta}^2)^{1/2} \varepsilon + (\mathbf{\Delta}^2)^{1/2}_p \zeta_{p,a} \\ &\sim (\mathbf{I}_m - \mathbf{\Delta}^2)^{1/2} \varepsilon + (\mathbf{\Delta}, \mathbf{0}_{m \times (p-m)}) \zeta_{p,a} \end{aligned}$$

Let $\eta \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$, $\xi \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p-m})$, and (η, ξ) be independent. From the definition of $\zeta_{p,a}$, for any $r^2 \geq 0$,

$$\begin{aligned} &P(\phi' \mathbf{B}^{-1} \phi \leq r^2) \\ &= P\left\{ \mathbf{\Gamma}' \mathbf{B}^{-1/2} \phi \in \mathcal{B}_m(r) \right\} \\ &= P\left\{ (\mathbf{I}_m - \mathbf{\Delta}^2)^{1/2} \varepsilon + (\mathbf{\Delta}, \mathbf{0}_{m \times (p-m)}) \zeta_{p,a} \in \mathcal{B}_m(r) \right\} \\ &= P\left\{ (\mathbf{I}_F - \mathbf{\Delta}^2)^{1/2} \varepsilon + (\mathbf{\Delta}, \mathbf{0}) \begin{pmatrix} \eta \\ \xi \end{pmatrix} \in \mathcal{B}_m(r) \mid \eta' \eta + \xi' \xi \leq a \right\} \\ &= P\left\{ (\mathbf{I}_m - \mathbf{\Delta}^2)^{1/2} \varepsilon + \mathbf{\Delta} \eta \in \mathcal{B}_m(r) \mid \eta' \eta + \xi' \xi \leq a \right\} \\ &\text{(A3)} \\ &= \mathbb{E} \left[P\left\{ (\mathbf{I}_m - \mathbf{\Delta}^2)^{1/2} \varepsilon + \mathbf{\Delta} \eta \in \mathcal{B}_m(r) \mid \eta' \eta \leq a - \xi' \xi, \xi \right\} \mid \eta' \eta + \xi' \xi \leq a \right]. \end{aligned}$$

From Lemma A15, for any given ξ , the conditional probability,

$$P\left\{ (\mathbf{I}_m - \mathbf{\Delta}^2)^{1/2} \varepsilon + \mathbf{\Delta} \eta \in \mathcal{B}_m(r) \mid \eta' \eta \leq a - \xi' \xi, \xi \right\}$$

is nondecreasing in the diagonal elements of $\mathbf{\Delta}^2$. Therefore, the quantity in (A3) depends only on (m, p, a) and the eigenvalues of $\mathbf{B}^{-1/2} \mathbf{B}_1 \mathbf{B}^{-1/2}$, and is nondecreasing in these eigenvalues. Therefore, Lemma A16 holds. $\square$

PROOF OF THEOREM 5. It follows directly from Lemma A16. $\square$

COMMENT ON LINEAR TRANSFORMATIONS OF $\hat{\tau}$ UNDER ReFM. From Theorem 4 and Lemma A11, for any $C \in \mathbb{R}^{p \times F}$ with $p \leq F$, the asymptotic sampling distribution of $C\hat{\tau}$ under ReFM is more peaked than that under the CRFE. From Theorem 2 and Lemma A9,

$$\begin{aligned} C(\hat{\tau} - \tau) \mid \mathcal{M} &\sim C \left(V_{\tau\tau}^\perp \right)^{1/2} \varepsilon + C \left(V_{\tau\tau}^\parallel \right)_{LF}^{1/2} \zeta_{LF,a} \\ &\sim \left(CV_{\tau\tau}^\perp C' \right)^{1/2} \xi + \left(CV_{\tau\tau}^\parallel C' \right)_{LF}^{1/2} \zeta_{LF,a}, \end{aligned}$$

where $\xi \sim \mathcal{N}(\mathbf{0}, I_p)$. From Lemma A16, if $CV_{\tau\tau}C'$ is invertible, then for any $\alpha \in (0, 1)$, the threshold $c_{1-\alpha}$ for $1-\alpha$ quantile region $\{\mu : \mu'(CV_{\tau\tau}C')^{-1}\mu \leq c_{1-\alpha}\}$ of the asymptotic sampling distribution of $C\hat{\tau}$ depends only on (p, LF, a) and the canonical correlation between $C\hat{\tau}$ and $\hat{\tau}_x$, and is nonincreasing in these canonical correlations. $\square$

PROOF OF COROLLARY 4. Because $\hat{\tau}_f$ is a one dimensional linear transformation of $\hat{\tau}$, Corollary 4 follows directly from the above comment on general lower dimensional linear transformations of $\hat{\tau}$ under ReFM. $\square$

A5.4. *Peakedness under* ReFMT_F.

PROOF OF THEOREM 9. For each $1 \leq h \leq H$, let $D_h \sim \mathcal{N}(\mathbf{0}, I_{LF_h})$, and $(\varepsilon, D_1, \dots, D_H)$ be jointly independent. From Proposition 2, both ε and ζ_{LF_h, a_h} are central convex unimodal. From Lemma A6, both $(V_{\tau\tau}^\perp)^{1/2}\varepsilon$ and $(W_{\tau\tau}^\parallel[h])_{LF_h}^{1/2}\zeta_{LF_h, a_h}$ are central convex unimodal. Thus, by Lemma A7,

$$\left(V_{\tau\tau}^\perp \right)^{1/2} \varepsilon + \sum_{h=2}^H (W_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h}$$

is central convex unimodal. From Lemma A13, $\zeta_{LF_1, a_1} \succ D_h$, which, based on Lemma A11, further implies that $(W_{\tau\tau}^\parallel[1])_{LF_1}^{1/2} \zeta_{LF_1, a_1} \succ (W_{\tau\tau}^\parallel[1])_{LF_1}^{1/2} D_1$. Thus, from Lemma A12,

$$\begin{aligned} &\left(V_{\tau\tau}^\perp \right)^{1/2} \varepsilon + \sum_{h=1}^H (W_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h} \\ &\succ \left(V_{\tau\tau}^\perp \right)^{1/2} \varepsilon + (W_{\tau\tau}^\parallel[1])_{LF_1}^{1/2} D_1 + \sum_{h=2}^H (W_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h} \end{aligned}$$

$$\sim \left(\mathbf{V}_{\tau\tau}^\perp + \mathbf{W}_{\tau\tau}^\parallel[1] \right)^{1/2} \varepsilon + \sum_{h=2}^H (\mathbf{W}_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h}.$$

Because $(\mathbf{V}_{\tau\tau}^\perp + \mathbf{W}_{\tau\tau}^\parallel[1])^{1/2} \varepsilon + \sum_{h=3}^H (\mathbf{W}_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h}$ is central convex unimodal, and $(\mathbf{W}_{\tau\tau}^\parallel[2])_{LF_2}^{1/2} \zeta_{LF_2, a_2} \succ (\mathbf{W}_{\tau\tau}^\parallel[2])_{LF_2}^{1/2} \mathbf{D}_2$, we have

$$\begin{aligned} & \left(\mathbf{V}_{\tau\tau}^\perp \right)^{1/2} \varepsilon + \sum_{h=1}^H (\mathbf{W}_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h} \\ & \succ \left(\mathbf{V}_{\tau\tau}^\perp + \mathbf{W}_{\tau\tau}^\parallel[1] \right)^{1/2} \varepsilon + \sum_{h=2}^H (\mathbf{W}_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h} \\ & \succ \left(\mathbf{V}_{\tau\tau}^\perp + \mathbf{W}_{\tau\tau}^\parallel[1] \right)^{1/2} \varepsilon + (\mathbf{W}_{\tau\tau}^\parallel[2])_{LF_2}^{1/2} \mathbf{D}_2 + \sum_{h=3}^H (\mathbf{W}_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h} \\ & \sim \left(\mathbf{V}_{\tau\tau}^\perp + \mathbf{W}_{\tau\tau}^\parallel[1] + \mathbf{W}_{\tau\tau}^\parallel[2] \right)^{1/2} \varepsilon + \sum_{h=3}^H (\mathbf{W}_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h}. \end{aligned}$$

Implementing the above procedure iteratively, we finally have

$$\left(\mathbf{V}_{\tau\tau}^\perp \right)^{1/2} \varepsilon + \sum_{h=1}^H (\mathbf{W}_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h} \succ \left(\mathbf{V}_{\tau\tau}^\perp + \sum_{h=1}^H \mathbf{W}_{\tau\tau}^\parallel[h] \right)^{1/2} \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_{\tau\tau}),$$

where the last formula follows from $\mathbf{V}_{\tau\tau}^\parallel = \sum_{h=1}^H \mathbf{W}_{\tau\tau}^\parallel[h]$ in the proof of Theorem 8. $\square$

To prove Theorem 10, we need the following three lemmas.

LEMMA A17. *Let $\varepsilon_0 \sim \mathcal{N}(0, 1)$, $\eta_{k_t, a_t} \sim D_{t1} \mid \mathbf{D}_t' \mathbf{D}_t \leq a_t$, where $\mathbf{D}_t = (D_{t1}, \dots, D_{tk_t}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k_t})$, a_t is a nonnegative constant that can be infinity, and $(\varepsilon_0, \eta_{k_1, a_1}, \eta_{k_2, a_2}, \dots, \eta_{k_T, a_T})$ are jointly independent. Let $\{\rho_t\}_{t=1}^{T+1}$ and $\{\tilde{\rho}_t\}_{t=1}^{T+1}$ be two nonnegative constant sequences satisfying $\sum_{t=1}^{T+1} \rho_t^2 = \sum_{t=1}^{T+1} \tilde{\rho}_t^2 = 1$. If $\rho_t \geq \tilde{\rho}_t$ for all $1 \leq t \leq T$, then for any $c \geq 0$,*

$$P \left(\left| \rho_{T+1} \varepsilon_0 + \sum_{t=1}^T \rho_t \eta_{k_t, a_t} \right| \leq c \right) \geq P \left(\left| \tilde{\rho}_{T+1} \varepsilon_0 + \sum_{t=1}^T \tilde{\rho}_t \eta_{k_t, a_t} \right| \leq c \right).$$

PROOF OF LEMMA A17. It follows from Li et al. [2018, Lemma A10]. $\square$

The following lemma extends the above lemma to the multivariate case. We introduce the following condition for a set of matrices with the same number of rows.

CONDITION A3. *The H matrices $\{\Delta_h \in \mathbb{R}^{m \times p_h}\}_{h=1}^H$ satisfy that*

- (a) *there is at most one nonzero element at each column and each row of Δ_h for $1 \leq h \leq H$;*
- (b) *the elements of Δ_h are all nonnegative for $1 \leq h \leq H$;*
- (c) *$\sum_{h=1}^H \Delta_h \Delta_h' \in \mathbb{R}^{m \times m}$ has all elements less than or equal to 1 (note that (a) implies that $\sum_{h=1}^H \Delta_h \Delta_h' \in \mathbb{R}^{m \times m}$ is a diagonal matrix).*

LEMMA A18. *Let $(\varepsilon, \zeta_{p_1, a_1}, \dots, \zeta_{p_H, a_H})$ be $H+1$ independent random vectors, where $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$, $\zeta_{p_h, a_h} \sim \mathbf{D}_h' \mathbf{D}_h \leq a_h$ is a truncated Gaussian random vector with $a_h \geq 0$ and $\mathbf{D}_h \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p_h})$. Assume $\{\Delta_h \in \mathbb{R}^{m \times p_h}\}_{h=1}^H$ satisfy Condition A3. For any $r \geq 0$, $1 \leq h \leq H$, and $1 \leq k \leq m$ the probability*

$$P \left\{ \left(\mathbf{I}_m - \sum_{h=1}^H \Delta_h \Delta_h' \right)^{1/2} \varepsilon + \sum_{h=1}^H \Delta_h \zeta_{p_h, a_h} \in \mathcal{B}_m(r) \right\}$$

is nondecreasing in the nonzero elements of the Δ_h 's, that is, for any two sets of matrices Δ_h 's and $\tilde{\Delta}_h$'s satisfying Condition A3 with the positions of possible nonzero elements being the same, if all elements of Δ_h 's are larger than or equal to $\tilde{\Delta}_h$'s, then

$$\begin{aligned} & P \left\{ \left(\mathbf{I}_m - \sum_{h=1}^H \Delta_h \Delta_h' \right)^{1/2} \varepsilon + \sum_{h=1}^H \Delta_h \zeta_{p_h, a_h} \in \mathcal{B}_m(r) \right\} \\ (A4) \quad & \geq P \left\{ \left(\mathbf{I}_m - \sum_{h=1}^H \tilde{\Delta}_h \tilde{\Delta}_h' \right)^{1/2} \varepsilon + \sum_{h=1}^H \tilde{\Delta}_h \zeta_{p_h, a_h} \in \mathcal{B}_m(r) \right\}. \end{aligned}$$

PROOF OF LEMMA A18. It suffices to prove that for any two sets of matrices Δ_h 's and $\tilde{\Delta}_h$'s satisfying Condition A3 with the positions of possible nonzero elements being the same, if there exists $1 \leq k \leq m$ such that for all $1 \leq h \leq H$, (a) Δ_h and $\tilde{\Delta}_h$ differ only in the k th row, and (b) the elements in the k th row of Δ_h are larger than or equal to that of $\tilde{\Delta}_h$, then (A4) holds for any $r \geq 0$. First, by symmetry, we consider only the case with $k = 1$. Second, without loss of generality, we assume that the possible nonzero elements in the first rows of Δ_h 's and $\tilde{\Delta}_h$'s are all in the first columns. This

is because permuting the columns of Δ_h will not change the distribution of $\Delta_h \zeta_{p_h, a_h}$, a fact implied by the spherical symmetry of ζ_{p_h, a_h} .

Let Δ_{h1} be the $(1, 1)$ th element of Δ_h , and $\Delta_{h,-1}$ be the submatrix of Δ_h excluding the first column and the first row. Define similarly $\tilde{\Delta}_{h1}$ and $\tilde{\Delta}_{h,-1}$. Let $\varepsilon_{-1} = (\varepsilon_2, \dots, \varepsilon_m)$ and $\mathbf{D}_{h,-1} = (D_{h2}, \dots, D_{hp_h})$ be the subvectors of ε and $\mathbf{D}_h$ excluding the first elements. We define a subset in $\mathbb{R}$ as follows:

$$\begin{aligned} & \mathcal{B}(r, \varepsilon_{-1}, \mathbf{D}_{h,-1}, \Delta_{h,-1}) \\ &= \left\{ x : \left(\left(\mathbf{I}_{m-1} - \sum_{h=1}^H \Delta_{h,-1} \Delta'_{h,-1} \right)^{1/2} \varepsilon_{-1} + \sum_{h=1}^H \Delta_{h,-1} \mathbf{D}_{h,-1} \right) \in \mathcal{B}_m(r) \right\}, \end{aligned}$$

which depends on ε_{-1} , $\mathbf{D}_{h,-1}$ and $\Delta_{h,-1}$ for all $1 \leq h \leq H$, and is either an empty set or a symmetric closed interval on the real line. For any $r \geq 0$ and $(\varepsilon_{-1}, \mathbf{D}_{1,-1}, \dots, \mathbf{D}_{H,-1})$,

$$\begin{aligned} & P \left\{ \left(\mathbf{I}_m - \sum_{h=1}^H \Delta_h \Delta'_h \right)^{1/2} \varepsilon + \sum_{h=1}^H \Delta_h \mathbf{D}_h \in \mathcal{B}_m(r) \mid \mathbf{D}'_h \mathbf{D}_h \leq a_h, \varepsilon_{-1}, \mathbf{D}_{h,-1}, \forall h \right\} \\ &= P \left\{ \left(1 - \sum_{h=1}^H \Delta_{h1}^2 \right)^{1/2} \varepsilon_1 + \sum_{h=1}^H \Delta_{h1} D_{h1} \in \mathcal{B}(r, \varepsilon_{-1}, \mathbf{D}_{h,-1}, \Delta_{h,-1}) \mid D_{h1}^2 \leq a_h - \mathbf{D}'_{h,-1} \mathbf{D}_{h,-1}, \varepsilon_{-1}, \mathbf{D}_{h,-1} \right\} \\ &\geq P \left\{ \left(1 - \sum_{h=1}^H \tilde{\Delta}_{h1}^2 \right)^{1/2} \varepsilon_1 + \sum_{h=1}^H \tilde{\Delta}_{h1} D_{h1} \in \mathcal{B}(r, \varepsilon_{-1}, \mathbf{D}_{h,-1}, \tilde{\Delta}_{h,-1}) \mid D_{h1}^2 \leq a_h - \mathbf{D}'_{h,-1} \mathbf{D}_{h,-1}, \varepsilon_{-1}, \mathbf{D}_{h,-1} \right\} \\ & \quad (\text{A5}) \\ &= P \left\{ \left(\mathbf{I}_m - \sum_{h=1}^H \tilde{\Delta}_h \tilde{\Delta}'_h \right)^{1/2} \varepsilon + \sum_{h=1}^H \tilde{\Delta}_h \mathbf{D}_h \in \mathcal{B}_m(r) \mid \mathbf{D}'_h \mathbf{D}_h \leq a_h, \varepsilon_{-1}, \mathbf{D}_{h,-1}, \forall h \right\}, \end{aligned}$$

where the second last inequality follows from Lemma A17. Taking expectations of both sides of (A5), we obtain (A4). $\square$

The following lemma extends Lemma A16 to general case with $H \geq 1$. Moreover, even when $H = 1$, Lemma A19 is still more general than Lemma A16 by only requiring $p_1 \geq \text{rank}(\mathbf{B}_1)$, instead of $p_1 \geq m$ in Lemma A16.

LEMMA A19. *Let $(\mathbf{B}_0, \mathbf{B}_1, \dots, \mathbf{B}_H)$ be $H+1$ positive semi-definite matrices in $\mathbb{R}^{m \times m}$ with ranks $(\gamma_0, \gamma_1, \dots, \gamma_H)$, $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ be a standard Gaussian random vector, and $\zeta_{p_h, a_h} \sim \mathbf{D}_h \mid \mathbf{D}'_h \mathbf{D}_h \leq a_h$ be a truncated Gaussian random vector, where $\mathbf{D}_h \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{p_h})$ and $p_h \geq \gamma_h$ ($h = 1, 2, \dots, H$). Define $\phi \sim (\mathbf{B}_0)^{1/2} \varepsilon + \sum_{h=1}^H (\mathbf{B}_h)^{1/2}_{p_h} \zeta_{p_h, a_h}$. If $\mathbf{B} = \sum_{h=0}^H \mathbf{B}_h$ is invertible, and there exists an orthogonal matrix $\mathbf{\Gamma}$ such that for all $1 \leq h \leq H$,*

$\Gamma' \mathbf{B}^{-1/2} \mathbf{B}_h \mathbf{B}^{-1/2} \Gamma = \mathbf{\Omega}_h^2$ is a diagonal matrix, then the threshold $c_{1-\alpha}$ for $1 - \alpha$ quantile region $\{\boldsymbol{\mu} : \boldsymbol{\mu}' \mathbf{B}^{-1} \boldsymbol{\mu} \leq c_{1-\alpha}\}$ of ϕ depends only on m , the p_h 's, the a_h 's, and the eigenvalues of $\mathbf{B}^{-1/2} \mathbf{B}_h \mathbf{B}^{-1/2}$, and is nonincreasing in these eigenvalues.

PROOF OF LEMMA A19. By definition,

$$\Gamma' \mathbf{B}^{-1/2} \mathbf{B}_0 \mathbf{B}^{-1/2} \Gamma = \Gamma' \left(\mathbf{I}_m - \sum_{h=1}^H \mathbf{B}^{-1/2} \mathbf{B}_h \mathbf{B}^{-1/2} \right) \Gamma = \mathbf{I}_m - \sum_{h=1}^H \mathbf{\Omega}_h^2.$$

Thus, from Lemma A9, the distribution of ϕ satisfies

$$\begin{aligned} \Gamma' \mathbf{B}^{-1/2} \phi &\sim \Gamma' \mathbf{B}^{-1/2} \left\{ (\mathbf{B}_0)^{1/2} \varepsilon + \sum_{h=1}^H (\mathbf{B}_h)^{1/2} \zeta_{p_h, a_h} \right\} \\ &\sim \left(\Gamma' \mathbf{B}^{-1/2} \mathbf{B}_0 \mathbf{B}^{-1/2} \Gamma \right)^{1/2} \varepsilon + \sum_{h=1}^H \left(\Gamma' \mathbf{B}^{-1/2} \mathbf{B}_h \mathbf{B}^{-1/2} \Gamma \right)^{1/2}_{p_h} \zeta_{p_h, a_h} \\ &\sim \left(\mathbf{I}_m - \sum_{h=1}^H \mathbf{\Omega}_h^2 \right)^{1/2} \varepsilon + \sum_{h=1}^H (\mathbf{\Omega}_h^2)^{1/2}_{p_h} \zeta_{p_h, a_h}. \end{aligned}$$

For each $1 \leq h \leq H$, if $p_h \geq m$, we further define $\tilde{\mathbf{\Omega}}_h = (\mathbf{\Omega}_h, \mathbf{0}_{m \times (p_h - m)})$; otherwise, $p_h < m$, $\mathbf{B}_h$ has rank at most p_h and thus $\mathbf{B}^{-1/2} \mathbf{B}_h \mathbf{B}^{-1/2}$ has at most p_h nonzero eigenvalues, we further define $\tilde{\mathbf{\Omega}}_h = \mathbf{\Omega}_h[\cdot, \mathcal{I}_h]$ as the submatrix of $\mathbf{\Omega}_h$ consisting of the $|\mathcal{I}_h| = p_h$ columns with possible nonzero eigenvalues. Thus, by the construction of $\tilde{\mathbf{\Omega}}_h$'s, we have $\tilde{\mathbf{\Omega}}_h \in \mathbb{R}^{m \times p_h}$ and $\tilde{\mathbf{\Omega}}_h \tilde{\mathbf{\Omega}}_h' = \mathbf{\Omega}_h^2$, i.e., $\tilde{\mathbf{\Omega}}_h = (\mathbf{\Omega}_h^2)^{1/2}_{p_h}$. Therefore, we can further simplify the distribution of $\Gamma' \mathbf{B}^{-1/2} \phi$ as

$$\Gamma' \mathbf{B}^{-1/2} \phi \sim \left(\mathbf{I}_m - \sum_{h=1}^H \tilde{\mathbf{\Omega}}_h \tilde{\mathbf{\Omega}}_h' \right)^{1/2} \varepsilon + \sum_{h=1}^H \tilde{\mathbf{\Omega}}_h \zeta_{p_h, a_h}.$$

For any $r \geq 0$, we have

$$\begin{aligned} P(\phi' \mathbf{B}^{-1} \phi \leq r^2) &= P \left\{ \Gamma' \mathbf{B}^{-1/2} \phi \in \mathcal{B}_m(r) \right\} \\ &= P \left\{ \left(\mathbf{I}_m - \sum_{h=1}^H \mathbf{\Omega}_h^2 \right)^{1/2} \varepsilon + \sum_{h=1}^H (\mathbf{\Omega}_h^2)^{1/2}_{p_h} \zeta_{p_h, a_h} \in \mathcal{B}_m(r) \right\} \\ (A6) \quad &= P \left\{ \left(\mathbf{I}_m - \sum_{h=1}^H \tilde{\mathbf{\Omega}}_h \tilde{\mathbf{\Omega}}_h' \right)^{1/2} \varepsilon + \sum_{h=1}^H \tilde{\mathbf{\Omega}}_h \zeta_{p_h, a_h} \in \mathcal{B}_m(r) \right\}. \end{aligned}$$

Because the matrices $\tilde{\Omega}_h$'s satisfy Condition A3, from Lemma A18, the quantity in (A6) depends only on m , p_h 's, a_h 's, and the eigenvalues of $\mathbf{B}^{-1/2}\mathbf{B}_h\mathbf{B}^{-1/2}$'s, and is nondecreasing in these eigenvalues. Therefore, Lemma A19 holds. $\square$

PROOF OF THEOREM 10. It follows directly from Lemma A19. $\square$

COMMENT ON LINEAR TRANSFORMATIONS OF $\hat{\tau}$ UNDER REFMT_F . From Theorem 9 and Lemma A11, for any $\mathbf{C} \in \mathbb{R}^{p \times F}$ with $p \leq F$, the asymptotic sampling distribution of $\mathbf{C}\hat{\tau}$ under REFMT_F is more peaked than that under the CRFE. From Theorem 7 and Lemma A9,

$$\begin{aligned} \mathbf{C}(\hat{\tau} - \tau) \mid \mathcal{T}_F &\sim \mathbf{C}(\mathbf{V}_{\tau\tau}^\perp)^{1/2} \varepsilon + \sum_{h=1}^H \mathbf{C}(\mathbf{W}_{\tau\tau}^\parallel[h])_{LF_h}^{1/2} \zeta_{LF_h, a_h} \\ &\sim (\mathbf{C}\mathbf{V}_{\tau\tau}^\perp \mathbf{C}')^{1/2} \boldsymbol{\xi} + \sum_{h=1}^H (\mathbf{C}\mathbf{W}_{\tau\tau}^\parallel[h] \mathbf{C}')_{LF_h}^{1/2} \zeta_{LF_h, a_h}, \end{aligned}$$

where $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. Based on Lemma A19, we can know that if the condition in Lemma A19 holds for $\mathbf{B}_h = \mathbf{C}\mathbf{W}_{\tau\tau}^\parallel[h] \mathbf{C}'$, then for any $\alpha \in (0, 1)$, the threshold $c_{1-\alpha}$ for $1 - \alpha$ quantile region $\{\boldsymbol{\mu} : \boldsymbol{\mu}'(\mathbf{C}\mathbf{V}_{\tau\tau}^\perp \mathbf{C}')^{-1} \boldsymbol{\mu} \leq c_{1-\alpha}\}$ of the asymptotic sampling distribution of $\mathbf{C}\hat{\tau}$ depends only on p , LF_h 's, a_h 's, and the canonical correlation between $\mathbf{C}\hat{\tau}$ and $\hat{\theta}_x[h]$'s, and is nonincreasing in these canonical correlations. $\square$

PROOF OF COROLLARY 6. Because $\hat{\tau}_f$ is a one dimensional linear transformation of $\hat{\tau}$, Corollary 6 follows directly from the above comment on general lower dimensional linear transformations of $\hat{\tau}$ under REFMT_F . $\square$

The proofs of Theorems A3, A4, and Corollary A2 under REFMT_{CF} , as well as the comment on lower dimensional linear transformations of $\hat{\tau}$, are almost the same as those under REFMT_F and thus omitted.

A6. Asymptotic conservativeness in inference.

A6.1. *Asymptotic conservativeness of sampling covariance estimators.* We need the following two lemmas to prove the asymptotic conservativeness of the sampling covariance estimators.

LEMMA A20. *Under either ReFM, ReFMT_F, or ReFMT_{CF}, if Condition 1 holds, then for any $1 \leq r, k \leq Q$, $1 \leq l, m \leq L$, and any (A_i, B_i) equal to $(Y_i(r), Y_i(k))$, $(Y_i(r), x_{il})$ or (x_{il}, x_{im}) ,*

$$s_{AB}(q) - S_{AB} = o_p(1), \quad (q = 1, 2, \dots, Q)$$

where $s_{AB}(q)$ is the sample covariance between the A_i 's and the B_i 's under treatment combination q , and S_{AB} is the corresponding finite population covariance.

PROOF OF LEMMA A20. The proof is similar to the proof of Lemma A15 in Li et al. [2018]. We omit it. $\square$

LEMMA A21. *Under either ReFM, ReFMT_F, or ReFMT_{CF}, if Condition 1 holds, then for each $1 \leq q \leq Q$,*

$$s_{qq}^\perp - S_{qq}^\perp = o_p(1), \quad s_{q,x} - S_{q,x} = o_p(1), \quad s_{xx}(q) - S_{xx} = o_p(1).$$

PROOF OF LEMMA A21. From Lemma A20, s_{qq} , $s_{q,x}$ and $s_{xx}(q)$ are consistent for S_{qq} , $S_{q,x}$ and S_{xx} , respectively. Because $s_{qq}^\perp = s_{qq} - s_{q,x} s_{xx}^{-1}(q) s_{x,q}$ and $S_{qq}^\perp = S_{qq} - S_{q,x} S_{xx}^{-1} S_{x,q}$, we know that $s_{qq}^\perp$ is also consistent for $S_{qq}^\perp$, and therefore Lemma A21 holds. $\square$

Define $\tilde{\mathbf{V}}_{\tau\tau}^\perp \equiv 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} \mathbf{b}_q \mathbf{b}_q' \cdot S_{qq}^\perp \geq \mathbf{V}_{\tau\tau}^\perp$. Under ReFM, from Lemma A21, $\hat{\mathbf{V}}_{\tau\tau}^\perp$ is consistent for $\tilde{\mathbf{V}}_{\tau\tau}^\perp$, and $\hat{\mathbf{V}}_{\tau\mathbf{x}} \mathbf{V}_{\mathbf{x}\mathbf{x}}^{-1/2}$ is consistent for $\mathbf{V}_{\tau\mathbf{x}} \mathbf{V}_{\mathbf{x}\mathbf{x}}^{-1/2}$. Therefore, the sampling covariance estimator is asymptotically conservative. Under ReFMT_F or ReFMT_{CF}, $\hat{\mathbf{V}}_{\tau\tau}^\perp$ is also consistent for $\tilde{\mathbf{V}}_{\tau\tau}^\perp$, and the estimated coefficients of the ζ_{LF_h, a_h} 's or ζ_{λ_j, a_j} 's are consistent for the true ones. Therefore, the sampling covariance estimators under ReFMT_F and ReFMT_{CF} are also asymptotically conservative.

A6.2. *Asymptotic conservativeness of the confidence sets.* We need the following lemma to prove Theorem 6,

LEMMA A22. *Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be two positive semi-definite matrices in $\mathbb{R}^{m \times m}$ satisfying that $\mathbf{V}_1 \leq \mathbf{V}_2$, and ε_1 and ε_2 be two Gaussian random vectors with mean zero and covariance matrices $\mathbf{V}_1$ and $\mathbf{V}_2$. Then $\varepsilon_1 \succ \varepsilon_2$.*

PROOF OF LEMMA A22. Let $\varepsilon_3 \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_2 - \mathbf{V}_1)$ be independent of ε_1 . From Proposition 2 and Lemma A6, ε_1 is central convex unimodal. Because $\mathbf{0} \succ \varepsilon_3$ and $\varepsilon_3 + \varepsilon_1 \sim \varepsilon_2$, from Lemma A12, $\varepsilon_1 \sim \mathbf{0} + \varepsilon_1 \succ \varepsilon_3 + \varepsilon_1 \sim \varepsilon_2$. Lemma A22 holds. $\square$

PROOF OF THEOREM 6. The proof of the asymptotic conservativeness of covariance estimator for $\mathbf{C}\hat{\boldsymbol{\tau}}$ follows directly from the discussion in A6.1, and thus we consider only the asymptotic conservativeness of confidence sets for $\mathbf{C}\boldsymbol{\tau}$ here. Let $\mathcal{L}$, $\tilde{\mathcal{L}}$ and $\hat{\mathcal{L}}$ be three F dimensional random vectors following the asymptotic sampling distribution of $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}$ under ReFM, the asymptotic sampling distribution with $\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp$ replaced by $\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp$, and the estimated asymptotic sampling distribution:

$$\begin{aligned}\mathcal{L} &\sim \left(\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp\right)^{1/2} \boldsymbol{\varepsilon} + \left(\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel\right)_{LF}^{1/2} \boldsymbol{\zeta}_{LF,a} \sim \left(\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp\right)^{1/2} \boldsymbol{\varepsilon} + \mathbf{V}_{\boldsymbol{\tau}\boldsymbol{x}} \mathbf{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1/2} \boldsymbol{\zeta}_{LF,a}, \\ \tilde{\mathcal{L}} &\sim \left(\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp\right)^{1/2} \boldsymbol{\varepsilon} + \left(\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\parallel\right)_{LF}^{1/2} \boldsymbol{\zeta}_{LF,a} \sim \left(\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp\right)^{1/2} \boldsymbol{\varepsilon} + \mathbf{V}_{\boldsymbol{\tau}\boldsymbol{x}} \mathbf{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1/2} \boldsymbol{\zeta}_{LF,a}, \\ \hat{\mathcal{L}} &\sim \left(\hat{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp\right)^{1/2} \boldsymbol{\varepsilon} + \hat{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{x}} \mathbf{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1/2} \boldsymbol{\zeta}_{LF,a}.\end{aligned}$$

Under Condition 1, from the discussion in Section A6.1, by Slutsky's theorem, $\hat{\mathcal{L}} \sim \tilde{\mathcal{L}}$. Note that $(\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp)^{1/2} \boldsymbol{\varepsilon} \succ (\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp)^{1/2} \boldsymbol{\varepsilon}$ from Lemma A22, and $\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{x}} \mathbf{V}_{\boldsymbol{x}\boldsymbol{x}}^{-1/2} \boldsymbol{\zeta}_{LF,a}$ is central convex unimodal from Proposition 2 and Lemma A6. From Lemma A12, $\mathcal{L} \succ \tilde{\mathcal{L}}$. Above all, $\mathcal{L} \succ \tilde{\mathcal{L}} \sim \hat{\mathcal{L}}$.

From Slutsky's theorem, $(\mathbf{C}\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp \mathbf{C}')^{-1/2} \mathbf{C}\tilde{\mathcal{L}} \sim (\mathbf{C}\hat{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp \mathbf{C}')^{-1/2} \mathbf{C}\hat{\mathcal{L}}$. From the continuous mapping theorem, $(\mathbf{C}\tilde{\mathcal{L}})'(\mathbf{C}\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp \mathbf{C}')^{-1} \mathbf{C}\tilde{\mathcal{L}} \sim (\mathbf{C}\hat{\mathcal{L}})'(\mathbf{C}\hat{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp \mathbf{C}')^{-1} \mathbf{C}\hat{\mathcal{L}}$. Thus, the $1 - \alpha$ quantile of $(\mathbf{C}\hat{\mathcal{L}})'(\mathbf{C}\hat{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp \mathbf{C}')^{-1} \mathbf{C}\hat{\mathcal{L}}$, $\hat{c}_{1-\alpha}$, is consistent for the $1 - \alpha$ quantile of $(\mathbf{C}\tilde{\mathcal{L}})'(\mathbf{C}\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp \mathbf{C}')^{-1} \mathbf{C}\tilde{\mathcal{L}}$, $\tilde{c}_{1-\alpha}$. Because $\mathcal{L} \succ \tilde{\mathcal{L}}$, and the set of form $\{\boldsymbol{\mu} : (\mathbf{C}\boldsymbol{\mu})'(\mathbf{C}\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp \mathbf{C}')^{-1} \mathbf{C}\boldsymbol{\mu} \leq c\}$ is symmetric convex, $c_{1-\alpha}$, the $1 - \alpha$ quantile of $(\mathbf{C}\mathcal{L})'(\mathbf{C}\tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp \mathbf{C}')^{-1} \mathbf{C}\mathcal{L}$, is smaller than or equal to $\tilde{c}_{1-\alpha}$. Above all, $\hat{c}_{1-\alpha}$ is consistent for $\tilde{c}_{1-\alpha} \geq c_{1-\alpha}$. Therefore, the $1 - \alpha$ confidence set for $\boldsymbol{\tau}$ is asymptotically conservative. When $\mathbf{S}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp = o(1)$, we have $\mathbf{V}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp - \tilde{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp = o(1)$, which implies $\mathcal{L} \sim \tilde{\mathcal{L}}$. Thus, $\tilde{c}_{1-\alpha} = c_{1-\alpha} + o(1)$, and the $1 - \alpha$ confidence set for $\boldsymbol{\tau}$ becomes asymptotically exact. $\square$

Note that both $\sum_{h=1}^H \mathbf{W}_{\boldsymbol{\tau}\boldsymbol{x}}[h](\mathbf{W}_{\boldsymbol{x}\boldsymbol{x}}[h])^{-1/2} \boldsymbol{\zeta}_{LF_h, a_h}$ and $\sum_{j=1}^J \mathbf{U}_{\boldsymbol{\tau}\boldsymbol{e}}[j](\mathbf{U}_{\boldsymbol{e}\boldsymbol{e}}[j])^{-1/2} \boldsymbol{\zeta}_{\lambda_j, a_j}$ are central convex unimodal. The proofs for the asymptotic conservativeness of symmetric convex confidence sets of form $\mathbf{C}\hat{\boldsymbol{\tau}} + \mathcal{O}(\mathbf{C}\hat{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp \mathbf{C}', c)$ for $\mathbf{C}\boldsymbol{\tau}$ under ReFMT_F and ReFMT_{CF} are almost the same as ReFM. Thus we omit the proofs of Theorems 11 and A5.

Moreover, under for ReFM, ReFMT_F, or ReFMT_{CF}, we consider $1 - \alpha$ confidence set $\mathbf{C}\hat{\boldsymbol{\tau}} + \tilde{\mathcal{O}}$ for $\mathbf{C}\boldsymbol{\tau}$, where $\tilde{\mathcal{O}}$ can depend on $(\hat{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{\tau}}^\perp, \hat{\mathbf{V}}_{\boldsymbol{\tau}\boldsymbol{x}}, \mathbf{V}_{\boldsymbol{x}\boldsymbol{x}})$ and satisfies that $P(\hat{\mathcal{L}} \in \tilde{\mathcal{O}}) = 1 - \alpha$. If $\tilde{\mathcal{O}}$ is a symmetric convex set, then the confidence set is generally asymptotically conservative, and the proof is similar to the proof of Theorem 6.

References.

- T. Dai. *On Multivariate Unimodal Distributions*. PhD thesis, University of British Columbia, 1989.
- S. Das Gupta, M. L. Eaton, I. Olkin, M. Perlman, L. J. Savage, and M. Sobel. Inequalities on the probability content of convex regions for elliptically contoured distributions. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, pages 241–265, Berkeley, California, 1972. University of California Press.
- S. Dharmadhikari and K. Joag-Dev. *Unimodality, Convexity, and Applications*. San Diego, CA: Academic Press, Inc., 1988.
- M. Kanter. Unimodality and dominance for symmetric random vectors. *Transactions of the American Mathematical Society*, 229:65–85, 1977.
- X. Li and P. Ding. General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112: 1759–1769, 2017.
- X. Li, P. Ding, and D. B. Rubin. Asymptotic theory of rerandomization in treatment-control experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 115:9157–9162, 2018.
- J. Lu. Covariate adjustment in randomization-based causal inference for factorial designs. *Statistics & Probability Letters*, 119:11–20, 2016.
- K. L. Morgan and D. B. Rubin. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40:1263–1282, 2012.
- K. L. Morgan and D. B. Rubin. Rerandomization to balance tiers of covariates. *Journal of the American Statistical Association*, 110:1412–1421, 2015.