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‡

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, Kempthorne 1952, Hinkelmann and Kempthorne 2007, Cox and Reid 2000) and engineering (see textbooks by Box et al. 2005, Wu and Hamada 2011). Recently, factorial experiments also become popular in social sciences [e.g., Angrist et al., 2009, Dasgupta et al., 2015, Branson et al., 2016]. The completely randomized factorial experiment (CRFE) balances covariates under different treatment combinations on average. However, with increasing numbers of pretreatment covariates and treatment factors, there will be an increasing chance to have unbalanced covariates with respect to multiple factorial effects. Many researchers have recognized this issue in different experimental designs [e.g., Fisher, 1926, Student, 1938, Hansen and Bowers, 2008, Bruhn and McKenzie, 2009]. To avoid this, we can force a treatment allocation to have covariate balance, which is sometimes called rerandomization [e.g., Cox, 1982, 2009, Morgan

Keywords and phrases: Covariate balance, Tiers of covariates, Tiers of factorial effects
and Rubin, 2012], restricted or constrained randomization [e.g., Yates, 1948, Grundy and Healy, 1950, Youden, 1972, Bailey, 1983].

Extending Morgan and Rubin [2012]'s proposal for treatment-control experiments, Branson et al. [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 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., Finney, 1943, Bose, 1947, Cochran and Cox, 1950, Cox and Reid, 2000, 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 et al., 2018a], 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 et al., 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 et al., 2018b] contains all technical details. We do not make any attempt to review the extensive literature of confounding and fractional replication in the context of factorial experiments. Instead, we focus on the repeated sampling properties of estimators under rerandomization in $2^K$ factorial experiments.

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 $\mathbf{u}(q) = (\nu_1(q), \nu_2(q), \ldots, \nu_K(q)) \in \{-1, +1\}^K$ be the levels of the $K$ factors. We use potential outcomes to define causal effects in factorial experiments [Neyman, 1923, Dasgupta et al., 2015, Branson et al., 2016]. For unit $i$, let $Y_i(q)$ be the potential outcome under treatment combination $q$, and $\mathbf{Y}_i = (Y_i(1), Y_i(2), \ldots, Y_i(Q))$ be the $Q$ dimensional row vector of all potential outcomes. Let $\bar{Y}(q) = \frac{1}{n} \sum_{i=1}^{n} Y_i(q)$ be the average potential outcome under treatment combination $q$, and $\bar{Y} = (\bar{Y}(1), \bar{Y}(2), \ldots, \bar{Y}(Q))$ be the $Q$ dimensional row vector of all average potential outcomes. Dasgupta et al. [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

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

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

where $\mathbf{g}_k = (g_{k1}, \ldots, g_{kQ})' = (\nu_k(1), \nu_k(2), \ldots, \nu_k(Q))'$ 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 element-wise 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 $\mathbf{g}_f = (g_{f1}, \ldots, g_{fQ})' \in \{-1, +1\}^Q$ be the generating vector for the $f$th factorial effect ($1 \leq f \leq F$). For unit $i$, $\tau_{if} =$
$2^{-(K-1)}Y_i^f$ is the $f$th individual factorial effect, and $\tau_i = (\tau_{i1}, \ldots, \tau_{iF})'$ is the $F$ dimension vector of all individual factorial effects. Let $\tau_f = 2^{-(K-1)}\bar{Y}^f$ be the $f$th average factorial effect, and $\tau = (\tau_1, \ldots, \tau_F)'$ be the $F$ dimensional column vector of all average factorial effects. The definitions of the factorial effects imply $\tau_i = 2^{-(K-1)}\sum_{Q=1}^Q b_q Y_i(q)$ and $\tau = 2^{-(K-1)}\sum_{q=1}^Q b_q \bar{Y}(q)$, with coefficient vectors

\[ b_1 = \begin{pmatrix} g_{11} \\ g_{21} \\ \vdots \\ g_{F1} \end{pmatrix}, \quad b_2 = \begin{pmatrix} g_{12} \\ g_{22} \\ \vdots \\ g_{F2} \end{pmatrix}, \quad \ldots \quad b_Q = \begin{pmatrix} g_{1Q} \\ g_{2Q} \\ \vdots \\ g_{FQ} \end{pmatrix}. \]

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 et al. [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 $g_f$'s and the $b_q$'s. Specifically, the first three columns $(g_1, g_2, g_3)$ represent the levels of three factors in all treatment combinations, and they generate the main effects of factors $(1, 2, 3)$. The remaining columns $(g_4, \ldots, g_7)$ are the element-wise multiplications of subsets of $(g_1, g_2, g_3)$ that generate the interaction effects. The coefficient vector $b_q$ consists of all the elements in the $q$th row of Table 1.

2.2. *Treatment assignment, covariate imbalance and rerandomization.* For each unit $i$, $x_i$ 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 GPA, 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 $Z = (Z_1, \ldots, Z_n)$ be the treatment assignment vector for all units. In the CRFE, the probability that $Z$ takes a particular value $z = (z_1, \ldots, z_n)$ is $n! \cdot n_q!/n!$, where $\sum_{i=1}^n 1\{z_i = q\} = n_q$ for all $q$. Let $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ be the finite population covariate mean vector; for
Table 1
gf’s and bq’s for 2^3 factorial experiments

| 1 | 2 | 3 | 12 | 13 | 23 | 123 |
|---|---|---|----|----|----|-----|
| −1 | −1 | −1 | +1 | +1 | −1 | b₁ |
| −1 | +1 | −1 | −1 | −1 | +1 | b₂ |
| +1 | −1 | +1 | −1 | +1 | −1 | b₃ |
| −1 | +1 | +1 | −1 | −1 | +1 | b₄ |
| +1 | −1 | −1 | +1 | +1 | −1 | b₅ |
| +1 | −1 | −1 | +1 | +1 | −1 | b₆ |
| +1 | +1 | +1 | −1 | −1 | +1 | b₇ |
| g₁ | g₂ | g₃ | g₄ | g₅ | g₆ | g₇ |

1 ≤ q ≤ Q, let \( \hat{x}(q) = n_q^{-1} \sum_{i: z_i = q} x_i \) be the covariate mean vector for units that receive treatment combination \( q \). For 1 ≤ f ≤ F, the \( L \) dimensional difference-in-means vector of covariates with respect to the \( f \)th factorial effect is

\[
(2.2) \quad \hat{\tau}_x,f = \frac{2}{Q} \sum_{q=1}^{Q} g_{fq} \hat{x}(q) = \frac{1}{2^{K-1}} \sum_{q: g_{fq} = 1} \hat{x}(q) - \frac{1}{2^{K-1}} \sum_{q: g_{fq} = -1} \hat{x}(q).
\]

Let \( \hat{\tau}_x = (\hat{\tau}_{x,1}', \ldots, \hat{\tau}_{x,F}')' \) be the \( LF \) dimensional column vector of the difference-in-means of covariates with respect to all factorial effects. Although \( \hat{\tau}_x \) has mean zero under the CRFE, for a realized value of \( Z \), 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 \( \hat{\tau}_x \) 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. The experimenter may observe some covariates, but s/he can not change their values. The treatments do not affect the covariates. 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] for a related discussion in the treatment-control experiment.

2.3. Notation. To facilitate the discussion, for a positive semi-definite matrix $A \in \mathbb{R}^{m \times m}$ with rank $p_0$, and a positive integer $p \geq p_0$, we use $A^1/2 \in \mathbb{R}^{m \times p}$ to denote a matrix such that $A_p^{1/2}(A_p^{1/2})' = A$. Specifically, if $A = \Gamma \Lambda^2 \Gamma'$ is the eigen-decomposition of $A$ where $\Gamma \in \mathbb{R}^{m \times p_0}$, $\Gamma' \Gamma = I_{p_0}$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{p_0})$, then we can choose $A_p^{1/2} = (\Gamma \Lambda, 0_{m \times (p - p_0)})$. The choice of $A^{1/2}$ is generally not unique. In the special case with $p = m$, we use $A^{1/2}$ to denote the unique positive-semidefinite matrix satisfying the definition of $A_{m}^{1/2}$. We use $\otimes$ for the Kronecker product of two matrices, and $\circ$ for element-wise multiplications of vectors. We say a matrix $M_1$ is smaller than or equal to $M_2$ and write as $M_1 \preceq M_2$, if $M_2 - M_1$ is positive semi-definite. We say a random vector $\phi$ (or its distribution) is symmetric, if $\phi \sim -\phi$ 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 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. 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, $S_{\tau\tau} = (n - 1)^{-1} \sum_{i=1}^{n} (\tau_i - \bar{\tau})(\tau_i - \bar{\tau})'$ for factorial effects, $S_{xx} = (n - 1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'$ for covariates, and $S_{q,x} = S'_{q,x} = (n - 1)^{-1} \sum_{i=1}^{n} \{Y_i(q) - \bar{Y}(q)\} (x_i - \bar{x})'$ for potential outcomes and covariates. The covariance $S_{xx}$ is known without any uncertainty. However, other variances or covariances (e.g., $S_{qk}, S_{\tau\tau}$ and $S_{q,x}$) involve potential outcomes or individual factorial effects and are thus
generally unknown.

3.1. Asymptotic sampling distribution under the CRFE. Let \( Y_{i, q}^{\text{obs}} = \sum_{q=1}^{Q} 1\{Z_i = q\} Y_i(q) \) be the observed outcome of unit \( i \), and \( \hat{Y}(q) = n_q^{-1} \sum_{i: Z_i = q} Y_{i, q}^{\text{obs}} \) 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 \( \hat{\tau} = (\hat{\tau}_1, \ldots, \hat{\tau}_F)' \) be the \( F \) dimensional column vector consisting of all factorial effect estimators.

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

**Condition 1.** As \( n \to \infty \), the sequence of finite populations satisfies that for each \( 1 \leq q \neq k \leq Q \),

(i) the proportion of units under treatment combination \( q \), \( n_q/n \), has a positive limit,

(ii) the finite population variance and covariances \( S_{qq}, S_{qk}, S_{xx} \) and \( S_{q, x} \) have limiting values, and \( S_{xx} \) and its limit are non-degenerate,

(iii) \( \max_{1 \leq i \leq n} |Y_i(q) - \hat{Y}(q)|^2/n \to 0 \) and \( \max_{1 \leq i \leq n} \|x_i - \bar{x}\|^2/n \to 0 \).

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

\[
V \equiv 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} \left( \begin{array}{cc} b_q b_q' S_{qq} & (b_q b_q') \otimes S_{q, x} \\ (b_q b_q') \otimes S_{x, q} & (b_q b_q') \otimes S_{x, x} \end{array} \right) - n^{-1} \left( \begin{array}{cc} S_{\tau \tau} & 0 \\ 0 & 0 \end{array} \right)
\]

\[
\equiv \begin{pmatrix} V_{\tau \tau} & V_{\tau x} \\ V_{x \tau} & V_{xx} \end{pmatrix}.
\]

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

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 et al., 2018b]. Proposition 1 immediately gives the sampling properties of any single factorial effect estimator. Let \( S_{\tau \tau_f} \) be the
\( f \)th diagonal element of \( S_{\tau \tau} \), and 

\[ V_{\tau f} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} S_{q q} - n^{-1} S_{\tau f \tau f} \]

be the \( f \)th diagonal element of \( V_{\tau \tau} \). Then \( \hat{\tau}_f \) is unbiased for \( \tau_f \) with sampling variance \( V_{\tau f} \), and \( \hat{\tau}_f - \tau_f \sim N(0, V_{\tau f}) \). Moreover, \( S_{\tau \tau} \) cannot be unbiasedly estimated from the observed data, and it equals 0 under the additivity defined below. Under the additivity, the individual treatment effect does not depend on covariates, i.e., there is no treatment-covariate interaction.

**Definition 1.** The factorial effects are additive if and only if the individual factorial effect \( \tau_i \) is a constant vector for all units, or, equivalently, \( S_{\tau \tau} = 0 \).

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

### 3.2. Linear projections.

First, we decompose the potential outcomes. Let 

\( Y_i^{\parallel}(q) = \bar{Y}(q) + S_q x_i S_{xx}^{-1}(x_i - \bar{x}) \) be the finite population linear projection of the \( Y_i(q) \)’s on the \( x_i \)’s, and \( Y_i^\perp(q) = Y_i(q) - Y_i^{\parallel}(q) \) be the corresponding residual. The finite population linear projection of \( \tau_i \) on \( x_i \) is then 

\[ \tau_i^{\parallel} = 2^{-(K-1)} \sum_{q=1}^{Q} b_q Y_i^{\parallel}(q), \]

and the corresponding residual is 

\[ \tau_i^\perp = 2^{-(K-1)} \sum_{q=1}^{Q} b_q Y_i^\perp(q). \]

Let \( S_{qq}^{\parallel}, S_{qq}^{\perp}, S_{\tau \tau}^{\parallel}, \) and \( S_{\tau \tau}^{\perp} \) be the finite population variances and covariances of \( Y_i^{\parallel}(q), Y_i^\perp(q), \tau_i^{\parallel}, \) and \( \tau_i^\perp \), respectively. Define 

\[ V_{\tau \tau}^{\parallel} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} b_q b_q' \cdot S_{qq}^{\parallel} - n^{-1} S_{\tau f \tau f}^{\parallel}, \]

\[ V_{\tau \tau}^{\perp} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} b_q b_q' \cdot S_{qq}^{\perp} - n^{-1} S_{\tau f \tau f}^{\perp} \]

as analogues of the sampling covariance \( V_{\tau \tau} \) 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 

\[ V_{\tau \tau} = V_{\tau \tau}^{\parallel} + V_{\tau \tau}^{\perp}. \]

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 \( V_{\tau x} V_{xx}^{-1} \hat{\tau}_x \), the corresponding residual is \( \hat{\tau} - \tau - V_{\tau x} V_{xx}^{-1} \hat{\tau}_x \), and they have sampling covariances:

\[
\text{Cov} \left( V_{\tau x} V_{xx}^{-1} \hat{\tau}_x \right) = V_{\tau \tau}, \quad \text{Cov} \left( \hat{\tau} - \tau - V_{\tau x} V_{xx}^{-1} \hat{\tau}_x \right) = V_{\tau \tau},
\]

\[
\text{Cov} \left( V_{\tau x} V_{xx}^{-1} \hat{\tau}_x, \hat{\tau} - \tau - V_{\tau x} V_{xx}^{-1} \hat{\tau}_x \right) = 0.
\]

Theorem 1 follows from Proposition 1 and some matrix calculations, with the proof in Appendix A2 of the Supplementary Material [Li et al., 2018b].

Let \( V_{\tau f} \) and \( S_{\tau f} \) be the \( f \)th diagonal elements of \( V_{\tau \tau} \) and \( S_{\tau \tau} \), 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:

\[
R^2_f = \text{Corr}^2(\hat{\tau}_f, \hat{\tau}_x) = \frac{V_{\tau f}^\tau V_{\tau f}}{V_{\tau f}^\tau V_{\tau f}} = \frac{2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} S_{qq}^\tau - n^{-1} S_{\tau f \tau f}}{2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} S_{qq} - n^{-1} S_{\tau f \tau f}}.
\]

It reduces to \( R^2_f = S_{11}^\tau / S_{11} \), 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 et al., 2018b].

4. Rerandomization using the Mahalanobis distance. As shown in Section 3.1, although \( \hat{\tau}_x \) has mean \( 0 \), its realized value can be very different from \( 0 \) 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 \( Z \) 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 \( M \) to denote the event that the treatment vector \( Z \) satisfies this criterion. From Proposition 1, \( M \) is asymptotically \( \chi^2_{LF} \), and therefore the asymptotic acceptance probability is \( p_a = P(\chi^2_{LF} \leq a) \) under ReFM. In practice, we usually choose a small
threshold $a$, or equivalently a small $p_a$, e.g., $p_a = 0.001$. 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 $\tau$. 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 $\varepsilon \sim \mathcal{N}(0, I_F)$ be an $F$ dimensional standard Gaussian random vector, and $\zeta_{LF,a} \sim D \mid D'D \leq a$ be an $LF$ dimensional truncated Gaussian random vector, where $D = (D_1, \ldots, D_{LF})' \sim \mathcal{N}(0, I_{LF})$. The following theorem shows the asymptotic sampling distribution of $\hat{\tau}$.

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

\[
\hat{\tau} - \tau \mid M \sim \left( V_{\tau \tau} \right)^{1/2} \varepsilon + \left( V_{\tau \tau} \right)^{1/2} \zeta_{LF,a},
\]

where $\varepsilon$ and $\zeta_{LF,a}$ are independent.

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 et al., 2018b]. We emphasize that, although the matrix $\left( V_{\tau \tau} \right)^{1/2}$ 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 $L, F, a, V_{\tau \tau}$, and $V_{\tau \tau}$. Theorem 2 immediately implies the asymptotic sampling distribution of a single factorial effect estimator. Let $\varepsilon_0 \sim \mathcal{N}(0, 1)$, and $\eta_{LF,a} \sim D_1 \mid D'D \leq a$ be the first coordinate of $\zeta_{LF,a}$.

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

\[
\hat{\tau}_f - \tau_f \mid M \sim \sqrt{V_{\tau f \tau f}} \left( \sqrt{1 - R_{f f}^2} \cdot \varepsilon_0 + \sqrt{R_{f f}^2} \cdot \eta_{LF,a} \right).
\]

The proof of Corollary 2 is in Appendix A3 of the Supplementary Material [Li et al., 2018b]. 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 et al., 2018a].
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 \( B \) of distributions on \( \mathbb{R}^m \), we say that \( B \) is closed convex if it satisfies two conditions: (i) for any distributions \( \nu_1, \nu_2 \in B \) and for any \( \lambda \in (0, 1) \), the distribution \( (1 - \lambda)\nu_1 + \lambda\nu_2 \) is in \( B \), and (ii) a distribution \( \nu \) is in \( B \) if there exists a sequence of distributions in \( B \) converging weakly to \( \nu \). For any set \( C \) of distributions, let the closed convex hull of \( C \) be the smallest closed convex set containing \( C \). A compact convex set in Euclidean space \( \mathbb{R}^m \) is called a convex body if it has a nonempty interior. A set \( K \subset \mathbb{R}^m \) is symmetric if \( K = \{-a : a \in K\} \). 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 \( U \), where \( U \) 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 \( \varepsilon \), 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 \( \varepsilon \) 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 et al., 2018b].

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 \( \chi^2_{LF,a} \sim \chi^2_{LF} \mid \chi^2_{LF} \leq a \) be a truncated \( \chi^2 \) random
variable, $\mathbf{S}$ be an $LF$ dimensional random vector whose coordinates are independent random signs with probability $1/2$ of being $\pm 1$, and $\mathbf{\beta}$ be an $LF$ dimensional Dirichlet random vector with parameters $(1/2, \ldots, 1/2)$. Let $\sqrt{\mathbf{\beta}}$ be the element-wise square root of the vector $\mathbf{\beta}$, and $v_{LF,a} = P(\chi_{LF+2}^2 \leq a)/P(\chi_{LF}^2 \leq a) \leq 1$.

**Proposition 3.** $\zeta_{LF,a}$ is spherically symmetric with covariance $v_{LF,a} \mathbf{I}_L$. It follows $\zeta_{LF,a} \sim \chi_{LF,a} \cdot \mathbf{S} \circ \sqrt{\mathbf{\beta}}$, where $(\chi_{LF,a}, \mathbf{S}, \mathbf{\beta})$ 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 et al., 2018b]. 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 et al., 2018b]. Let $\mathbf{R} = \mathbf{V}_{\tau \tau}^{-1/2} \mathbf{V}_{\tau \tau} \mathbf{V}_{\tau \tau}^{-1/2}$ 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 $\mathbf{\Pi}^2 = \text{diag}(\pi_1^2, \ldots, \pi_F^2) \in \mathbb{R}^{F \times F}$ is a diagonal matrix with nonnegative elements. The eigenvalues $(\pi_1^2, \ldots, \pi_F^2)$ 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, $\pi_1^2 = \cdots = \pi_F^2 = S_{11}/S_{11}$. The following corollary gives an equivalent form of (4.1) highlighting the dependence on the canonical correlations $(\pi_1^2, \ldots, \pi_F^2)$.

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

\[
\hat{\tau} - \tau \mid \mathcal{M} \sim \mathbf{V}_{\tau \tau}^{-1/2} \mathbf{\Gamma} \left\{ (\mathbf{I}_F - \mathbf{\Pi}^2)^{1/2} \epsilon + (\mathbf{\Pi} \mathbf{0}_{F \times (L-1)F}) \zeta_{LF,a} \right\}.
\]

The proof of Corollary 3 is in Appendix A3 of the Supplementary Material [Li et al., 2018b]. The second term in (4.3), affected by rerandomization, depends on the canonical correlations $(\pi_1^2, \ldots, \pi_F^2)$ 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 $(\mathbf{I}_3 - \mathbf{\Pi}^2)^{1/2} \epsilon + \mathbf{\Pi} \zeta_{3,a}$, 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 $\zeta_{3,a}$ for different $\pi_1^2$. As $\pi_1^2$ 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 $\zeta_{3,a}$ 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 $\varepsilon$ component in the asymptotic sampling distribution (4.3) does not depend on $p_a$ and is usually nonzero. For example, when $V_{\tau\tau}^\perp$ is positive definite, $I - R = V_{\tau\tau}^{-1/2}V_{\tau\tau}^\perp V_{\tau\tau}^{-1/2}$ is positive definite, as well as the coefficient of $\varepsilon$ in (4.3). Therefore, the gain of ReFM by decreasing $p_a$ usually becomes smaller as $p_a$ decreases.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{(a) $(\pi_1^2, \pi_2^2, p_a) = (0.5, 0.5, 0.001)$}
\caption{(b) $(\pi_1^2, \pi_2^2, \pi_3^2) = (0.5, 0.5, 0.5)$}
\end{figure}

4.5. Asymptotic unbiasedness, sampling covariance and peakedness. In this subsection, we further study the asymptotic properties of $\hat{\tau}$ under ReFM. First, the factorial effects estimator $\hat{\tau}$ is consistent for $\tau$. 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 \( \hat{\tau} \) 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 \( \hat{\tau} \) under ReFM is smaller than or equal to that under the CRFE, and the reduction in asymptotic sampling covariance is \( (1 - v_{LF,a})nV_{\tau\tau}^{II} \). Specifically, the percentage reduction in asymptotic sampling variance (PRIASV) of \( \hat{\tau}_f \) is \( (1 - v_{LF,a})R^2_f \).

Theorem 3 follows from Theorem 2 and Proposition 3, with the proof in Appendix A4 of the Supplementary Material [Li et al., 2018b]. Rigorously, the reductions in Theorem 3 should be \( (1 - v_{LF,a})\lim_{n \to \infty} nV_{\tau\tau}^{II} \) and \( (1 - v_{LF,a})\lim_{n \to \infty} R^2_f \). However, for descriptive simplicity, we omit the limit signs. From Theorem 3, the larger the squared multiple correlation \( R^2_f \) is, the more PRIASV of the factorial effect estimator is through ReFM. When \( a \) is close to zero, or equivalently the asymptotic acceptance probability \( p_a \) is small, the asymptotic sampling variance of \( \hat{\tau}_f \) reduces to \( V_{\tau_f\tau_f}(1 - R^2_f) \), 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 \( \hat{\tau} \) under ReFM and the CRFE, because of its close connection to the volumes of confidence sets for \( \tau \). 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 et al. [2009] discussed some related properties. Here we use the definition in Sherman [1955].

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

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 $\phi$ and $\psi$ 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 et al., 2018b]. For two Gaussian vectors $\phi$ and $\psi$, $\text{Cov}(\phi) \leq \text{Cov}(\psi)$ 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 $\varepsilon_0$ 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, $\xi_0$ is not more peaked than $\varepsilon_0$, 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 $\zeta_{LF,a}$ is more peaked than the standard Gaussian random vector. Its proof is in Appendix A5 of the Supplementary Material [Li et al., 2018b]. 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 $\pi_{2,k}$'s.

Theorem 5. Under ReFM, assume Condition 1. Let $c_{1-\alpha}$ be the solution of $\lim_{n \to \infty} P\{\hat{\tau} - \tau \in \mathcal{O}(V_{\tau\tau}, c_{1-\alpha}) \mid M\} = 1 - \alpha$ for any $\alpha \in (0, 1)$. It depends only on $(L, K, a)$ and the canonical correlation $\pi_{2,k}$'s, and is non-increasing in these canonical correlations for fixed $(L, K, a)$.

Theorem 5 is a multivariate extension of Li et al. [2018a, Theorem 2], with the proof in Appendix A5 of the Supplementary Material [Li et al., 2018b]. 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 $\pi_{2,k}$’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], i.e., 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 \( \hat{\tau} \), i.e., a linear combination of a standard Gaussian random vector and a truncated Gaussian random vector. For conciseness, we relegate the discussion to the Supplementary Material [Li et al., 2018b], 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 et al., 2018a].

**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 \( [-c_{1-\alpha}V_{\tau \tau}^{1/2}, c_{1-\alpha}V_{\tau \tau}^{1/2}] \) of \( \hat{\tau}_f - \tau_f \) under ReFM is smaller than or equal to that under the CRFE, and is nonincreasing in \( R^2_f \).

The proof of Corollary 4 is in Appendix A5 of the Supplementary Material [Li et al., 2018b]. From Corollary 4, with larger squared multiple correlation \( R^2_f \), 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 \( (V_{\tau \tau})_L^{-1/2} = V_{\tau x}V_{xx}^{-1/2} \), 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 \( s_{qq} = s_{qq} - s_{q,x}s_{xx}^{-1}(q)s_{x,q} \) as the sample variance of the residuals from the linear projection of observed outcomes on covariates. We estimate \( V_{\tau \tau}^{\perp} \) by

\[
\hat{V}_{\tau \tau}^{\perp} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1}s_{qq}b_qb_q',
\]

\( V_{\tau x} \) by \( \hat{V}_{\tau x} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1}(b_qb_q') \otimes \{ s_{q,x}s_{xx}^{-1/2}(q)S_{xx}^{1/2} \} \), and \( (V_{\tau \tau})_L^{-1/2} \) by \( \hat{V}_{\tau x}V_{xx}^{-1/2} \). We can then obtain a covariance estimator and construct confidence sets for \( \tau \) or its linear transformations. When the threshold \( a \) is small, \( \zeta_{LF,a} \) is close to zero, and the distribution (4.1) of \( \hat{\tau} \) is close to
the Gaussian distribution with mean $\tau$ and covariance matrix $V_{\tau\tau}$. Therefore, for a parameter of interest $C\tau$, we recommend confidence sets of the form $C\hat{\tau} + O(CV_{\tau\tau}^{-1}C')$, where $\hat{\tau}$ is the estimated sample mean and $V_{\tau\tau}$ is the estimated covariance matrix.

Theorem 6. Under ReFM and Condition 1, consider inferring $C\tau$, where $C$ has full row rank. The probability limit of the covariance estimator $C\hat{\tau} + O(CV_{\tau\tau}^{-1}C')$, is larger than or equal to the sampling covariance, and the $1 - \alpha$ confidence set, $C\hat{\tau} + O(CV_{\tau\tau}^{-1}C')$, has asymptotic coverage rate $\geq 1 - \alpha$, with equality holding if $S_{\tau\tau}^\perp \to 0$ as $n \to \infty$.

Theorem 6 holds because the ordering of peakedness still holds by adding an independent central convex unimodal random vector. Its proof is in Appendix A6 of the Supplementary Material [Li et al., 2018b]. 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 the Supplementary Material [Li et al., 2018b].

5. Rerandomization with tiers of factorial effects. From Corollary 1, under the additivity, the squared multiple correlations between $\hat{\tau}_f$ and $\hat{\tau}_x$ are the same for all $f$: $R^2_f = \cdots = R^2_{11} = S_{11}^\perp / S_{11}$. From Section 4.5, under the additivity, 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 $F = \{1, 2, \ldots, F\}$ be the set of all factorial effects. We partition $F$ into $H$ tiers $(F_1, \ldots, F_H)$ with decreasing importance, where the $F_h$'s are disjoint and $F = \bigcup_{h=1}^H F_h$. The cardinality $F_h \equiv |F_h|$ represents the number of factorial effects in tier $h$. For example, we can partition $F$ into three tiers: $F_1$ contains the $K$ main effects, $F_2$ contains the $(K)$ interaction effects between two factors, and $F_3$ contains the remaining factorial effects with higher-order interactions.

Define $\gamma^2_{jk} = \text{Corr}^2(\hat{\tau}_f, \hat{\tau}_{x,k})$. When the $f$th factorial effect is more important, we would like to put more restriction on the difference-in-means vector...
with larger squared multiple correlation $\gamma_{fk}^2$. Although general results for the relative magnitudes of the $\gamma_{fk}^2$’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 $\max_{1 \leq k \leq F} \gamma_{fk}^2 = \gamma_{fF}^2 = R_{fF}^2 = S_{11F}/S_{11}$ for $1 \leq f \leq F$. The squared multiple partial correlation between $\hat{\tau}_f$ and $\hat{\tau}_x$ given $\tau_x$ is zero, i.e., the residuals from the linear projections of $\hat{\tau}_f$ and $\hat{\tau}_x$ on $\hat{\tau}_{x,f}$ are uncorrelated. If further $n_1 = \ldots = n_Q = n/Q$, 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 et al., 2018b]. 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 $\hat{\tau}_x[F_h]$ be the subvector of $\hat{\tau}_x$, consisting of the difference-in-means of covariates $\hat{\tau}_{x,f}$ with respect to factorial effect $f \in F_h$. From Section 5.1, the smaller the $h$ is, the more restriction we want to impose on $\hat{\tau}_x[F_h]$. However, due to the correlations among the $\hat{\tau}_x[F_h]$’s, restrictions on one also restrict others. For example, balancing $\hat{\tau}_x[F_1]$ partially balances $\hat{\tau}_x[F_2]$. 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 $\tilde{B} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1} b_q b_q'$. From Proposition 1, the sampling covariance of $\hat{\tau}_x$ under the CRFE, $V_{xx} = \tilde{B} \otimes S_{xx}$, contains two components: $\tilde{B}$ determined by the coefficient vector $b_q$’s and $S_{xx}$ determined by the covariates. Below we introduce a block-wise Gram–Schmidt orthogonalization of the coefficient vector $b_q$’s, taking into account the tiers of factorial effects.

Let $F_{\tilde{h}} = \bigcup_{h=1}^\tilde{h} F_h$ be the factorial effects in the first $\tilde{h}$ tiers. We use $b_q[F_{\tilde{h}}]$ and $b_q[F_{\tilde{h}}]$ to denote the subvectors of $b_q$ with indices in $F_{\tilde{h}}$ and $F_{\tilde{h}}$, and $\tilde{B}[F_{\tilde{h}},F_{\tilde{h}}]$ and $\tilde{B}[F_{\tilde{h}},F_{\tilde{h}}]$ to denote the submatrices of $\tilde{B}$ with indices in $F_{\tilde{h}} \times F_{\tilde{h}}$ and $F_{\tilde{h}} \times F_{\tilde{h}}$. For each $1 \leq q \leq Q$, we define the orthogonalized coefficient vector $c_q = (c_q[1], \ldots, c_q[H])'$ as $c_q[1] = b_q[F_1]$, and

$$c_q[h] = b_q[F_{\tilde{h}}] - \tilde{B}[F_{\tilde{h}},F_{\tilde{h}}-1] \left\{ \tilde{B}[F_{\tilde{h}-1},F_{\tilde{h}-1}] \right\}^{-1} b_q[F_{\tilde{h}-1}], \quad (2 \leq h \leq H).$$
The difference-in-means vector of covariates with respect to orthogonalized coefficient vectors is

\[
\hat{\theta}_x \equiv \left( \begin{array}{c}
\hat{\theta}_x[1] \\
\vdots \\
\hat{\theta}_x[H]
\end{array} \right) = 2^{-(K-1)\sum_{q=1}^{Q} n_q^{-1}(c_q[1])} \otimes \hat{\bar{x}}(q).
\]

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

**Proposition 6.** Under the CRFE, \((\hat{\tau}' - \tau', \hat{\theta}_x')'\) has mean zero and sampling covariance:

\[
\text{Cov} \left( \hat{\tau} - \tau, \hat{\theta}_x[h] \right) \equiv W_{\tau x}[h] = 2^{-2(K-1)\sum_{q=1}^{Q} n_q^{-1}(b_q c_q'[h])} \otimes S_{q,x},
\]

\[
\text{Cov} \left( \hat{\theta}_x[h] \right) \equiv W_{xx}[h] = 2^{-2(K-1)\sum_{q=1}^{Q} n_q^{-1}(c_q[h] c_q'[h])} \otimes S_{xx},
\]

and \( \text{Cov}(\hat{\theta}_x[h], \hat{\theta}_x[\tilde{h}]) = 0 \) if \( h \neq \tilde{h} \).

**Proposition 6** follows from some algebra, with the proof in Appendix A2 of the Supplementary Material [Li et al., 2018b]. From Proposition 6, \((\hat{\theta}_x[1], \ldots, \hat{\theta}_x[H])\) are mutually uncorrelated under the CRFE, and thus are essentially from a block-wise Gram–Schmidt orthogonalization of \( (\hat{\tau}_x[F_1], \ldots, \hat{\tau}_x[F_H]) \).

We define the Mahalanobis distance in tier \( h \) as

\[
M_h = \hat{\theta}_x'[h] (W_{xx}[h])^{-1} \hat{\theta}_x[h], \quad (1 \leq h \leq H).
\]

Let \((a_1, \ldots, a_H)\) be \( H \) positive constants predetermined in the design stage. Under rerandomization with tiers of factorial effects, denoted by ReFMTF, a randomization is acceptable if and only if \( M_h \leq a_h \) for all \( 1 \leq h \leq H \). Below we use \( F_F \) to denote the event that the treatment vector \( Z \) satisfies this criterion. From the finite population central limit theorem, asymptotically, \( M_h \) is \( \chi^2_{L_{F_h}} \), and \((M_1, \ldots, M_H)\) are jointly independent. Therefore, the asymptotic acceptance probability under ReFMTF is \( p_a = \prod_{h=1}^{H} P(\chi^2_{L_{F_h}} \leq a_h) \).

We usually choose small \( a_h \)'s. The relative magnitude of \( a_h \)'s depend on our prior knowledge or belief on the relative importance of covariates in all tiers. See Morgan and Rubin [2015] for a more detailed discussion.

With equal treatment groups sizes, \( M_h \) has simpler form.
Proposition 7. When \( n_1 = \cdots = n_Q = n/Q \), the coefficient \( c_q[h] \) in (5.1) reduces to \( b_q[F_h] \), the difference-in-means of covariates \( \theta_x[h] \) in (5.2) reduces to \( \hat{\tau}_x[F_h] \), and the Mahalanobis distance \( M_h \) in (5.3) reduces to \( M_h = n/4 \cdot \sum_{f \in F_h} \hat{\tau}_{x,f} S_{x,x}^{-1} \hat{\tau}_{x,f} \).

Proposition 7 follows from some algebra, with the proof in Appendix A2 of the Supplementary Material [Li et al., 2018b]. In Proposition 7, if further each tier contains exactly one factorial effect, ReFMT reduces to the rerandomization scheme discussed in Branson et al. [2016].

5.3. Asymptotic sampling distribution of \( \hat{\tau} \). In this subsection, we study the asymptotic sampling distribution of \( \hat{\tau} \) under ReFMT. Let \( W_{\tau\tau}^\perp[h] = W_{\tau x}[h](W_{xx}[h])^{-1} W_{x\tau}[h] \) be the sampling covariance matrix of \( \hat{\tau} \) explained by \( \theta_x[h] \) in the linear projection under the CRFE. Extending earlier notation, let \( \zeta_{LF_h,a_h} \sim D_h \mid D_h' D_h \leq a_h \) be a truncated Gaussian vector with \( LF_h \) dimensions, where \( D_h = (D_{h1}, \ldots, D_{hLF_h})' \sim N(0, I_{LF_h}) \).

Theorem 7. Under ReFMT_f and Condition 1,

\[
\hat{\tau} - \tau \mid \mathcal{T}_F \sim \left(V_{\tau\tau}^\perp\right)^{1/2} \varepsilon + \sum_{h=1}^{H} \left(W_{\tau\tau}^I[h]\right)^{1/2} \zeta_{LF_h,a_h},
\]

where \( (\varepsilon, \zeta_{LF_1,a_1}, \ldots, \zeta_{LF_H,a_H}) \) are jointly independent.

The proof of Theorem 7, similar to that of Theorem 2, is in Appendix A3 of the Supplementary Material [Li et al., 2018b].

Let \( W_{\tau_f\tau_f}^\perp[h] \) be the \( f \)th diagonal element of \( W_{\tau\tau}^\perp[h] \). The squared multiple correlation between \( \hat{\tau}_f \) and \( \hat{\theta}_x[h] \) under the CRFE is then \( \rho_f^2[h] = W_{\tau_f\tau_f}^\perp[h]/V_{\tau_f\tau_f} \). When treatment group sizes are equal, \( \rho_f^2[h] \) reduces to \( \rho_f^2[h] = \sum_{k:h \in F_h} \rho_{f_k}^2 \) for all \( f \); if further the additivity holds, \( \rho_f^2[h] \) reduces to \( \sum_{i=1}^{H} \rho_f^2 \) if \( f \in F_h \), and zero otherwise. Because the \( \hat{\theta}_x[h] \)'s are from a block-wise 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 \( \eta_{LF_h,a_h} \sim D_h \mid D_h' D_h \leq a_h \) be the first coordinate of \( \zeta_{LF_h,a_h} \).

Corollary 5. Under ReFMT_f and Condition 1, for \( 1 \leq f \leq F \),

\[
\hat{\tau}_f - \tau_f \mid \mathcal{T}_F \sim \sqrt{V_{\tau_f\tau_f}^\perp} \left(\sqrt{1 - R_f^2} \cdot \varepsilon_0 + \sum_{h=1}^{H} \sqrt{\rho_f^2[h]} \cdot \eta_{LF_h,a_h}\right),
\]
where \((\epsilon_0, \eta_{LF_1,a_1}, \ldots, \eta_{LF_H,a_H})\) are jointly independent.

The proof of Corollary 5 is in Appendix A3 of the Supplementary Material [Li et al., 2018b].

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, \((\epsilon, \zeta_{LF_1,a_1}, \ldots, \zeta_{LF_H,a_H})\) are all central convex unimodal from Proposition 2, and thus the asymptotic sampling distribution (5.4) of \(\hat{\tau}\) under ReFMTF is also central convex unimodal. The symmetry of the asymptotic sampling distributions ensures that the factorial effect estimator \(\hat{\tau}\) is consistent for \(\tau\) under ReFMTF, 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 ReFMTF and the CRFE. For each \(1 \leq h \leq H\), let \(v_{LF_h,a_h} = P(\chi^2_{LF_h} + 2 \leq a_h)/P(\chi^2_{LF_h} \leq a_h) \leq 1\).

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

Theorem 8 follows from Theorem 7 and Proposition 3, with the proof in Appendix A4 of the Supplementary Material [Li et al., 2018b]. When the threshold \(a_h\)'s are close to zero, the asymptotic sampling variance of \(\hat{\tau}_f\) reduces to \(V_{\tau_f}\tau_f(1 - R^2_f/\hat{\rho}^2_f)\), 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 ReFMTF and the CRFE.

**Theorem 9.** Under Condition 1, the asymptotic sampling distribution of \(\hat{\tau} - \tau\) under ReFMTF 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 et al., 2018b]. We then consider the specific symmetric convex set \(O(V_{\tau\tau}, c)\). Unfortunately, considering joint quantile region for \(\tau\) is technically challenging in general, and we consider the case where the following condition holds.


**Condition 2.** There exists an orthogonal matrix $\Gamma \in \mathbb{R}^{F \times F}$ such that
\[
\Gamma'V_{\tau\tau}^{-1/2}W_{\tau\tau}^1[h]V_{\tau\tau}^{-1/2}\Gamma = \text{diag}(\omega^2_{h1}, \ldots, \omega^2_{hF}), \quad (1 \leq h \leq H)
\]
where $(\omega^2_{h1}, \ldots, \omega^2_{hF})$ 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. Let $\Psi \in \mathbb{R}^{F \times F}$ be the common linear transformation matrix from $c_q$ to $b_q$, i.e., $b_q = \Psi c_q$ for all $1 \leq q \leq Q$. Recall that $\tilde{B} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} b_q b_q'$, and $\tilde{C} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} c_q c_q'$.

**Proposition 8.** Under the additivity in Definition 1, Condition 2 holds with orthogonal matrix $\Gamma = \tilde{B}^{-1/2} \Psi \tilde{C}^{-1/2}$, and the canonical correlations between $\hat{\tau}$ and $\hat{\theta}_x[h]$ have exactly $F_h$ nonzero elements, which are all equal to $S_1^1/S_{11}$.

Proposition 8 follows from some algebra, with the proof in Appendix A5 of the Supplementary Material [Li et al., 2018b].

**Theorem 10.** Under ReFMT$_F$, assume that Conditions 1 and 2 hold. Let $c_{1-\alpha}$ be the solution of $\lim_{n \to \infty} P \{\hat{\tau} - \tau \in \mathcal{O}(V_{\tau\tau}, c_{1-\alpha}) | T_F\} = 1 - \alpha$. It depends only on $L$, $F_h$’s, $a_h$’s, and $(\omega^2_{h1}, \ldots, \omega^2_{hF})$’s, and is nonincreasing in $\omega^2_{hf}$ 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 et al., 2018b].

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 the Supplementary Material [Li et al., 2018b], and consider only the asymptotic sampling distribution of a single factorial effect estimator below.

**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 $[c_{1-\alpha} V_{\tau f}^{1/2}, c_{1-\alpha} V_{\tau f}^{1/2}]$ of $\hat{\tau}_f - \tau_f$ under ReFMT$_F$ is smaller than or equal to that under the CRFE, and is nonincreasing in $\rho^2_{hf}[h]$ for $1 \leq h \leq H$.

The proof of Corollary 6 is in Appendix A5 of the Supplementary Material [Li et al., 2018b]. From Corollary 6, with larger squared multiple correlation $\rho^2_{hf}[h]$, 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, which implies that \(R^2_f\) 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 \(\rho^2_k[1] = R^2_f\) and \(\rho^2_k[2] = 0\) 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 \(P(\chi^2_{LF_1} \leq a_1) = 0.002\) and \(P(\chi^2_{LF_2} \leq a_2) = 0.5\). Figure 2 shows the PRIASV, divided by \(R^2_f\), 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. □

![Figure 2: PRIASV of main effect estimators divided by \(R^2_f\)](image)

5.5. **Conservative covariance estimator and confidence sets under ReFMT\(_F\).**

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

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

The proof of Theorem 11, similar to that of Theorem 6, is in Appendix A6 of the Supplementary Material [Li et al., 2018b]. 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 et al., 2018b].

6. An education example. We illustrate the theory of rerandomization using a dataset from the Student Achievement and Retention Project [Angrist et al., 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 1006 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.
RERANDOMIZATION IN $2^K$ FACTORIAL EXPERIMENTS

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 quantile ranges.

| Factorial effect       | Reduction in variance | Reduction in 95% 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%        |

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 $(R^2_1, R^2_2, R^2_3) = (0.247, 0.244, 0.245)$.

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

We then consider confidence sets for the two average main effects $(\tau_1, \tau_2)$ 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 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 $p_{ah} \equiv P(\chi^2_{LF_h} \leq a_h)$ 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 non-monotone 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.

Fig 3: PRIASVs of all factorial effects as $p_{a1}$, with the overall acceptance probability $p_a = p_{a1}p_{a2}$ fixed at 0.001.

7. Extensions. When covariates have varying importance for the potential outcomes, we can further consider balance criterion using tiers of covariates, i.e., rerandomized factorial experiments with tiers of both covariates and factorial effects. We discuss in detail this balance criterion in the Supplementary Material [Li et al., 2018b]. There we study the asymptotic sampling distribution and properties of the difference-in-means estimator, and the repeated sampling inference for the average factorial effects. We demonstrate that the advantage of further considering tiers of covariates can increase with the numbers of covariates and factors.

We focused on improving completely randomized factorial experiments using rerandomization. Our future work will focus on using rerandomization to ensure covariate balance in more complex experiments, for example, randomized block, Latin squares and split-plot designs [Cox and Reid, 2000, Hinkelmann and Kempthorne, 2007, Zhao et al., 2018].
Acknowledgments. We thank the Associate Editor and two reviewers for helpful comments. Zach Branson at Harvard made constructive suggestions on an early version of our paper.

SUPPLEMENTARY MATERIAL

Supplementary Material to “Rerandomization in $2^K$ Factorial Experiments” (DOI: 10.1214/00-AOSXXXXSUPP; .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 in the Supplementary Material [Li et al., 2018b].

References.

J. Angrist, D. Lang, and P. Oreopoulos. Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1:136–163, 2009.

R. A. Bailey. Restricted randomization. *Biometrika*, 70:183–198, 1983.

A. Baldi Antognini and M. Zagoraiou. The covariate-adaptive biased coin design for balancing clinical trials in the presence of prognostic factors. *Biometrika*, 98, 2011.

P. J. Bickel and E. L. Lehmann. Descriptive statistics for nonparametric models. iii. dispersion. *The Annals of Statistics*, 4:1139–1158, 1976.

Z. W. Birnbaum. On Random Variables with Comparable Peakedness. *The Annals of Mathematical Statistics*, 19:76–81, 1948.

R. C. Bose. Mathematical theory of the symmetrical factorial design. *Sankhyā: The Indian Journal of Statistics*, 8:107–166, 1947.

G. E. P. Box, J. S. Hunter, and W. G. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery*. Hoboken, NJ: Wiley-Interscience, 2 edition, 2005.

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

M. Bruhn and D. McKenzie. In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1:200–232, 2009.

W. G. Cochran and G. M. Cox. *Experimental Designs*. New York: Wiley, 1950.

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

D. R. Cox. Randomization in the design of experiments. *International Statistical Review*, 77:415–429, 2009.

D. R. Cox and N. Reid. *The Theory of the Design of Experiments*. New York: Chapman and Hall/CRC, 2000.

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 Probabil*
X. Li, P. Ding and D. B. Rubin

ity, Volume 2: Probability Theory, pages 241–265, Berkeley, California, 1972. University of California Press.
T. Dasgupta, N. S. Pillai, and D. B. Rubin. Causal inference from $2^K$ factorial designs by using potential outcomes. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77:727–753, 2015.
S. Dharmadhikari and K. Joag-Dev. Unimodality, Convexity, and Applications. San Diego, CA: Academic Press, Inc., 1988.
S. W. Dharmadhikari and K. Jogdeo. Multivariate unimodality. The Annals of Statistics, 4:607–613, 1976.
D. J. Finney. The fractional replication of factorial arrangements. Annals of Human Genetics, 12:291–301, 1943.
R. A. Fisher. The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain, 33:503–513, 1926.
R. A. Fisher. The Design of Experiments, 1st Edition. Edinburgh, London: Oliver and Boyd, 1935.
A. Giovagnoli and H. P. Wynn. Multivariate dispersion orderings. Statistics & Probability Letters, 22:325–332, 1995.
P. M. Grundy and M. J. R. Healy. Restricted randomization and quasi-latin squares. Journal of the Royal Statistical Society, Series B (Methodological), 12:286–291, 1950.
B. B. Hansen and J. Bowers. Covariate balance in simple, stratified and clustered comparative studies. Statistical Science, 23:219–236, 2008.
K. Hinkelmann and O. Kempthorne. Design and Analysis of Experiments, Volume 1, Introduction to Experimental Design, 2nd Edition. New York: John Wiley & Sons, 2007.
M. Kanter. Unimodality and dominance for symmetric random vectors. Transactions of the American Mathematical Society, 229:65–85, 1977.
O. Kempthorne. The Design and Analysis of Experiments. New York: Chapman and Hall/CRC, 1952.
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, 2018a.
X. Li, P. Ding, and D. B. Rubin. Supplementary Material to “Rerandomization in $2^K$ Factorial Experiments”. https://doi.org/10.1214/00-AOSXXXXSUPP, 2018b.
J. Lu. Covariate adjustment in randomization-based causal inference for factorial designs. Statistics & Probability Letters, 119:11–20, 2016.
A. W. Marshall, I. Olkin, and B. C. Arnold. Inequalities: Theory of Majorization and Its Applications. Springer New York Dordrecht Heidelberg London, 2 edition, 2009.
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.
J. Neyman. On the application of probability theory to agricultural experiments. essay on principles (with discussion). section 9 (translated). reprinted ed. Statistical Science, 5:465–472, 1923.
M. Shaked. Ordering distributions by dispersion. In Johnson and Kotz, editors, Encyclopedia of Statistical Sciences, volume 6, pages 485–490. 1985.
S. Sherman. A theorem on convex sets with applications. The Annals of Mathematical
Student. Comparison between balanced and random arrangements of field plots. *Biometrika*, 29:363–378, 1938.

C. F. J. Wu. Post-Fisherian experimentation: from physical to virtual. *Journal of the American Statistical Association*, 110:612–620, 2015.

C. F. J. Wu and M. S. Hamada. *Experiments: Planning, Analysis, and Optimization*. Hoboken, NJ: John Wiley & Sons, Inc., 2011.

F. Yates. The design and analysis of factorial experiments. Technical report, Imperial Bureau of Soil Sciences—Technical Communication. No. 35, Harpenden, 1937.

F. Yates. Comment to F. J. Anscombe. *Journal of the Royal Statistical Society, Series A (General)*, 111:204–205, 1948.

W. J. Youden. Randomization and experimentation. *Technometrics*, 14:13–22, 1972.

A. Zhao, P. Ding, and T. Dasgupta. Randomization-based causal inference from unbalanced $2^2$ split-plot designs. *The Annals of Statistics*, page in press, 2018.
Supplementary Material

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, 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 $x_i[t]$ to denote the $L_t$ dimensional covariates in tier $t$. Let $x_i[\bar{t}] = (x_i[1], \ldots, x_i[t])$ be the covariates in the first $t$ tiers, $S_{x_i[t],x_i[\bar{t}]}$ be the finite population covariance of the covariates in the first $t$ tiers, and $S_{x_i[t],x_i[\bar{t}]}^{-1}$ be the finite population covariance between $x_i[t]$ and $x_i[\bar{t} - 1]$. We then apply a block-wise Gram–Schmidt orthogonalization to the $x_i[t]$’s: $e_i[1] = x_i[1]$, and

$$e_i[t] = x_i[t] - S_{x_i[t],x_i[\bar{t}]}^{-1} S_{x_i[\bar{t} - 1],x_i[\bar{t} - 1]}^{-1} x_i[\bar{t} - 1], \quad (2 \leq t \leq T)$$

where $e_i[t]$ is the residual from the linear projection of the covariates $x_i[t]$ in tier $t$ onto the space spanned by the covariates in previous tiers. Li et al. [2018a] call $e_i[t]$ 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., $\mathcal{F} = \bigcup_{h=1}^{H} \mathcal{F}_{t}$, and the corresponding block-wise Gram–Schmidt orthogonalization of $(b_{q}^{T}[\mathcal{F}_{1}], \ldots, b_{q}^{T}[\mathcal{F}_{H}])$ is $(c_{q}^{[1]}, \ldots, c_{q}^{[H]})$ defined in (5.1). Let $\hat{e}_{t}(q) = n_{q}^{-1} \sum_{i: Z_{i}=q} e_{i}[t]$ 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 $c_{q}$’s is

$$(A1) \quad \theta_{e[t]} = \left( \hat{\theta}_{e[t]}^{[1]} \vdots \hat{\theta}_{e[t]}^{[H]} \right) = 2^{-(K-1)} \sum_{q=1}^{Q} \left( c_{q}^{[1]} \vdots c_{q}^{[H]} \right) \otimes \hat{e}_{t}(q), \quad (1 \leq t \leq T).$$

Let $S_{e[t]e[t]}$ be the finite population covariance of the orthogonalized covariates in tier $t$, and $S_{q.e[t]}$ be the finite population covariance between the $Y_{t}(q)$’s and $e_{i}[t]$’s.

**Proposition A1.** Under the CRFE, $(\hat{\tau}' - \tau', \hat{\theta}_{e[t]}^{[1]}, \ldots, \hat{\theta}_{e[T]}^{[1]})'$ has mean zero and sampling covariance:

$$\text{Cov} \left( \hat{\tau} - \tau, \hat{\theta}_{e[t]}^{[h]} \right) = W_{e[t]e[t]}[h] = 2^{-2(K-1)} \sum_{q=1}^{Q} n_{q}^{-1} (b_{q} c_{q}^{[h]}) \otimes S_{q.e[t]},$$

$$\text{Cov} \left( \hat{\theta}_{e[t]}^{[h]} \right) = W_{e[t]e[t]}[h] = 2^{-2(K-1)} \sum_{q=1}^{Q} n_{q}^{-1} (c_{q}[h] c_{q}^{[h]}) \otimes S_{e[t]e[t]},$$

and $\text{Cov}(\hat{\theta}_{e[t]}^{[h]}, \hat{\theta}_{e[t]}^{[\tilde{h}]}) = 0$ if $t \neq \tilde{t}$ or $h \neq \tilde{h}$, i.e., $\hat{\theta}_{e[t]}^{[h]}$’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

$$M_{t,h} = \hat{\theta}_{e[t]}^{[h]} (W_{e[t]e[t]}[h])^{-1} \hat{\theta}_{e[t]}^{[h]}, \quad (1 \leq h \leq H, 1 \leq t \leq T).$$

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 $M_{t,1}, M_{t,2}, \ldots, M_{t,H}$ 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 $M_{1,h}, M_{2,h}, \ldots, M_{T,h}$ 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 \( h < \tilde{h} \) and \( t < \tilde{t} \), 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 \( S = \{(t,h) : 1 \leq t \leq T, 1 \leq h \leq H\} \) of all combinations of tiers of covariates and factorial effects. In practice, we can divide the set \( S \) into \( J \) tiers \( (S_1, \ldots, S_J) \) with decreasing importance, which are coherent with the partial order on \( S \).

**Example A1.** A choice of the \( S_j \)'s is the triangular tiers, where \( J = \min\{T, H\} \), and

\[
S_1 = \{(t,h) : h + t = j + 1, 1 \leq h \leq H, 1 \leq t \leq T\}, \quad (1 \leq j \leq J - 1)
\]

\[
S_J = \{(t,h) : h + t > J, 1 \leq h \leq H, 1 \leq t \leq T\}.
\]

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

\[
p_a = \prod_{j=1}^J P(\chi^2_{\lambda_j} \leq a_j).
\]

By the same logic as Proposition 7, with equal treatment groups sizes, \( M_{t,h} \) reduces to \( M_{t,h} = n/4 \cdot \sum_{f \in F_h} \hat{\tau}_{e[t],f} S_{e[t],e[t]}^{-1} \hat{\tau}_{e[t],f} \), where \( \hat{\tau}_{e[t],f} = 2^{-(K-1)} \sum_{q=1}^Q g_{fq} \hat{e}_{[t]}(q) \) is the difference-in-means of orthogonalized covariate \( e[t] \) with respect to the \( f \)th factorial effect.

**A1.3. Asymptotic sampling distribution of \( \hat{\tau} \).** For each tier \( S_j \), let \( U_{\tau e[j]} \in \mathbb{R}^{F \times \lambda} \) be a matrix consisting of the columns of matrices \( \{W_{\tau e[j][h]}(t,h) \}_{(t,h) \in S_j} \); \( U_{ee[j]} \in \mathbb{R}^{\lambda_j \times \lambda} \) be a block diagonal matrix with \( \{W_{e[t]e[t][h]} \}_{(t,h) \in S_j} \) as the diagonal components, and \( U_{\tau \tau[j]} = U_{\tau e[j]}(U_{ee[j]})^{-1}U_{e\tau[j]} \) be the
sampling covariance matrix of \( \hat{\tau} \) explained by \( \{ \hat{\theta}_{e[t]}[h] \}_{(t,h) \in S_j} \) in the linear projection under the CRFE. Recall \( \zeta_{\lambda_j,a_j} \sim D_j | D_j D_j' \leq a_j \), where \( D_j = (D_{j1}, \ldots, D_{j,\lambda_j})' \sim N(0, I_{\lambda_j}) \).

**Theorem A1.** Under ReFMT$_{CF}$ and Condition 1,

\[
(A2) \quad \hat{\tau} - \tau \mid T_{CF} \sim \left( V_{\tau\tau}^{-1} \right)^{1/2} \varepsilon + \sum_{j=1}^{J} \left( U_{\tau\tau}^\pi[j] \right)^{1/2} \zeta_{\lambda_j,a_j},
\]

where \( (\varepsilon, \zeta_{\lambda_1,a_1}, \ldots, \zeta_{\lambda_J,a_J}) \) are jointly independent.

Let \( U_{\tau\tau}^\pi[j] \) be the fth diagonal element of \( U_{\tau\tau}^\pi[j] \), and \( \beta_f^2[j] = \beta^2_f[j] / V_{\tau\tau}^\pi \) be the proportion of variance of \( \hat{\tau}_f \) explained by \( \{ \hat{\theta}_{e[t]}[h] \}_{(t,h) \in S_j} \) in the linear projection. Because the \( \{ \hat{\theta}_{e[t]}[h] \}_{(t,h) \in S_j} \)'s are essentially from a block-wise Gram–Schmidt orthogonalization of \( \hat{\tau}_x \), we have \( \sum_{j=1}^{J} \beta_f^2[j] = R_f^2 \). Let \( \eta_{\lambda_j,a_j} \sim D_{j1} | D_j D_j' \leq a_j \) be the first coordinate of \( \zeta_{\lambda_j,a_j} \). Theorem A1 immediately implies the following asymptotic sampling distribution of a single factorial effect estimator.

**Corollary A1.** Under ReFMT$_{CF}$ and Condition 1, for \( 1 \leq f \leq F \),

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

where \( (\varepsilon_0, \eta_{\lambda_1,a_1}, \ldots, \eta_{\lambda_J,a_J}) \) are jointly independent.

A1.4. *Asymptotic unbiasedness, sampling covariance and peakedness.* First, the asymptotic sampling distribution \((A2)\) is central convex unimodal. Therefore, the factorial effect estimator \( \hat{\tau} \) is asymptotically unbiased for \( \tau \) 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 \( \hat{\tau} \) under ReFMT$_{CF}$.

**Theorem A2.** Under Condition 1, \( \hat{\tau} \) has smaller asymptotic sampling covariance matrix under ReFMT$_{CF}$ than that under the CRFE, and the reduction in asymptotic sampling covariance is \( n \sum_{j=1}^{J} (1 - v_{\lambda_j,a_j}) U_{\tau\tau}^\pi[j] \).

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 $O(V_{\tau\tau}, c)$. Similar to Section 5.4, due to some technical difficulties, we consider only the case under the following condition.

Condition A3. There exists an orthogonal matrix $\Gamma \in \mathbb{R}^{F \times F}$ such that

$$\Gamma^TV_{\tau\tau}^{-1/2}U_{\tau\tau}^1[j]V_{\tau\tau}^{-1/2}\Gamma = \text{diag}(\kappa^2_{j1}, \ldots, \kappa^2_{jF}), \quad (1 \leq j \leq J)$$

where $(\kappa^2_{j1}, \ldots, \kappa^2_{jF})$ are the canonical correlations between $\hat{\tau}$ and $(\hat{\theta}_{e[t][h]}{(t,h)} \in S_j)$ under the CRFE.

Condition A3 holds automatically when $J = 1$. Moreover, the following proposition shows that the additivity is a sufficient condition for Condition A3. Recall that $\Psi$ is the linear transformation from $c_q$ to $b_q$, $\tilde{B} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1}b_qb_q'$, and $\tilde{C} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1}c_qc_q'$.

Proposition A2. Under the additivity in Definition 1, Condition A3 holds with orthogonal matrix $\Gamma = \tilde{B}^{-1/2}\Psi\tilde{C}^{1/2}$.

Theorem A4. Under ReFMT$_{CF}$, assume that Conditions 1 and A3 hold. Let $c_{1-\alpha}$ be the solution of $\lim_{n \to \infty} P \{ \hat{\tau} - \tau \in O(V_{\tau\tau}, c_{1-\alpha}) | \mathcal{T}_{CF} \} = 1 - \alpha$. It depends only on $K$, $\lambda_j$'s, $a_j$'s, and $(\kappa^2_{j1}, \ldots, \kappa^2_{jF})$'s, and is nonincreasing in $\beta^2_{jf}$ 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 $\beta^2_{jf}(j)$, 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 $S_1$ consists of the combination of main effects and first covariate with $\lambda_1 = K$, and $S_2$ 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 $P(\chi^2_{\lambda_1} \leq a_1) = 0.002$ and $P(\chi^2_{\lambda_2} \leq a_2) = 0.5$. Then for the main effect $1 \leq k \leq K, \beta_2^k[1] = R^2_k$ and $\beta_2^k[2] = 0$. Figure A1 shows the PRIASV, divided by $R^2_f$, of the main effect estimators for ReFM and ReFMCF. It shows that the advantage of further using tiers of covariates increases as numbers of factors and covariates increase.

\[ \text{Fig A1: PRIASV of main effect estimators under ReFM and ReFMCF, divided by } R^2_f \]

A1.5. Conservative covariance estimator and confidence set under ReFMCF.

We define $s_{q, e[t]}$ and $s_{e[t], e[t]}(q)$ as the sample covariances of the observed outcomes and orthogonalized covariates under treatment combination $q$. We estimate $V_{\tau\tau}$ by $\hat{V}_{\tau\tau}$ in (4.4), $W_{\tau e[t]}[h]$ by $\hat{W}_{\tau e[t]}[h] = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1}(b_q c'_q[h]) \otimes \{s_{q, e[t]} s_{e[t], e[t]}(q) s_{e[t], e[t]}(q)^{1/2}\} U_{\tau e[e]}[j]$ by $\hat{U}_{\tau e[e]}[j]$ consisting of the corresponding $W_{\tau e[t]}[h]$s, and $\left(U_{\tau \tau}[j]\right)^{1/2} = U_{\tau e[e]}[j](U_{ee}[e][j]^{-1/2} \hat{U}_{\tau e[e]}[j](U_{ee}[e][j]^{-1/2})$. We can then obtain a covariance estimator and construct confidence sets for $\tau$ 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 $C\tau + \mathcal{O}(CV_{\tau\tau}^{-1}C',c)$, and choose the threshold $c$ by simulating the estimated asymptotic distribution. Let $\hat{c}_{1-\alpha}$ be the $1-\alpha$ quantile of $(C\phi)'(CV_{\tau\tau}^{-1}C')^{-1}(C\phi)$ with $\phi$ following the estimated asymptotic sampling distribution of $\tau - \tau$ 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 $C\hat{\tau}$, $CV_{\tau\tau}^{-1}C' + \sum_{j=1}^{J} v_{\lambda_{j},a_{j}} C\hat{U}_{\tau\epsilon}[j](U_{\epsilon\epsilon}[j])^{-1}\hat{U}_{\epsilon\tau}[j]C'$, is larger than or equal to the actual sampling covariance, and the $1-\alpha$ confidence set, $C\hat{\tau} + \mathcal{O}(CV_{\tau\tau}^{-1}C',\hat{c}_{1-\alpha})$, has asymptotic coverage rate $\geq 1-\alpha$, with equality holding if $S_{\tau\tau}^{2} \rightarrow 0$ 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 $1_{m}$ to denote an $m$ dimensional column vector with all elements one, $I_{m}$ to denote an $m \times m$ identity matrix, and $J_{m}$ to denote an $m \times m$ matrix with all elements one.

**Lemma A1.** Let $A, B, C$ and $D$ be four matrices.

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

Recall that $g_{f} = (g_{f1}, \ldots, g_{fQ})'$ is the generating vector for the $f$th factorial effect, $b_{q} = (g_{1q}, \ldots, g_{Fq})'$ is the coefficient vector for the treatment combination $q$, and $B = 2^{-2(K-1)} \sum_{q=1}^{Q} n_{q}^{-1} b_{q} b_{q}'$.

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

(A1) $2^{-2(K-1)} n_{q}^{-1} n_{k}^{-1} b_{q}' B^{-1} b_{k} = n_{q}^{-1} \times 1\{q = k\} - n^{-1}$.

**Proof of Lemma A2.** Let $B = (b_{1}, \ldots, b_{Q}) = (g_{1}, \ldots, g_{F})' \in \mathbb{R}^{F \times Q}$, and $N = \text{diag}(n_{1}, \ldots, n_{Q}) \in \mathbb{R}^{Q \times Q}$. Then $B = 2^{-2(K-1)} \sum_{q=1}^{Q} n_{q}^{-1} b_{q} b_{q}' =
2^{-2(K-1)}BN^{-1}B', and 2^{-2(K-1)}n_q^{-1}n_k^{-1}b_q'B^{-1}b_k is the \((q,k)\)th element of matrix \(C = N^{-1}B'(BN^{-1}B')^{-1}BN^{-1}\).

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

\[
0 = B(C - N^{-1}) = (g_1, \ldots, g_F)'(C - N^{-1}).
\]

By the properties of generating vectors, \((1_Q, g_1, \ldots, g_F)\) constitute an orthogonal basis of \(\mathbb{R}^Q\). Equation (A2) implies that each column of \(C - N^{-1}\) is orthogonal to \(g_1, \ldots, g_F\), and thus has to be \(c1_{Q \times 1}\) for some constant \(c\). Because \(C - N^{-1}\) is a symmetric matrix, \(C - N^{-1}\) must be \(cJ_Q\) for some constant \(c\).

Second, we show that \(c = -n^{-1}\) and \(C = N^{-1} - n^{-1}J_Q\). On the one hand, \(C = N^{-1} + cJ_Q\) implies

\[
\text{tr}(CN) = \text{tr}(I) + c \cdot \text{tr}(J_QN) = Q + c \sum_{q=1}^Q n_q = Q + cn;
\]
on the other hand, the definition of \(C\) implies

\[
\text{tr}(CN) = \text{tr}\left\{N^{-1}B'(BN^{-1}B')^{-1}B\right\} = \text{tr}\left\{BN^{-1}B'(BN^{-1}B')^{-1}\right\} = F = Q - 1.
\]

Therefore, \(c = -n^{-1}, C = N^{-1} - n^{-1}J_Q\), and Lemma A2 holds.

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

**Proof of Lemma A3.** First, we consider the case with \(p = m\), i.e., \(A\) and \(B\) are square matrices. From the polar decomposition, \(A = (AA')^{1/2}\Gamma_1\) and \(B = (BB')^{1/2}\Gamma_2\), where \(\Gamma_1\) and \(\Gamma_2\) are orthogonal matrices. Therefore,

\[
B\Gamma_2'\Gamma_1 = (BB')^{1/2}\Gamma_2'\Gamma_1 = (AA')^{1/2}\Gamma_1 = A,
\]

where \(\Gamma_2'\Gamma_1\) is an orthogonal matrix. Thus, Lemma A3 holds when \(p = m\).

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

\[
A_1 = \begin{pmatrix} A \\ 0_{(m-p) \times m} \end{pmatrix} \in \mathbb{R}^{m \times m}, \quad B_1 = \begin{pmatrix} B \\ 0_{(m-p) \times m} \end{pmatrix} \in \mathbb{R}^{m \times m}.
\]

We can verify \(A_1A_1' = B_1B_1'\). From the first case for square matrices, there exists an orthogonal matrix \(\Gamma \in \mathbb{R}^{m \times m}\) such that \(A_1 = B_1\Gamma\), which immediately implies \(A = B\Gamma\). Thus, Lemma A3 holds when \(p < m\).
Third, we consider the case with \( p > m \). Let \( W_0 \in \mathbb{R}^{p \times p_0} \) be a matrix whose columns are an orthonormal basis of \( \{ x : A'x = 0 \} \subset \mathbb{R}^p \), and \( W_1 \in \mathbb{R}^{p \times (p - p_0)} \) be the matrix such that \( W = (W_1, W_0) \in \mathbb{R}^{p \times p} \) is an orthogonal matrix. We can verify \( W_0' A = 0 \).

Because \( W_0' B = 0 \), we have \( p_0 \geq p - m \) and therefore \( p - p_0 \leq m \). Because the two matrices \( W_1' A \) and \( W_1' B \) in \( \mathbb{R}^{(p - p_0) \times m} \) satisfy \( W_1' AA' W_1 = W_1' BB' W_1 \), from the second case, there exists an orthogonal matrix \( \Gamma \in \mathbb{R}^{m \times m} \) such that \( W_1' A = W_1' B \Gamma \). Thus,

\[
W' B \Gamma = \frac{W_1'}{W_0'} BT = \left( \frac{W_1'}{W_0'} B \Gamma \right) = \left( \frac{W_1' A}{W_0' A} \right) = W'A.
\]

Because \( W \) is an orthogonal matrix, we have \( B \Gamma = A \). Thus, Lemma A3 holds when \( p > m \).

**A2.2. Covariances between \( \hat{\tau} \) and \( \hat{\tau}_x \).**

**Proof of Proposition 1.** We first write the vector \((\hat{\tau}', \hat{\tau}_x')'\) as a linear combination of average observed outcomes and covariates:

\[
\begin{pmatrix}
\hat{\tau} \\
\hat{\tau}_x
\end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^{Q} \begin{pmatrix}
b_q \hat{Y}(q) \\
g_{1q} \hat{x}(q) \\
\vdots \\
g_{Fq} \hat{x}(q)
\end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^{Q} \begin{pmatrix}
b_q & 0_{F \times L} \\
0_{F \times L} & b_q \otimes I_L
\end{pmatrix} \begin{pmatrix}
\hat{Y}(q) \\
\hat{x}(q)
\end{pmatrix}.
\]

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

\[
2^{-(K-1)} \sum_{q=1}^{Q} \begin{pmatrix}
b_q & 0_{F \times L} \\
0_{F \times L} & b_q \otimes I_L
\end{pmatrix} \begin{pmatrix}
\bar{Y}(q) \\
\bar{x}
\end{pmatrix} = \begin{pmatrix}
\tau \\
0_{F \times L}
\end{pmatrix},
\]
and sampling covariance matrix

\[
2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} \begin{pmatrix}
  b_q & 0_{FL} \\
  0_{FL} & b_q \otimes I_L
\end{pmatrix}
\begin{pmatrix}
  S_{qq} & S_{q,x} \\
  S_{x,q} & S_{xx}
\end{pmatrix}
\begin{pmatrix}
  b_q & 0_{FL} \\
  0_{FL} & b_q \otimes I_L
\end{pmatrix}'
\]
\[= 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} \begin{pmatrix}
  b_q b_q' S_{qq} & b_q S_{q,x} \cdot (b_q \otimes I_L)' \\
  (b_q \otimes I_L) \cdot S_{x,q} b_q' & (b_q \otimes I_L) S_{xx} (b_q \otimes I_L)'
\end{pmatrix}
\[-n^{-1} \begin{pmatrix}
  S_{\tau \tau} & 0 \\
  0 & 0
\end{pmatrix}.
\]

From Lemma A1,

\[
b_q S_{q,x} \cdot (b_q \otimes I_L)' = (b_q \otimes S_{q,x}) (b_q' \otimes I_L) = (b_q b_q') \otimes S_{q,x},
\]

\[
(b_q \otimes I_L) S_{xx} (b_q \otimes I_L)' = (b_q \otimes I_L) (1 \otimes S_{xx}) (b_q' \otimes I_L) = (b_q b_q') \otimes S_{xx}.
\]

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

**Proof of Proposition 6.** We first rewrite

\[
\begin{pmatrix}
  \hat{\tau} - \tau \\
  \hat{\theta}_x
\end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^{Q} \begin{pmatrix}
  b_q \hat{Y}(q) - b_q \bar{Y}(q) \\
  c_q \otimes \hat{x}(q)
\end{pmatrix} = 2^{-(K-1)} \sum_{q=1}^{Q} \begin{pmatrix}
  b_q & 0 \\
  0 & c_q \otimes I_L
\end{pmatrix}
\begin{pmatrix}
  \hat{Y}(q) - \bar{Y}(q) \\
  \hat{x}(q)
\end{pmatrix}.
\]

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

\[
2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} \begin{pmatrix}
  b_q & 0 \\
  c_q \otimes I_L
\end{pmatrix}
\begin{pmatrix}
  S_{qq} & S_{q,x} \\
  S_{x,q} & S_{xx}
\end{pmatrix}
\begin{pmatrix}
  b_q & 0 \\
  0 & c_q \otimes I_L
\end{pmatrix}' - n^{-1} \begin{pmatrix}
  S_{\tau \tau} & 0 \\
  0 & 0
\end{pmatrix}
\]
\[= 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} \begin{pmatrix}
  b_q b_q' S_{qq} & (b_q c_q') \otimes S_{q,x} \\
  (c_q b_q') \otimes S_{x,q} & (c_q c_q') \otimes S_{xx}
\end{pmatrix}
\]-n^{-1} \begin{pmatrix}
  S_{\tau \tau} & 0 \\
  0 & 0
\end{pmatrix}.
\]

From the Gram–Schmidt orthogonalization of the \(c_q\)'s,

\[
\sum_{q=1}^{Q} n_q^{-1} c_q c_q' = \sum_{q=1}^{Q} n_q^{-1} \begin{pmatrix}
  c_q[1] & 0 & \cdots & 0 \\
  0 & c_q[2] c_q'[2] & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & c_q[H] c_q'[H]
\end{pmatrix}.
\]

Therefore, Proposition 6 holds. 

\[\square\]
Proof of Proposition A1. We first rewrite

\[
\begin{pmatrix}
\hat{\tau} - \tau \\
\hat{\theta}_{e[1]} \\
\vdots \\
\hat{\theta}_{e[T]}
\end{pmatrix}
= 2^{-(K-1)} \sum_{q=1}^{Q} \begin{pmatrix}
b_q \\
0 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
\hat{Y}(q) - b_q \bar{Y}(q) \\
c_q \otimes \hat{e}[1](q) \\
\vdots \\
c_q \otimes \hat{e}[T](q)
\end{pmatrix}
= 2^{-(K-1)} \sum_{q=1}^{Q} \begin{pmatrix}
b_q \\
0 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
0 \\
c_q \otimes I_{L_1} \\
\vdots \\
c_q \otimes I_{L_T}
\end{pmatrix}
\begin{pmatrix}
\hat{Y}(q) - \bar{Y}(q) \\
\hat{e}[1](q) \\
\vdots \\
\hat{e}[T](q)
\end{pmatrix}.
\]

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

\[
2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} \begin{pmatrix}
S_{qq} & S_{q,e[1]} & \cdots & S_{q,e[T]}
\end{pmatrix}
\begin{pmatrix}
b_q \\
0 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
S_{e[1],q} & S_{e[1],e[1]} & \cdots & S_{e[1],e[T]}
\end{pmatrix}
\begin{pmatrix}
0 \\
0 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
0 \\
0 \\
\vdots \\
0
\end{pmatrix}
- n^{-1}
\begin{pmatrix}
S_{\tau\tau} & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{pmatrix}
= 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} \begin{pmatrix}
(c_q b_q') \otimes S_{e[1],q} & (c_q b_q') \otimes S_{q,e[1]} & \cdots & (c_q b_q') \otimes S_{q,e[T]}
\end{pmatrix}
\begin{pmatrix}
b_q' S_{qq} \\
0 \\
\vdots \\
0
\end{pmatrix}
\begin{pmatrix}
0 \\
0 \\
\vdots \\
0
\end{pmatrix}
- n^{-1}
\begin{pmatrix}
S_{\tau\tau} & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{pmatrix}.
\]

Therefore, Proposition A1 holds.
\qed
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

\[
V_{\tau x} V_{xx}^{-1} V_{x\tau} = 2^{4(K-1)} \left\{ \sum_{q=1}^{Q} n_q^{-1}(b_q b'_q) \otimes S_{q, x} \right\} \left( \tilde{B} \otimes S_{xx} \right)^{-1} \left\{ \sum_{k=1}^{Q} n_k^{-1}(b_k b'_k) \otimes S_{x,k} \right\}
\]

\[
= 2^{4(K-1)} \sum_{q=1}^{Q} \sum_{k=1}^{Q} \left\{ n_q^{-1}(b_q b'_q) \otimes S_{q, x} \right\} \left( \tilde{B} \otimes S_{xx} \right)^{-1} \left\{ n_k^{-1}(b_k b'_k) \otimes S_{x,k} \right\}
\]

\[
= 2^{-4(K-1)} \sum_{q=1}^{Q} \sum_{k=1}^{Q} \left\{ n_q^{-1}(b_q b'_q) \cdot S_{x, q} \right\} \left( \tilde{B} \otimes S_{xx} \right)^{-1} \left\{ n_k^{-1}(b_k b'_k) \cdot S_{x, k} \right\}
\]

\[
= 2^{-2(K-1)} \sum_{q=1}^{Q} \sum_{k=1}^{Q} b_q \left( 2^{-2(K-1)} n_q^{-1} n_k^{-1} b_q b'_q \tilde{B}^{-1} b_k \right) \cdot S_{q, x} S_{x, q}^{-1} S_{x, k}
\]

where in the last equality the Kronecker product reduces to the matrix product because \( S_{q, x} S_{x, q}^{-1} S_{x, k} \) is a scalar. Using Lemma A2, we have

\[
V_{\tau x} V_{xx}^{-1} V_{x\tau} = 2^{-4(K-1)} \left[ \sum_{q=1}^{Q} \left( n_q^{-1} b_q b'_q \right) \cdot S_{q, x} S_{x, q}^{-1} S_{x, q} \right] - \sum_{q=1}^{Q} \sum_{k=1}^{Q} n_q^{-1} b_q b'_q \cdot S_{q, x} S_{x, q}^{-1} S_{x, k}
\]

\[
= 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} b_q b'_q \cdot S_{q, q} - n^{-1} \left( 2^{-(K-1)} \sum_{q=1}^{Q} b_q S_{q, x} \right) S_{x, x}^{-1} \left( 2^{-(K-1)} \sum_{k=1}^{Q} S_{x, k} b'_k \right)
\]

\[
= 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} b_q b'_q \cdot S_{q, q} - n^{-1} S_{x, x} S_{x, q}^{-1} S_{x, x} S_{x, x}
\]

\[
= 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} b_q b'_q \cdot S_{q, q} - n^{-1} S_{x, x}^{-1} S_{x, x} \equiv V_{\tau, x}^2
\]

Therefore, the sampling covariance of the residual from the linear projection
of \( \hat{\tau} \) on \( \hat{\tau}_x \) satisfies
\[
V_{\tau \tau} - V_{\tau x} V_{xx}^{-1} V_{\tau x} = V_{\tau \tau} - V_{\tau \tau}^\| = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} b_q b_q' \left( S_{qq} - S_{qq}^1 \right) - n^{-1} \left( \mathbf{S}_{\tau \tau} - \mathbf{S}_{\tau \tau}^\| \right)
\]
\[
= 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} b_q b_q' \cdot S_{qq}^1 - n^{-1} \mathbf{S}_{\tau \tau} = 2^{-2(K-1)} \mathbf{S}_{\tau \tau} = \mathbf{V}_{\tau \tau}^\perp.
\]

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 \( \hat{\tau}_x \) in the linear projection is \( V_{\tau_f \tau_f}^\| \). Therefore, the squared multiple correlation between \( \hat{\tau}_f \) and \( \hat{\tau}_x \) is
\[
R^2(f) = \frac{V_{\tau_f \tau_f}}{V_{\tau_f \tau_f}^\ell} = \frac{2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} S_{qq}^\ell - n^{-1} \mathbf{S}_{\tau_f \tau_f}^\ell}{2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} S_{qq} - n^{-1} \mathbf{S}_{\tau_f \tau_f}}.
\]

Under the additivity, \( S_{11} = \cdots = S_{QQ} \), \( S_{11}^\| = \cdots = S_{QQ}^\| \), and \( S_{\tau_f \tau_f} = S_{\tau_f \tau_f}^\ell = 0 \), which further imply \( R^2_f = S_{11}^\ell / S_{11} \). \( \square \)

**Proof of Proposition 5.** Under the additivity, from Proposition 1, the sampling variance and covariances are \( \text{Var}(\hat{\tau}_f) = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} S_{11} \), \( \text{Cov}(\hat{\tau}_{x,k}) = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} S_{xx} \), and \( \text{Cov}(\hat{\tau}_f, \hat{\tau}_{x,k}) = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} g_{f q} g_{k q} S_{1,1} \).

Let \( \mathbf{g}_0 = \mathbf{1}_Q \). By the property of \( \mathbf{2}^K \) 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 \( f \circ k \) as the index such that \( \mathbf{g}_f \circ \mathbf{g}_k = \mathbf{g}_{f \circ k} \).

Let \( w_q = n_q^{-1} / \sum_{k=1}^{K} n_k^{-1} \) 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 \( \hat{\tau}_{x,k} \) under the CRFE as
\[
\text{Cov}(\hat{\tau}_f, \hat{\tau}_{x,k}) \text{Cov}^{-1}(\hat{\tau}_{x,k}) \text{Cov}(\hat{\tau}_{x,k}, \hat{\tau}_f) = 2^{-2(K-1)} \left( \sum_{q=1}^{Q} n_q^{-1} g_{f \circ k, q} S_{1,1} \right) \left( \sum_{q=1}^{Q} n_q^{-1} S_{xx} \right)^{-1} \left( \sum_{q=1}^{Q} n_q^{-1} g_{f \circ k, q} S_{1,1} \right)
\]
\[
= 2^{-2(K-1)} \left( \sum_{q=1}^{Q} n_q^{-1} \right) \left( \sum_{q=1}^{Q} w_q g_{f \circ k, q} \right)^2 S_{11}^\ell.
\]
Therefore, the squared multiple correlation between $\hat{\tau}_f$ and $\hat{\tau}_{x,k}$ satisfies

$$\gamma_{fk}^2 = \frac{\text{Cov}(\hat{\tau}_f, \hat{\tau}_{x,k})\text{Cov}^{-1}(\hat{\tau}_{x,k}, \hat{\tau}_f)}{\text{Var}(\hat{\tau}_f)}$$

$$= \frac{2^{-2(K-1)} \left( \sum_{q=1}^{Q} n_q^{-1} \right) \left( \sum_{q=1}^{Q} w_q g_{f,k,q} \right)^2 S_{11}^I}{2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} S_{11}}$$

$$= \left( \sum_{q=1}^{Q} w_q g_{f,k,q} \right)^2 S_{11}^I / S_{11} = \left( \sum_{q=1}^{Q} w_q - \sum_{q \neq f,k,q} w_q \right)^2 S_{11}^I / S_{11}$$

$$\leq \left( \sum_{q=1}^{Q} w_q \right)^2 S_{11}^I / S_{11} = S_{11}^I / S_{11},$$

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

**Proof of Proposition 7.** With equal treatment group sizes, from the definition of coefficient vector $b_q$, we have

$$\tilde{B} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} b_q b_q' = (4/Q^2)(n/Q)^{-1} \cdot \sum_{q=1}^{Q} b_q b_q'$$

$$= 4/(nQ) \cdot (b_1, \ldots, b_Q) \cdot (b_1, \ldots, b_Q)'$$

$$= 4/(nQ) \cdot (g_1, \ldots, g_F)', (g_1, \ldots, g_F).$$

By the property of $2^K$ factorial experiments, these generating vectors satisfy that, for $1 \leq f \neq k \leq F$, $g_{j}^{f} g_{k} = 0$ and $g_{j}^{f} g_{f} = Q$. Thus, $\tilde{B}$ further reduces to $\tilde{B} = 4/(nQ) \cdot Q I_F = (4/n) I_F$, which is a diagonal matrix. Therefore, in (5.1), $\tilde{B}[F_h, F_{h-1}] = 0$, and $c_q[h]$ reduces to $b_q[F_h]$. Consequently, $\theta_x[h]$ in
(5.2) reduces to $\hat{\tau}_x[\mathcal{F}_h]$. Because

$$W_{xx}[h] = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1}(c_q[h]c_q'[h]) \otimes S_{xx}$$

$$= 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1}(b_q[\mathcal{F}_h]b_q'[\mathcal{F}_h]) \otimes S_{xx} = \tilde{B}[\mathcal{F}_h, \mathcal{F}_h] \otimes S_{xx}$$

$$= \frac{4}{n} \cdot I_{F_h} \otimes S_{xx},$$

we can simplify $M_h$ as

$$M_h = \hat{\theta}_x'[h] (W_{xx}[h])^{-1} \hat{\theta}_x[h] = \hat{\tau}_x'[\mathcal{F}_h] (4/n \cdot I_{F_h} \otimes S_{xx})^{-1} \hat{\tau}_x[\mathcal{F}_h]$$

$$= \frac{n}{4} \cdot \sum_{f \in \mathcal{F}_h} \tilde{\tau}_{x,f} S_{xx}^{-1} \tilde{\tau}_{x,f}.$$ 

Therefore, Proposition 7 holds. $\square$

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

A3.1. **Lemmas for central convex unimodality.** We use $B_m(r)$ 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 $p \geq 1$, $(\psi', 0_{p \times 1})' \in \mathbb{R}^{m+p}$ is central convex unimodal.

**Proof of Lemma A5.** For any random vector $\phi$ 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 $\mathcal{K} \times B_p(r) \subset \mathbb{R}^{m+p}$. Because $\mathcal{K} \times B_p(r)$ is a symmetric convex set, $\phi_r$ is central convex unimodal. As $r$ goes to zero, $\phi_r$ converges weakly to $(\phi', 0_{p \times 1})'$. Therefore, $(\phi', 0_{p \times 1})'$ 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 $C \in \mathbb{R}^{p \times m}$, $C\psi \in \mathbb{R}^p$ is central convex unimodal.

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

Second, we consider the case where $C$ has full row rank, i.e., $\text{rank}(C) = p \leq m$. In this case, there exists an $(m-p) \times m$ matrix $\bar{C}$ such that $(C', \bar{C}')$ is an $m \times m$ invertible matrix. From the first case, we know for any central convex unimodal random vector $\psi \in \mathbb{R}^m$, $(C', \bar{C}')\psi$ is also central convex unimodal. From Lemma A4, its subvector $C\psi$ is also central convex unimodal.

Third, we consider a general matrix $C$ of rank $p_1$. Let $W$ be a matrix permuting the rows of $C$ such that the first $p_1$ rows of $WC$ are linearly independent, and the remaining $p_2 = p - p_1$ rows of $WC$ are all linear combinations of the first $p_1$ rows, i.e., $(WC)' = (C_1', C_2')$, where $C_1 \in \mathbb{R}^{p_1 \times m}$, $C_2 \in \mathbb{R}^{p_2 \times m}$, $\text{rank}(C_1) = p_1$, and $C_2 = \Gamma C_1$. Thus,

$$C\psi = W^{-1}(WC)\psi = W^{-1}\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} \psi = W^{-1}\begin{pmatrix} C_1 \\ \Gamma C_1 \end{pmatrix} \psi = W^{-1}\begin{pmatrix} C_1 \psi \\ \Gamma C_1 \psi \end{pmatrix}$$

$$= W^{-1}\begin{pmatrix} I_{p_1} & 0_{p_1 \times p_2} \\ \Gamma & I_{p_2} \end{pmatrix} \begin{pmatrix} C_1 \psi \\ 0_{p_2 \times 1} \end{pmatrix} \equiv \tilde{\Gamma} \begin{pmatrix} C_1 \psi \\ 0_{p_2 \times 1} \end{pmatrix}.$$  

From the second case, $C_1 \psi$ is central convex unimodal. From Lemma A5, $((C_1 \psi)', 0_{p_2 \times 1})$ is also central convex unimodal. Because the $p \times p$ matrix $\tilde{\Gamma}$ is invertible, from the first case, we know $C\psi$ is central convex unimodal.  

**Lemma A7.** Let $\psi_1$ and $\psi_2$ be two independent $m$ dimensional random vectors. If both $\psi_1$ and $\psi_2$ 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].  

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

**Proof of Proposition 2.** For any \( m \), the densities of \( \varepsilon \) and \( \zeta_{m,a} \) are

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

and the log-densities of \( \varepsilon \) and \( \zeta_{m,a} \) are

\[
\log f(x) = -(m/2) \log(2\pi) - x' x/2,
\]

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

It is straightforward to show that both \( \log f(\cdot) \) and \( \log g(\cdot) \) are concave functions. From Lemma A8, both \( \varepsilon \) and \( \zeta_{m,a} \) are central convex unimodal. Because both \( \varepsilon \) and \( \zeta_{LF,a} \) are central convex unimodal, using Lemma A6, both \( (V_{\tau^T})^{1/2} \varepsilon \) and \( (V_{\tau^T})^{1/2} \zeta_{LF,a} \) 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 \( D \sim \mathcal{N}(0, I_{LF}) \). For any orthogonal matrix \( \Gamma \), \( \Gamma D \sim D \), and thus,

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

Second, from Morgan and Rubin [2012, Theorem 3.1], \( \text{Cov}(\zeta_{LF,a}) = v_{LF,a} I_{LF} \).

Third, we show the representation for \( \zeta_{LF,a} \). By the spherical symmetry of the standard Gaussian random vector \( D \), \( (D' D)^{-1/2} U = (U_1, \ldots, U_{LF})' \) is uniformly distributed on the \( LF \) dimensional unit sphere, \( D' D \) follows \( \chi_{LF}^2 \), and they are independent. Therefore,

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

Let \( \text{sign}(U_i) \) be the sign of \( U_i \). Given \( (|U_1|, \ldots, |U_{LF}|) \), the sign vector \( (\text{sign}(U_1), \ldots, \text{sign}(U_{LF})) \) has the same probability to be any value in \( \{-1, 1\}^m \). Thus, \( (\text{sign}(U_1), \ldots, \text{sign}(U_{LF})) \sim S \), and it is independent of \( (|U_1|, \ldots, |U_{LF}|) \).

Because \( U \sim (D' D)^{-1/2} I_{LF} \), and \( (D_1^2, D_{LF}^2) \) are independent and identically Gamma distributed with shape parameter 1/2 and rate parameter 1/2, we have

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

Therefore, \( U = (\text{sign}(U_1), \ldots, \text{sign}(U_{LF}))' \circ (|U_1|, \ldots, |U_{LF}|)' \sim S \circ \sqrt{\beta} \), which further implies \( \zeta_{LF,a} \sim \chi_{LF,a} \cdot S \circ \sqrt{\beta} \). \( \square \)
The following lemma helps to simplify the asymptotic sampling distributions of \( \hat{\tau} \) under the CRFE, ReFMT\(_F\) and ReFMT\(_{CF}\).

**Lemma A9.** Let \( \zeta_{m,a} \sim D \mid D' D \leq a \) be an \( m \) dimensional truncated Gaussian random vector, where \( D \sim N(0, I_m) \). If two matrices \( A \) and \( B \) in \( \mathbb{R}^{p \times m} \) satisfy \( AA' = BB' \), then \( AC_{m,a} \sim BC_{m,a} \).

**Proof of Lemma A9.** From Lemma A3, there exists an orthogonal matrix \( \Gamma \in \mathbb{R}^{m \times m} \) such that \( A = B \Gamma \). From Proposition 3, by the spherical symmetry of \( \zeta_{m,a} \), \( A\zeta_{m,a} \sim B\Gamma \zeta_{m,a} \sim B\zeta_{m,a} \).

**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 A4.** Let \( B \sim N(0, \Lambda) \). For any \( \Lambda > 0 \), the 0-1 function \( \kappa(\sqrt{n}\hat{\tau}_x, nV_{xx}) \equiv \kappa(\hat{\tau}_x, V_{xx}) \) satisfies

(a) \( \kappa \) is almost surely continuous,

(b) \( \text{Var}\{B \mid \kappa(B, \Lambda) = 1\} \), as a function of \( \Lambda \), is continuous,

(c) \( P\{\kappa(B, \Lambda) = 1\} > 0 \),

(d) \( \kappa(\mu, \Lambda) = \kappa(-\mu, \Lambda) \), for all \( \mu \).

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 A4.

**Lemma A10.** Let \( (\sqrt{n}\hat{\tau}', \sqrt{n}\hat{\tau}'_x) \) be a random vector following \( N(0, nV) \). Then as \( n \to \infty \), under Conditions 1 and A4, the two conditional distributions, \( \sqrt{n}(\hat{\tau}' - \tau', \hat{\tau}'_x) \mid \kappa(\hat{\tau}_x, V_{xx}) = 1 \) and \( \sqrt{n}(\hat{\tau}', \hat{\tau}'_x) \mid \kappa(\hat{\tau}_x, V_{xx}) = 1 \), converge weakly to the same distribution, i.e.,

\[
\begin{align*}
\left(\frac{\hat{\tau} - \tau}{\hat{\tau}_x}\right) \mid \kappa(\hat{\tau}_x, V_{xx}) = 1 & \; \sim \; \left(\frac{\hat{\tau}}{\hat{\tau}_x}\right) \mid \kappa(\hat{\tau}_x, V_{xx}) = 1.
\end{align*}
\]

**Proof of Lemma A10.** The proof of this lemma is almost identical to Li et al. [2018a, Proposition A1 and Corollary A1]. We omit it.

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

$$\hat{\tau} - \tau \mid \mathcal{M} \sim \hat{\tau} \mid \mathcal{M} \sim \hat{\tau} - \tau \mid \mathcal{M} \sim \begin{cases} 
\hat{\tau}' \mathbf{V}_{xx}^{-1} \mathbf{x} \leq a & \sim \hat{\tau}' + \mathbf{V}_{xx}^{-1/2} \mathbf{D} \mid \mathbf{D}' \mathbf{D} \leq a \\
\mathbf{V}_{xx}^{-1/2} \mathbf{D} \mid \mathbf{D}' \mathbf{D} \leq a & \sim \left( \mathbf{V}_{xx}^{-1/2} \right)^{1/2} \mathbf{\varepsilon} + \mathbf{V}_{xx}^{-1/2} \mathbf{\zeta}_{LF,a},
\end{cases}$$

where $\mathbf{\varepsilon} \sim \mathcal{N}(0, I_{LF})$ is independent of $\mathbf{\zeta}_{LF,a}$. From Theorem 1, $
\mathbf{V}_{\tau \tau} = \mathbf{V}_{\tau \tau} = \mathbf{V}_{\tau \tau} \mathbf{V}_{xx}^{-1/2} \mathbf{V}_{xx}^{-1/2} \mathbf{\zeta}_{LF,a} \sim (\mathbf{V}_{\tau \tau})_{LF,a}^{1/2} \mathbf{\zeta}_{LF,a}$. Therefore, Theorem 2 holds.

**Proof of Corollary 2.** Let $\mathbf{e}_f = (0, \ldots, 0, 1, 0, \ldots, 0)'$ be an $F$ dimensional unit vector with $f$th element one. From Theorem 2,

$$\hat{\tau}_f - \tau_f \mid \mathcal{M} \sim \mathbf{e}_f' \left( \mathbf{V}_{\tau \tau}^{-1/2} \right)_{LF,a}^{1/2} \mathbf{\varepsilon} + \mathbf{e}_f' \left( \mathbf{V}_{\tau \tau}^{-1/2} \right)_{LF,a}^{1/2} \mathbf{\zeta}_{LF,a} \sim \sqrt{\mathbf{e}_f' \mathbf{V}_{\tau \tau} \mathbf{e}_f - \mathbf{e}_f' \mathbf{V}_{\tau \tau} \mathbf{e}_f \cdot \mathbf{\eta}_{LF,a}} + \sqrt{\mathbf{e}_f' \mathbf{V}_{\tau \tau} \mathbf{e}_f \cdot \mathbf{\eta}_{LF,a}},$$

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

$$\hat{\tau}_f - \tau_f \mid \mathcal{M} \sim \sqrt{\mathbf{e}_f' \mathbf{V}_{\tau \tau} \mathbf{e}_f - \mathbf{e}_f' \mathbf{V}_{\tau \tau} \mathbf{e}_f \cdot \mathbf{\eta}_{LF,a}} + \sqrt{\mathbf{e}_f' \mathbf{V}_{\tau \tau} \mathbf{e}_f \cdot \mathbf{\eta}_{LF,a}}.$$

Therefore, Corollary 2 holds.

**Proof of Corollary 3.** Recall that $\mathbf{R} = \mathbf{V}_{\tau \tau}^{-1/2} \mathbf{V}_{\tau \tau} \mathbf{V}_{xx}^{-1/2} = \mathbf{\Pi}^2 \mathbf{\Gamma}'$ 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 $\mathbf{\Omega} = \mathbf{\Pi}^{-1} \mathbf{\Gamma}' \mathbf{V}_{\tau \tau}^{-1/2} \mathbf{V}_{xx} \mathbf{V}_{xx}^{-1/2}$. Then $\mathbf{\Omega} = \mathbf{\Pi}^{-1} \mathbf{\Gamma}' \mathbf{R} \mathbf{\Pi}^{-1} = \mathbf{I}_F$, and $\mathbf{V}_{\tau \tau}^{-1/2} \mathbf{V}_{xx} \mathbf{V}_{xx}^{-1/2} = \mathbf{\Gamma} \mathbf{\Pi} \mathbf{\Omega}$ is the singular value decomposition of $\mathbf{V}_{\tau \tau}^{-1/2} \mathbf{V}_{xx} \mathbf{V}_{xx}^{-1/2}$. 



Note that \( V_{\tau\tau}^{-1/2}V_{\tau\tau}^\perp V_{\tau\tau}^{-1/2} = I_F - R = \Gamma(I_F - \Pi^2)\Gamma' \). We can simplify the asymptotic distribution (4.1) as

\[
\hat{\tau} - \tau \mid M \sim \left(V_{\tau\tau}^\perp\right)^{1/2} \varepsilon + V_{\tau\tau}^\perp V_{\tau\tau}^{-1/2} \tilde{\zeta}_{LF,a}
\]

\[
= V_{\tau\tau}^{-1/2} \left(V_{\tau\tau}^\perp \right)^{1/2} \varepsilon + V_{\tau\tau}^{-1/2} V_{\tau\tau}^\perp V_{\tau\tau}^{-1/2} \tilde{\zeta}_{LF,a}
\]

\[
\sim V_{\tau\tau}^{-1/2} \left(V_{\tau\tau}^\perp \right)^{1/2} \varepsilon + V_{\tau\tau}^{-1/2} V_{\tau\tau}^\perp V_{\tau\tau}^{-1/2} \tilde{\zeta}_{LF,a}
\]

\[
= V_{\tau\tau}^{1/2} \left\{ \Gamma(I_F - \Pi^2)\Gamma' \right\}^{1/2} \varepsilon + \Gamma \Pi \Omega \tilde{\zeta}_{LF,a}
\]

\[
\sim V_{\tau\tau}^{1/2} \left\{ \Gamma(I_F - \Pi^2)\Gamma' \right\}^{1/2} \varepsilon + \Gamma \Pi \Omega \tilde{\zeta}_{LF,a}
\]
which, by the identical covariance structure between \((\hat{\tau}', \hat{\theta}'_x)\)' and \((\hat{\tau} - \tau', \hat{\theta}_x')\)', has the same covariance as the sampling covariance of \(\hat{\tau}_e\), the residual from the linear projection of \(\hat{\tau}\) on \(\theta_x\). Because \(\hat{\theta}_x\) and \(\hat{\tau}_x\) are linear transformations of each other, \(\hat{\tau}_e\) is the same as the residual from the linear projection of \(\hat{\tau}\) on \(\theta_x\). Thus, from Theorem 1, \(\text{Cov}(\hat{\tau}_e) = \text{Cov}(\hat{\tau}_x) = V_{\hat{\tau} \tau}^{1/2}\). Let \(\varepsilon = (V_{\hat{\tau} \tau}^{1/2})^{-1/2}\hat{\tau}_e \sim N(0, I_{F})\) be the standardization of \(\hat{\tau}_e\), and \(D_h = (W_{xx}[h])^{-1/2}\hat{\theta}_x[h]\) be the standardization of \(\hat{\theta}_x[h]\). We have

\[
\hat{\tau} - \tau | T_F
\]

\[
\sim \tilde{\varepsilon} \left\{ \hat{\theta}'_x[h] (W_{xx} [h])^{-1} \hat{\theta}_x[h] \leq a_h, h = 1, \ldots, H \right\}
\]

\[
\sim \hat{\tau}_e + \sum_{h=1}^{H} W_{\tau x}[h] (W_{xx} [h])^{-1} \hat{\theta}_x[h] \left\{ \hat{\theta}'_x[h] (W_{xx} [h])^{-1} \hat{\theta}_x[h] \leq a_h, 1 \leq h \leq H \right\}
\]

\[
\sim (V_{\hat{\tau} \tau}^{1/2})^{1/2} \varepsilon + \sum_{h=1}^{H} W_{\tau x}[h] (W_{xx} [h])^{-1/2} D_h \left\{ D'_h D_h \leq a_h, 1 \leq h \leq H \right\}
\]

\[
\sim (V_{\hat{\tau} \tau}^{1/2})^{1/2} \varepsilon + \sum_{h=1}^{H} W_{\tau x}[h] (W_{xx} [h])^{-1/2} \zeta_{LF_h, a_h}.
\]

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

**Proof of Corollary 5.** Recall that \(e_f\) 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

\[
\hat{\tau}_f - \tau_f | T_F
\]

\[
\sim e'_f \left( V_{\hat{\tau} \tau}^{1/2} \right)^{1/2} \varepsilon + \sum_{h=1}^{H} e'_f \left( W_{\hat{\tau} \tau}^{1/2} \right)^{1/2} _{LF_h} \zeta_{LF_h, a_h}
\]

\[
\sim \sqrt{e'_f V_{\hat{\tau} \tau}^{1/2} e_f \cdot \varepsilon_0 + \sum_{h=1}^{H} \sqrt{e'_f W_{\hat{\tau} \tau}^{1/2} e_f \cdot c_h \zeta_{LF_h, a_h}}}
\]

\[
\sim \sqrt{V_{\tau f \tau f}} \left( \sqrt{1 - V_{\tau f \tau f}^{1/2} / V_{\tau f \tau f}^{1/2} \cdot \varepsilon_0 + \sum_{h=1}^{H} \sqrt{W_{\tau f \tau f}^{1/2} [h] / V_{\tau f \tau f}^{1/2} \cdot c'_h \zeta_{LF_h, a_h}}} \right)
\]

where \(c'_h = (e'_f W_{\tau f}^{1/2} e_f)^{-1/2} e'_f \left( W_{\hat{\tau} \tau}^{1/2} \right)^{1/2} \) is a unit vector with length one. By the definitions of \(R^2_f\) and the \(\rho^2_f[h]\)'s, and the spherical symmetry
of the $\zeta_{LF_h,a_h}$'s from Proposition 3,

$$\hat{\tau}_f - \tau_f \mid T_F \sim \sqrt{V_{\hat{\tau}_f \tau_f}} \left( 1 - R^2_f \cdot \varepsilon_0 + \sum_{h=1}^H \sqrt{\rho^2_f[h]} \cdot \eta_{LF_h,a_h} \right).$$

Therefore, Corollary 5 holds. 

\[ \begin{align*}
\text{A3.5. Asymptotic sampling distribution of $\hat{\tau}$ under ReFMT}_{CF}. \\
\text{Proof of Theorem A1.} \quad \text{For each tier } S_j, \text{ let } \hat{\delta}'_e[j] \text{ be the concatenation of } \hat{\theta}[t|h] \text{ with } (t,h) \in S_j, \text{ and } \hat{\delta}'_e = (\hat{\delta}'_e[1], \ldots, \hat{\delta}'_e[J]). \text{ From Proposition A1, under the CRFE, } (\hat{\tau}' - \tau', \hat{\delta}'_e)^\prime \text{ has sampling mean zero and sampling covariance matrix}
\end{align*} \]

\[ \begin{pmatrix}
V_{\tau\tau} & U_{\tau e[1]} & \cdots & U_{\tau e[J]} \\
U_{e\tau[1]} & U_{ee[1]} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
U_{e\tau[J]} & 0 & \cdots & U_{ee[J]}
\end{pmatrix} \]  

\[ (A1) \]

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

$$\hat{\tau} - \tau \mid T_{CF} \sim \hat{\tau} \mid \left\{ \hat{\delta}'_e[j](U_{ee[j]})^{-1/2}\hat{\delta}_e[j] \leq a_j, j = 1, \ldots, J \right\}.$$

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

$$\hat{\tau} - \tau \mid T_{CF} \sim \left(V_{\hat{\tau}\hat{\tau}} \right)^{1/2} \varepsilon + \sum_{j=1}^J U_{\tau e[j]}(U_{ee[j]})^{-1/2} \zeta_{\lambda_j,a_j}.$$ 

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

\[ \begin{align*}
\text{Proof of Corollary A1.} \quad \text{Recall that } e_f \text{ is a unit vector with the } f\text{th coordinate being one. From Theorem A1, the asymptotic sampling distribu-}
\end{align*} \]
tion of $\hat{\tau}_f$ under ReFMT$_{\text{CF}}$ is

$$\hat{\tau}_f - \tau_f \mid T_{\text{CF}} 
\sim e'_f \left( V_{\tau\tau}^\perp \right)^{1/2} \varepsilon + \sum_{j=1}^J e'_j \left( U_{\tau\tau}^\| [j] \right)^{1/2} \zeta_{\lambda_j, a_j}
\sim \sqrt{e'_f V_{\tau\tau}^\perp} e_f \cdot \varepsilon_0 + \sum_{j=1}^J \sqrt{e'_j U_{\tau\tau}^\| [j]} e_f \cdot e'_j \zeta_{\lambda_j, a_j}
\sim \sqrt{V_{\tau\tau}^\perp} \left( \sqrt{1 - V_{\tau\tau}^\| / V_{\tau\tau}^\perp} \cdot \varepsilon_0 + \sum_{j=1}^J \sqrt{U_{\tau\tau}^\| [j]} / V_{\tau\tau}^\perp \cdot e'_j \zeta_{\lambda_j, a_j} \right),$$

where $e'_j = (e'_f U_{\tau\tau}^\| [j] e_f)^{-1/2} e'_f \left( U_{\tau\tau}^\| [j] \right)^{1/2}$ is a unit vector with length one.

By the definitions of $R_{\tau}^2$ and $\beta_{\tau}^2 [j]$'s, and the spherical symmetry of $\zeta_{\lambda_j, a_j}$'s from Proposition 3,

$$\hat{\tau}_f - \tau_f \mid T_{\text{CF}} \sim \sqrt{V_{\tau\tau}^\perp} \left( \sqrt{1 - R_{\tau}^2} \cdot \varepsilon_0 + \sum_{j=1}^J \sqrt{\beta_{\tau}^2 [j]} \cdot \eta_{\lambda_j, a_j} \right).$$

Therefore, Corollary A1 holds.

\[\square\]

**A4. Reduction in asymptotic sampling covariances.** We use $\text{Var}_a$ and $\text{Cov}_a$ 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, $\text{Cov}_a\{\sqrt{n}(\hat{\tau} - \tau)\} = \lim_{n \to \infty} n V_{\tau\tau}^\perp = \lim_{n \to \infty} n V_{\tau\tau}^\perp + \lim_{n \to \infty} n V_{\tau\tau}^\|$. For notational simplicity, we omit the limiting signs. From Theorem 2 and Proposition 3,

$$\text{Cov}_a\{\sqrt{n}(\hat{\tau} - \tau) \mid M\} = \text{Cov}\left\{\left(n V_{\tau\tau}^\perp\right)^{1/2} \varepsilon\right\} + n \left(V_{\tau\tau}^\|\right)^{1/2} \text{Cov} \left(\zeta_{\lambda_F, a}\right) \left(\left(V_{\tau\tau}^\|\right)^{1/2}\right)' = n V_{\tau\tau}^\perp + v_{\lambda_F, a} \cdot n \left(V_{\tau\tau}^\|\right)^{1/2} \left(\left(V_{\tau\tau}^\|\right)^{1/2}\right)' = n V_{\tau\tau}^\perp + v_{\lambda_F, a} \cdot n V_{\tau\tau}^\|.$$

Therefore, the reduction in asymptotic sampling covariance is $(1 - v_{\lambda_F, a}) n V_{\tau\tau}^\|.$
Second, we consider the PRIASV of $\hat{\tau}_f$. From Proposition 1 and Corollary 2, the asymptotic sampling variance of $\hat{\tau}_f$ are

$$\text{Var}_a \left\{ \sqrt{n} (\hat{\tau}_f - \tau_f) \right\} = n V_{\tau_f\tau_f},$$

$$\text{Var}_a \left\{ \sqrt{n} (\hat{\tau}_f - \tau_f) \mid M \right\} = n V_{\tau_f\tau_f} \left( 1 - R_f^2 + R_f^2 \cdot v_{LF,a} \right)$$

$$= n V_{\tau_f\tau_f} \left\{ 1 - (1 - v_{LF,a}) R_f^2 \right\}.$$ 

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

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

$$\text{Cov}_a \left\{ \sqrt{n} (\hat{\tau} - \tau) \mid T_F \right\}$$

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

$$= n V_{\tau\tau}^\perp + n \sum_{h=1}^H v_{LF_h,a_h} W_{\tau\tau}^\parallel [h].$$

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., $V_{\tau\tau}^\parallel = \sum_{h=1}^H W_{\tau\tau}^\parallel [h]$. Thus,

$$\text{Cov}_a \left\{ \sqrt{n} (\hat{\tau} - \tau) \right\} = n V_{\tau\tau}^\perp + n V_{\tau\tau}^\parallel = n \sum_{h=1}^H W_{\tau\tau}^\parallel [h].$$

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

$$\text{Cov}_a \left\{ \sqrt{n} (\hat{\tau} - \tau) \right\} - \text{Cov}_a \left\{ \sqrt{n} (\hat{\tau} - \tau) \mid T_F \right\} = n \sum_{h=1}^H (1 - v_{LF_h,a_h}) W_{\tau\tau}^\parallel [h].$$

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

$$\text{Var}_a \left\{ \sqrt{n} (\hat{\tau}_f - \tau_f) \right\} = n V_{\tau_f\tau_f},$$

$$\text{Var}_a \left\{ \sqrt{n} (\hat{\tau}_f - \tau_f) \mid T_F \right\} = n V_{\tau_f\tau_f} \left( 1 - R_f^2 + \sum_{h=1}^H \rho_f^2 [h] v_{LF_h,a_h} \right).$$
By definition of $\rho_f^2[h]$ and the fact that $V_{\tau\tau}^0 = \sum_{h=1}^H W_{\tau\tau}^0[h]$, we have $R_f^2 = \sum_{h=1}^H \rho_f^2[h]$. Therefore, the PRIASV of $\hat{\tau}_f$ under ReFMT is

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

**Proof of Theorem A2.** Note that the $\{\tilde{\theta}_{e[t]}[h]\}_{(t,h)\in S_j}$'s are from a block-wise Gram–Schmidt orthogonalization of $\tilde{\tau}_x$. 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 $\phi \in \mathbb{R}^m$ is more peaked than another random vector $\psi \in \mathbb{R}^m$ and write as $\phi \succ \psi$, if $P(\phi \in K) \geq P(\psi \in K)$ for every symmetric convex set $K \subset \mathbb{R}^m$.

**Lemma A11.** If two $m$ dimensional symmetric random vectors $\phi_1$ and $\phi_2$ satisfy $\phi_1 > \phi_2$, then for any matrix $C \in \mathbb{R}^{p \times m}$, $C\phi_1 > C\phi_2$.

**Lemma A12.** Let $\psi, \phi_1$ and $\phi_2$ be three independent $m$ dimensional symmetric random vectors. If $\psi$ is central symmetric unimodal and $\phi_1 > \phi_2$, then $\psi + \phi_1 > \psi + \phi_2$.  

**Proof of Lemmas A11 and A12.** See Dharmadhikari and Joag-Dev [1988, Lemma 7.2 and Theorem 7.5].

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 $\epsilon \sim \mathcal{N}(0, I_m)$ and $\zeta_{m,a} \sim \epsilon | \epsilon^T \epsilon \leq a$. Then $\zeta_{m,a} \succ \epsilon$.

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

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

where $r^{-1}K = \{r^{-1}x : x \in K\}$. For any $r_1 \geq r_2 \geq 0$, if $\tilde{x} \in r^{-1}_1K$, then $r_1\tilde{x} \in K$. By the symmetric convexity of $K$, we then have $r_2\tilde{x} = r_2/r_1 \cdot (r_1\tilde{x}) \in K$, i.e., $\tilde{x} \in r^{-1}_2K$. Thus, $r^{-1}_1K \subset r^{-1}_2K$, 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 $\zeta \in K$, respectively, as follows:

$$P(\zeta_{m,a} \in K) = P(\varepsilon \in K \mid \varepsilon \varepsilon \leq a) = E \{ P(\varepsilon \in K \mid \varepsilon \varepsilon \leq a, \|\varepsilon\|_2) \mid \varepsilon \varepsilon \leq a \}$$

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

$$P(\varepsilon \in K) = E \{ G(\chi_m) \} = P(\chi_m \leq a)E \{ G(\chi_m) \mid \chi_m^2 \leq a \} + P(\chi_m > a)E \{ G(\chi_m) \mid \chi_m^2 > a \} \leq P(\chi_m \leq a)E \{ G(\chi_m) \mid \chi_m^2 \leq a \} + P(\chi_m > a)G(\sqrt{a})$$

$$\leq P(\chi_m \leq a)E \{ G(\chi_m) \mid \chi_m^2 \leq a \} + P(\chi_m > a)E \{ G(\chi_m) \mid \chi_m^2 \leq a \} = E \{ G(\chi_m) \mid \chi_m^2 \leq a \} = P(\zeta_{m,a} \in K).$$

Therefore, Lemma A13 holds.

**Proof of Proposition 4.** From Lemma A11, for any vector $c \in \mathbb{R}^m$, $c'\phi \succ c'\psi$. Thus,

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

$$= \int_0^\infty \{ 1 - P(c'\phi \in [-t,t]) \} dt$$

$$\leq \int_0^\infty \{ 1 - P(c'\psi \in [-t,t]) \} = \text{Var}(c'\psi).$$

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

**A5.2. Propositions for simultaneous diagonalization.**
Proof of Proposition 8. Recall that $\Psi$ is the matrix such that $b_q = \Psi c_q$. By the construction of the $c_q$'s,

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

is a block diagonal matrix, and $\tilde{B} = \Psi \tilde{C} \Psi'$. Partition $\tilde{C}$ into $\tilde{C} = (\tilde{C}[1,1], \ldots, \tilde{C}[H,H])$.

From Proposition 6, under the additivity, we have

$$
W_{\tau x|h} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} (b_q c_q' |h|) \otimes S_{1x} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} (\Psi c_q c_q' |h|) \otimes S_{1x}
$$

$$
= \left( \Psi \tilde{C}[.,h] \right) \otimes S_{1x},
$$

$$
W_{xx|h} = 2^{-2(K-1)} \sum_{q=1}^{Q} n_q^{-1} (c_q |h| c_q' |h|) \otimes S_{xx} = \tilde{C}[h,h] \otimes S_{xx}.
$$

Thus, under the additivity, for each $1 \leq h \leq H$, $W^{|I|}_{\tau \tau|h}$ reduces to

$$
W^{|I|}_{\tau \tau|h} = W_{\tau x|h} W^{-1}_{xx|h} W_{xx|h} W^{I}_{\tau \tau|h}
$$

$$
= \left\{ \Psi \tilde{C}[.,h] \left( \tilde{C}[h,h] \right)^{-1} (\tilde{C}[.,h])' \Psi' \right\} \otimes (S_{1x} S^{-1}_{xx} S_{x,1})
$$

$$
= S_{11}^{I} \cdot \Psi \cdot \text{diag}(0, \ldots, 0, \tilde{C}[h,h], 0, \ldots, 0) \cdot \Psi'
$$

$$
= S_{11}^{I} \cdot \Psi \cdot \tilde{C}^{1/2} \cdot \text{diag}(0, \ldots, 0, I_{F_h}, 0, \ldots, 0) \cdot \tilde{C}^{1/2} \cdot \Psi'
$$

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

$$
V^{-1/2}_{\tau \tau} W^{|I|}_{\tau \tau|h} V^{-1/2}_{\tau \tau}
$$

$$
= S_{11}^{I} \cdot \tilde{B}^{-1/2} \left\{ S_{11}^{I} \cdot \Psi \cdot \tilde{C}^{1/2} \cdot \text{diag}(0, \ldots, 0, I_{F_h}, 0, \ldots, 0) \cdot \tilde{C}^{1/2} \cdot \Psi' \right\} \cdot S_{11}^{-1/2} \tilde{B}^{-1/2}
$$

$$
= S_{11}^{I} / S_{11} \cdot \tilde{B}^{-1/2} \Psi \tilde{C}^{1/2} \cdot \text{diag}(0, \ldots, 0, I_{F_h}, 0, \ldots, 0) \cdot \tilde{C}^{1/2} \Psi' \tilde{B}^{-1/2}
$$

$$
= \Gamma \cdot \text{diag}(0, \ldots, 0, S_{11}^{I} / S_{11} \cdot I_{F_h}, 0, \ldots, 0) \cdot \Gamma'
$$

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

$$
\Gamma \Gamma' = \tilde{B}^{-1/2} \Psi \tilde{C}^{1/2} \tilde{C}^{1/2} \Psi' \tilde{B}^{-1/2} = \tilde{B}^{-1/2} \left( \Psi \tilde{C} \Psi' \right) \tilde{B}^{-1/2} = \tilde{B}^{-1/2} \tilde{B} \tilde{B}^{-1/2} = I_F,
$$

we have that $\Gamma$ is an orthogonal matrix. Note that $\text{diag}(0, \ldots, 0, S_{11}^{I} / S_{11} \cdot I_{F_h}, 0, \ldots, 0)$ is a diagonal matrix with exactly $F_h$ nonzero elements, which are all equal to $S_{11}^{I} / S_{11}$. Therefore, Condition 2 and Proposition 8 hold. \(\Box\)
Proof of Proposition A2. Recall $\Psi$ is the matrix such that $b_q = \Psi c_q$. From the proof of Proposition 8, $\tilde{C} = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1}(c_q c_q')$ is a block diagonal matrix, and $\tilde{B} = \Psi \tilde{C} \Psi'$. From Proposition A1, under the additivity, $W_{\tau e[t]}[h]$ and $W_{e[t]e[t]}[h]$ reduce to

$$W_{\tau e[t]}[h] = 2^{-2(K-1)} \left\{ \Psi \sum_{q=1}^Q n_q^{-1}(c_q c_q'[h]) \right\} \otimes S_{1,e[t]} = \left( \Psi \tilde{C}[h] \right) \otimes S_{1,e[t]}$$

$$W_{e[t]e[t]}[h] = 2^{-2(K-1)} \sum_{q=1}^Q n_q^{-1}(c_q[h] c_q'[h]) \otimes S_{e[t]e[t]} = \tilde{C}[h, h] \otimes S_{e[t]e[t]},$$

which then implies

$$W_{e[t]e[t]}[h] = W_{\tau e[t]}[h] W_{e[t]e[t]}[h] W_{\tau e[t]}[h]^{-1}$$

$$= S_{1,e[t]} S_{e[t]}^{-1} S_{e[t]} S_{e[t]}^{-1} \Psi \cdot \text{diag}(0, \ldots, 0, \tilde{C}[h, h], 0, \ldots, 0) \cdot \Psi'$$

$$= S_{11}[t] \cdot \Psi \cdot \text{diag}(0, \ldots, 0, \tilde{C}[h, h], 0, \ldots, 0) \cdot \Psi'$$

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

$$V_{\tau\tau}^{-1/2} W_{e[t]e[t]}[h] V_{\tau\tau}^{-1/2}$$

$$= S_{11}[t] / S_{11} \cdot \tilde{B}^{-1/2} \left\{ \Psi \tilde{C}^{1/2} \cdot \text{diag}(0, \ldots, 0, I_{F_h}, 0, \ldots, 0) \cdot \tilde{C}^{1/2} \Psi' \right\} \tilde{B}^{-1/2}$$

$$= S_{11}[t] / S_{11} \cdot \tilde{B}^{-1/2} \Psi \tilde{C}^{1/2} \cdot \text{diag}(0, \ldots, 0, I_{F_h}, 0, \ldots, 0) \cdot \tilde{C}^{1/2} \Psi' \tilde{B}^{-1/2}$$

$$= \Gamma \Omega_{[t]}[h] \Gamma',$$
and $\Gamma = \hat{B}^{-1/2} \Psi \hat{C}^{1/2}$ is an orthogonal matrix. By definition, we then have

$$V^{-1/2}_\tau U^{\|}_\tau[j] V^{-1/2}_\tau = V^{-1/2}_\tau \left( \sum_{(t,h) \in S_j} W^{\|}_{e[t]e[t]}[h] \right) V^{-1/2}_\tau = \sum_{(t,h) \in S_j} V^{-1/2}_\tau W^{\|}_{e[t]e[t]}[h] V^{-1/2}_\tau$$

$$= \sum_{(t,h) \in S_j} \Gamma \Omega_{[t]}[h] \Gamma' = \Gamma \left( \sum_{(t,h) \in S_j} \Omega_{[t]}[h] \right) \Gamma', \quad (1 \leq j \leq J)$$

where $\sum_{(t,h) \in S_j} \Omega_{[t]}[h]$ is a diagonal matrix. Therefore, Condition A3 and Proposition A2 hold.

A5.3. Peakedness under ReFM.

Proof of Theorem 4. Let $D \sim N(0, I_{LF})$. From Lemma A13, $\zeta_{LF,a} \succ D$. From Lemma A11, $(V^{\|}_{\tau \tau})^{-1/2}_{LF} \zeta_{LF,a} \succ (V^{\|}_{\tau \tau})^{-1/2}_{LF} D$. From Proposition 2, $\varepsilon$ is central convex unimodal. From Lemma A6, $(V^{\perp}_{\tau \tau})^{-1/2}\varepsilon$ is also central convex unimodal. Then, from Lemma A12,

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

Therefore, Theorem 4 holds.

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

Lemma A14. Let $\varepsilon, \eta \sim N(0, 1)$ 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. [2018a, Lemma A3], and is also a special case of Das Gupta et al. [1972, Theorem 2.1].

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

Lemma A15. Let $\varepsilon$ and $\eta$ be two independent $m$ dimensional standard Gaussian random vectors, $(\rho_1, \ldots, \rho_m)$ be $m$ constants in $[0, 1]$, and $\Delta$ be a diagonal matrix with diagonal elements $(\rho_1, \ldots, \rho_m)$. For any $r \geq 0$, the probability

$$P \left\{ (I_m - \Delta^2)^{-1/2} \varepsilon + \Delta \eta \in B_m(r) \mid \eta' \eta \leq a \right\}$$
is nondecreasing in \((\rho_1, \ldots, \rho_m)\). Specifically, for any constants \((\rho_1, \ldots, \rho_m)\) and \((\tilde{\rho}_1, \ldots, \tilde{\rho}_m)\) in \([0, 1]\), if \(\rho_j \geq \tilde{\rho}_j\) for \(1 \leq j \leq m\), then for any \(r \geq 0\),

\[
P \left\{ (I_F - \Delta^2)^{1/2} \varepsilon + \Delta \eta \in B_m(r) \mid \eta' \eta \leq a \right\}
\]

(A1)

\[
\geq P \left\{ (I_F - \Delta^2)^{1/2} \varepsilon + \Delta \eta \in B_m(r) \mid \eta' \eta \leq a \right\}.
\]

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

**Proof of Lemma A15.** To prove Lemma A15, it suffices to prove that for any constants \((\rho_1, \ldots, \rho_m)\) and \((\tilde{\rho}_1, \ldots, \tilde{\rho}_m)\) 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 \(\eta \sim \mathcal{N}(0, I_m)\) independent of \(\varepsilon\), and \(\varepsilon_1 = (\varepsilon_2, \ldots, \varepsilon_m)\) and \(\eta_1 = (\eta_2, \ldots, \eta_m)\) be the subvectors of \(\varepsilon\) and \(\eta\) excluding the first coordinates. Define

\[
B(r, \varepsilon_{-1}, \eta_{-1}, \rho_2, \ldots, \rho_m) = \left\{ x : \begin{pmatrix} x \\ \sqrt{1 - \rho_1^2} \cdot \varepsilon_2 + \rho_2 \eta_2 \\ \vdots \\ \sqrt{1 - \rho_m^2} \cdot \varepsilon_m + \rho_m \eta_m \end{pmatrix} \in B_m(r) \right\} \subset \mathbb{R},
\]

which is either an empty set or a symmetric closed interval on the real line. For any \(r \geq 0\), \(\varepsilon_{-1}\) and \(\eta_{-1}\),

\[
P \left\{ (I_F - \Delta^2)^{1/2} \varepsilon + \Delta \eta \in B_m(r) \mid \eta' \eta \leq a, \varepsilon_{-1}, \eta_{-1} \right\}
\]

\[
= P \left\{ \sqrt{1 - \rho_1^2} \cdot \varepsilon_1 + \rho_1 \cdot \eta_1 \in B(r, \varepsilon_{-1}, \eta_{-1}, \rho_2, \ldots, \rho_m) \mid \eta_1^2 \leq a - \eta_{-1}' \eta_{-1}, \varepsilon_{-1}, \eta_{-1} \right\}
\]

\[
\geq P \left\{ \sqrt{1 - \tilde{\rho}_1^2} \cdot \varepsilon_1 + \tilde{\rho}_1 \cdot \eta_1 \in B(r, \varepsilon_{-1}, \eta_{-1}, \tilde{\rho}_2, \ldots, \tilde{\rho}_m) \mid \eta_1^2 \leq a - \eta_{-1}' \eta_{-1}, \varepsilon_{-1}, \eta_{-1} \right\}
\]

(A2)

\[
= P \left\{ (I_F - \tilde{\Delta}^2)^{1/2} \varepsilon + \tilde{\Delta} \eta \in B_m(r) \mid \eta' \eta \leq a, \varepsilon_{-1}, \eta_{-1} \right\},
\]

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

**Lemma A16.** Let \(B_0\) and \(B_1\) be two \(m \times m\) positive semi-definite matrices, \(\varepsilon \sim \mathcal{N}(0, I_m)\) be a standard Gaussian random vector, and \(\zeta_{p,a} \sim D \mid D' D \leq a\) be a truncated Gaussian random vector, where \(p \geq m\) and \(a \geq 0\). Define \(\phi \sim (B_0)^{1/2} \varepsilon + (B_1)^{1/2} \zeta_{p,a}\). If \(B = B_0 + B_1\) is positive definite, then for any \(\alpha \in (0, 1)\), the threshold \(c_{1-\alpha}\) for \(1 - \alpha\) quantile region \(\{\mu : \mu' B^{-1} \mu \leq c_{1-\alpha}\}\) of \(\phi\) depends only on \((m, p, a)\) and the eigenvalues of \(B^{-1/2} B_1 B^{-1/2}\), and is nonincreasing in these eigenvalues.
Proof of Lemma A16. Let \( B^{-1/2}B_1B^{-1/2} = \Gamma \Delta^2 \Gamma' \) be the eigendecomposition of \( B^{-1/2}B_1B^{-1/2} \), where \( \Gamma \in \mathbb{R}^{m \times m} \) is an orthogonal matrix, and \( \Delta^2 = \text{diag}(\rho_1^2, \ldots, \rho_m^2) \) is a diagonal matrix. From Lemma A9,

\[
\Gamma' B^{-1/2} \phi \sim \Gamma' B^{-1/2} (B_0)^{1/2} \xi + \Gamma' B^{-1/2} (B_1)^{1/2} \zeta_{p,a} \\
\sim \left( \Gamma' B^{-1/2} B_0 B^{-1/2} \Gamma \right)^{1/2} \xi + \left( \Gamma' B^{-1/2} B_1 B^{-1/2} \Gamma \right)^{1/2} \zeta_{p,a} \\
\sim \left\{ \Gamma' (I_m - \Gamma \Delta^2 \Gamma') \Gamma \right\}^{1/2} \xi + \left( \Gamma' \Gamma \Delta^2 \Gamma' \Gamma \right)^{1/2} \zeta_{p,a} \\
\sim \left( I_m - \Delta^2 \right)^{1/2} \xi + \left( \Delta^2 \right)^{1/2} \zeta_{p,a} \\
\sim \left( I_m - \Delta^2 \right)^{1/2} \xi + \left( \Delta, 0_{m \times (p-m)} \right) \zeta_{p,a}
\]

Let \( \eta \sim \mathcal{N}(0, I_m) \), \( \xi \sim \mathcal{N}(0, I_{p-m}) \), and \( (\eta, \xi) \) be independent. From the definition of \( \zeta_{p,a} \), for any \( r^2 \geq 0 \),

\[
P\left( \phi' B^{-1} \phi \leq r^2 \right) \\
= P \left\{ \Gamma' B^{-1/2} \phi \in B_m(r) \right\} \\
= P \left\{ (I_m - \Delta^2)^{1/2} \xi + \left( \Delta, 0_{m \times (p-m)} \right) \zeta_{p,a} \in B_m(r) \right\} \\
= P \left\{ (I_F - \Delta^2)^{1/2} \xi + \left( \Delta, 0 \right) \left( \begin{array}{c} \eta \\ \xi \end{array} \right) \in B_m(r) \left| \eta' \eta + \xi' \xi \leq a \right. \right\} \\
= P \left\{ (I_m - \Delta^2)^{1/2} \xi + \Delta \eta \in B_m(r) \left| \eta' \eta + \xi' \xi \leq a \right. \right\} \\
= \mathbb{E} \left[ P \left\{ (I_m - \Delta^2)^{1/2} \xi + \Delta \eta \in B_m(r) \left| \eta' \eta \leq a - \xi' \xi, \xi \right. \right. \left| \eta' \eta + \xi' \xi \leq a \right. \right] .
\]

From Lemma A15, for any given \( \xi \), the conditional probability,

\[
P \left\{ (I_m - \Delta^2)^{1/2} \xi + \Delta \eta \in B_m(r) \left| \eta' \eta \leq a - \xi' \xi, \xi \right. \right. \leq a \right. \right. \}
\]

is nondecreasing in the diagonal elements of \( \Delta^2 \). Therefore, the quantity in (A3) depends only on \((m, p, a)\) and the eigenvalues of \( B^{-1/2}B_1B^{-1/2} \), 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,

$$C(\hat{\tau} - \tau) | M \sim C \left( V_{\tau \tau}^\perp \right)^{1/2} \varepsilon + C \left( V_{\tau \tau}^\parallel \right)^{1/2}_{LF} \zeta_{LF, a}$$

$$\sim \left( CV_{\tau \tau}^\perp C' \right)^{1/2} \xi + \left( CV_{\tau \tau}^\parallel C' \right)^{1/2}_{LF} \zeta_{LF, a},$$

where $\xi \sim \mathcal{N}(0, I_p)$. 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 $\{ \mu : \mu'(CV_{\tau \tau} C')^{-1} \mu \leq c_{1-\alpha} \}$ and the canonical correlation between $C \hat{\tau}$ and $\hat{\tau}_x$, and is nonincreasing in these canonical correlations.

**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. \qed

### A5.4. Peakedness under ReFMT

**Proof of Theorem 9.** For each $1 \leq h \leq H$, let $D_h \sim \mathcal{N}(0, I_{LF_h})$, and $(\varepsilon, D_1, \ldots, D_H)$ be jointly independent. From Proposition 2, both $\varepsilon$ and $\zeta_{LF, a}$ are central convex unimodal. From Lemma A6, both $(V_{\tau \tau}^\perp)^{1/2} \varepsilon$ and $(W_{\tau \tau}^\parallel [h])_{LF, a}$ are central convex unimodal. Thus, by Lemma A7,

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

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

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

$$\succ \left( V_{\tau \tau}^\perp \right)^{1/2} \varepsilon + \left( W_{\tau \tau}^\parallel [1] \right)_{LF_1} D_1 + \sum_{h=2}^H (W_{\tau \tau}^\parallel [h])_{LF, a} \zeta_{LF, a}$$

$$\sim \left( V_{\tau \tau}^\perp + W_{\tau \tau}^\parallel [1] \right)^{1/2} \varepsilon + \sum_{h=2}^H (W_{\tau \tau}^\parallel [h])_{LF, a} \zeta_{LF, a}.$$
Because \((V_{\tau \tau}^\perp + W_{\tau \tau}^\parallel [1])^{1/2} e + \sum_{h=3}^{H} (W_{\tau \tau}^\parallel [h])^{1/2} \zeta_{LFh,a_t} \) is central convex unimodal, and \((W_{\tau \tau}^\parallel [2])^{1/2} \zeta_{LF2,a_2} \succ (W_{\tau \tau}^\parallel [2])^{1/2} D_2\), we have

\[
\left( V_{\tau \tau}^\perp \right)^{1/2} e + \sum_{h=1}^{H} (W_{\tau \tau}^\parallel [h])^{1/2} \zeta_{LFh,a_t} \succ \left( V_{\tau \tau}^\perp + \sum_{h=1}^{H} W_{\tau \tau}^\parallel [h] \right)^{1/2} e \sim \mathcal{N}(0, V_{\tau \tau}),
\]

where the last formula follows from \(V_{\tau \tau}^\parallel = \sum_{h=1}^{H} W_{\tau \tau}^\parallel [h] \) in the proof of Theorem 8.

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

**Lemma A17.** Let \(\varepsilon_0 \sim \mathcal{N}(0, 1)\), \(\eta_{kt,a_t} \sim D_{t1} \mid D_t \) where \(D_t = (D_{t1,1}, \ldots, D_{t1,k_t}) \sim \mathcal{N}(0, I_{k_t})\), \(a_t\) is a nonnegative constant that can be infinity, and \((\varepsilon_0, \eta_{k1,a1}, \eta_{k2,a2}, \ldots, \eta_{kT,aT})\) 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 \(c \geq 0\),

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

**Proof of Lemma A17.** It follows from Li et al. [2018a, Lemma A10].

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 A5. 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 $\Delta_h$ for $1 \leq h \leq H$;

(b) the elements of $\Delta_h$ 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 $(\varepsilon, \zeta_{p_h,a_h}, \ldots, \eta_{p_h,a_h})$ be $H + 1$ independent random vectors, where $\varepsilon \sim N(0, I_m)$, $\zeta_{p_h,a_h} \sim D_h$ and $D_h \sim \mathcal{N}(0, I_{p_h})$. Assume $\{\Delta_h \in \mathbb{R}^{m \times p_h}\}_{h=1}^{H}$ satisfy Condition A5. For any $r \geq 0$, $1 \leq h \leq H$, and $1 \leq k \leq m$ the probability

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

is nondecreasing in the nonzero elements of the $\Delta_h$'s, that is, for any two sets of matrices $\Delta_h$'s and $\tilde{\Delta}_h$'s satisfying Condition A5 with the positions of possible nonzero elements being the same, if all elements of $\Delta_h$'s are larger than or equal to $\tilde{\Delta}_h$'s, then

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

(A4)

PROOF OF LEMMA A18. It suffices to prove that for any two sets of matrices $\Delta_h$'s and $\tilde{\Delta}_h$'s satisfying Condition A5 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) $\Delta_h$ and $\tilde{\Delta}_h$ differ only in the $k$th row, and (b) the elements in the $k$th row of $\Delta_h$ 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 $\Delta_h$'s and $\tilde{\Delta}_h$'s are all in the first columns. This is because permuting the columns of $\Delta_h$ will not change the distribution of $\Delta_h \zeta_{p_h,a_h}$, a fact implied by the spherical symmetry of $\zeta_{p_h,a_h}$.

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

\[
B(r, \varepsilon_{-1}, D_{h,-1}, \Delta_{h,-1}) = \left\{ x : \left( I_{m-1} - \sum_{h=1}^{H} \Delta_{h,-1} \Delta'_{h,-1} \right)^{1/2} x \varepsilon_{-1} + \sum_{h=1}^{H} \Delta_{h,-1} D_{h,-1} \right\} \in B_m(r),
\]

which depends on \( \varepsilon_{-1}, D_{h,-1} \) 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 \( (\varepsilon_{-1}, D_{1,-1}, \ldots, D_{H,-1}) \),

\[
P \left\{ \left( I_m - \sum_{h=1}^{H} \Delta_h \Delta_h' \right)^{1/2} x + \sum_{h=1}^{H} \Delta_h D_h \in B_m(r) \mid D_h D_h \leq a_h, \varepsilon_{-1}, D_{h,-1}, \forall h \right\}
\]

\[
= P \left\{ \left( 1 - \sum_{h=1}^{H} \Delta^2_{h1} \right)^{1/2} x_{\varepsilon_{-1} h} + \sum_{h=1}^{H} \tilde{\Delta}_{h1} D_{h1} \in B(r, \varepsilon_{-1}, D_{h1,-1}, \Delta_{h1,-1}) \mid D^2_{h1} \leq a_h, D^2_{h1,-1} D_{h1,-1}, \varepsilon_{-1}, D_{h1,-1} \right\}
\]

\[
\geq P \left\{ \left( 1 - \sum_{h=1}^{H} \tilde{\Delta}^2_{h1} \right)^{1/2} x_{\varepsilon_{-1} h} + \sum_{h=1}^{H} \tilde{\Delta}_{h1} D_{h1} \in B(r, \varepsilon_{-1}, D_{h1,-1}, \tilde{\Delta}_{h1,-1}) \mid D^2_{h1} \leq a_h, D^2_{h1,-1} D_{h1,-1}, \varepsilon_{-1}, D_{h1,-1} \right\}
\]

(A5)

\[
= P \left\{ \left( I_m - \sum_{h=1}^{H} \tilde{\Delta}_h \tilde{\Delta}_h' \right)^{1/2} x + \sum_{h=1}^{H} \tilde{\Delta}_h D_h \in B_m(r) \mid D_h D_h \leq a_h, \varepsilon_{-1}, D_{h,-1}, \forall h \right\},
\]

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

The following lemma extends Lemma A16 to general case with \( H \geq 1 \). Moreover, even when \( H = 1 \), Lemma A19 is still more general then Lemma A16 by only requiring \( p_1 \geq \text{rank}(B_1) \), instead of \( p_1 \geq m \) in Lemma A16.

**Lemma A19.** Let \( (B_0, B_1, \ldots, B_H) \) be \( H + 1 \) positive semi-definite matrices in \( \mathbb{R}^{m \times m} \) with ranks \( (\gamma_0, \gamma_1, \ldots, \gamma_H) \), \( \varepsilon \sim \mathcal{N}(0, I_m) \) be a standard Gaussian random vector, and \( \zeta_{ph,a} \sim D_h \mid D_h D_h \leq a_h \) be a truncated Gaussian random vector, where \( D_h \sim \mathcal{N}(0, I_{ph}) \) and \( p_h \geq \gamma_h \) (\( h = 1, 2, \ldots, H \)).

Define \( \phi \sim (B_0)^{1/2} \varepsilon + \sum_{h=1}^{H} (B_h)^{1/2} \zeta_{ph,a} \). If \( B = \sum_{h=0}^{H} B_h \) is invertible, and there exists an orthogonal matrix \( \Gamma \) such that for all \( 1 \leq h \leq H \), \( \Gamma^T B^{-1/2} B_h B^{-1/2} \Gamma = \Omega_h^2 \) is a diagonal matrix, then the threshold \( c_{1-\alpha} \) for \( 1 - \alpha \) quantile region \( \{ \mu : \mu^T B^{-1/2} B^{-1/2} \mu \leq c_{1-\alpha} \} \) of \( \phi \) depends only on \( m \), the \( p_h \)’s, the \( a_h \)’s, and the eigenvalues of \( B^{-1/2} B_h B^{-1/2} \), and is nonincreasing in these eigenvalues.
Proof of Lemma A19. By definition,
\[ \Gamma' B^{-1/2} B_0 B^{-1/2} \Gamma = \Gamma' \left( I_m - \sum_{h=1}^{H} B^{-1/2} B_h B^{-1/2} \right) \Gamma = I_m - \sum_{h=1}^{H} \Omega_h^2. \]
Thus, from Lemma A9, the distribution of \( \phi \) satisfies
\[ \Gamma' B^{-1/2} \phi \sim \Gamma' B^{-1/2} \left\{ (B_0)^{1/2} \varepsilon + \sum_{h=1}^{H} (B_h)^{1/2} \xi_{p_h, a_h} \right\} \]
\[ \sim \left( \Gamma' B^{-1/2} B_0 B^{-1/2} \Gamma \right)^{1/2} \varepsilon + \sum_{h=1}^{H} \left( \Gamma' B^{-1/2} B_h B^{-1/2} \Gamma \right)^{1/2} \xi_{p_h, a_h} \]
\[ \sim \left( I_m - \sum_{h=1}^{H} \Omega_h^2 \right)^{1/2} \varepsilon + \sum_{h=1}^{H} \left( \Omega_h^2 \right)^{1/2} \xi_{p_h, a_h}. \]
For each \( 1 \leq h \leq H \), if \( p_h \geq m \), we further define \( \tilde{\Omega}_h = (\Omega_h, 0_{m \times (p_h - m)}) \); otherwise, \( p_h < m \), \( B_h \) has rank at most \( p_h \) and thus \( B^{-1/2} B_h B^{-1/2} \) has at most \( p_h \) nonzero eigenvalues, we further define \( \tilde{\Omega}_h = \Omega_h \left[ I_{h} \right] \) as the submatrix of \( \Omega_h \) consisting of the \( |I_h| = p_h \) columns with possible nonzero eigenvalues. Thus, by the construction of \( \Omega_h \)'s, we have \( \tilde{\Omega}_h \in \mathbb{R}^{m \times p_h} \) and \( \tilde{\Omega}_h \Gamma_{\tilde{h}} = \Omega_h^{(2)} \), i.e., \( \tilde{\Omega}_h = (\Omega_h^{(2)})^{1/2} \). Therefore, we can further simply the distribution of \( \Gamma' B^{-1/2} \phi \) as
\[ \Gamma' B^{-1/2} \phi \sim \left( I_m - \sum_{h=1}^{H} \tilde{\Omega}_h \tilde{\Omega}_h' \right)^{1/2} \varepsilon + \sum_{h=1}^{H} \tilde{\Omega}_h \xi_{p_h, a_h}. \]
For any \( r \geq 0 \), we have
\[ P(\phi' B^{-1} \phi \leq r^2) = P \left\{ \Gamma' B^{-1/2} \phi \in B_m(r) \right\} \]
\[ = P \left\{ \left( I_m - \sum_{h=1}^{H} \Omega_h^2 \right)^{1/2} \varepsilon + \sum_{h=1}^{H} \left( \Omega_h^2 \right)^{1/2} \xi_{p_h, a_h} \in B_m(r) \right\} \]
\[ = P \left\{ \left( I_m - \sum_{h=1}^{H} \tilde{\Omega}_h \tilde{\Omega}_h' \right)^{1/2} \varepsilon + \sum_{h=1}^{H} \tilde{\Omega}_h \xi_{p_h, a_h} \in B_m(r) \right\}. \]
Because the matrices \( \tilde{\Omega}_h \)'s satisfy Condition A5, from Lemma A18, the quantity in (A6) depends only on \( m \), \( p_h \)'s, \( a_h \)'s, and the eigenvalues of \( B^{-1/2} B_h B^{-1/2} \)'s, and is nondecreasing in these eigenvalues. Therefore, Lemma A19 holds. \( \square \)
Proof of Theorem 10. It follows directly from Lemma A19.

Comment on linear transformations of \( \hat{\tau} \) under ReFMT\(_F\). From Theorem 9 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 ReFMT\(_F\) is more peaked than that under the CRFE. From Theorem 7 and Lemma A9,

\[
C(\hat{\tau} - \tau) | T_F \sim C\left(V_{\tau \tau}^{1/2}\right) \varepsilon + \sum_{h=1}^{H} C\left(W_{\tau \tau}^{\|}[h]\right)_{LF_h} \xi_{LF_h,a_h}
\]

\[
\sim \left(CV_{\tau \tau}^{1/2}C'\right) \xi + \sum_{h=1}^{H} \left(CW_{\tau \tau}^{\|}[h]C'\right)_{LF_h} \xi_{LF_h,a_h},
\]

where \( \xi \sim \mathcal{N}(0, I_p) \). Based on Lemma A19, we can know that if the condition in Lemma A19 holds for \( B_h = CW_{\tau \tau}^{\|}[h]C' \), 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 \( p, LF_h \)’s, \( a_h \)’s, and the canonical correlation between \( C\hat{\tau} \) and \( \hat{\theta}_x[h] \)’s, and is nonincreasing in these canonical correlations.

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\(_F\).

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 \((Y_i(r),Y_i(k)),(Y_i(r),x_{il}),(x_{il},x_{im})\),

\[s_{AB}(q) - S_{AB} = o_p(1), \quad (q = 1, 2, \ldots, Q)\]

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. [2018a]. We omit it. \qed

**Lemma A21.** Under either ReFM, ReFMT\(_F\), or ReFMT\(_CF\), if Condition 1 holds, then for each \(1 \leq q \leq Q\),

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

**Proof of Lemma A21.** From Lemma A20, \(s_{\perp qq}, s_{q,x}\) and \(s_{xx}(q)\) are consistent for \(S_{\perp qq}, S_{q,x}\) and \(S_{xx}\), respectively. Because \(s_{\perp qq} = s_{qq} - s_{q,x} S_{xx}^{-1}(q) S_{x,q}\) and \(S_{\perp qq} = S_{qq} - S_{q,x} S_{xx}^{-1} S_{x,q}\), we know that \(s_{\perp qq}\) is also consistent for \(S_{\perp qq}\), and therefore Lemma A21 holds. \qed

Define \(\tilde{V}_{\perp \tau \tau} \equiv 2^{-2(K-1)} \sum_{q=1}^{Q} q^{-1} b_q b_q' \cdot S_{\perp \tau \tau} \geq V_{\perp \tau \tau}\). Under ReFM, from Lemma A21, \(V_{\perp \tau \tau}\) is consistent for \(\tilde{V}_{\perp \tau \tau}\), and \(V_{\tau \tau} V_{xx}^{-1/2}\) is consistent for \(V_{\tau \tau} V_{xx}^{-1/2}\). Therefore, the sampling covariance estimator is asymptotically conservative. Under ReFMT\(_F\) or ReFMT\(_CF\), \(V_{\perp \tau \tau}\) is also consistent for \(\tilde{V}_{\perp \tau \tau}\), and the estimated coefficients of the \(\zeta_{LFh,a_h}'s\) or \(\zeta_{\lambda_j,a_j}'s\) 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 \(V_1 \) and \(V_2\) be two positive semi-definite matrices in \(\mathbb{R}^{m \times m}\) satisfying that \(V_1 \preceq V_2\), and \(\varepsilon_1\) and \(\varepsilon_2\) be two Gaussian random vectors with mean zero and covariance matrices \(V_1\) and \(V_2\). Then \(\varepsilon_1 \succeq \varepsilon_2\).

**Proof of Lemma A22.** Let \(\varepsilon_3 \sim \mathcal{N}(0, V_2 - V_1)\) be independent of \(\varepsilon_1\). From Proposition 2 and Lemma A6, \(\varepsilon_1\) is central convex unimodal. Because \(0 \succeq \varepsilon_3\) and \(\varepsilon_3 + \varepsilon_1 \sim \varepsilon_2\), from Lemma A12, \(\varepsilon_1 \sim 0 + \varepsilon_1 \succeq \varepsilon_3 + \varepsilon_1 \sim \varepsilon_2\). Lemma A22 holds. \qed

**Proof of Theorem 6.** The proof of the asymptotic conservativeness of covariance estimator for \(\hat{C}\tau\) follows directly from the discussion in A6.1, and thus we consider only the asymptotic conservativeness of confidence sets for \(\hat{C}\tau\) here. Let \(\mathcal{L}, \tilde{\mathcal{L}}\) and \(\hat{\mathcal{L}}\) be three \(F\) dimensional random vectors following the asymptotic sampling distribution of \(\hat{\tau} - \tau\) under ReFM, the asymptotic sampling distribution with \(\hat{V}_{\perp \tau \tau}\) replaced by \(\tilde{V}_{\perp \tau \tau}\), and the esti-
mated asymptotic sampling distribution:
\[
\mathcal{L} \sim \left( V_{\tau \tau} \right)^{1/2} \varepsilon + \left( V_{\tau \tau} \right)^{1/2} \tilde{\zeta}_{LF,a} \sim \left( V_{\tau \tau} \right)^{1/2} \varepsilon + V_{\tau \tau} V_{xx}^{-1/2} \tilde{\zeta}_{LF,a},
\]
\[
\tilde{\mathcal{L}} \sim \left( V_{\tau \tau} \right)^{1/2} \varepsilon + \left( V_{\tau \tau} \right)^{1/2} \zeta_{LF,a} \sim \left( V_{\tau \tau} \right)^{1/2} \varepsilon + V_{\tau \tau} V_{xx}^{-1/2} \zeta_{LF,a},
\]
\[
\hat{\mathcal{L}} \sim \left( V_{\tau \tau} \right)^{1/2} \varepsilon + \hat{V}_{\tau \tau} V_{xx}^{-1/2} \zeta_{LF,a}.
\]

Under Condition 1, from the discussion in Section A6.1, by Slutsky’s theorem, \( \hat{\mathcal{L}} \sim \hat{\mathcal{L}} \). Note that \( (V_{\tau \tau})^{1/2} \varepsilon \sim (V_{\tau \tau})^{1/2} \varepsilon \) from Lemma A22, and \( V_{\tau \tau} V_{xx}^{-1/2} \zeta_{LF,a} \) is central convex unimodal from Proposition 2 and Lemma A6. From Lemma A12, \( \mathcal{L} \supset \hat{\mathcal{L}} \). Above all, \( \mathcal{L} \supset \hat{\mathcal{L}} \sim \hat{\mathcal{L}} \).

From Slutsky’s theorem, \( (C \tilde{V}_{\tau \tau} C')^{-1/2} C \tilde{\mathcal{L}} \sim (C \tilde{V}_{\tau \tau} C')^{-1/2} C \hat{\mathcal{L}} \). From the continuous mapping theorem, \( (C \mathcal{L})'(C \tilde{V}_{\tau \tau} C')^{-1} C \mathcal{L} \sim (C \mathcal{L})'(C \tilde{V}_{\tau \tau} C')^{-1} C \hat{\mathcal{L}} \).

Thus, the \( 1 - \alpha \) quantile of \( (C \mathcal{L})'(C \tilde{V}_{\tau \tau} C')^{-1} C \mathcal{L}, \hat{c}_{1-\alpha}, \) is consistent for the \( 1 - \alpha \) quantile of \( (C \mathcal{L})'(C \tilde{V}_{\tau \tau} C')^{-1} C \mathcal{L}, \tilde{c}_{1-\alpha} \). Because \( \mathcal{L} \supset \hat{\mathcal{L}} \), and the set of form \( \{ \mu : (C \mu)'(C \tilde{V}_{\tau \tau} C')^{-1} C \mu \leq c \} \) is symmetric convex, \( c_{1-\alpha} \), the \( 1 - \alpha \) quantile of \( (C \mathcal{L})'(C \tilde{V}_{\tau \tau} C')^{-1} C \mathcal{L} \), is smaller than or equal to \( \tilde{c}_{1-\alpha} \).

Above all, \( \tilde{c}_{1-\alpha} \) consistent for \( \tilde{c}_{1-\alpha} \leq c_{1-\alpha} \). Therefore, the \( 1 - \alpha \) confidence set for \( \tau \) is asymptotically conservative. When \( S_{\gamma \gamma} = o(1) \), we have \( V_{\tau \tau} - \tilde{V}_{\tau \tau} = o(1) \), which implies \( \mathcal{L} \sim \hat{\mathcal{L}} \). Thus, \( \tilde{c}_{1-\alpha} = c_{1-\alpha} + o(1) \), and the \( 1 - \alpha \) confidence set for \( \tau \) becomes asymptotically exact.

Note that both \( \sum_{h=1}^{H} W_{\tau \tau}[h](W_{xx}[h])^{-1/2} \zeta_{LF,a,h} \) and \( \sum_{j=1}^{j} U_{\tau \tau}[j](U_{ee}[j])^{-1/2} \zeta_{\lambda,j,a,j} \) are central convex unimodal. The proofs for the asymptotic conservativeness of symmetric convex confidence sets of form \( C \mathcal{L} + O(C \tilde{V}_{\tau \tau} C', c) \) for \( C \tau \) under ReFMTp and ReFMTCF are almost the same as ReFM. Thus we omit the proofs of Theorems 11 and A5.

Moreover, under for ReF, ReFMTp, or ReFMTCF, we consider \( 1 - \alpha \) confidence set \( C \mathcal{L} + \hat{\mathcal{L}} \) for \( C \tau \), where \( \hat{\mathcal{L}} \) can depend on \( (V_{\tau \tau}, \tilde{V}_{\tau \tau}, V_{xx}) \) and satisfies that \( P(\mathcal{L} \in \hat{\mathcal{L}}) = 1 - \alpha \). If \( \hat{\mathcal{L}} \) is a symmetric convex set, then the confidence set is generally asymptotically conservative, and the proof is similar to the proof of Theorem 6.
Department of Statistical Science
Temple University
Philadelphia, Pennsylvania 19122
USA
E-mail: dbrubin@me.com