

Covariate adjustment in multiarmed, possibly factorial experiments

Anqi Zhao¹ and Peng Ding² 

¹Department of Statistics and Data Science, National University of Singapore, Singapore, Singapore

²Department of Statistics, University of California, Berkeley, CA, USA

Address for correspondence: Peng Ding, Department of Statistics, University of California, Berkeley, CA 94720, USA.

Email: pengding@berkeley.edu

Abstract

Randomized experiments are the gold standard for causal inference and enable unbiased estimation of treatment effects. Regression adjustment provides a convenient way to incorporate covariate information for additional efficiency. This article provides a unified account of its utility for improving estimation efficiency in multiarmed experiments. We start with the commonly used additive and fully interacted models for regression adjustment in estimating average treatment effects (ATE), and clarify the trade-offs between the resulting ordinary least squares (OLS) estimators in terms of finite sample performance and asymptotic efficiency. We then move on to regression adjustment based on restricted least squares (RLS), and establish for the first time its properties for inferring ATE from the design-based perspective. The resulting inference has multiple guarantees. First, it is asymptotically efficient when the restriction is correctly specified. Second, it remains consistent as long as the restriction on the coefficients of the treatment indicators, if any, is correctly specified and separate from that on the coefficients of the treatment-covariate interactions. Third, it can have better finite sample performance than the unrestricted counterpart even when the restriction is moderately misspecified. It is thus our recommendation when the OLS fit of the fully interacted regression risks large finite sample variability in case of many covariates, many treatments, yet a moderate sample size. In addition, the newly established theory of RLS also provides a unified way of studying OLS-based inference from general regression specifications. As an illustration, we demonstrate its value for studying OLS-based regression adjustment in factorial experiments. Importantly, although we analyse inferential procedures that are motivated by OLS, we do not invoke any assumptions required by the underlying linear models.

Keywords: causal inference, design-based inference, potential outcomes, regression adjustment, rerandomization, restricted least squares

1 Introduction

1.1 Multiarmed experiment and covariate adjustment

Multiarmed experiments enable comparisons of more than two treatment levels simultaneously, and are intrinsic to applications with multiple factors of interest (see, e.g., Chakraborty et al., 2009; Collins et al., 2009; Mukerjee et al., 2018). They have been extensively used in agricultural and industrial settings (see, e.g., Box et al., 2005; Wu & Hamada, 2009), and are becoming increasingly popular in social and biomedical applications (see, e.g., Blackwell & Pashley, 2021; Duflo et al., 2007; Egami & Imai, 2019; Hainmueller et al., 2014). Ordinary least squares (OLS) regression remains the dominant strategy for subsequent inference of treatment effects, delivering not only point estimates but also their standard errors. The flexibility of model specifications further provides a convenient way to incorporate covariates for additional efficiency.

The additive and fully interacted regressions are two commonly used strategies for covariate adjustment by OLS. In particular, the additive regression regresses the outcome on the treatment and covariates to adjust for covariate imbalance across treatment groups (Fisher, 1935). The fully

Received: January 18, 2022. Revised: August 29, 2022. Accepted: September 12, 2022

© (RSS) Royal Statistical Society 2023. All rights reserved. For permissions, please e-mail:
journals.permissions@oup.com

interacted regression regresses the outcome on not only the treatment and covariates but also their interactions to further accommodate treatment effect heterogeneity (Lin, 2013). The theoretical superiority of the fully interacted regression is well established under treatment-control experiments (see, e.g., Li & Ding, 2020; Lin, 2013). Similar discussion, however, is largely missing for experiments with more than two treatment arms, except for some preliminary results in Freedman (2008b), Lin (2013), Lu (2016b), and Schochet (2018). To fill this gap, we clarify the validity and relative efficiency of the additive and fully interacted regressions for estimating average treatment effects (ATE) under multiarmed experiments from the *design-based* perspective, which conditions on the potential outcomes and evaluates the sampling properties of estimators over the distribution of the treatment assignments (Dasgupta et al., 2015; Freedman, 2008a; Imbens & Rubin, 2015; Lin, 2013; Neyman, 1923). The results are similar to those under the treatment-control experiment. The OLS estimator from the fully interacted regression is consistent and ensures efficiency gains over the unadjusted regression asymptotically. The additive regression, on the other hand, is always consistent yet only ensures efficiency gains when the correlations between potential outcomes and covariates are constant across treatment levels. In addition, we establish the asymptotic conservativeness of the associated Eicker–Huber–White (EHW) robust covariance estimators for estimating the true sampling covariances under both the additive and fully interacted regressions. This constitutes our first contribution on the design-based justification of regression adjustment.

Despite being theoretically superior, the fully interacted specification includes all interactions between treatment indicators and covariates, and can incur substantial finite sample variability when there are many treatment levels, many covariates, yet only a moderate sample size. The additive regression, in contrast, can have better finite sample performance than both the unadjusted and fully interacted regressions under such circumstances as long as the treatment effects are not too heterogeneous across covariate levels. The choice between the additive and fully interacted specifications is thus a trade-off between finite sample performance and asymptotic efficiency. Of interest is whether there can be some better middle ground.

To this end, we propose restricted least squares (RLS) as an alternative way to infer ATE from the fully interacted regression, and establish its sampling properties under the design-based framework. The resulting inference has multiple guarantees. First, it is asymptotically efficient when the restriction is correctly specified. Second, it remains consistent as long as the restriction on the coefficients of the treatment indicators, if any, is correctly specified and separate from that on the coefficients of the treatment-covariate interactions. Third, it can have better finite sample performance than its unrestricted counterpart even when the restriction is moderately misspecified. It is thus our recommendation for regression adjustment in multiarmed experiments when the OLS estimator from the fully interacted regression risks large finite sample variability. We also prove a novel design-based Gauss–Markov theorem for RLS, clarifying the asymptotic bias-variance trade-off between OLS and RLS under constant treatment effects. These design-based results on RLS are not only of theoretical interest in themselves but also provide a unified framework for studying regression-based inference from general specifications. This constitutes our second contribution on the design-based theory of RLS for estimating ATE. The classical theory of RLS assumes a correct linear model with homoskedastic errors under correct restriction. Our theory, in contrast, is design-based and allows for not only heteroskedastic errors but also misspecification of both the linear model and the restriction.

We then move on to factorial experiments (Box et al., 2005; Dasgupta et al., 2015; Pashley & Bind, 2022; Wu & Hamada, 2009; Zhao & Ding, 2022a), and illustrate the value of our theory for studying covariate adjustment under this special type of multiarmed experiments. Specifically, factorial experiments concern multiple factors of interest, and assign experimental units to all possible levels of their combinations. The special structure of the treatment levels enables convenient factor-based regression analysis (Lu, 2016a, 2016b; Wu & Hamada, 2009; Zhao & Ding, 2022a), which fits the observed outcome on indicators of the factor levels by OLS and interprets the resulting coefficients as the factorial effects of interest. The design-based theory on covariate adjustment under factorial experiments has so far focused on *factor-saturated* regressions that include all possible interactions between the factor indicators (Lu, 2016b; Zhao & Ding, 2022b). The resulting specifications have model complexity that increases exponentially with

the number of factors, risking substantial finite sample variability even with a moderate number of factors and covariates.

To address this issue, we consider *factor-unsaturated* specifications that include only a subset of the interactions between the factor indicators when conducting regression adjustment, and clarify the properties of the resulting OLS estimators from the design-based perspective. The choice between the factor-saturated and factor-unsaturated specifications, as it turns out, boils down to a trade-off between asymptotic bias and variance, extending the result in covariate-free settings (Zhao & Ding, 2022a). The resulting theory includes most commonly used specifications in practice as special cases, and offers guidelines on the causal interpretations of their results. This constitutes our third contribution on causal inference with factorial experiments.

Importantly, although the regression-based covariate adjustment was originally motivated by linear models, we do not invoke any of the underlying assumptions but view regression as a purely numeric procedure based on OLS or RLS. We evaluate the sampling properties of the resulting point estimators and standard errors over the distribution of the treatment assignments. All our theories are as such design-based and hold regardless of how well the regression equations represent the true data-generating process (see, e.g., Abadie et al., 2020; Bloniarz et al., 2016; Fogarty, 2018; Freedman, 2008a, 2008b; Guo & Basse, 2021; Imbens & Rubin, 2015; Lin, 2013; Liu & Yang, 2020; Miratrix et al., 2013; Schochet, 2010, 2018).

1.2 Notation and definitions

Let 0_m and $0_{m \times n}$ denote the $m \times 1$ vector and $m \times n$ matrix of zeros, respectively. Let 1_m and $1_{m \times n}$ denote the $m \times 1$ vector and $m \times n$ matrix of ones, respectively. Let I_m denote the $m \times m$ identity matrix. We suppress the dimensions when they are clear from the context. Let $1(\cdot)$ denote the indicator function. Let \otimes denote the Kronecker product of matrices. For a set of real numbers $\{u_q : q \in \mathcal{T}\}$, let $\text{diag}(u_q)_{q \in \mathcal{T}}$ denote the diagonal matrix with u_q 's on the diagonal. For two $m \times m$ square matrices M_1 and M_2 , write $M_1 \leq M_2$ or $M_2 \geq M_1$ if $M_2 - M_1$ is positive semi-definite.

For a finite population $(Y_i, u_i)_{i=1}^N$ with $Y_i \in \mathbb{R}$ and $u_i \in \mathbb{R}^m$, let $Y_i \sim u_i$ denote the linear regression of Y_i on u_i free of any modelling assumptions. Let $\hat{\epsilon}_i$ denote the fitted residual of unit i from OLS. The EHW covariance estimator of the coefficient vector of u_i equals $(U^\top U)^{-1} U^\top \text{diag}(\hat{\epsilon}_i^2)_{i=1}^N U (U^\top U)^{-1}$, where $U = (u_1, \dots, u_N)^\top$.

Let plim denote the probability limit of a sequence of random elements, and let \rightsquigarrow denote convergence in distribution. For $\hat{\theta}_1$ and $\hat{\theta}_2$ that are both consistent and asymptotically normal for estimating parameter $\theta \in \mathbb{R}^m$ as the sample size N tends to infinity: $\sqrt{N}(\hat{\theta}_k - \theta) \rightsquigarrow \mathcal{N}(0_m, V_k)$ for $k = 1, 2$, we say

- (i) $\hat{\theta}_1$ is *asymptotically more efficient* than $\hat{\theta}_2$ if $V_1 \leq V_2$ and the inequality is strict for some data-generating process;
- (ii) $\hat{\theta}_1$ and $\hat{\theta}_2$ are *asymptotically equally efficient* if $V_1 = V_2$.

2 Causal inference with multiarmed experiments

2.1 Potential outcomes and treatment effects

Consider an experiment with $Q \geq 2$ treatment levels, $q \in \mathcal{T} = \{1, \dots, Q\}$, and a study population of N units, $i = 1, \dots, N$. Let $Y_i(q)$ be the potential outcome of unit i if assigned to treatment level q (Neyman, 1923). Let $\bar{Y}(q) = N^{-1} \sum_{i=1}^N Y_i(q)$ be the finite population average, vectorized as $\bar{Y} = (\bar{Y}(1), \dots, \bar{Y}(Q))^\top$. The goal is to estimate the finite population ATE

$$\tau = C \bar{Y} \tag{1}$$

for some prespecified contrast matrix C with rows orthogonal to 1_Q .

The above setting is general and includes many commonly used designs as special cases. We use the treatment-control experiment, the three-armed experiment, the 2^2 factorial experiment, the 2^K factorial experiment with general $K \geq 2$, and the general $Q_1 \times \dots \times Q_K$ factorial experiment as running examples to illustrate the main ideas. We give their definitions in Examples 1–5 for concreteness. The theory in Sections 2–4 applies to general experiments with arbitrary structures of

treatment levels. The special structures of factorial experiments further allow for the more flexible factor-based inference, which we detail in Section 5.

With a slight abuse of notation, we use customized indexes for treatment levels in Examples 1–5 to match the conventions in the literature. They have a one-to-one mapping with the $q \in \{1, \dots, Q\}$ indexing in lexicographical order.

Example 1 The treatment-control experiment has $Q = 2$ treatment levels, indexed by $q \in \mathcal{T} = \{0, 1\}$. The individual treatment effect is $\tau_i = Y_i(1) - Y_i(0)$, and the ATE is $\tau = N^{-1} \sum_{i=1}^N \tau_i = \bar{Y}(1) - \bar{Y}(0) = C\bar{Y}$, with $C = (-1, 1)$ and $\bar{Y} = (\bar{Y}(0), \bar{Y}(1))^T$.

Example 2 Consider a three-armed experiment with one control level and two active levels, indexed by $q \in \mathcal{T} = \{0 \text{ (control)}, 1 \text{ (active)}, 2 \text{ (active)}\}$. The ATE of the two active levels relative to control are $\tau_q = \bar{Y}(q) - \bar{Y}(0)$ for $q = 1, 2$. We vectorize them as

$$\tau = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} = \begin{pmatrix} \bar{Y}(1) - \bar{Y}(0) \\ \bar{Y}(2) - \bar{Y}(0) \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{Y}(0) \\ \bar{Y}(1) \\ \bar{Y}(2) \end{pmatrix} = C\bar{Y},$$

where $C = (-1_2, I_2)$ and $\bar{Y} = (\bar{Y}(0), \bar{Y}(1), \bar{Y}(2))^T$.

The setting of Example 2 extends immediately to general multiarmed experiments with a single factor of $Q > 2$ levels, as in the setting of the one-way analysis of variance. This is among the most common types of multiarmed experiments in practice, featuring no special structure of the treatment levels. Factorial designs, on the other hand, accommodate multiple factors of interest in one single experiment and define treatment levels as their combinations.

Example 3 The 2^2 factorial experiment is a special type of the four-armed experiment with two binary factors of interest, $A, B \in \{-1, +1\}$, and in total $Q = 2^2 = 4$ treatment levels as their combinations: $q \in \mathcal{T} = \{(-1, -1), (-1, +1), (+1, -1), (+1, +1)\}$. We abbreviate $(-1, -1)$ as $(--)$, etc. when no confusion would arise. Let $\bar{Y} = (\bar{Y}(--), \bar{Y}(-+), \bar{Y}(+-), \bar{Y}(++)^T$ be the vector of average potential outcomes in lexicographical order of the treatment combinations. The standard main effects and interaction effect equal

$$\begin{aligned} \tau_A &= 2^{-1}\{\bar{Y}(++) + \bar{Y}(+-)\} - 2^{-1}\{\bar{Y}(--) + \bar{Y}(-+)\} = c_A^T \bar{Y}, \\ \tau_B &= 2^{-1}\{\bar{Y}(-+) + \bar{Y}(++)\} - 2^{-1}\{\bar{Y}(--) + \bar{Y}(+-)\} = c_B^T \bar{Y}, \\ \tau_{AB} &= 2^{-1}\{\bar{Y}(--) + \bar{Y}(++)\} - 2^{-1}\{\bar{Y}(-+) + \bar{Y}(+-)\} = c_{AB}^T \bar{Y}, \end{aligned}$$

respectively, with $c_A = 2^{-1}(-1, -1, 1, 1)^T$, $c_B = 2^{-1}(-1, 1, -1, 1)^T$, and $c_{AB} = 2^{-1}(1, -1, -1, 1)^T$ (Dasgupta et al., 2015). The main effect of a factor compares the average potential outcomes when the factor is at level $+1$ and level -1 , respectively. The interaction effect compares the average potential outcomes when the two factors take the same and different levels, respectively.

Example 4 The 2^K factorial experiment is a special type of the 2^K -armed experiment with K binary factors of interest, $k = 1, \dots, K$, and in total $Q = 2^K$ treatment levels as their combinations: $q \in \mathcal{T} = \{(z_1, \dots, z_K) : z_k = -1, +1\}$, where z_k indicates the level of factor k . The treatment-control and 2^2 factorial experiments in Examples 1 and 3 are both special cases with $K = 1$ and $K = 2$, respectively. There are $2^K - 1$ standard factorial effects corresponding to the main effects and two- to K -way interaction effects, respectively (Wu & Hamada, 2009).

Example 5 The general $Q_1 \times \cdots \times Q_K$ factorial experiment has $Q = \prod_{k=1}^K Q_k$ treatment levels, as the combinations of $K \geq 2$ factors of interest, $k = 1, \dots, K$, each of $Q_k \geq 2$ levels. It is a special type of the Q -armed experiment, and includes the 2^K factorial experiment in Example 4 as a special case with $Q_k = 2$ for all k .

2.2 Assignment mechanism, estimators, and regression formulation

We focus on complete randomization defined below.

Definition 1 (Complete randomization). For prespecified, fixed $N_q > 0$ ($q \in \mathcal{T}$) with $\sum_{q \in \mathcal{T}} N_q = N$, the experimenter randomly draws an allocation from all possible allocations that have N_q units receiving treatment level q ($q \in \mathcal{T}$).

Let $Z_i \in \mathcal{T}$ denote the treatment level received by unit i . The observed outcome equals

$$Y_i = Y_i(Z_i) = \sum_{q \in \mathcal{T}} 1(Z_i = q) Y_i(q) \quad (2)$$

for unit i . Let $\hat{Y}(q) = N_q^{-1} \sum_{i: Z_i=q} Y_i$ be the average observed outcome under treatment level q . The sample-mean estimator of \bar{Y} equals $\hat{Y}_N = (\hat{Y}(1), \dots, \hat{Y}(Q))^T$ and suggests

$$\hat{\tau}_N = C \hat{Y}_N \quad (3)$$

as an intuitive choice for estimating τ .

Let $e_q = N_q/N$ denote the proportion of units under treatment level q . Let $S = (S_{qq'})_{q,q' \in \mathcal{T}}$ be the finite population covariance matrix of $\{Y_i(q) : q \in \mathcal{T}\}_{i=1}^N$ with $S_{qq'} = (N-1)^{-1} \sum_{i=1}^N \{Y_i(q) - \bar{Y}(q)\} \{Y_i(q') - \bar{Y}(q')\}$. Under complete randomization, $\hat{\tau}_N$ is unbiased for τ with sampling covariance $\text{cov}(\hat{\tau}_N) = N^{-1} C V_N C^T$, where $V_N = N \text{cov}(\hat{Y}_N) = \text{diag}(S_{qq}/e_q)_{q \in \mathcal{T}} - S$. Define

$$t_i = (1(Z_i = 1), \dots, 1(Z_i = Q))^T \in \mathbb{R}^Q$$

as the treatment indicator vector for unit i . Then \hat{Y}_N equals the coefficient vector of t_i from the OLS fit of

$$Y_i \sim 1(Z_i = 1) + \cdots + 1(Z_i = Q) \iff Y_i \sim t_i \quad (4)$$

over $i = 1, \dots, N$ without an intercept. We call (4) the *unadjusted treatment-based regression*, taking the treatment indicators as regressors.

The presence of covariates allows for opportunities to further improve estimation efficiency. Let $x_i = (x_{i1}, \dots, x_{ij})^T$ denote the $J \times 1$ covariate vector for unit i . To simplify the presentation, assume centred $(x_i)_{i=1}^N$ throughout the paper with mean $\bar{x} = N^{-1} \sum_{i=1}^N x_i = 0_J$ and finite population covariance $S_x^2 = (N-1)^{-1} \sum_{i=1}^N x_i x_i^T$. Specifications

$$Y_i \sim t_i + x_i, \quad (5)$$

$$Y_i \sim t_i + \sum_{q \in \mathcal{T}} 1(Z_i = q) x_i \iff Y_i \sim t_i + t_i \otimes x_i \quad (6)$$

give two intuitive ways to adjust for covariates on the basis of Equation (4). Refer to them as the *additive* and *fully interacted treatment-based regressions*, respectively, depending on whether the regression equations include the interactions between t_i and x_i or not. We denote by $\hat{Y}_F \in \mathbb{R}^Q$ and $\hat{Y}_L \in \mathbb{R}^Q$ the coefficient vectors of t_i from the OLS fits of Equations (5) and (6), respectively, as two

covariate-adjusted variants of \hat{Y}_N from Equation (4) for estimating \bar{Y} . As a convention, we use the subscripts ‘N’, ‘F’, and ‘L’ to signify quantities associated with the unadjusted, additive, and fully interacted regressions, respectively. Example 6 in Section 3 will clarify their respective connections with Neyman (1923), Fisher (1935), and Lin (2013). Replacing \hat{Y}_N with \hat{Y}_* in Equation (3) yields

$$\hat{\tau}_* = C\hat{Y}_* \quad (* = F, L)$$

as two covariate-adjusted estimators of τ . Of interest is their validity and efficiency relative to $\hat{\tau}_N$ from the design-based perspective. We address this question in Section 3 after clarifying the intuition behind the regression formulation in Section 2.3.

2.3 Derived linear models and target parameters

Derived linear models (Hinkelmann & Kempthorne, 2008; Kempthorne, 1952) provide the intuition for using the coefficient vectors of t_i from Equations (4)–(6) to estimate \bar{Y} .

In particular, consider the OLS fit of $Y_i(q) \sim 1$ over $i = 1, \dots, N$ for $q \in \mathcal{T}$. This is a theoretical OLS fit with the $Y_i(q)$ ’s only partially observable, and yields the fitted model $Y_i(q) = \bar{Y}(q) + \epsilon_{N,i}(q)$ with $\epsilon_{N,i}(q) = Y_i(q) - \bar{Y}(q)$. Plugging this fitted model in Equation (2) implies the *unadjusted derived linear model* of the observed outcome:

$$Y_i = \sum_{q \in \mathcal{T}} 1(Z_i = q) \bar{Y}(q) + \epsilon_{N,i} = t_i^T \bar{Y} + \epsilon_{N,i} \quad (7)$$

with $\epsilon_{N,i} = \sum_{q \in \mathcal{T}} 1(Z_i = q) \epsilon_{N,i}(q)$. This motivates the unadjusted regression (4).

The presence of covariates motivates extending Equation (7) to include regression adjustment. Consider the OLS fit of $Y_i(q) \sim 1 + x_i$ over $i = 1, \dots, N$ for $q \in \mathcal{T}$. This is also a theoretical OLS fit, and yields the fitted model $Y_i(q) = \bar{Y}(q) + x_i^T \gamma_q + \epsilon_{L,i}(q)$. We have $\gamma_q = (S_x^2)^{-1} S_{xY(q)}$, where $S_{xY(q)} = (N-1)^{-1} \sum_{i=1}^N x_i \{Y_i(q) - \bar{Y}(q)\}$ denotes the finite population covariance of $\{x_i, Y_i(q)\}_{i=1}^N$. For simplicity, we call γ_q the *population correlation* of $Y_i(q)$ and x_i . Plugging this fitted model in Equation (2) implies the *covariate-adjusted derived linear model* of the observed outcome:

$$\begin{aligned} Y_i &= \sum_{q \in \mathcal{T}} 1(Z_i = q) \bar{Y}(q) + \sum_{q \in \mathcal{T}} \{1(Z_i = q) x_i\}^T \gamma_q + \epsilon_{L,i} \\ &= t_i^T \bar{Y} + (t_i \otimes x_i)^T \gamma + \epsilon_{L,i} \end{aligned} \quad (8)$$

with $\epsilon_{L,i} = \sum_{q \in \mathcal{T}} 1(Z_i = q) \epsilon_{L,i}(q)$ and $\gamma = (\gamma_1^T, \dots, \gamma_Q^T)^T$. This motivates the fully interacted regression (6). We call \bar{Y} and γ the *target parameters* of t_i and $t_i \otimes x_i$ in Equation (6), respectively, and refer to $\theta_L = (\bar{Y}^T, \gamma^T)^T$ as the target parameter of Equation (6) as a whole.

The unadjusted regression (4) can be viewed as a *restricted variant* of Equation (6), assuming that the coefficient vectors of the $1(Z_i = q) x_i$ ’s are all zero. Likewise can we view the additive regression (5) as a restricted variant of Equation (6), assuming that the coefficient vectors of the $1(Z_i = q) x_i$ ’s are all equal. This motivates the definitions of the zero and equal correlation conditions below. They provide not only the heuristics for the functional forms of Equations (4) and (5) relative to Equation (6) but also the sufficient conditions for the asymptotic efficiency of $\hat{\tau}_N$ and $\hat{\tau}_F$, respectively (see Section 3, Lemma 1).

Condition 1 (Zero correlation). $\gamma_1 = \dots = \gamma_Q = 0_J$.

Condition 2 (Equal correlation). $\gamma_1 = \dots = \gamma_Q$.

The variation in $\{\gamma_q : q \in \mathcal{T}\}$ measures the heterogeneity in treatment effects that is explained by covariates (Ding et al., 2019). Condition 2 stipulates that the population correlations between potential outcomes and covariates are constant across treatment levels, implying homogeneous

treatment effects. Condition 1 is stronger than Condition 2 and stipulates uncorrelatedness for all levels. Condition 3 gives a sufficient condition for Condition 2.

Condition 3 (Constant treatment effects). For all $q, q' \in \mathcal{T}$, the individual treatment effects $Y_i(q) - Y_i(q')$ are constant across $i = 1, \dots, N$.

Importantly, the derived linear models (7) and (8) are purely numeric decompositions of the observed outcome without any assumptions on the data-generating process. We use them as props to motivate the zero and equal correlation conditions and introduce the target parameters. The design-based framework conditions on the potential outcomes and covariates, and attributes the randomness in Y_i solely to the randomness in the treatment assignments Z_i 's. The covariances of $(\epsilon_{*,i})_{i=1}^N$ ($* = N, F, L$) are accordingly fully determined by the joint distribution of the Z_i 's, and are in general neither jointly independent nor homoskedastic under complete randomization.

The classical Gauss–Markov model, on the other hand, conditions on the Z_i 's, and attributes the randomness in Y_i 's to the sampling errors due to the study population being a random sample from some hypothetical superpopulation. The covariance of the error terms is specified by model assumptions, with joint independence and homoskedasticity being two commonly invoked options. Freedman (2008a) pointed out that randomization does not justify these assumptions. The theory we are about to present, nevertheless, suggests that \hat{Y}_* ($* = N, F, L$), as purely numeric outputs from OLS, can still deliver valid design-based inferences when coupled with the EHW covariance estimators. We elaborate on the details in Section 3.

3 Regression adjustment by ordinary least squares

3.1 Asymptotic efficiency of $\hat{\tau}_L$ over $\hat{\tau}_N$ and $\hat{\tau}_F$

We establish in this subsection the validity and asymptotic relative efficiency of $\hat{\tau}_*$ ($* = N, F, L$) for inferring τ . The result extends Fisher (1935), Freedman (2008a), and Lin (2013) to multiarmed experiments, and complements Freedman (2008b) on the asymptotics of the fully interacted regression. See Tsiatis et al. (2008), Bugni et al. (2018, 2019), Negi and Wooldridge (2021), and Ye et al. (2022) for analogous results under alternative superpopulation frameworks.

Condition 4 is standard for finite population asymptotics under complete randomization (Li & Ding, 2017).

Condition 4. As $N \rightarrow \infty$, for $q \in \mathcal{T}$, (i) $e_q = N_q/N$ has a limit in $(0, 1)$; (ii) the first two finite population moments of $\{Y_i(q), x_i, x_i Y_i(q) : q \in \mathcal{T}\}$ have finite limits; both S_x^2 and its limit are nonsingular; (iii) $N^{-1} \sum_{i=1}^N Y_i^4(q) = O(1)$, $N^{-1} \sum_{i=1}^N \|x_i\|_4^4 = O(1)$, and $N^{-1} \sum_{i=1}^N \|x_i Y_i(q)\|_4^4 = O(1)$.

Recall that γ_q denotes the population correlation of $Y_i(q)$ and x_i . Let $\bar{\gamma} = \sum_{q \in \mathcal{T}} e_q \gamma_q$ be a weighted average. Let $S_N = (S_{N,qq'})_{q,q' \in \mathcal{T}}$, $S_F = (S_{F,qq'})_{q,q' \in \mathcal{T}}$, and $S_L = (S_{L,qq'})_{q,q' \in \mathcal{T}}$ be the finite population covariance matrices of $\{Y_i(q) : q \in \mathcal{T}\}_{i=1}^N$, $\{Y_i(q) - \bar{\gamma}^T x_i : q \in \mathcal{T}\}_{i=1}^N$, and $\{Y_i(q) - \gamma_q^T x_i : q \in \mathcal{T}\}_{i=1}^N$, respectively, with $S_N = S$. Let

$$V_* = \text{diag}(S_{*,qq}/e_q)_{q \in \mathcal{T}} - S_* \quad (* = N, F, L), \quad (9)$$

which have finite limits under Condition 4. To simplify the presentation, we will also use the same symbols to denote their respective limiting values when no confusion would arise.

Let $\hat{\Psi}_*$ be the EHW covariance estimator of \hat{Y}_* from the same OLS fit for $* = N, F, L$. We first make a moderate contribution by providing a unified theory for the design-based properties of \hat{Y}_* and $\hat{\Psi}_*$ ($* = N, F, L$) in Lemma 1. The result clarifies the consistency and asymptotic normality of \hat{Y}_* for estimating \bar{Y} , and ensures the asymptotic conservativeness of $\hat{\Psi}_*$ for estimating the true sampling covariance.

Lemma 1. Assume complete randomization and Condition 4. Then

- (i) $\sqrt{N}(\hat{Y}_* - \bar{Y}) \rightsquigarrow \mathcal{N}(0, V_*)$ for $* = N, F, L$ with $N\hat{\Psi}_* - V_* = S_* + o_p(1)$, where $S_* \geq 0$;
- (ii) $V_L \leq V_N$ and $V_L \leq V_F$;
- (iii) Under Condition 2, $V_F = V_L \leq V_N$; under Condition 1, $V_N = V_F = V_L$.

Lemma 1(i) justifies the large-sample Wald-type inference of τ based on $\hat{\tau}_* = C\hat{Y}_*$ and $C\hat{Y}_*C^T$ as the point estimator and estimated covariance, respectively, for $* = N, F, L$. Lemma 1(ii) ensures the asymptotic efficiency of $\hat{\tau}_L$ over $\hat{\tau}_N$ and $\hat{\tau}_F$. The additively adjusted $\hat{\tau}_F$, on the other hand, may be less efficient than the unadjusted $\hat{\tau}_N$, especially when the experiments have unequal group sizes and heterogeneous treatment effects with respect to the covariates (Freedman, 2008a). Lemma 1(iii) gives two exceptions. First, the equal correlation condition ensures that $\hat{\tau}_F$ is asymptotically equally efficient as $\hat{\tau}_L$, rendering the inclusion of interaction terms unnecessary. Second, the zero correlation condition ensures that $\hat{\tau}_N$ is asymptotically equally efficient as $\hat{\tau}_F$ and $\hat{\tau}_L$, rendering regression adjustment unnecessary.

Lemma 1 unifies the existing theory on regression adjustment under the treatment-control experiment as a special case. We review in Example 6 the results from Neyman (1923), Fisher (1935), and Lin (2013), and clarify their connections with Equations (4)–(6).

Example 6 Consider the treatment-control experiment from Example 1. We have $\hat{\tau}_* = (-1, 1)\hat{Y}_*$ for $* = N, F, L$, where the \hat{Y}_* 's are the coefficient vectors of $t_i = (1(Z_i = 0), 1(Z_i = 1))^T = (1 - Z_{i2}, Z_i)^T$ from the OLS fits of Equations (4)–(6), respectively. Then $\hat{\tau}_N = \hat{Y}(1) - \hat{Y}(0)$ equals the difference-in-means estimator and can also be computed as the coefficient of Z_i from the OLS fit of $Y_i \sim 1 + Z_i$. Neyman (1923) showed that $\hat{\tau}_N$ is unbiased for τ .

In the presence of covariate information, Fisher (1935) suggested to estimate τ by the coefficient of Z_i from the OLS fit of $Y_i \sim 1 + Z_i + x_i$. Lin (2013) proposed to include also the interactions between x_i and Z_i , and estimate τ by the coefficient of Z_i from the OLS fit of $Y_i \sim 1 + Z_i + x_i + Z_i x_i$. The one-to-one correspondence between $(1, Z_i)^T$ and t_i ensures that the resulting estimators equal $\hat{\tau}_F$ and $\hat{\tau}_L$, respectively. This clarifies the connections of Equations (4)–(6) to Neyman (1923), Fisher (1935), and Lin (2013), respectively. The properties of $\hat{\tau}_*$ ($* = N, F, L$) follow readily from Lemma 1.

As it turns out, the asymptotic efficiency of $\hat{\tau}_L$ extends beyond Lemma 1 to a class of general linear estimators of τ that includes $\hat{\tau}_*$ ($* = N, F, L$) as special cases. We relegate the details to the online supplementary material.

3.2 Trade-off between $\hat{\tau}_L$ and $\hat{\tau}_F$

Despite the guaranteed gains in asymptotic efficiency, the fully interacted regression (6) involves $Q + JQ$ estimated coefficients, subjecting $\hat{\tau}_L$ to large finite sample variability when the sample size N is moderate relative to $Q + JQ$. Simulation evidence, on the other hand, suggests that $\hat{\tau}_F$ can be more efficient than $\hat{\tau}_L$ in finite samples as long as the equal correlation condition is not severely violated. The choice between the additive and fully interacted regressions is thus a trade-off between finite sample performance and asymptotic efficiency. This, together with the asymptotic efficiency of $\hat{\tau}_F$ under the equal correlation condition, grants $\hat{\tau}_F$ a triple guarantee: it is consistent and asymptotically normal for general potential outcomes, ensures asymptotic efficiency if the γ_q 's are all equal, and can have better finite sample performance than $\hat{\tau}_L$ as long as the γ_q 's are not too different. Fisher (1935)'s analysis of covariance, as a result, can be viewed as a compromise between unadjusted inference and the fully interacted adjustment in moderate samples. See Schochet (2010) for empirical evidence based on eight large social policy experiments.

Recall from Section 2.3 that the additive regression (5) can be viewed as a restricted variant of Equation (6), assuming that the coefficient vectors of $1(Z_i = q)x_i$'s are all equal. This motivates a more general approach to regression adjustment by restricted least squares, which estimates the coefficients of Equation (6) subject to some prespecified linear restrictions. We establish the design-based properties of the resulting inference in Section 4.

4 Regression adjustment by restricted least squares

4.1 Restricted least squares

Restricted least squares is a standard tool for fitting linear models, enabling convenient encoding of prior knowledge on model parameters. Its theoretical properties are well studied under the

classical Gauss–Markov model (Greene & Seaks, 1991; Rao, 1973; Theil, 1971). The corresponding theory, however, is so far missing under the design-based framework, where the errors are intrinsically dependent and heteroskedastic (cf. Section 2.3). This section fills this gap and clarifies the design-based properties of RLS for regression adjustment. The resulting theory is not only of theoretical interest in itself but also provides a unified way to study OLS-based inference from general regressions. We focus on RLS-based inference for general multiarmed experiments in this section, and demonstrate its value for studying OLS-based regression adjustment in factorial experiments in Section 5. The classical theory of RLS assumes a correct linear model with homoskedastic errors under correct restriction. Our theory, in contrast, is design-based and allows for not only heteroskedastic errors but also misspecification of both the linear model and the restriction.

Recall $\theta_L = (\bar{Y}^T, \gamma^T)^T$ as the target parameter of the fully interacted regression (6), motivated by the derived linear model (8). Let $\chi_{L,i} = (t_i^T, t_i^T \otimes x_i^T)^T$ denote the regressor vector of Equation (6). The OLS fit of Equation (6) can be viewed as estimating θ_L by

$$\hat{\theta}_L = (\hat{Y}_L^T, \hat{\gamma}_L^T)^T = \operatorname{argmin}_{\theta} \sum_{i=1}^N (Y_i - \chi_{L,i}^T \theta)^2,$$

where \hat{Y}_L and $\hat{\gamma}_L = (\hat{\gamma}_{L,1}^T, \dots, \hat{\gamma}_{L,Q}^T)^T$ denote the coefficient vectors of t_i and $t_i \otimes x_i$, respectively, with $\hat{\gamma}_{L,q}$ corresponding to $1(Z_i = q)x_i$.

The RLS fit, on the other hand, estimates θ_L subject to some prespecified linear restrictions.

Definition 2 (Restricted least squares). For some prespecified restriction matrix $R \in \mathbb{R}^{m \times (Q+JQ)}$ that has full row rank $m \leq Q + JQ$ and vector $r \in \mathbb{R}^m$, the RLS fit of Equation (6) yields

$$\hat{\theta}_r = (\hat{Y}_r^T, \hat{\gamma}_r^T)^T = \operatorname{argmin}_{\theta} \sum_{i=1}^N (Y_i - \chi_{L,i}^T \theta)^2 \quad \text{subjectto } R\theta = r, \quad (10)$$

where \hat{Y}_r and $\hat{\gamma}_r = (\hat{\gamma}_{r,1}^T, \dots, \hat{\gamma}_{r,Q}^T)^T$ denote the coefficient vectors of t_i and $t_i \otimes x_i$, respectively, with $\hat{\gamma}_{r,q}$ corresponding to $1(Z_i = q)x_i$.

By Equation (10), we can view $\hat{\theta}_r$ as an alternative estimator of θ_L that satisfies $R\hat{\theta}_r = r$. This is as if we estimate θ_L subject to the prior belief of

$$R\theta_L = r, \quad (11)$$

and suggests $\hat{\tau}_r = C\hat{Y}_r$ as an alternative option for inferring τ . The goal is to quantify its sampling properties relative to $\hat{\tau}_L$.

To simplify the presentation, we will use Equation (11) to represent the restriction in Equation (10) in terms of the target parameter θ_L , and refer to Equation (10) as the RLS fit subject to *working restriction* (11). The working restriction (11) is accordingly a purely numeric input for RLS that may or may not match the truth. We say Equation (11) is *correctly specified* if it indeed matches the truth.

Examples 7 and 8 establish both \hat{Y}_N and \hat{Y}_F as special cases of \hat{Y}_r with specific choices of (R, r) .

Example 7 $\hat{Y}_N = \hat{Y}_r$, where \hat{Y}_r is the coefficient vector of t_i from the RLS fit of Equation (6) subject to the *zero correlation restriction* that $\gamma_q = 0_J$ for all q . A matrix representation of this restriction is $R\theta_L = \gamma = 0_{JQ}$ with $R = (0_{JQ \times Q}, I_{JQ})$.

Example 8 $\hat{Y}_F = \hat{Y}_r$, where \hat{Y}_r is the coefficient vector of t_i from the RLS fit of Equation (6) subject to the *equal correlation restriction* that $\gamma_1 = \dots = \gamma_Q$. A matrix representation of this restriction is $R\theta_L = (\gamma_2^T - \gamma_1^T, \dots, \gamma_Q^T - \gamma_1^T)^T = 0_{J(Q-1)}$ with $R = (0_{J(Q-1) \times Q}, (-1_{Q-1}, I_{Q-1}) \otimes I_J)$.

Recall from Section 2.3 that regressions (4) and (5) can be viewed as restricted variants of Equation (6), assuming the zero and equal correlation restrictions, respectively. Examples 7 and 8 illustrate the numeric correspondence between the RLS fit and the OLS fit of the corresponding restricted specification, establishing $\hat{\tau}_N$ and $\hat{\tau}_F$ as special cases of $\hat{\tau}_r$. See Lemma S3 in the [online supplementary material](#) for a more general result.

Importantly, the equal correlation *restriction* differs from the equal correlation *condition* in Condition 2. Echoing the comments after Equation (11), we view the equal correlation restriction as a purely numeric input for RLS that may or may not match the truth. Condition 2, in contrast, represents our assumption about the true data-generating process. The equal correlation restriction is hence correctly specified if and only if Condition 2 holds. Likewise for the correspondence between the zero correlation restriction and Condition 1.

The restrictions in Examples 7 and 8 involve only γ . Example 9 imposes restrictions on both \bar{Y} and γ .

Example 9 Consider the 2^2 factorial experiment from Example 3. Define $\tau_{AB,i} = 2^{-1}\{Y_i(- -) + Y_i(++)\} - 2^{-1}\{Y_i(-+) + Y_i(+ -)\}$ as the individual interaction effect for unit i . Assume that prior knowledge suggests that $\tau_{AB,i} = 0$ for all i . This implies

$$\tau_{AB} = 0, \quad (\gamma_{--} + \gamma_{++}) - (\gamma_{-+} + \gamma_{+-}) = 0_J, \quad (12)$$

which can in turn be used as a working restriction for fitting (6) by RLS. A matrix representation of the restriction (12) is $R\theta_L = 0_{J+1}$ with $R = \text{diag}(c_{AB}^T, c_{AB}^T \otimes I_J)$.

4.2 Design-based theory of restricted least squares

Write Equation (8) in matrix form as $Y = \chi_L \theta_L + \epsilon_L$, where $Y = (Y_1, \dots, Y_N)^T$, $\chi_L = (\chi_{L,1}, \dots, \chi_{L,N})^T$, and $\epsilon_L = (\epsilon_{L,1}, \dots, \epsilon_{L,N})^T$. The Gauss–Markov model assumes that Y has expectation $\chi_L \theta_L$ and covariance $\sigma^2 I_N$, and ensures the efficiency of \hat{Y}_r among all linear unbiased estimators when the restriction is correctly specified (Rao, 1973; Theil, 1971). The design-based framework violates these assumptions and leaves the sampling properties of \hat{Y}_r unclear. This is our focus for this subsection.

To this end, we first review in Lemma 2 the numeric expression of $\hat{\theta}_r$ free of any modelling assumptions. For simplicity, we assume that $\chi_L^T \chi_L$ is nonsingular throughout; see Greene and Seaks (1991) for more general formulas.

Lemma 2. $\hat{\theta}_r = (I - M_r R) \hat{\theta}_L + M_r r$ with $M_r = (\chi_L^T \chi_L)^{-1} R^T \{R(\chi_L^T \chi_L)^{-1} R^T\}^{-1}$.

The rest of this section is organized as follows. Section 4.2.1 introduces two types of restrictions as our focus for the main paper. Sections 4.2.2 and 4.2.3 give the sampling properties of \hat{Y}_r under the two restriction types, respectively. Section 4.2.4 then proposes a novel estimator of the sampling covariance of \hat{Y}_r . Throughout Section 4, we focus on not only the consistency of \hat{Y}_r but also its asymptotic normality and robust covariance estimation for large-sample Wald-type inference.

Recall that $\hat{\gamma}_{r,q}$ denotes the coefficient vector of $1(Z_i = q)x_i$ from the RLS fit of Equation (6). Theorem S3 in the [online supplementary material](#) ensures that $(\hat{\gamma}_{r,q})_{q \in \mathcal{T}}$ all have finite probability limits, denoted by $\gamma_{r,q} = \text{plim } \hat{\gamma}_{r,q}$, under complete randomization and Condition 4 regardless of whether the working restriction $R\theta_L = r$ is correctly specified or not. Let $S_r = (S_{r,qq'})_{q,q' \in \mathcal{T}}$ be the finite population covariance matrix of $\{Y_i(q) - \hat{\gamma}_{r,q}^T x_i : q \in \mathcal{T}\}_{i=1}^N$, analogous to S_* ($*$ = N, F, L). Let $V_r = \text{diag}(S_{r,qq}/e_q)_{q \in \mathcal{T}} - S_r$, analogous to V_* ($*$ = N, F, L). Lemma 3 gives a numeric result on V_r that underlies the asymptotic efficiency of \hat{Y}_r relative to \hat{Y}_L .

Lemma 3. $V_r \geq V_L$, where the equality holds if $\gamma_{r,q} = \gamma_q$ for all $q \in \mathcal{T}$.

4.2.1 Two types of restrictions

To simplify the presentation, we focus on the following two types of restrictions in the main paper due to their prevalence in practice and their ability to convey all main points of the general theory. We relegate the general theory to the [online supplementary material](#).

Definition 3 (i) A *separable* restriction restricts \bar{Y} separately from γ , with Equation (11) reduced to

$$\rho_Y \bar{Y} = r_Y, \quad \rho_\gamma \gamma = r_\gamma \quad (13)$$

for some prespecified ρ_Y and ρ_γ . Without loss of generality, assume that $(\rho_Y, r_Y) = (0_Q^T, 0)$ if there is no restriction on \bar{Y} , and that ρ_Y has full row rank if otherwise; likewise for (ρ_γ, r_γ) .

(ii) A *correlation-only* restriction restricts only γ , with Equation (13) reduced to

$$\rho_\gamma \gamma = r_\gamma. \quad (14)$$

The correlation-only restriction is a special type of the separable restriction with $(\rho_Y, r_Y) = (0_Q^T, 0)$. In Definition 3, $R = \text{diag}(\rho_Y, \rho_\gamma)$ and $r = (r_Y^T, r_\gamma^T)^T$ for a separable restriction with restrictions on both \bar{Y} and γ ; $R = (\rho_Y, 0)$ and $r = r_Y$ for a separable restriction with only restriction on \bar{Y} ; $R = (0, \rho_\gamma)$ and $r = r_\gamma$ for a correlation-only restriction with nonempty restriction on γ . An empty restriction is always correctly specified by definition.

Examples 7 and 8 are special cases of the correlation-only restriction, whereas Example 9 exemplifies the more general separable restriction. A common choice of ρ_Y is a contrast matrix with rows orthogonal to 1_Q , imposing restrictions on a set of finite population ATE, namely $\rho_Y \bar{Y}$. The $\tau_{AB} = 0$ in Example 9 is an example with $\rho_Y = c_{AB}^T$. More generally, the assumptions of no higher-order interactions for analysing factorial experiments all fall into this category; see Zhao and Ding (2022a). We will give more examples in Section 5.

4.2.2 \hat{Y}_r subject to the correlation-only restriction

Theorem 1. Assume complete randomization, Condition 4, and RLS subject to Equation (14). Then

- (i) $\sqrt{N}(\hat{Y}_r - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, V_r)$, where $V_r \geq V_L$ by Lemma 3;
- (ii) $V_r = V_L$ if (14) is correctly specified.

Theorem 1 establishes two theoretical guarantees of \hat{Y}_r from the correlation-only restriction. Theorem 1(i) ensures that it is always consistent and asymptotically normal regardless of whether Equation (14) is correctly specified or not. Theorem 1(ii) ensures that it attains the same asymptotic efficiency as \hat{Y}_L when Equation (14) is indeed correct. Simulation in Section 6 further suggests that \hat{Y}_r can have better finite sample performance than \hat{Y}_L as long as the restriction is not severely misspecified. Intuitively, the restriction on $\hat{\gamma}_r$, namely $\rho_\gamma \hat{\gamma}_r = r_\gamma$, reduces its variability relative to $\hat{\gamma}_L$ from OLS in finite samples. Such reduction in variability, despite having no effect on the asymptotic efficiency of \hat{Y}_r when Equation (14) is correctly specified, improves its precision in finite samples. This gives the third guarantee of \hat{Y}_r from the correlation-only restriction. It is our recommendation for mitigating the conundrum of many covariates and many treatments when the sample size is moderate.

Additional restriction on \bar{Y} , on the other hand, promises the possibility of additional efficiency over \hat{Y}_L . We elaborate on the details below.

4.2.3 \hat{Y}_r subject to the separable restriction with $\rho_Y \neq 0$

Let $\Pi = \text{diag}(e_q)_{q \in \mathcal{T}}$, and let

$$U = I_Q - \Pi^{-1} \rho_Y^T (\rho_Y \Pi^{-1} \rho_Y^T)^{-1} \rho_Y, \quad \mu_r = -\Pi^{-1} \rho_Y^T (\rho_Y \Pi^{-1} \rho_Y^T)^{-1} (\rho_Y \bar{Y} - r_Y) \quad (15)$$

for $\rho_Y \neq 0$.

Theorem 2. Assume complete randomization, Condition 4, and RLS subject to Equation (13) with $\rho_Y \neq 0$. Then

$$\sqrt{N}(\hat{Y}_r - \bar{Y} - \mu_r) \rightsquigarrow \mathcal{N}(0_Q, UV_r U^T), \quad (16)$$

where $V_r \geq V_L$ by Lemma 3. In particular, (i) $\mu_r = 0$ if $\rho_Y \bar{Y} = r_Y$ is correctly specified; (ii) $V_r = V_L$ if $\rho_\gamma \gamma = r_\gamma$ is correctly specified.

The asymptotic normality in Equation (16) implies $\hat{Y}_r - \bar{Y} = \mu_r + o_p(1)$ such that \hat{Y}_r is consistent if and only if $\mu_r = o(1)$. This is in general not true unless the restriction on \bar{Y} is correctly specified. Given that we can never verify the correctness of $\rho_Y \bar{Y} = r_Y$ exactly when $\rho_Y \neq 0$, this illustrates one advantage of imposing restriction on only γ .

Theorem 2(i) ensures the consistency and asymptotic normality of \hat{Y}_r when the restriction on \bar{Y} is indeed correctly specified. Theorem 2(ii) ensures that $UV_L U^T$ gives a lower bound of the asymptotic covariance of \hat{Y}_r when the restriction on γ is correctly specified. Whereas there is in general no definite order between $UV_L U^T$ and V_L , one exception is under the constant treatment effects condition which ensures $UV_L U^T \leq V_L$. Theorem 3 builds on this intuition, and establishes the asymptotic bias-variance trade-off between \hat{Y}_r and \hat{Y}_L when the restriction on \bar{Y} is nonempty. The result extends Zhao and Ding (2022a, Theorem A5) on unadjusted estimators to the covariate-adjusted variants, and ensures the asymptotic efficiency of \hat{Y}_r over \hat{Y}_L under constant treatment effects when the restriction is correctly specified.

Theorem 3. Assume complete randomization, Conditions 3–4, and RLS subject to Equation (13) with $\rho_Y \neq 0$ being a contrast matrix with rows orthogonal to 1_Q .

- (i) If $\rho_\gamma \gamma = r_\gamma$ is correctly specified, then $\hat{Y}_r - \bar{Y} = \mu_r + o_p(1)$ and \hat{Y}_r has smaller asymptotic covariance than \hat{Y}_L .
- (ii) If both $\rho_Y \bar{Y} = r_Y$ and $\rho_\gamma \gamma = r_\gamma$ are correctly specified, then \hat{Y}_r is consistent, asymptotically normal, and asymptotically more efficient than \hat{Y}_L . That is, $\sqrt{N}(\hat{Y}_r - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, UV_L U^T)$ with $UV_L U^T \leq V_L$.

Assume constant treatment effects and correctly specified restriction on γ . Theorem 3(i) states the reduction in asymptotic covariance by arbitrary restriction on contrasts of \bar{Y} at the cost of possibly nondiminishing bias. Further assume that the restriction on \bar{Y} is also correctly specified. Theorem 3(ii) ensures the asymptotic efficiency of \hat{Y}_r over \hat{Y}_L . As it turns out, the efficiency of \hat{Y}_r extends to a class of general linear consistent estimators of \bar{Y} . This gives the design-based counterpart of the classical Gauss–Markov theorem for RLS. We relegate the formal statement to Theorem S2 in the [online supplementary material](#).

Juxtapose Theorems 1–3: The restrictions on \bar{Y} and γ have distinct consequences on the sampling properties of \hat{Y}_r . The restriction on \bar{Y} , on the one hand, incurs nondiminishing bias when misspecified, yet promises lower asymptotic covariance than \hat{Y}_L under constant treatment effects. The choice of whether to restrict \bar{Y} is thus a trade-off between asymptotic bias and variance. The restriction on γ , on the other hand, retains consistency regardless of whether correctly specified or not, but undermines asymptotic efficiency when misspecified. Simulation studies further suggest that it can improve the finite sample performance of \hat{Y}_r . The choice of whether to restrict γ is thus a trade-off between finite sample performance and asymptotic efficiency.

4.2.4 Robust covariance for restricted least squares

Recall that $\hat{\theta}_r = (\hat{Y}_r^T, \hat{\gamma}_r^T)^T = (I - M_r R)\hat{\theta}_L + M_r r$ from Lemma 2. Let

$$\hat{\Sigma}_r = (\chi_L^T \chi_L)^{-1} \{ \chi_L^T \text{diag}(\hat{\epsilon}_{r,1}^2, \dots, \hat{\epsilon}_{r,N}^2) \} \{ \chi_L^T \chi_L \}^{-1}$$

be a variant of the EHW, also known as the *sandwich*, robust covariance estimator of $\hat{\theta}_L$, where we use the RLS residuals $\hat{\epsilon}_{r,i} = Y_i - t_i^T \hat{Y}_r - (t_i \otimes x_i)^T \hat{\gamma}_r$ in the middle. We define

$$(I - M_r R) \hat{\Sigma}_r (I - M_r R)^T = (I - M_r R) (\chi_L^T \chi_L)^{-1} \{ \chi_L^T \text{diag}(\hat{\epsilon}_{r,1}^2, \dots, \hat{\epsilon}_{r,N}^2) \} \{ \chi_L^T \chi_L \}^{-1} (I - M_r R)^T$$

as the *double-decker-taco robust* covariance estimator of $\hat{\theta}_r$ and use its upper-left $Q \times Q$ submatrix, denoted by $\hat{\Psi}_r$, to estimate the sampling covariance of \hat{Y}_r .

Recall the definition of S_r from the beginning of Section 4.2. Theorem 4 gives the probability limit of $\hat{\Psi}_r$.

Theorem 4. Assume complete randomization and Condition 4.

- (i) Under RLS subject to Equation (14), we have $N\hat{\Psi}_r = V_r + S_r + o_p(1)$ with $S_r \geq 0$.
- (ii) Under RLS subject to Equation (13) with $\rho_Y \neq 0$, we have

$$N\hat{\Psi}_r = UV_r U^T + U\{S_r + \text{diag}(\mu_{r,q}^2/e_q)_{q \in T}\}U^T + o_p(1),$$

where $\mu_{r,q}$ denotes the q th element of μ_r with $U\{S_r + \text{diag}(\mu_{r,q}^2/e_q)_{q \in T}\}U^T \geq 0$.

Further assume that $\rho_Y \bar{Y} = r_Y$ is correctly specified. Then $N\hat{\Psi}_r = UV_r U^T + US_r U^T + o_p(1)$ with $US_r U^T \geq 0$.

Recall that $\sqrt{N}(\hat{Y}_r - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, V_r)$ under RLS subject to Equation (14) from Theorem 1, and $\sqrt{N}(\hat{Y}_r - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, UV_r U^T)$ under RLS subject to Equation (13) when $\rho_Y \neq 0$ and $\rho_Y \bar{Y} = r_Y$ is correctly specified from Theorem 2. Theorem 4 thus ensures that $\hat{\Psi}_r$ is asymptotically conservative for estimating the true sampling covariance of \hat{Y}_r under both scenarios. This justifies the large-sample Wald-type inference of τ based on $\hat{\tau}_r = C\hat{Y}_r$ and $C\hat{\Psi}_r C^T$ as the point estimator and estimated covariance, respectively, when $\rho_Y \bar{Y} = r_Y$ is correctly specified.

In addition, recall that the unadjusted and additive regressions can be viewed as restricted variants of Equation (6), assuming the zero and equal correlation restrictions, respectively. The $\hat{\Psi}_r$ introduced above provides an alternative way to estimate the covariance of \hat{Y}_* ($*$ = N, F) in addition to the $\hat{\Psi}_*$'s from the OLS fits in Lemma 1. Let $\hat{\Psi}_{N,r}$ and $\hat{\Psi}_{F,r}$ be the values of $\hat{\Psi}_r$ from fitting (6) subject to the zero and equal correlation restrictions, respectively. Theorem 4(i) and Lemma 1(i) together ensure that $\hat{\Psi}_*$ and $\hat{\Psi}_{*,r}$ are asymptotically equivalent for $* = N, F$. Proposition S4 in the [online supplementary material](#) gives a stronger result on their numeric equivalence, i.e., $\hat{\Psi}_* = \hat{\Psi}_{*,r}$ ($*$ = N, F), free of any assumptions. This illustrates the equivalence between the double-decker-taco covariance estimator from the RLS fit and the EHW covariance estimator from the OLS fit of the corresponding restricted specification. See Lemma S3 in the [online supplementary material](#) for a general result.

4.3 Concluding remarks and an extension to rerandomization

Theorems 1–4 summarize our main results on the design-based properties of RLS for regression adjustment in multiarmed experiments. The RLS estimator $\hat{\tau}_r = C\hat{Y}_r$ includes $\hat{\tau}_*$ ($*$ = N, F) as special cases and ensures a triple guarantee for inferring τ : it is (i) asymptotically efficient when the restriction is correctly specified, (ii) consistent when the restriction on \bar{Y} is correctly specified and separate from that on γ , and (iii) can have better finite sample performance than the unrestricted $\hat{\tau}_l$ when the restriction is not severely misspecified. These, together with the asymptotic conservativeness of $C\hat{\Psi}_r C^T$ for estimating the true sampling covariance, suggest the advantage of RLS in finite samples.

The correctness of the restriction is central to the theoretical guarantees of $\hat{\tau}_r$. This can be assessed by testing the null hypothesis of $H_0: R\theta_L = r$. Let $\hat{\Sigma}_L$ be the EHW covariance estimator of $\hat{\theta}_L$ from the OLS fit of Equation (6). We propose to test H_0 with $W = (R\hat{\theta}_L - r)^T(R\hat{\Sigma}_L R^T)^{-1}(R\hat{\theta}_L - r)$ as the test statistic and compute a one-sided p -value by comparing W to χ_m^2 with $m = \text{rank}(R)$. The resulting test preserves the nominal type one error rates asymptotically. We give the details in the [online supplementary material](#).

Rerandomization gives another way to incorporate covariate information in the design stage of experiments (Morgan & Rubin, 2012). It accepts a treatment allocation if and only if it satisfies some prespecified covariate balance criterion. Let $\hat{x}(q) = N_q^{-1} \sum_{i: Z_i=q} x_i$ denote the sample mean of covariates under treatment level $q \in T$. Contrasts of $\hat{x}(q)$'s provide intuitive measures of covariate balance across the Q treatment groups, and allow us to form covariate balance criterion based on the Mahalanobis distance of their concatenation to the origin (Branson et al., 2016;

(Li et al., 2020). The resulting inference based on $\hat{\tau}_r$ inherits all guarantees from inference under complete randomization and, in addition, ensures less loss in asymptotic efficiency when the restriction is misspecified. Due to space limitations, we relegate the theory to the [online supplementary material](#). The result extends the existing literature on rerandomization and highlights its value for additional protection against model misspecification.

5 Regression adjustment in factorial experiments

5.1 Overview of factorial experiments

Factorial experiments are a special type of multiarmed experiments, featuring treatments as combinations of two or more factors, each of two or more levels. The treatment-based regressions provide a principled way of studying factorial experiments, enabling inference of arbitrary ATE by least squares. Despite their generality and nice theoretical guarantees, however, they are not the dominant choice for analysing factorial data in practice. Factor-based regressions, as a more popular approach, regress the observed outcome directly on the factors themselves and interpret the coefficients as the corresponding factorial effects of interest. This enables not only direct inference of the treatment effects based on regression outputs but also flexible unsaturated specifications to reduce model complexity.

The existing literature on covariate adjustment for factor-based regressions focuses on *factor-saturated* specifications that include all possible interactions between the factors (Lu, 2016b; Zhao & Ding, 2022b). Recall from Example 5 that a general factorial experiment has in total $Q = \prod_{k=1}^K Q_k$ treatment levels, with $Q \geq 2^K$. The resulting regressions contain $Q + J$ and $Q(1 + J)$ estimated coefficients under the additive and fully interacted specifications, respectively, subjecting subsequent inference to large finite sample variability even when K is moderate. We extend the discussion to *factor-unsaturated* regressions and establish their properties for covariate adjustment from the design-based perspective. The resulting theory is not only of practical relevance in itself given the rising popularity of factorial experiments but also illustrates the value of RLS for studying estimators from general OLS regressions. We will focus on 2^K factorial experiments for notational simplicity. The result extends to the general factorial experiments in Example 5 with minimal modification; see also Zhao and Ding (2022a, Section A).

5.2 Standard factorial effects for 2^K factorial experiments

Consider a 2^K factorial experiment with K factors of interest, $k = 1, \dots, K$, each of two levels. Inherit the notation from Example 4. The $Q = 2^K$ treatment levels are $q = (z_1, \dots, z_K) \in \mathcal{T} = \{-1, +1\}^K$, where $z_k \in \{-1, +1\}$ indicates the level of factor k .

Let $\mathcal{P}_K = \{K : K \subseteq [K] \text{ and } K \neq \emptyset\}$ denote the set of the $2^K - 1$ nonempty subsets of $[K] = \{1, \dots, K\}$. There are $2^K - 1$ standard factorial effects under the 2^K factorial experiment, one for each $\mathcal{K} \in \mathcal{P}_K$, characterizing the main effect or $|\mathcal{K}|$ -way interaction of the factor(s) in \mathcal{K} for $|\mathcal{K}| = 1$ and $|\mathcal{K}| \geq 2$, respectively. The $(\tau_A, \tau_B, \tau_{AB})$ in Example 3 are special cases with $\mathcal{P}_2 = \{\{A\}, \{B\}, \{A, B\}\}$. Let $\tau_{\mathcal{K}} = c_{\mathcal{K}}^T \bar{Y}$ denote the standard factorial effect corresponding to $\mathcal{K} \in \mathcal{P}_K$. Lemma 4 gives the numeric properties of $c_{\mathcal{K}}$'s (Wu & Hamada, 2009).

Lemma 4. For all $\mathcal{K} \neq \mathcal{K}' \in \mathcal{P}_K$, we have (i) $c_{\mathcal{K}}^T \mathbf{1}_Q = 0$; (ii) $2^{K-1} c_{\mathcal{K}} \in \{-1, +1\}^Q$; (iii) $c_{\mathcal{K}}^T c_{\mathcal{K}'} = 0$.

By Lemma 4, the standard factorial effects are *orthogonal* in terms of the contrast vectors that define them. Example 10 gives the formulas for the main effects and two-way interactions; see Dasgupta et al. (2015) and Li et al. (2020) for formulas for the higher-order interactions.

Example 10 The standard main effect of factor k equals

$$\tau_{\{k\}} = \frac{1}{2^{K-1}} \sum_{q: z_k=+1} \bar{Y}(q) - \frac{1}{2^{K-1}} \sum_{q: z_k=-1} \bar{Y}(q),$$

comparing the average potential outcomes when factor k is at level $+1$ and level -1 , respectively. The standard interaction effect between factors k

and k' equals

$$\tau_{[k,k']} = \frac{1}{2^{K-1}} \sum_{q:z_k z_{k'}=+1} \bar{Y}(q) - \frac{1}{2^{K-1}} \sum_{q:z_k z_{k'}=-1} \bar{Y}(q),$$

comparing the average potential outcomes when the two factors take the same and different levels, respectively. The $(\tau_A, \tau_B, \tau_{AB})$ in Example 3 are special cases at $K = 2$.

5.3 Factor-saturated regressions for 2^K factorial experiments

Let $Z_{ik} \in \{-1, +1\}$ indicate the level of factor k received by unit i . For $\mathcal{K} \in \mathcal{P}_K$, let $Z_{i,\mathcal{K}} = \prod_{k \in \mathcal{K}} Z_{ik}$ represent the interaction between the factors in \mathcal{K} . The classical experimental design literature takes

$$Y_i \sim 1 + \sum_{k=1}^K Z_{ik} + \sum_{k \neq k'} Z_{ik} Z_{ik'} + \cdots + \prod_{k=1}^K Z_{ik} \iff Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} \quad (17)$$

as the standard specification for factor-based regression analysis, and estimates $\tau_{\mathcal{K}}$ by 2 times the OLS coefficient of $Z_{i,\mathcal{K}}$ (Lu, 2016a; Wu & Hamada, 2009). The $Y_i \sim 1 + Z_i$ under the treatment-control experiment and the $Y_i \sim 1 + Z_{i1} + Z_{i2} + Z_{i1}Z_{i2}$ under the 2^2 factorial experiment are both special cases with $K = 1$ and $K = 2$, respectively. We call (17) the *factor-saturated* unadjusted regression, which includes all possible interactions between elements of $(Z_{ik})_{k=1}^K$. The presence of covariates further motivates

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} + x_i, \quad (18)$$

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} + x_i + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} \cdot x_i \quad (19)$$

as the additive and fully interacted variants, respectively.

Regressions (17)–(19) define three factor-saturated specifications for factor-based analysis of the 2^K factorial experiment, paralleling (4)–(6) under the treatment-based formulation. The upper panel of Table 1 summarizes them. Let $c_{\emptyset} = 2^{-(K-1)} 1_Q$ be an orthogonal complement of the linear span of $\{c_{\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\}$ by Lemma 4. The one-to-one correspondence between t_i and $\{1, Z_{i,\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\}$ ensures that $2^{-1}\tau_{\mathcal{K}}$, $2^{-1}(c_{\emptyset}^T \otimes I_J)\gamma$, and $2^{-1}(c_{\mathcal{K}}^T \otimes I_J)\gamma$ give the target parameters of $Z_{i,\mathcal{K}}$, x_i , and $Z_{i,\mathcal{K}} \cdot x_i$ for $\mathcal{K} \in \mathcal{P}_K$ in Equation (19), respectively; see the proof of Proposition S5 in the [online supplementary material](#). This gives the intuition behind using 2 times the OLS coefficient of $Z_{i,\mathcal{K}}$ for estimating $\tau_{\mathcal{K}}$.

Let $\tilde{\tau}_{*,\mathcal{K}}$ ($*$ = N, F, L) be 2 times the coefficients of $Z_{i,\mathcal{K}}$ from the OLS fits of Equations (17)–(19), respectively, vectorized as

$$\tilde{\tau}_* = \{\tilde{\tau}_{*,\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\}.$$

Let $\tilde{\Omega}_*$ be the EHW covariance estimator of $\tilde{\tau}_*$ from the corresponding OLS fit. As a convention, we use $* = N, F, L$ to signify the unadjusted, additive, and fully interacted specifications, respectively, and use the tilde (\sim) to signify outputs from factor-based regressions. Let

$$\tau_s = \{\tau_{\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\} = C_s \bar{Y}$$

be the vectorization of $\tau_{\mathcal{K}}$'s in the same order of \mathcal{K} as in $\tilde{\tau}_*$. Then C_s is a $(Q-1) \times Q$ contrast matrix with $\{c_{\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\}$ as its row vectors. Proposition 1 follows from the invariance of OLS to nondegenerate linear transformation of the regressors, and justifies the large-sample Wald-type inference

Table 1. Six factor-based regressions under the 2^K factorial experiment

| Base model | Regression equation | OLS estimator of τ_K | OLS estimator of $\tau_{s,+}$ |
|---|---|---------------------------|-------------------------------|
| $Y_i \sim 1 + \sum_{K \in \mathcal{P}_K} Z_{i,K}$ (factor-saturated) | (17): $Y_i \sim 1 + \sum_{K \in \mathcal{P}_K} Z_{i,K}$ (18): $Y_i \sim 1 + \sum_{K \in \mathcal{P}_K} Z_{i,K} + x_i$ | $\tilde{\tau}_{N,K}$ | $\tilde{\tau}_{N,+}$ |
| $Y_i \sim 1 + \sum_{K \in \mathcal{F}_+} Z_{i,K}$ (factor-unsaturated) | (19): $Y_i \sim 1 + \sum_{K \in \mathcal{P}_K} Z_{i,K} + x_i + \sum_{K \in \mathcal{P}_K} Z_{i,K} \cdot x_i$ (20): $Y_i \sim 1 + \sum_{K \in \mathcal{F}_+} Z_{i,K}$ | $\tilde{\tau}_{F,K}$ | $\tilde{\tau}_{F,+}$ |
| | (21): $Y_i \sim 1 + \sum_{K \in \mathcal{F}_+} Z_{i,K} + x_i$ (22): $Y_i \sim 1 + \sum_{K \in \mathcal{F}_+} Z_{i,K} + x_i + \sum_{K \in \mathcal{F}_+} Z_{i,K} \cdot x_i$ | $\tilde{\tau}_{L,K}$ | $\tilde{\tau}_{L,+}$ |
| | | $\tilde{\tau}_{N,u,K}$ | $\tilde{\tau}_{N,u,+}$ |
| | | $\tilde{\tau}_{F,u,K}$ | $\tilde{\tau}_{F,u,+}$ |
| | | $\tilde{\tau}_{L,u,K}$ | $\tilde{\tau}_{L,u,+}$ |

of τ_s based on $(\tilde{\tau}_*, \tilde{\Omega}_*)$. The results on $\tilde{\tau}_N$ and $\tilde{\tau}_L$ are not new (Lu, 2016b; Zhao & Ding, 2022a), whereas that on $\tilde{\tau}_F$ is.

Recall $\hat{\Psi}_*$ as the EHW covariance estimator of \hat{Y}_* ($*$ = N, F, L).

Proposition 1. $\tilde{\tau}_* = C_s \hat{Y}_*$ and $\tilde{\Omega}_* = C_s \hat{\Psi}_* C_s^T$ for $*$ = N, F, L.

Proposition 1 is numeric, and ensures the equivalence between the factor-saturated specifications and their treatment-based counterparts for estimating τ_s . The relative efficiency between $\tilde{\tau}_*$ ($*$ = N, F, L) follows immediately from Lemma 1, with $\tilde{\tau}_L$ being asymptotically the most efficient.

5.4 Factor-unsaturated regressions

Despite the conceptual straightforwardness and nice theoretical guarantees, the factor-saturated regressions (17)–(19) involve 2^K , $2^K + J$, and $2^K(1 + J)$ estimated coefficients, respectively, subjecting subsequent inference to large finite sample variability even when K is moderate (Zhao & Ding, 2022a). Often, only the main effects and two-way interactions are of interest, with the higher-order effects believed to be small or nonexistent. A common practice is to include only the relevant terms, and estimate the effects of interest from first- or second-order specifications like $Y_i \sim 1 + \sum_{k=1}^K Z_{ik}$ or $Y_i \sim 1 + \sum_{k=1}^K Z_{ik} + \sum_{k \neq k'} Z_{ik} Z_{ik'}$.

More generally, assume that we are interested in only a subset of the $2^K - 1$ standard factorial effects, summarized by

$$\tau_{s,+} = \{\tau_K : K \in \mathcal{F}_+\} = C_{s,+} \bar{Y}$$

for some $\mathcal{F}_+ \subsetneq \mathcal{P}_K$ with $\mathcal{F}_- = \mathcal{P}_K \setminus \mathcal{F}_+ \neq \emptyset$. The $C_{s,+}$ is a $|\mathcal{F}_+| \times Q$ submatrix of C_s with $\{c_K : K \in \mathcal{F}_+\}$ as its row vectors. As a convention, we use + and – in the subscripts to indicate *effects of interest* and *nuisance effects*, respectively. The discussion from Section 5.3 ensures that the vectors

$$\tilde{\tau}_{*,+} = \{\tilde{\tau}_{*,K} : K \in \mathcal{F}_+\} \quad (* = N, F, L)$$

give three estimators of $\tau_{s,+}$ from the factor-saturated regressions (17)–(19). They are all consistent by Proposition 1 but may have substantial finite sample variability when K is large.

Alternatively, recall that we estimate τ_K by 2 times the OLS coefficient of $Z_{i,K}$. Now that we are interested in only the factorial combinations in \mathcal{F}_+ , we can define

$$Y_i \sim 1 + \sum_{K \in \mathcal{F}_+} Z_{i,K}, \tag{20}$$

$$Y_i \sim 1 + \sum_{K \in \mathcal{F}_+} Z_{i,K} + x_i, \tag{21}$$

$$Y_i \sim 1 + \sum_{K \in \mathcal{F}_+} Z_{i,K} + x_i + \sum_{K \in \mathcal{F}_+} Z_{i,K} \cdot x_i \tag{22}$$

as the *factor-unsaturated* variants of Equations (17)–(19), respectively, to include only the terms relevant to $\mathcal{K} \in \mathcal{F}_+$, and use the resulting coefficients to estimate $\tau_{s,+}$. This gives three additional regression specifications for estimating $\tau_{s,+}$, with the first- and second-order specifications above being special cases with $\mathcal{F}_+ = \{\{k\} : k \in [K]\}$ and $\mathcal{F}_+ = \{\{k\}, \{k, k'\} : k \neq k' \in [K]\}$, respectively. They are arguably the more commonly used approach in practice, allowing for considerable reduction in model complexity. Of interest is the implications of such reduction on subsequent inference. We address this question in Section 5.5.

Let

$$\tau_{s,-} = \{\tau_{\mathcal{K}} : \mathcal{K} \in \mathcal{F}_-\} = C_{s,-}\bar{Y}$$

denote the vector of nuisance effects that complements $\tau_{s,+}$. Then $C_{s,-}$ is a $(Q - 1 - |\mathcal{F}_+|) \times Q$ submatrix of C_s with $\{c_{\mathcal{K}} : \mathcal{K} \in \mathcal{F}_-\}$ as its row vectors. The intuition on target parameters after (19) allows us to view (20)–(22) as restricted variants of Equation (19) subject to working restrictions

$$N: \quad \tau_{s,-} = 0, \quad \gamma = 0; \quad (23)$$

$$F: \quad \tau_{s,-} = 0, \quad \gamma_1 = \dots = \gamma_Q; \quad (24)$$

$$L: \quad \tau_{s,-} = 0, \quad (C_{s,-} \otimes I_J)\gamma = 0, \quad (25)$$

respectively. In particular, Lemma 4 ensures that the $\gamma = 0$ in Equation (23) gives the concise form of $(c_{\mathcal{K}}^T \otimes I_J)\gamma = 0$ for all $\mathcal{K} \in \{\emptyset\} \cup \mathcal{P}_K$, corresponding to Equation (20). Likewise for the $\gamma_1 = \dots = \gamma_Q$ in Equation (24) and $(C_{s,-} \otimes I_J)\gamma = 0$ in Equation (25) to give the concise forms of $(c_{\mathcal{K}}^T \otimes I_J)\gamma = 0$ for all $\mathcal{K} \in \mathcal{P}_K$ and $(c_{\mathcal{K}}^T \otimes I_J)\gamma = 0$ for all $\mathcal{K} \in \mathcal{F}_-$, respectively, corresponding to Equations (21) and (22).

5.5 Design-based properties of the factor-unsaturated regressions

Let $\tilde{\tau}_{*,u,\mathcal{K}}$ ($* = N, F, L$) denote 2 times the coefficients of $Z_{i,\mathcal{K}}$ from the OLS fits of Equations (20)–(22), respectively. This gives three additional estimators of $\tau_{\mathcal{K}}$ for $\mathcal{K} \in \mathcal{P}_K$, summarized in the lower panel of Table 1. We use the subscript ‘ u ’ to signify their origins from the factor-unsaturated regressions. Let

$$\tilde{\tau}_{*,u,+} = \{\tilde{\tau}_{*,u,\mathcal{K}} : \mathcal{K} \in \mathcal{F}_+\} \quad (* = N, F, L)$$

be the corresponding estimators of $\tau_{s,+}$. This, together with the $\tilde{\tau}_{*,+}$ ’s from the factor-saturated regressions, gives in total six estimators of $\tau_{s,+}$, summarized in the last column of Table 1.

Let $\tilde{\Omega}_{*,u,+}$ be the EHW covariance estimator of $\tilde{\tau}_{*,u,+}$ from the corresponding OLS fits of Equations (20)–(22). Let $\hat{Y}_{N,r}$, $\hat{Y}_{F,r}$, and $\hat{Y}_{L,r}$ be the coefficient vectors of t_i from the RLS fits of Equation (6) subject to working restrictions (23)–(25), respectively. They are all special cases of \hat{Y}_r with separate restrictions on \bar{Y} and γ . Let $\hat{\Psi}_{*,r}$ be the corresponding double-decker-taco covariance estimator of $\hat{Y}_{*,r}$ ($* = N, F, L$) from Section 4.2.4. Proposition 2 states the numeric correspondence between $(\tilde{\tau}_{*,u,+}, \tilde{\Omega}_{*,u,+})$ and $(\hat{Y}_{*,r}, \hat{\Psi}_{*,r})$.

Proposition 2. $\tilde{\tau}_{*,u,+} = C_{s,+}\hat{Y}_{*,r}$ and $\tilde{\Omega}_{*,u,+} = C_{s,+}\hat{\Psi}_{*,r}C_{s,+}^T$ for $* = N, F, L$.

The design-based properties of $(\tilde{\tau}_{*,u,+}, \tilde{\Omega}_{*,u,+})$ then follow from those of $(\hat{Y}_{*,r}, \hat{\Psi}_{*,r})$ in Theorems 2–4. In particular, Theorem 2 implies that $\tilde{\tau}_{*,u,+}$ is in general not consistent for estimating $\tau_{s,+}$ unless $\tau_{s,-} = 0$ is correctly specified. When $\tau_{s,-}$ is indeed 0, Corollary 1 follows from Theorems 2 and 4, and justifies the large-sample Wald-type inference of $\tau_{s,+}$ based on $(\tilde{\tau}_{*,u,+}, \tilde{\Omega}_{*,u,+})$.

Recall the definition of V_r from the beginning of Section 4.2. Let $V_{*,r}$ be the value of V_r associated with $\hat{Y}_{*,r}$ for $* = N, F, L$. Let $(U_s, \mu_{r,s})$ be the values of (U, μ_r) at $\rho_Y = C_{s,-}$ from Equation (15).

Corollary 1. Assume complete randomization, Condition 4, and $\tau_{s,-} = 0$. For $* = N, F, L$, we have $\sqrt{N}(\tilde{\tau}_{*,u,+} - \tau_{s,+}) \rightsquigarrow \mathcal{N}(0, C_{s,+} U_s V_{*,r} U_s^T C_{s,+}^T)$, with $\tilde{\Omega}_{*,u,+}$ being asymptotically conservative for estimating the true sampling covariance of $\tilde{\tau}_{*,u,+}$.

Corollary 2 follows from Theorem 3(ii) and gives the asymptotic relative efficiency of $\{\tilde{\tau}_{*,+}, \tilde{\tau}_{*,u,+} : * = N, F, L\}$ under constant treatment effects.

Corollary 2. Assume complete randomization, Conditions 3–4, and $\tau_{s,-} = 0$.

- (i) Among $\tilde{\tau}_{*,+}$ ($* = N, F, L$) from the factor-saturated regressions (17)–(19), $\tilde{\tau}_{F,+}$ and $\tilde{\tau}_{L,+}$ are asymptotically equally efficient, and are both more efficient than $\tilde{\tau}_{N,+}$.
- (ii) Among $\tilde{\tau}_{*,u,+}$ ($* = N, F, L$) from the factor-unsaturated regressions (20)–(22), $\tilde{\tau}_{F,u,+}$ and $\tilde{\tau}_{L,u,+}$ are asymptotically equally efficient, and are both more efficient than $\tilde{\tau}_{N,u,+}$.
- (iii) The $\tilde{\tau}_{*,u,+}$ from the factor-unsaturated regression is asymptotically more efficient than $\tilde{\tau}_{*,+}$ from the factor-saturated regression for $* = N, F, L$.

Assume Condition 3 of constant treatment effects. Corollary 2 establishes the asymptotic efficiency of $\tilde{\tau}_{F,u,+}$ and $\tilde{\tau}_{L,u,+}$ among $\{\tilde{\tau}_{*,+}, \tilde{\tau}_{*,u,+} : * = N, F, L\}$ for estimating $\tau_{s,+}$. Intuitively, Condition 3 implies the equal correlation condition, which ensures that the additive adjustment is asymptotically as efficient as the fully interacted adjustment to begin with. This underlies the asymptotic equivalence between $\tilde{\tau}_{F,+}$ and $\tilde{\tau}_{L,+}$ in Corollary 2(i), and that between $\tilde{\tau}_{F,u,+}$ and $\tilde{\tau}_{L,u,+}$ in Corollary 2(ii). The additional, correct knowledge on the nuisance effects further secures extra precision of $\tilde{\tau}_{F,u,+}$ and $\tilde{\tau}_{L,u,+}$ over $\tilde{\tau}_{F,+}$ and $\tilde{\tau}_{L,+}$, as formalized by Corollary 2(iii). This illustrates the value of factor-unsaturated regressions in combination with covariate adjustment for improving efficiency.

A key limitation of the factor-unsaturated regressions is that the consistency of $\tilde{\tau}_{*,u,+}$ depends critically on the actual absence of the nuisance effects. This can never be verified exactly in reality and subjects subsequent inference to possibly nondiminishing biases. This suggests the merit of the factor-saturated additive regression (18) as a compromise between asymptotic bias, asymptotic efficiency, and finite sample performance. The resulting inference is always consistent, and ensures asymptotic efficiency under equal correlations. Simulation in Section 6 further demonstrates its finite sample advantage over the asymptotically more efficient fully interacted counterpart (19).

One exception is when the treatment groups are of equal size. The resulting $\tilde{\tau}_{*,u,+}$ is consistent even when $\tau_{s,-} \neq 0$ thanks to the orthogonality of c_K 's from Lemma 4. We formalize the intuition in Proposition 3.

Condition 5 (Equal-sized design). $e_q = Q^{-1}$ for all $q \in \mathcal{T}$.

Proposition 3. Assume complete randomization and Conditions 4–5. Then

- (i) $\sqrt{N}(\tilde{\tau}_{*,u,+} - \tau_{s,+}) \rightsquigarrow \mathcal{N}(0, C_{s,+} V_{*,r} C_{s,+}^T)$ for $* = N, F, L$, with $\tilde{\Omega}_{*,u,+}$ being asymptotically conservative for estimating the true sampling covariance of $\tilde{\tau}_{*,u,+}$;
- (ii) $\tilde{\tau}_{L,+}$ is asymptotically more efficient than $\tilde{\tau}_{*,u,+}$ for $* = N, F, L$;
- (iii) $\tilde{\tau}_{L,u,+}$ is asymptotically as efficient as $\tilde{\tau}_{L,+}$ if $(C_{s,-} \otimes I_J)\gamma = 0$;
 $\tilde{\tau}_{F,u,+}$ and $\tilde{\tau}_{L,u,+}$ are asymptotically as efficient as $\tilde{\tau}_{L,+}$ under Condition 2;
 $\tilde{\tau}_{*,u,+}$ ($* = N, F, L$) are asymptotically as efficient as $\tilde{\tau}_{L,+}$ under Condition 1.

Proposition 3(i) justifies the large-sample Wald-type inference of $\tau_{s,+}$ based on $(\tilde{\tau}_{*,u,+}, \tilde{\Omega}_{*,u,+})$ in equal-sized designs regardless of whether $\tau_{s,-} = 0$ or not. The result for $\tilde{\tau}_{N,u,+}$ is a direct consequence of the numeric equivalence between $\tilde{\tau}_{N,u,+}$ and $\tilde{\tau}_{N,+}$ under equal-sized designs: $\tilde{\tau}_{N,u,+} = C_{s,+} \hat{Y}_N = \tilde{\tau}_{N,+}$; see Zhao and Ding (2022a). We give similar results on $\tilde{\tau}_{F,u,+}$ and $\tilde{\tau}_{L,u,+}$ in the [online supplementary material](#). This additional protection against model misspecification echoes the emphasis of the classical experimental design literature on equal treatment group sizes and orthogonality of the factorial effects.

Table 2. Six factor-based regressions and their respective restrictions relative to Equation (19)

| | Regression equation | Restriction relative to Equation (19) | If correctly specified |
|------|--|--|------------------------|
| N | (17): $Y_i \sim 1 + A_i + B_i + A_i B_i$ | $\gamma_q = 0_J$ for all $q \in \mathcal{T}$ | No |
| F | (18): $Y_i \sim 1 + A_i + B_i + A_i B_i + x_i$ | γ_q 's are all equal | No |
| L | (19): $Y_i \sim 1 + A_i + B_i + A_i B_i + x_i + A_i x_i + B_i x_i + A_i B_i x_i$ | no restriction | Yes |
| N_us | (20): $Y_i \sim 1 + A_i + B_i$ | $\tau_{AB} = 0$ and $\gamma_q = 0_J$ for all $q \in \mathcal{T}$ | No |
| F_us | (21): $Y_i \sim 1 + A_i + B_i + x_i$ | $\tau_{AB} = 0$ and γ_q 's are all equal | No |
| L_us | (22): $Y_i \sim 1 + A_i + B_i + x_i + A_i x_i + B_i x_i$ | $\tau_{AB} = 0$ and $(\gamma_{--} + \gamma_{++}) - (\gamma_{-+} + \gamma_{+-}) = 0_J$ | Yes |

Note. We use 'N', 'F', and 'L' to indicate the unadjusted, additive, and fully interacted adjustment schemes, respectively, and use the suffix '_us' to indicate the factor-unsaturated variants.

The asymptotic efficiency of $\tilde{\tau}_{*,u,+}$ nevertheless still requires the associated restriction on γ in Equations (23)–(25) to be correctly specified, and can no longer exceed that of $\tilde{\tau}_{l,+}$, as demonstrated by Proposition 3(ii)–(iii). In particular, the condition $(C_{s,-} \otimes I_J)\gamma = 0$ implies that the restriction on γ in Equation (25) is correctly specified, and ensures that $\tilde{\tau}_{l,u,+}$ attains the same asymptotic efficiency as $\tilde{\tau}_{l,+}$. Condition 2 implies that the restrictions on γ in Equations (24) and (25) are both correctly specified, and ensures that $\tilde{\tau}_{f,u,+}$ and $\tilde{\tau}_{l,u,+}$ attain the same asymptotic efficiency as $\tilde{\tau}_{l,+}$. Condition 1 implies that the restrictions on γ in Equations (23)–(25) are all correctly specified, and ensures that all three $\tilde{\tau}_{*,u,+}$ ($*$ = N, F, L) attain the same asymptotic efficiency as $\tilde{\tau}_{l,+}$. This illustrates the asymptotic bias-variance trade-off regarding the use of equal-sized designs, which ensure the consistency of $\tilde{\tau}_{*,u,+}$ ($*$ = N, F, L) but preclude the possibility of additional asymptotic efficiency over $\tilde{\tau}_{l,+}$ (cf. Theorem 3).

5.6 Extensions

We focused on specifications (17)–(22) because of their intuitiveness. The same results extend to general specifications like $Y_i \sim 1 + \sum_{K \in \mathcal{F}_+} Z_{i,K} + x_i + \sum_{K \in \mathcal{F}'_+} Z_{i,K} \cdot x_i$ for possibly different $\mathcal{F}_+, \mathcal{F}'_+ \subseteq \mathcal{P}_K$ with minimal modification. We relegate the details to the [Supplementary material](#).

In addition, we focused on specifications under the {−1, +1} coding system for inference of the standard factorial effects. Practitioners instead often encode the factor levels by {0, 1}. The resulting specifications, despite prevalent in practice, cannot recover the standard factorial effects directly as regression coefficients (Zhao & Ding, 2022a). The invariance of OLS to nondegenerate linear transformation of the regressors, on the other hand, ensures that all results so far extend to the {0, 1}-coded regressions for estimating a different set of factorial effects with minimal modification. We relegate the details to the [online supplementary material](#).

6 Simulation

We now illustrate the finite sample properties of the proposed method by simulation. Consider a 2^2 factorial experiment with $Q = 4$ treatment levels, $q \in \mathcal{T} = \{(-), (-+), (+-), (++)\}$, and a study population of $N = 100$ units, $i = 1, \dots, N$. We choose the moderate sample size on purpose to illustrate the limitation of the factor-saturated fully interacted regression in finite samples.

Inherit the notation from Example 3. For each i , we draw a $J = 20$ dimensional covariate vector as $x_i \sim \mathcal{N}(0_J, I_J)$, and generate the potential outcomes as $Y_i(q) \sim \mathcal{N}(x_i^\top \beta_q, 1)$ for $q = (-+), (+-)$, and $(++)$ and $Y_i(--) = Y_i(-+) + Y_i(-) - Y_i(++)$. The resulting $Y_i(q)$'s satisfy the condition of zero individual interaction effects from Example 9, and ensure that working restriction (12) is correctly specified with $\tau_{AB} = 0$ and $(\gamma_{--} + \gamma_{++}) - (\gamma_{-+} + \gamma_{+-}) = 0_J$. We fix $\{Y_i(q), x_i : q \in \mathcal{T}\}_{i=1}^N$ in the simulation and generate the treatment assignments from complete randomization with $(N_{--}, N_{-+}, N_{+-}, N_{++}) = (22, 23, 24, 31)$.

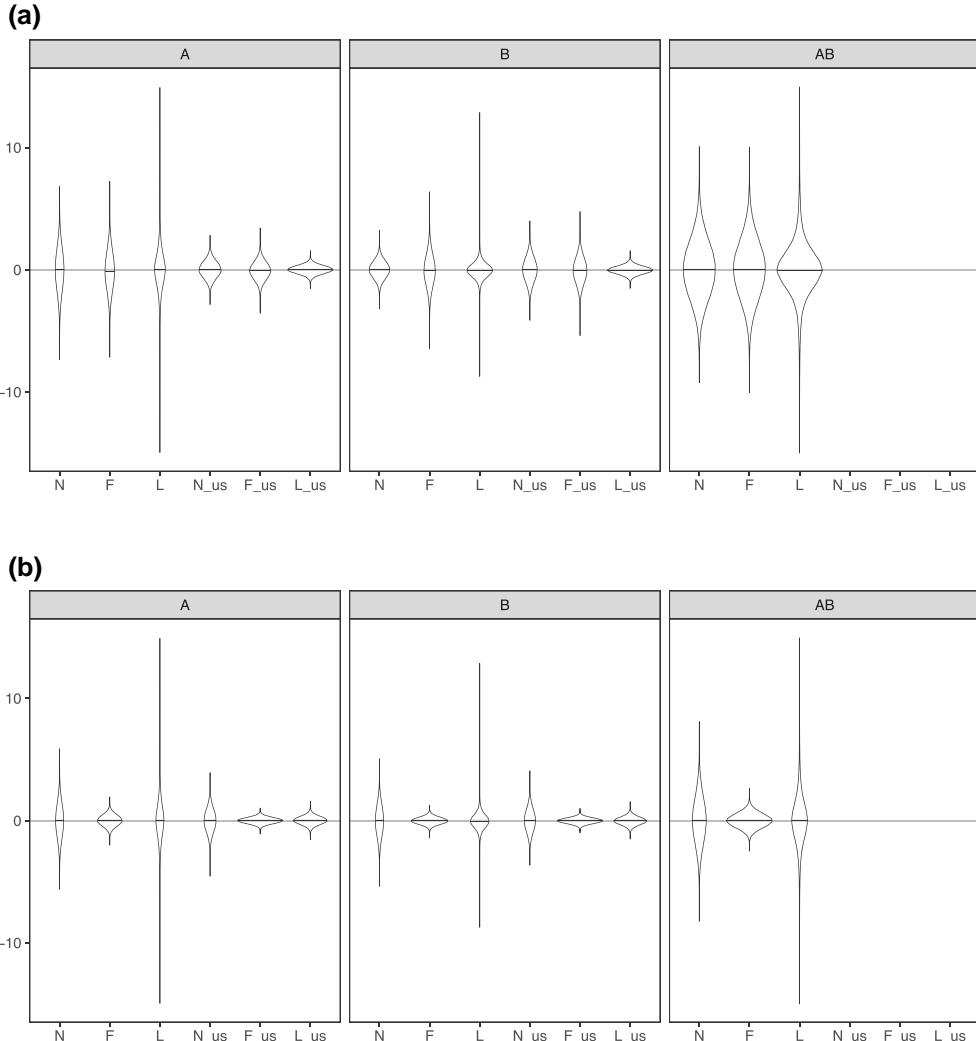


Figure 1. Violin plot of the differences between 2 times the ordinary least square coefficients of (A_i, B_i, A_iB_i) and the true values of $(\tau_A, \tau_B, \tau_{AB})$ over 100,000 independent complete randomizations. We use 'N', 'F', and 'L' to represent the factor-saturated unadjusted, additive, and fully interacted regressions, respectively, and use 'N_us', 'F_us', and 'L_us' to represent their respective factor-unsaturated variants. See Table 2. (a) $(\beta_{-+}, \beta_{+-}, \beta_{++}) = (1_J, 0_J, -1_J)$ such that the γ_q 's differ considerably. (b) $\beta_{-+} = \beta_{+-} = \beta_{++} = 1_J$ such that the γ_q 's are fairly close.

Consider six models for estimating $\tau_s = (\tau_A, \tau_B, \tau_{AB})^T$ by OLS, summarized in Table 2. They are special cases of regressions (17)–(22), respectively, with $K = 2$ and $\mathcal{F}_+ = \{\{A\}, \{B\}\}$. We use A_i and B_i to indicate the levels of factors A and B for unit i for intuitiveness. The equivalence between RLS and OLS on the corresponding restricted specification ensures that the OLS fit of Equation (22) is equivalent to the RLS fit of Equation (19) subject to Equation (12), which is correctly specified. The OLS fits of Equations (17), (18), (20), and (21), on the other hand, are equivalent to fitting (19) subject to restrictions that are misspecified, summarized in the last column of Table 2. Observe that β_q gives a good approximation to the finite population correlation γ_q by construction for $q = (-+), (+-), (++)$. The variability in $(\beta_{-+}, \beta_{+-}, \beta_{++})$ hence reflects the closeness of the ground truth to the equal correlation condition that justifies the additive regression (18) and its factor-unsaturated variant (21).

Figure 1 shows the violin plot of the differences between 2 times the OLS coefficients of (A_i, B_i, A_iB_i) and the true values of $(\tau_A, \tau_B, \tau_{AB})$ over 100,000 independent complete randomizations. **Figure 1a** corresponds to potential outcomes generated from $(\beta_{-+}, \beta_{+-}, \beta_{++}) = (1_J, 0_J, -1_J)$. The heterogeneity among β_q 's suggests considerable deviation from the equal correlation condition such that regressions (18) ('F') and (21) ('F_us') are both substantially misspecified. **Figure 1b** corresponds to potential outcomes generated from $\beta_{-+} = \beta_{+-} = \beta_{++} = 1_J$. The equality of β_q 's suggests reasonable closeness to the equal correlation condition such that both (18) and (21) are approximately correctly specified.

The message is coherent across different values of β_q 's and in line with the theory. The correctly specified factor-unsaturated regression (22) ('L_us') shows the smallest variability in **Figure 1a**. The factor-unsaturated additive regression (21) ('F_us') is only approximately correctly specified in **Figure 1b**, but already delivers even better finite sample performance than the correctly specified (22) thanks to the more parsimonious model. The fully interacted regression ('L'), despite being asymptotically the most efficient among the three factor-saturated specifications, shows substantial variability in all cases. The four misspecified regressions, namely 'N', 'F', 'N_us', and 'F_us', on the other hand, show much stabler performance even in **Figure 1a**, where the equal and zero correlation conditions are considerably violated. This illustrates the robustness of the proposed method to misspecification of the restriction.

7 Discussion

We recommend using restricted least squares on the fully interacted regression for covariate adjustment in multiarmed experiments when the sample size is moderate relative to the number of covariates or treatment levels. Assume that the restriction on the average potential outcomes is correctly specified and separate from that on the correlations between potential outcomes and covariates. The resulting inference is consistent for estimating the ATE regardless of whether the restriction on the correlations is correctly specified or not, and ensures additional efficiency over the OLS counterpart if the restriction on the correlations is indeed correct under constant treatment effects. Simulation studies further show that it can have better finite sample performance than the OLS counterpart even when the restriction is moderately misspecified.

When prior knowledge on the average potential outcomes is less reliable, we recommend imposing restriction on only the correlations. The resulting estimator ensures consistency yet can be at most as efficient as the OLS counterpart asymptotically. Importantly, all results are design-based and hold without assuming any stochastic models for the potential outcomes.

In case of experiments with multiple factors of interest, we assumed a *full factorial design* that includes all possible combinations of the different levels of the factors, and used factor-unsaturated specifications to incorporate prior knowledge of the absence of certain factorial effects. Alternatively, we can accommodate prior knowledge of negligible effects by a *fractional factorial design* in the design stage. A natural question is then how these two strategies compare to each other in terms of the validity and efficiency of subsequent inference. Due to space limitations, we illustrate intuitions about their relative strengths and weaknesses by simulation in the [online supplementary material](#), and leave the complete theory to future research.

Our theory is asymptotic. While we have also evaluated the finite sample properties of the proposed method by simulation, it is of great interest to derive nonasymptotic results under the design-based framework. This is a nontrivial task because of the lack of probability tools in this framework. We leave it to future research.

Supplementary material

Supplementary data is available online at *Journal of the Royal Statistical Society* (<http://mtp.oxfordjournals.org/>).

Conflicts of interest: None declared.

Funding

Zhao was funded by the Start-Up grant R-155-000-216-133 from the National University of Singapore. Ding was partially funded by the U.S. National Science Foundation (grant # 1945136).

Data availability

This paper does not contain data analysis.

References

- Abadie A., Athey S., Imbens G. W., & Wooldridge J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88(1), 265–296. <https://doi.org/10.3982/ECTA12675>
- Blackwell M., & Pashley N. E. (2021). Noncompliance and instrumental variables for 2^K factorial experiments. *Journal of the American Statistical Association*, in press.
- Bloniarz A., Liu H. Z., Zhang C., Sekhon J., & Yu B. (2016). Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), 7383–7390. <https://doi.org/10.1073/pnas.1510506113>
- Box G. E. P., Hunter J. S., & Hunter G. W. (2005). *Statistics for experimenters: Design, innovation, and discovery* (2nd ed.). Wiley.
- Branson Z., Dasgupta T., & Rubin D. B. (2016). Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education High School Study. *The Annals of Applied Statistics*, 10(4), 1958–1976. <https://doi.org/10.1214/16-AOAS959>
- Bugni F. A., Canay I. A., & Shaikh A. M. (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association*, 113(524), 1784–1796. <https://doi.org/10.1080/01621459.2017.1375934>
- Bugni F. A., Canay I. A., & Shaikh A. M. (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics*, 10(4), 1747–1785. <https://doi.org/10.3982/QE1150>
- Chakraborty B., Collins L. M., Strecher V. J., & Murphy S. A. (2009). Developing multicomponent interventions using fractional factorial designs. *Statistics in Medicine*, 28(21), 2687–2708. <https://doi.org/10.1002/sim.3643>
- Collins L. M., Dziak J. J., & Li R. Z. (2009). Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychological Methods*, 14(3), 202–224. <https://doi.org/10.1037/a0015826>
- Dasgupta T., Pillai N., & Rubin D. B. (2015). Causal inference from 2^K factorial designs by using potential outcomes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4), 727–753. <https://doi.org/10.1111/rssb.2015.77.issue-4>
- Ding P., Feller A., & Miratrix L. (2019). Decomposing treatment effect variation. *Journal of the American Statistical Association*, 114(525), 304–317. <https://doi.org/10.1080/01621459.2017.1407322>
- Duflo E., Glennerster R., & Kremer M. (2007). Using randomization in development economics research: A toolkit. In T. P. Schultz & J. A. Strauss (Eds.), *Handbook of development economics* (Vol. 4, Chap. 61, pp. 3895–3962). Elsevier.
- Egami N., & Imai K. (2019). Causal interaction in factorial experiments: Application to conjoint analysis. *Journal of the American Statistical Association*, 114(526), 529–540. <https://doi.org/10.1080/01621459.2018.1476246>
- Fisher R. A. (1935). *The design of experiments* (1st ed.). Oliver and Boyd.
- Fogarty C. B. (2018). Regression-assisted inference for the average treatment effect in paired experiments. *Biometrika*, 105(4), 994–1000. <https://doi.org/10.1093/biomet/asy034>
- Freedman D. A. (2008a). On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40(2), 180–193. <https://doi.org/10.1016/j.aam.2006.12.003>
- Freedman D. A. (2008b). On regression adjustments in experiments with several treatments. *Annals of Applied Statistics*, 2(1), 176–196. <https://doi.org/10.1214/07-AOAS143>
- Greene W. H., & Seaks T. G. (1991). The restricted least squares estimator: A pedagogical note. *The Review of Economics and Statistics*, 73(3), 563–567. <https://doi.org/10.2307/2109587>
- Guo K., & Basse G. (2021). The generalized Oaxaca–Blinder estimator. *Journal of the American Statistical Association*. <https://doi.org/10.1080/01621459.2021.1941053>
- Hainmueller J., Hopkins D. J., & Yamamoto T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1), 1–30. <https://doi.org/10.1093/pan/mpt024>
- Hinkelmann K., & Kempthorne O. (2008). *Design and analysis of experiments: Introduction to experimental design* (2nd ed., Vol. 1). John Wiley & Sons.
- Imbens G. W., & Rubin D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.
- Kempthorne O. (1952). *The design and analysis of experiments*. John Wiley & Sons.
- Li X., & Ding P. (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112(520), 1759–1769. <https://doi.org/10.1080/01621459.2017.1295865>

- Li X., & Ding P. (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1), 241–268. <https://doi.org/10.1111/rssb.12353>
- Li X., Ding P., & Rubin D. B. (2020). Rerandomization in 2^K factorial experiments. *The Annals of Statistics*, 48(1), 43–63. <https://doi.org/10.1214/18-AOS1790>
- Lin W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *The Annals of Applied Statistics*, 7(1), 295–318. <https://doi.org/10.1214/12-AOAS583>
- Liu H., & Yang Y. (2020). Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika*, 107(4), 935–948. <https://doi.org/10.1093/biomet/asaa038>
- Lu J. (2016a). On randomization-based and regression-based inferences for 2^K factorial designs. *Statistics and Probability Letters*, 112(May), 72–78. <https://doi.org/10.1016/j.spl.2016.01.010>
- Lu J. (2016b). Covariate adjustment in randomization-based causal inference for 2^K factorial designs. *Statistics and Probability Letters*, 119(December), 11–20. <https://doi.org/10.1016/j.spl.2016.07.010>
- Miratrix L., Sekhon J., & Yu B. (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(2), 369–396. <https://doi.org/10.1111/rssb.2013.75.issue-2>
- Morgan K. L., & Rubin D. B. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40(2), 1263–1282. <https://doi.org/10.1214/12-AOS1008>
- Mukerjee R., Dasgupta T., & Rubin D. B. (2018). Using standard tools from finite population sampling to improve causal inference for complex experiments. *Journal of the American Statistical Association*, 113(522), 868–881. <https://doi.org/10.1080/01621459.2017.1294076>
- Negi A., & Wooldridge J. M. (2021). Revisiting regression adjustment in experiments with heterogeneous treatment effects. *Econometric Reviews*, 40(5), 504–534. <https://doi.org/10.1080/07474938.2020.1824732>
- Neyman J. (1923). On the application of probability theory to agricultural experiments. *Statistical Science*, 5(4), 465–472. <https://doi.org/10.1214/ss/1177012031>
- Pashley N. E., & Bind M.-A. C. (2022). Causal inference for multiple treatments using fractional factorial designs. *The Canadian Journal of Statistics*, in press. <https://doi.org/10.1002/cjs.11734>
- Rao C. R. (1973). *Linear statistical inference and its applications* (2nd ed.). John Wiley & Sons.
- Schochet P. Z. (2010). Is regression adjustment supported by the Neyman model for causal inference? *Journal of Statistical Planning and Inference*, 140(1), 246–259. <https://doi.org/10.1016/j.jspi.2009.07.008>
- Schochet P. Z. (2018). Design-based estimators for average treatment effects for multi-armed RCTs. *Journal of Educational and Behavioral Statistics*, 43(5), 568–593. <https://doi.org/10.3102/1076998618786968>
- Theil H. (1971). *Principles of econometrics*. John Wiley & Sons.
- Tsiatis A. A., Davidian M., Zhang M., & Lu X. M. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine*, 27(23), 4658–4677. <https://doi.org/10.1002/sim.v27:23>
- Wu C. F. J., & Hamada M. (2009). *Experiments: Planning, analysis, and optimization*. John Wiley & Sons.
- Ye T., Shao J., Yi Y. Y., & Zhao Q. Y. (2022). Toward better practice of covariate adjustment in analyzing randomized clinical trials. *Journal of the American Statistical Association*, 1–13. <https://doi.org/10.1080/01621459.2022.2049278>
- Zhao A., & Ding P. (2022a). Regression-based causal inference with factorial experiments: estimands, model specifications and design-based properties. *Biometrika*, 109(3), 799–815. <https://doi.org/10.1093/biomet/asab051>
- Zhao A., & Ding P. (2022b). Reconciling design-based and model-based causal inferences for split-plot experiments. *The Annals of Statistics*, 50(2), 1170–1192. <https://doi.org/10.1214/21-AOS2144>

Supplementary Material

Section S1 uses simulation to compare full and fractional factorial designs.

Section S2 discusses the properties of a class of general linear estimators of \bar{Y} . It forms the basis of all theoretical results.

Section S3 presents additional results on restricted least squares.

Section S4 presents additional results on 2^K factorial experiments.

Section S5 presents the results on rerandomization using the Mahalanobis distance.

Section S6 gives the proofs of the results in Section S2.

Section S7 gives the proofs of the results on treatment-based regressions. In particular, Section S7.3 gives the proof of the results from ordinary least squares, and Section S7.4 gives the proof of the results from restricted least squares.

Section S8 gives the proofs of the results on factor-based regressions.

Contents

| | |
|---|------------|
| S1 Simulation comparison of full and fractional factorial designs | S3 |
| S1.1 Scenario I: all interaction effects are zero | S4 |
| S1.2 Scenario II: no two-way interaction between A and B | S5 |
| S2 A class of linear estimators | S7 |
| S3 Additional results on restricted least squares | S9 |
| S3.1 Additional results under the correlation-only and separable restrictions | S9 |
| S3.2 Properties under general restriction | S9 |
| S3.3 Numeric properties of restricted least squares | S10 |
| S4 Additional results on 2^K factorial experiments | S12 |
| S4.1 Design-based theory for general regression specifications | S13 |
| S4.2 Factorial effects under $\{0, 1\}$ -coded regressions | S16 |
| S5 Rerandomization using the Mahalanobis distance | S17 |
| S5.1 Overview | S17 |
| S5.2 ReM under multi-armed experiments | S18 |
| S5.3 Asymptotic properties | S19 |
| S5.4 ReM-specific inference | S21 |
| S6 Proof of the results on $\hat{Y}\langle b \rangle \in \mathcal{Y}$ | S22 |
| S6.1 Lemmas | S22 |
| S6.2 Results under complete randomization | S23 |
| S6.3 Results under ReM | S23 |

| | |
|--|------------|
| S7 Proof of the results on treatment-based regressions | S25 |
| S7.1 Notation and useful facts | S25 |
| S7.2 Asymptotics of $\hat{\theta}_L$ | S27 |
| S7.3 Results on outputs from ordinary least squares | S29 |
| S7.4 Results on outputs from restricted least squares | S31 |
| S7.4.1 RLS subject to general restriction | S31 |
| S7.4.2 RLS subject to correlation-only restriction | S33 |
| S7.4.3 RLS subject to separable restriction with $\rho_Y \neq 0$ | S33 |
| S7.4.4 Robust covariance estimator | S35 |
| S8 Proof of the results on factor-based regressions | S38 |
| S8.1 Standard factorial effects | S38 |
| S8.2 Factorial effects under {0, 1}-coded regressions | S40 |

Notation and definition

Assume centered covariates with $\bar{x} = 0_J$ throughout to simplify the presentation.

For a sequence of random vectors $(A_N)_{N=1}^\infty$ with $A_N \rightsquigarrow A$, let $E_\infty(A_N) = E(A)$ and $\text{cov}_\infty(A_N) = \text{cov}(A)$ denote the expectation and covariance with respect to the asymptotic distribution. For two sequences of random vectors $\{A_N\}_{N=1}^\infty$ and $\{B_N\}_{N=1}^\infty$ where N denotes the sample size, write $A_N \approx B_N$ if $\sqrt{N}(A_N - B_N) = o_P(1)$.

Recall that we quantify the relative efficiency of consistent and asymptotically normal estimators by their asymptotic covariances in the main paper. Following Li et al. (2020), we use the concept of peakedness (Sherman 1955) below to generalize the comparison to estimators with nonnormal asymptotic distributions.

Definition S1 (peakedness). For two symmetric random vectors A and B in \mathbb{R}^m , we say A is *more peaked* than B if $\mathbb{P}(A \in \mathcal{C}) \geq \mathbb{P}(B \in \mathcal{C})$ for every symmetric convex set $\mathcal{C} \in \mathbb{R}^m$, denoted by $A \succeq B$.

For $m = 1$, a more peaked random variable has narrower central quantile ranges. For A and B with finite second moments, $A \succeq B$ implies $\text{cov}(A) \leq \text{cov}(B)$ (Li et al. 2020, Proposition 4). For A and B that are both normal with zero means, $A \succeq B$ is equivalent to $\text{cov}(A) \leq \text{cov}(B)$. This relation between peakedness and covariance suggests that peakedness gives a more refined measure of asymptotic efficiency than covariance for estimators with nonnormal asymptotic distributions. An asymptotically more peaked estimator has not only a smaller asymptotic covariance but also narrower central quantile ranges. We formalize the idea in Definition S2 below.

Definition S2. Assume that $\hat{\theta}_1$ and $\hat{\theta}_2$ are two consistent estimators for parameter $\theta \in \mathbb{R}^m$ as the sample size N tends to infinity, with $\sqrt{N}(\hat{\theta}_1 - \theta) \rightsquigarrow A_1$ and $\sqrt{N}(\hat{\theta}_2 - \theta) \rightsquigarrow A_2$ for some symmetric random vectors A_1 and A_2 . We say

- (i) $\hat{\theta}_1$ and $\hat{\theta}_2$ are *asymptotically equally efficient* if A_1 and A_2 have the same distribution, denoted by $\hat{\theta}_1 \sim \hat{\theta}_2$;
- (ii) $\hat{\theta}_1$ is *asymptotically more efficient* $\hat{\theta}_2$ if $A_1 \succeq A_2$, denoted by $\hat{\theta}_1 \succeq_{\infty} \hat{\theta}_2$ or $\hat{\theta}_2 \preceq_{\infty} \hat{\theta}_1$.

For two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ that are both consistent for parameter θ , we have $\hat{\theta}_1 \approx \hat{\theta}_2$ implies that $\hat{\theta}_1 \sim \hat{\theta}_2$ by Slutsky's theorem.

S1. Simulation comparison of full and fractional factorial designs

The discussion on factorial experiments in the main paper assumed a *full factorial design* that includes all possible combinations of the different levels of the factors, and used factor-unsaturated specifications to incorporate prior knowledge of negligible factorial effects. Alternatively, we can accommodate such knowledge by a *fractional factorial design* in the design stage.

Consider an experiment with three binary factors A, B, and C. Assume that prior knowledge suggests that the interaction between A and B is negligible. This motivates adopting a 2^{3-1} fractional factorial design with *defining relation* $C = AB$ (Wu and Hamada 2009). Index by $\{-1, +1\}$ the two levels of each factor, abbreviated as $\{-, +\}$. Instead of considering all eight possible treatment combinations as under a 2^3 full factorial design, the 2^{3-1} fractional factorial design includes only $2^{3-1} = 4$ combinations of A, B, and C, featuring a 2^2 full factorial design of factors A and B, and the level of factor C determined by the defining relation as the product of the levels of A and B:

| A | B | C |
|---|---|---|
| - | - | + |
| - | + | - |
| + | - | - |
| + | + | + |

We can then randomize the units to these four treatment combinations by complete randomization, and estimate the factorial effects of interest by OLS. A natural question is how these two strategies compare to each other in terms of the validity and efficiency of subsequent inference.

On the one hand, the fractional factorial design includes only half of the treatment combinations in the full factorial design, effectively doubling the sizes of the treatment groups. This promises the opportunity to improve estimation efficiency.

On the other hand, the defining relation $C = AB$ implies $A = BC$ and $B = AC$. This suggests that the main effects of factors A and B are aliased with the interaction between factors B and C and that between factors A and C, respectively, such that they are not identifiable under the fractional design unless the corresponding interactions are zero. Likewise for the three-way interaction between A, B, and C to be indistinguishable from the intercept given $ABC = AB(AB) = I$.

The same limitation does not exist under the full factorial design, where we can use the restricted specification $Y_i \sim 1 + A_i + B_i + C_i + B_i C_i + A_i C_i + A_i B_i C_i$ to encode the prior knowledge of no

interaction between A and B without affecting the estimation of the main effects. The full factorial design is in this light more flexible than the fractional factorial design for accommodating a specific subset of negligible effects.

The choice between the full and fractional factorial designs is therefore a trade-off between estimation and exploration. The fractional factorial design promises improved efficiency for estimating the factorial effects of interest if and only if the assumption of negligible higher-order interactions is correctly specified. In contrast, the full factorial design allows us to explore the absence or presence of higher-order interactions. We illustrate these intuitions by simulation in the following two subsections, and leave the complete theory to future research. See Pashley and Bind (2022) for related discussion.

S1.1. Scenario I: all interaction effects are zero

Consider an experiment with $K = 3$ binary factors, A, B, and C, and a study population of $N = 80$ units, indexed by $i = 1, \dots, N$. Index the $2^3 = 8$ possible treatment combinations by $\mathcal{T} = \{(abc) : a, b, c = -1, +1\}$. Building on Example 3 in the main paper, let τ_A , τ_{AB} , and τ_{ABC} denote the main effect of factor A, the interaction effect between factors A and B, and the interaction effect between factors A, B, and C, respectively; likewise define τ_B , τ_C , τ_{AC} , and τ_{BC} .

Let $(x_i)_{i=1}^N$ be independent $\mathcal{N}(0, 4)$ and $(\epsilon_i)_{i=1}^N$ be independent $\mathcal{N}(0, 1)$. We generate the potential outcomes as $Y_i(abc) = 4 + 2a + 2b + 2c + ax_i + \epsilon_i$, and then center $\{Y_i(abc)\}_{i=1}^N$ for each $(abc) \in \mathcal{T}$ to ensure $(\tau_A, \tau_B, \tau_C, \tau_{AB}, \tau_{BC}, \tau_{AC}, \tau_{ABC}) = (4, 4, 4, 0, 0, 0, 0)$. This ensures that all interaction effects are absent in the ground truth.

Fix the potential outcomes in the simulation. The 2^3 full factorial design assigns $N_{abc} = 10$ units to each of the eight treatment combinations in \mathcal{T} . We consider the following three specifications for estimating the standard factorial effects by OLS:

$$\begin{aligned}\text{m1 : } & Y_i \sim 1 + A_i + B_i + C_i + x_i, \\ \text{m2 : } & Y_i \sim 1 + A_i + B_i + C_i + A_i C_i + B_i C_i + A_i B_i C_i + x_i, \\ \text{m3 : } & Y_i \sim 1 + A_i + B_i + C_i + A_i B_i + A_i C_i + B_i C_i + A_i B_i C_i + x_i.\end{aligned}$$

Specification m1 assumes the absence of all two-way and three-way interactions, matching exactly the ground truth. Specification m2 assumes the absence of the interaction effect between factors A and B. Specification m3 is factor-saturated and makes no assumptions about the absence or presence of the factorial effects. They are all correctly specified given the ground truth. We use the additive specification for covariate adjustment given the moderate sample size.

The 2^{3-1} fractional factorial design, on the other hand, assigns $N'_{abc} = 20$ units to each of the four treatment combinations in $\mathcal{T}' = \{(abc) : a, b = -1, +1; c = ab\} = \{(- - +), (- + -), (+ - -), (+ + +)\}$ subject to the defining relation $C = AB$, and runs OLS using the same specification as m1 under the full design. We index the resulting regression model by m4 to distinguish between

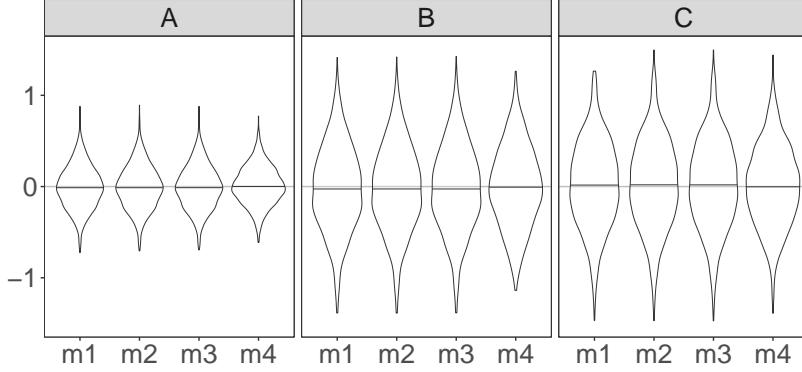


Figure S1: Violin plot of the differences between the estimated and true values for the three main effects over 1,000 independent complete randomizations from regressions m_1 – m_4 under Scenario I. Regressions m_1 – m_3 correspond to the 2^3 full factorial design, and regression m_4 corresponds to the 2^{3-1} fractional factorial design. All four regressions are correctly specified.

the full and fractional designs:

$$m_4 : \quad Y_i \sim 1 + A_i + B_i + C_i + x_i \quad \text{under the fractional factorial design.}$$

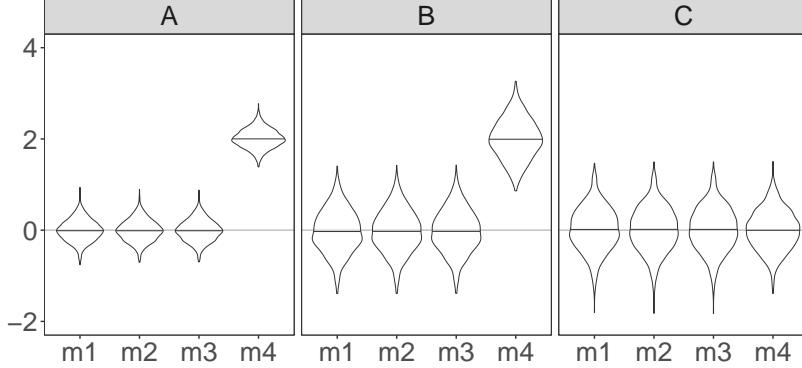
It assumes the absence of all two-way and three-way interactions as specification m_1 , matching exactly the ground truth.

Figure S1 shows the violin plot of the differences between the estimated and true values for the three main effects over 1,000 independent complete randomizations under the full and fractional factorial designs, respectively. Specification m_4 under the fractional factorial design is on average the most efficient, yielding visible improvement over specifications m_1 – m_3 for estimating the main effects of factors A and B. This illustrates the gain in efficiency by the fractional factorial design when its assumption on negligible higher-order interactions is correctly specified.

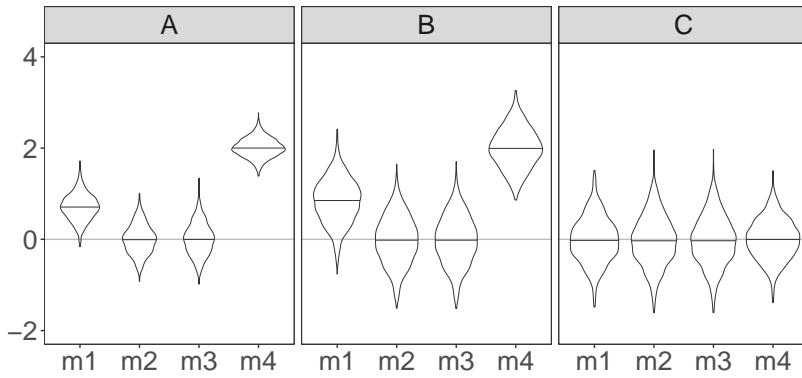
S1.2. Scenario II: no two-way interaction between A and B

Inherit most of the setting from Scenario I except that we now generate the potential outcomes as $Y_i(abc) = 4 + 2a + 2b + 2c + ac + bc + 2^{-1}abc + ax_i + \epsilon_i$, and then center $\{Y_i(abc)\}_{i=1}^N$ for each $(abc) \in \mathcal{T}$ to ensure $(\tau_A, \tau_B, \tau_C, \tau_{AB}, \tau_{BC}, \tau_{AC}, \tau_{ABC}) = (4, 4, 4, 0, 2, 2, 1)$. This ensures that only the interaction effect between factors A and B is absent in the ground truth, such that regressions m_2 and m_3 are correctly specified whereas regressions m_1 and m_4 are not.

Figure S2(a) shows the violin plot of the differences between the estimated and true values for the three main effects over 1,000 independent complete randomizations under the full and fractional factorial designs, respectively. Specification m_4 under the fractional factorial design is now considerably biased for estimating the main effects of factors A and B. This is coherent with the fact that the main effects of factors A and B are aliased with the interaction between B and C and that between A and C, respectively, and illustrates the potential problems with the fractional



(a) Equal-sized full factorial design with $N_q = 10$ for all $q \in \mathcal{T}$.



(b) Unequal-sized full factorial design with $(N_q)_{q \in \mathcal{T}} = (5, 5, 15, 15, 5, 15, 5, 15)$.

Figure S2: Violin plot of the differences between the estimated and true values for the three main effects over 1,000 independent complete randomizations from regressions m1-m4 under Scenario II. Regressions m1-m3 correspond to the 2^3 full factorial design, and regression m4 corresponds to the 2^{3-1} fractional factorial design. Regressions m2 and m3 are correctly specified, whereas regressions m1 and m4 are not.

factorial design when its assumption is misspecified.

Specifications m1-m3 under the full factorial design, on the other hand, show no visible empirical biases. This illustrates the advantage of the full factorial design when only a subset of the interaction effects is absent. Importantly, specification m1 shows no visible empirical biases despite being misspecified. This is a direct consequence of Proposition 3, and illustrates the robustness of factor-unsaturated specifications under full factorial designs when the treatment groups are of equal sizes. Figure S2(b) shows the corresponding results when we change the treatment group sizes under the full factorial design to $(N_q)_{q \in \mathcal{T}} = (5, 5, 15, 15, 5, 15, 5, 15)$ in lexicographical order of $(abc) \in \mathcal{T}$. Specification m1 now shows visible empirical biases for estimating the main effects of factors A and B. The biases are nevertheless still smaller than that of specification m4 under the fractional design. The full factorial design as such allows for more robust inference when the prior knowledge on negligible effects is less certain.

S2. A class of linear estimators

We introduce in this section a class of linear estimators of \bar{Y} that includes \hat{Y}_* ($*$ = N, F, L) as special cases. The results ensure the asymptotic efficiency of $\hat{\tau}_L$ over a broader range, and provide the basis for all theoretical results in the main paper.

Recall $\hat{x}(q) = N_q^{-1} \sum_{i:Z_i=q} x_i$ as the sample mean of covariates under treatment level $q \in \mathcal{T}$. For some prespecified $b_q \in \mathbb{R}^J$ that may depend on the data, define

$$\hat{Y}(q; b_q) = \hat{Y}(q) - \hat{x}(q)^T b_q$$

as a covariate-adjusted variant of $\hat{Y}(q)$ for estimating $\bar{Y}(q)$. Intuitively, it is the sample mean of the covariate-adjusted potential outcomes $Y_i(q; b_q) = Y_i(q) - x_i^T b_q$. A common choice is to take b_q as the coefficient vector of x_i from the OLS fit of $Y_i \sim 1 + x_i$ over $\{i : Z_i = q\}$, denoted by $\hat{\gamma}_q$ (Lin 2013).

Define

$$\hat{Y}\langle b \rangle = (\hat{Y}(1; b_1), \dots, \hat{Y}(Q; b_Q))^T$$

as the corresponding covariate-adjusted estimator of \bar{Y} , with $b = (b_1^T, \dots, b_Q^T)^T \in \mathbb{R}^{JQ}$. We focus on the set of b 's that have finite probability limits under complete randomization:

$$\mathcal{B} = \{b \in \mathbb{R}^{JQ} : \text{plim } b \text{ exists and is finite under complete randomization and Condition 4}\}.$$

By definition, \mathcal{B} contains all constant vectors in \mathbb{R}^{JQ} . The random vector $\hat{\gamma} = (\hat{\gamma}_1^T, \dots, \hat{\gamma}_Q^T)^T$ is also in \mathcal{B} with $\hat{\gamma}_q = \gamma_q + o_P(1)$ for all $q \in \mathcal{T}$. Let

$$\mathcal{Y} = \{\hat{Y}\langle b \rangle : b \in \mathcal{B}\} \tag{S1}$$

summarize the corresponding covariate-adjusted linear estimators of \bar{Y} .

Recall that $\gamma = (\gamma_1^T, \dots, \gamma_Q^T)^T$ and $\bar{\gamma} = \sum_{q \in \mathcal{T}} e_q \gamma_q$. Recall $\hat{\gamma}_L$ as the coefficient vector of $t_i \otimes x_i$ from the OLS fit of (6), with $\hat{\gamma}_L = \hat{\gamma}$ by standard properties of OLS. Further let $\hat{\beta}_F$ denote the coefficient vector of x_i from the OLS fit of (5). Proposition S1 below ensures that \hat{Y}_* ($*$ = N, F, L) are all elements in \mathcal{Y} .

Proposition S1. $\hat{Y}_* = \hat{Y}\langle b_* \rangle$ for $* = N, F, L$ with $b_N = 0_{JQ}$, $b_F = 1_Q \otimes \hat{\beta}_F$, and $b_L = \hat{\gamma}_L$. Further assume complete randomization and Condition 4. Then $\hat{\beta}_F = \bar{\gamma} + o_P(1)$ and $\hat{\gamma}_L = \gamma + o_P(1)$ with $\text{plim } b_F = 1_Q \otimes \bar{\gamma}$ and $\text{plim } b_L = \gamma$.

For $b \in \mathcal{B}$ with $\text{plim } b = b_\infty = (b_{1,\infty}^T, \dots, b_{Q,\infty}^T)^T$, define

$$V_{b,\infty} = \text{diag}(S_{b,\infty,qq}/e_q)_{q \in \mathcal{T}} - S_{b,\infty}, \tag{S2}$$

where $S_{b,\infty} = (S_{b,\infty,qq'})_{q,q' \in \mathcal{T}}$ denotes the finite population covariance matrix of $\{Y_i(q; b_{q,\infty}) : q \in \mathcal{T}\}$.

$\mathcal{T}_{i=1}^N$. By Proposition S1, the V_* ($*$ = N, F, L) in (9) are all special cases of $V_{b,\infty}$ with $b = b_*$. Lemma S1 below ensures that the estimators in \mathcal{Y} are all consistent and asymptotic normal for estimating \bar{Y} , and establishes the asymptotic efficiency of \hat{Y}_L over \mathcal{Y} .

Lemma S1. For all $b \in \mathcal{B}$ with $\text{plim } b = b_\infty$, we have

- (i) $V_{b,\infty} \geq V_L$, where the equality holds if $b_\infty = \gamma$;
- (ii) $\sqrt{N}(\hat{Y}\langle b \rangle - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, V_{b,\infty})$ under complete randomization and Condition 4.

Lemma S1 includes Lemmas 1 and 3 as special cases, and establishes the asymptotic efficiency of \hat{Y}_L over all linear estimators in \mathcal{Y} for estimating \bar{Y} . This ensures the asymptotic efficiency of $\hat{\gamma}_L$ over $\{C\hat{Y}\langle b \rangle : \hat{Y}\langle b \rangle \in \mathcal{Y}\}$ for estimating $\tau = C\bar{Y}$, extending Li and Ding (2020, Example 9) to multi-armed experiments. In addition, $\hat{Y}\langle b \rangle \sim \hat{Y}_L$ if $b_\infty = \gamma$.

Moreover, let $\hat{Y}\langle \gamma \rangle = (\hat{Y}(1; \gamma_1), \dots, \hat{Y}(Q; \gamma_Q))^T$ be the linear estimator defined by $\gamma = (\gamma_1^T, \dots, \gamma_Q^T)^T$. Lemma S1 establishes $\hat{Y}\langle \gamma \rangle$ as the oracle estimator with fixed adjustment coefficients $b_q = \gamma_q$ ($q \in \mathcal{T}$). Intuitively, the properties of OLS ensure full reduction of the variability due to the covariates, and guarantee the efficiency of $\hat{Y}\langle \gamma \rangle$ over \mathcal{Y} . Recall that γ_q gives the target parameter of $1(Z_i = q)x_i$ in (6) from the derived linear model perspective. The additive and fully interacted regressions can thus also be viewed as two ways to estimate the optimal adjustment coefficients γ_q ($q \in \mathcal{T}$). In particular, we can estimate γ_q by the OLS coefficient of $1(Z_i = q)x_i$, namely $\hat{\gamma}_{L,q}$, from the fully interacted regression (6), or by the OLS coefficient of x_i , namely $\hat{\beta}_F$, from the additive regression (5). The resulting $\hat{Y}\langle b \rangle$ equals \hat{Y}_L and \hat{Y}_F , respectively, by Proposition S1. Proposition S1 further ensures that $\hat{\gamma}_{L,q}$ is consistent for γ_q , whereas $\hat{\beta}_F$ in general is not. This gives the intuition behind the asymptotic efficiency of \hat{Y}_L .

Theorem S1 below states a stronger result than Lemma S1, characterizing the asymptotic distance of $\hat{Y}\langle b \rangle \in \mathcal{Y}$ from the oracle $\hat{Y}\langle \gamma \rangle$. Let $D_b = \text{diag}\{(b_q - \gamma_q)^T\}_{q \in \mathcal{T}} \in \mathbb{R}^{Q \times JQ}$ with

$$\hat{Y}\langle b \rangle = \hat{Y}_N - \{\text{diag}(b_q^T)_{q \in \mathcal{T}}\}\hat{x} = \hat{Y}\langle \gamma \rangle - D_b\hat{x},$$

where $\hat{x} = (\hat{x}(1)^T, \dots, \hat{x}(Q)^T)^T$. Let $D_{b,\infty} = \text{plim } D_b = \text{diag}\{(b_{q,\infty} - \gamma_q)^T\}_{q \in \mathcal{T}}$.

Theorem S1. Assume complete randomization and Condition 4. For $\hat{Y}\langle b \rangle \in \mathcal{Y}$ with $\text{plim } b = b_\infty$, we have

$$\hat{Y}\langle b \rangle \underset{\approx}{\sim} \hat{Y}\langle \gamma \rangle - D_{b,\infty}\hat{x} \underset{\preceq_\infty}{\preceq} \hat{Y}\langle \gamma \rangle$$

with $V_{b,\infty} = N\text{cov}_\infty\{\hat{Y}\langle b \rangle\} = V_L + D_{b,\infty}V_xD_{b,\infty}^T$. In particular, $\hat{Y}\langle b \rangle \underset{\approx}{\sim} \hat{Y}\langle \gamma \rangle$ if

$$D_{b,\infty}V_xD_{b,\infty}^T = 0. \tag{S3}$$

A sufficient condition for (S3) is $b_\infty = \gamma$.

Theorem S1 states the asymptotic equivalence of $\hat{Y}\langle b \rangle$ and $\hat{Y}\langle \gamma \rangle - D_{b,\infty}\hat{x} = \hat{Y}\langle b_\infty \rangle$, suggesting $D_{b,\infty}\hat{x}$ as the distance of $\hat{Y}\langle b \rangle$ from the optimally adjusted $\hat{Y}\langle \gamma \rangle$.

S3. Additional results on restricted least squares

S3.1. Additional results under the correlation-only and separable restrictions

Recall the definition of \mathcal{B} and \mathcal{Y} from Section S2. Proposition S2 below states the numeric equivalence between \hat{Y}_r and $\hat{Y}\langle\hat{\gamma}_r\rangle$ under RLS subject to the correlation-only restriction, and ensures that \hat{Y}_r is an element of \mathcal{Y} . The design-based properties of \hat{Y}_r in Theorem 1 then follow immediately from Lemma S1.

Proposition S2. Assume RLS subject to the correlation-only restriction (14). Then $\hat{Y}_r = \hat{Y}\langle\hat{\gamma}_r\rangle$, where $\hat{\gamma}_r$ belongs to \mathcal{B} and satisfies $\text{plim } \hat{\gamma}_r = \gamma$ if (14) is correctly specified.

Proposition S3 below ensures that \hat{Y}_r is a linear function of $\hat{Y}\langle\hat{\gamma}_r\rangle \in \mathcal{Y}$ under RLS subject to the separable restriction. This, together with Lemma S1, leads to the design-based properties of \hat{Y}_r in Theorem 2. Recall U and μ_r in (15).

Proposition S3. Assume RLS subject to the separable restriction (13) with $\rho_Y \neq 0$. Then

$$\hat{Y}_r - \bar{Y} = U(\hat{Y}\langle\hat{\gamma}_r\rangle - \bar{Y}) + \mu_r,$$

where $\hat{\gamma}_r$ belongs to \mathcal{B} and satisfies $\text{plim } \hat{\gamma}_r = \gamma$ if $\rho_\gamma\gamma = r_\gamma$ is correctly specified.

Theorem S2 below extends Theorem 3(ii) and gives the design-based counterpart of the classical Gauss–Markov theorem for RLS subject to the separable restriction. The result extends Zhao and Ding (2022, Theorem A5) on unadjusted estimators to the covariate-adjusted settings.

Let

$$\mathcal{Y}' = \left\{ L\hat{Y}\langle b \rangle + a : \hat{Y}\langle b \rangle \in \mathcal{Y}, L \in \mathbb{R}^{Q \times Q}, \text{ and } a \in \mathbb{R}^Q \text{ with } L\hat{Y}\langle b \rangle + a = \bar{Y} + o_p(1) \right\}$$

denote a class of linear consistent estimators of \bar{Y} , which contains all elements in \mathcal{Y} and hence \hat{Y}_* ($*$ = N, F, L) as special cases. By Lemma S1, all elements in \mathcal{Y}' are consistent and asymptotically normal such that we can compare their asymptotic efficiency by comparing their asymptotic covariances.

Theorem S2. Assume complete randomization, Conditions 3–4, and RLS subject to the separable restriction (13) with $\rho_Y \neq 0$ being a contrast matrix with rows orthogonal to 1_Q . If both $\rho_Y\bar{Y} = r_Y$ and $\rho_\gamma\gamma = r_\gamma$ are correctly specified, then \hat{Y}_r is the best linear consistent estimator in \mathcal{Y}' , in the sense that \hat{Y}_r is an element of \mathcal{Y}' and has the smallest asymptotic covariance over all estimators in \mathcal{Y}' .

S3.2. Properties under general restriction

Write $R = (R_Y, R_\gamma)$, where $R_Y \in \mathbb{R}^{m \times Q}$ and $R_\gamma \in \mathbb{R}^{m \times JQ}$ denote the columns of R corresponding to \bar{Y} and γ , respectively. Theorem S3 below quantifies the asymptotic behaviors of \hat{Y}_r for general

R , and suggests that \hat{Y}_r is in general not consistent for estimating \bar{Y} unless $R_Y = 0$ or (11) is correctly specified.

Recall $M_r = (\chi_L^T \chi_L)^{-1} R^T \{R(\chi_L^T \chi_L)^{-1} R^T\}^{-1}$ from Lemma 2. Let ξ_r be the first Q rows of $-M_r(R\theta_L - r)$. Let $M_{r,\infty} = \text{plim } M_r$ be the finite probability limit of M_r under complete randomization and Condition 4. Let Σ_r be the upper-left $Q \times Q$ submatrix of $(I - M_{r,\infty} R)\{N\text{cov}_{\infty}(\hat{\theta}_L)\}(I - M_{r,\infty} R)^T$. We give the explicit forms of $M_{r,\infty}$ and $N\text{cov}_{\infty}(\hat{\theta}_L)$ in (S33) and Lemma S9. Recall $\Pi = \text{diag}(e_q)_{q \in \mathcal{T}}$ with $e_q = N_q/N$. Let $\Delta_0 = [R \text{ diag}\{\Pi^{-1}, (\Pi \otimes S_x^2)^{-1}\} R^T]^{-1}$.

Theorem S3. Assume RLS subject to (11). Then

$$\hat{Y}_r = \hat{Y} \langle \hat{\gamma}_r \rangle - \Pi^{-1} R_Y^T [R \{N(\chi_L^T \chi_L)^{-1}\} R^T]^{-1} (R\hat{\theta}_L - r).$$

Further assume complete randomization and Condition 4. Then

- (i) $\hat{\gamma}_r = \gamma - (\Pi \otimes S_x^2)^{-1} R_\gamma^T \Delta_0 (R\theta_L - r) + o_P(1)$;
- (ii) $\sqrt{N}(\hat{Y}_r - \bar{Y} - \xi_r) \rightsquigarrow \mathcal{N}(0_Q, \Sigma_r)$ with $\xi_r = -\Pi^{-1} R_Y^T \Delta_0 (R\theta_L - r) + o_P(1)$. In particular, $\xi_r = 0$, and hence $\sqrt{N}(\hat{Y}_r - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, \Sigma_r)$, if (11) is correctly specified.

Theorem S3(i) ensures that $\hat{\gamma}_r$ has a finite probability limit under complete randomization and Condition 4 regardless of whether the restriction $R\theta_L = r$ is correctly specified or not. Theorem S3(ii) implies $\text{plim } \xi_r = -\Pi^{-1} R_Y^T \Delta_0 (R\theta_L - r)$ as the asymptotic bias of \hat{Y}_r . Juxtapose Theorem S3 with Theorems 1 and 2. The consistency of \hat{Y}_r in general requires the whole restriction (11) to be correctly specified, whereas special structures like $R_Y = 0$ or $R = \text{diag}(\rho_Y, \rho_{\gamma})$ promise weaker sufficient conditions. In particular, the correlation-only restriction ensures consistency regardless of whether the restriction on γ is correctly specified or not; the separable restriction ensures consistency as long as the restriction on \bar{Y} is correctly specified.

S3.3. Numeric properties of restricted least squares

We present in this subsection some useful numeric properties about RLS. The results are stated in terms of general regression formulation to highlight their generality.

For $Y \in \mathbb{R}^N$, $X \in \mathbb{R}^{N \times p}$, $R \in \mathbb{R}^{m \times p}$, and $r \in \mathbb{R}^m$, consider the RLS fit of

$$Y = X\hat{\beta} + \hat{\epsilon} \quad \text{subject to } R\hat{\beta} = r,$$

where $\hat{\beta}$ and $\hat{\epsilon} = (\hat{\epsilon}_1, \dots, \hat{\epsilon}_N)^T$ denote the RLS coefficient vector and residuals, respectively. To simplify the presentation, we suppress the subscript “r” for RLS in this subsection when no confusion would arise. We propose to estimate the sampling covariance of $\hat{\beta}$ by

$$\hat{V} = (I_p - M_r R)(X^T X)^{-1} X^T \{ \text{diag}(\hat{\epsilon}_i^2)_{i=1}^N \} X (X^T X)^{-1} (I_p - M_r R)^T, \quad (\text{S4})$$

where $M_r = (X^T X)^{-1} R^T \{R(X^T X)^{-1} R^T\}^{-1}$. Refer to \hat{V} as the *double-decker-taco robust* covariance estimator from RLS.

Lemma S2 below states the invariance of RLS to nondegenerate linear transformation of the regressors.

Lemma S2. For $Y \in \mathbb{R}^N$, $X \in \mathbb{R}^{N \times p}$, and $X' = X\Gamma \in \mathbb{R}^{N \times p}$ for some nonsingular $\Gamma \in \mathbb{R}^{p \times p}$, consider the RLS fits of

$$\begin{aligned} Y &= X\hat{\beta} + \hat{\epsilon} && \text{subject to } R\hat{\beta} = r, \\ Y &= X'\hat{\beta}' + \hat{\epsilon}' && \text{subject to } (R\Gamma)\hat{\beta}' = r, \end{aligned}$$

where $(\hat{\beta}, \hat{\epsilon})$ and $(\hat{\beta}', \hat{\epsilon}')$ denote the RLS coefficient vectors and residuals. Further let \hat{V} and \hat{V}' denote the robust covariance estimators of $\hat{\beta}$ and $\hat{\beta}'$ by (S4), respectively. Then

$$\hat{\beta} = \Gamma\hat{\beta}', \quad \hat{\epsilon} = \hat{\epsilon}', \quad \hat{V} = \Gamma\hat{V}'\Gamma^T.$$

Lemma S2 holds for arbitrary choices of R and r . Lemma S3 below presents a novel result on the numeric equivalence between the robust covariance (S4) from the RLS fit and that from the OLS fit of a corresponding restricted specification when $r = 0$. The proof follows from direct linear algebra and is omitted.

Lemma S3. For $Y \in \mathbb{R}^N$, $X \in \mathbb{R}^{N \times p}$, and $R \in \mathbb{R}^{m \times p}$, consider the RLS fit of

$$Y = X\hat{\beta} + \hat{\epsilon} \quad \text{subject to } R\hat{\beta} = 0, \tag{S5}$$

where $\hat{\beta} \in \mathbb{R}^p$ and $\hat{\epsilon} \in \mathbb{R}^N$ denote the RLS coefficient vector and residuals, respectively. Then

(i) The corresponding restricted specification can be formed as

$$Y = \{XR_{\perp}^T(R_{\perp}R_{\perp}^T)^{-1}\}\hat{\beta}_{\text{OLS}} + \hat{\epsilon}_{\text{OLS}}, \tag{S6}$$

where $R_{\perp} \in \mathbb{R}^{(p-m) \times p}$ is an orthogonal complement of R in the sense that (R_{\perp}^T, R^T) is non-singular with $R_{\perp}R^T = 0$. Let $\hat{\beta}_{\text{OLS}} \in \mathbb{R}^{p-m}$ and $\hat{\epsilon}_{\text{OLS}} \in \mathbb{R}^N$ denote the coefficient vector and residuals from the OLS fit of (S6). Then

$$\hat{\beta}_{\text{OLS}} = R_{\perp}\hat{\beta}, \quad \hat{\epsilon}_{\text{OLS}} = \hat{\epsilon}.$$

(ii) Let $\hat{V}_{\text{RLS}} = R_{\perp}\hat{V}R_{\perp}^T$ denote the robust covariance estimator of $R_{\perp}\hat{\beta}$ from the RLS fit of (S5), with \hat{V} given by (S4). Let \hat{V}_{OLS} denote the EHW covariance estimator of $\hat{\beta}_{\text{OLS}}$ from the OLS fit of (S6). Then

$$\hat{V}_{\text{RLS}} = \hat{V}_{\text{OLS}}.$$

Lemma S3 includes Examples S1 and S2 below as special cases, which correspond to the unadjusted and additive regressions, respectively.

Example S1. For $Y \in \mathbb{R}^N$, $X_1 \in \mathbb{R}^{N \times k}$, and $X_2 \in \mathbb{R}^{N \times l}$, consider the RLS fit of

$$Y = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + \hat{\epsilon} \quad \text{subject to } \hat{\beta}_2 = 0,$$

where $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\epsilon}$ denote the RLS coefficient vectors and residuals, respectively. Let \hat{V}_{RLS} denote the robust covariance estimator of $\hat{\beta}_1$ by (S4). Alternatively, consider the OLS fit of

$$Y = X_1 \hat{\beta}_{1,\text{OLS}} + \hat{\epsilon}_{\text{OLS}}$$

as the corresponding restricted specification, where $\hat{\beta}_{1,\text{OLS}}$ and $\hat{\epsilon}_{\text{OLS}}$ denote the OLS coefficient vector and residuals, respectively. Let \hat{V}_{OLS} denote the EHW covariance estimator of $\hat{\beta}_{1,\text{OLS}}$. Then

$$\hat{\beta}_1 = \hat{\beta}_{1,\text{OLS}}, \quad \hat{\epsilon} = \hat{\epsilon}_{\text{OLS}}, \quad \hat{V}_{\text{RLS}} = \hat{V}_{\text{OLS}}.$$

Example S2. For $Y \in \mathbb{R}^N$, $X_0 \in \mathbb{R}^{N \times k}$, and $X_1, \dots, X_Q \in \mathbb{R}^{N \times l}$, consider the RLS fit of

$$Y = X_0 \hat{\beta}_0 + X_1 \hat{\beta}_1 + \dots + X_Q \hat{\beta}_Q + \hat{\epsilon} \quad \text{subject to } \hat{\beta}_1 = \dots = \hat{\beta}_Q,$$

where $\hat{\beta}_q$ ($q = 0, \dots, Q$) and $\hat{\epsilon}$ denote the RLS coefficient vectors and residuals, respectively. Let \hat{V}_{RLS} denote the robust covariance estimator of $\hat{\beta}_0$ by (S4). Alternatively, consider the OLS fit of

$$Y = X_0 \hat{\beta}_{0,\text{OLS}} + (X_1 + \dots + X_Q) \hat{\beta}_{\text{OLS}} + \hat{\epsilon}_{\text{OLS}}$$

as the corresponding restricted specification, where $\hat{\beta}_{0,\text{OLS}}$, $\hat{\beta}_{\text{OLS}}$, and $\hat{\epsilon}_{\text{OLS}}$ denote the OLS coefficient vectors and residuals, respectively. Let \hat{V}_{OLS} denote the EHW covariance estimator of $\hat{\beta}_{0,\text{OLS}}$. Then

$$\hat{\beta}_0 = \hat{\beta}_{0,\text{OLS}}, \quad \hat{\beta}_q = \hat{\beta}_{\text{OLS}} \quad (q = 1, \dots, Q), \quad \hat{\epsilon} = \hat{\epsilon}_{\text{OLS}}, \quad \hat{V}_{\text{RLS}} = \hat{V}_{\text{OLS}}.$$

Recall that $\hat{\Psi}_*$ and $\hat{\Psi}_{*,r}$ denote the robust covariances of \hat{Y}_* ($*$ = N, F) from the OLS fits of (4)–(5) and the RLS fits of (6), respectively. The numeric equivalence between $\hat{\Psi}_*$ and $\hat{\Psi}_{*,r}$ follows immediately from Examples S1 and S2.

Proposition S4. $\hat{\Psi}_* = \hat{\Psi}_{*,r}$ for $* = \text{N}, \text{F}$.

S4. Additional results on 2^K factorial experiments

Recall that $\tau_{\mathcal{K}} = c_{\mathcal{K}}^T \bar{Y}$ denotes the standard factorial effect corresponding to $\mathcal{K} \in \mathcal{P}_K$ under the 2^K factorial experiment. We will write $(\tau_{\mathcal{K}}, c_{\mathcal{K}})$ as $(\tau_{s,\mathcal{K}}, c_{s,\mathcal{K}})$ throughout this section to differentiate the standard factorial effects from variants under the {0,1} coding system. Let $c_{s,\emptyset} = c_{\emptyset} = 2^{-(K-1)} 1_Q$. Let $\mathcal{P}'_K = \{\emptyset\} \cup \mathcal{P}_K$, and let $Z_{i,\emptyset} = 1$ for all i .

Table S1: The choices of $(\mathcal{F}_+, \mathcal{F}'_+)$ in (S7) for the six factor-based regressions in Table 1. We use \emptyset to denote the empty set, and use $\{\emptyset\}$ to denote the set that contains the empty set as the only element.

| base model | regression equation | \mathcal{F}_+ | \mathcal{F}'_+ |
|---|--|-----------------|------------------------------------|
| $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}}$ (factor-saturated) | (17): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}}$ | \mathcal{P}_K | \emptyset |
| | (18): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} + x_i$ | | $\{\emptyset\}$ |
| | (19): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} + x_i + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} x_i$ | | \mathcal{P}'_K |
| $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}}$ (factor-unsaturated) | (20): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}}$ | \mathcal{F}_+ | \emptyset |
| | (21): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} + x_i$ | | $\{\emptyset\}$ |
| | (22): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} + x_i + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} x_i$ | | $\{\emptyset\} \cup \mathcal{F}_+$ |

S4.1. Design-based theory for general regression specifications

For $\mathcal{F}_+ \subseteq \mathcal{P}_K$ and $\mathcal{F}'_+ \subseteq \mathcal{P}'_K$, consider the regression specification

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} + \sum_{\mathcal{K} \in \mathcal{F}'_+} Z_{i,\mathcal{K}} \cdot x_i, \quad (\text{S7})$$

where $Z_{i,\emptyset} \cdot x_i = x_i$ if $\mathcal{K} = \emptyset \in \mathcal{F}'_+$. It is a general specification and includes (17)–(22) as special cases with different choices of $(\mathcal{F}_+, \mathcal{F}'_+)$, summarized in Table S1. The resulting regression is factor-saturated if $\mathcal{F}_+ = \mathcal{P}_K$, and factor-unsaturated if $\mathcal{F}_+ \subsetneq \mathcal{P}_K$ with $\mathcal{F}_- = \mathcal{P}_K \setminus \mathcal{F}_+ \neq \emptyset$.

Assume that \mathcal{F}_+ is non-empty throughout this section. We state below the design-based properties of the OLS outputs from (S7). The result includes those of (17)–(22) as special cases.

Recall $\tau_{s,\mathcal{K}} = c_{s,\mathcal{K}}^T \bar{Y}$ as the standard factorial effect corresponding to $\mathcal{K} \in \mathcal{P}_K$. Let

$$\tau_{s,+} = \{\tau_{s,\mathcal{K}} : \mathcal{K} \in \mathcal{F}_+\} = C_{s,+} \bar{Y}, \quad \tau_{s,-} = \{\tau_{s,\mathcal{K}} : \mathcal{K} \in \mathcal{F}_-\} = C_{s,-} \bar{Y}$$

concatenate the effects of interest and nuisance effects corresponding to \mathcal{F}_+ and $\mathcal{F}_- = \mathcal{P}_K \setminus \mathcal{F}_+$, respectively. To simplify the presentation, define $C_{s,-} = 0_J^T$ if \mathcal{F}_- is empty, with $\tau_{s,-} = 0$.

Let $\mathcal{F}'_- = \mathcal{P}'_K \setminus \mathcal{F}'_+$ analogous to \mathcal{F}_- , and let $C'_{s,-}$ concatenate rows of $\{c_{s,\mathcal{K}} : \mathcal{K} \in \mathcal{F}'_-\}$ with $C'_{s,-} = 0_J^T$ if \mathcal{F}'_- is empty.

Recall that $2^{-1}\tau_{s,\mathcal{K}}$, $2^{-1}(c_{s,\emptyset}^T \otimes I_J)\gamma$, and $2^{-1}(c_{s,\mathcal{K}}^T \otimes I_J)\gamma$ give the target parameters of $Z_{i,\mathcal{K}}$, x_i , and $Z_{i,\mathcal{K}} \cdot x_i$ for $\mathcal{K} \in \mathcal{P}_K$ in (19), respectively; see the proof of Proposition S5 in Section S8. Regression (S7) is thus a restricted variant of (19) subject to the separable restriction

$$C_{s,-} \bar{Y} = 0, \quad (C'_{s,-} \otimes I_J)\gamma = 0, \quad (\text{S8})$$

and allows us to estimate $\tau_{s,\mathcal{K}}$ by 2 times the OLS coefficient of $Z_{i,\mathcal{K}}$ for $\mathcal{K} \in \mathcal{P}_K$. In particular,

we say that there is no restriction on \bar{Y} if $\mathcal{F}_+ = \mathcal{P}_K$, with $C_{s,-}\bar{Y} = 0$ in (S8) correctly specified by definition. Likewise for there to be no restriction on γ if $\mathcal{F}'_+ = \mathcal{P}'_K$, with $(C'_{s,-} \otimes I_J)\gamma = 0$ in (S8) correctly specified by definition.

Let $\tilde{\tau}_{r,K}$ be 2 times the coefficient of $Z_{i,K}$ from the OLS fit of (S7), vectorized as

$$\tilde{\tau}_{r,+} = \{\tilde{\tau}_{r,K} : K \in \mathcal{F}_+\}.$$

Let $\tilde{\Omega}_{r,+}$ be the EHW covariance estimator of $\tilde{\tau}_{r,+}$ from the same OLS fit. Let $\hat{Y}_{r,s}$ be the coefficient vector of t_i from the RLS fit of (6) subject to (S8), and let $\hat{\Psi}_{r,s}$ be the double-decker-taco robust covariance estimator of $\hat{Y}_{r,s}$ from Section 4.2.4. Proposition S5 below states the numeric correspondence between $(\tilde{\tau}_{r,+}, \tilde{\Omega}_{r,+})$ and $(\hat{Y}_{r,s}, \hat{\Psi}_{r,s})$.

Proposition S5. $\tilde{\tau}_{r,+} = C_{s,+}\hat{Y}_{r,s}$ and $\tilde{\Omega}_{r,+} = C_{s,+}\hat{\Psi}_{r,s}C_{s,+}^T$.

The design-based properties of $(\tilde{\tau}_{r,+}, \tilde{\Omega}_{r,+})$ then follow from those of $(\hat{Y}_{r,s}, \hat{\Psi}_{r,s})$ in Lemma S1 and Theorems 1–4. Let $\hat{\gamma}_{r,s}$ and $V_{r,s}$ be the values of $\hat{\gamma}_r$ and V_r associated with $\hat{Y}_{r,s}$. By Propositions S2 and S3, $\hat{\gamma}_{r,s}$ belongs to \mathcal{B} regardless of whether the restriction is correctly specified or not, and satisfies $\text{plim } \hat{\gamma}_{r,s} = \gamma$ if $(C'_{s,-} \otimes I_J)\gamma = 0$. Corollary S1 below justifies the large-sample Wald-type inference of τ_s based on factor-saturated specifications for arbitrary choice of \mathcal{F}'_+ .

Corollary S1. For (S7) that is factor-saturated with $\mathcal{F}_+ = \mathcal{P}_K$, we have

$$\tilde{\tau}_{r,+} = C_s \hat{Y} \langle \hat{\gamma}_{r,s} \rangle.$$

Further assume complete randomization and Condition 4. Then

- (i) $\sqrt{N}(\tilde{\tau}_{r,+} - \tau_s) \rightsquigarrow \mathcal{N}(0, C_s V_{r,s} C_s^T)$ with $\tilde{\tau}_{r,+} \preceq_\infty \tilde{\tau}_L$;
- (ii) $\tilde{\Omega}_{r,+}$ is asymptotically conservative for estimating the true sampling covariance of $\tilde{\tau}_{r,+}$;
- (iii) $\tilde{\tau}_{r,+} \stackrel{\sim}{\rightarrow} \tilde{\tau}_L$ if

$$(C'_{s,-} \otimes I_J)\gamma = 0. \tag{S9}$$

The condition (S9) holds if Condition 2 holds and (S7) includes x_i .

Next, Corollary S2 below justifies the large-sample Wald-type inference of $\tau_{s,+}$ based on factor-unsaturated specifications when the nuisance effects excluded are indeed zero. From (15), let

$$U_s = I_Q - \Pi^{-1} C_{s,-}^T (C_{s,-} \Pi^{-1} C_{s,-}^T)^{-1} C_{s,-}, \quad \mu_{r,s} = -\Pi^{-1} C_{s,-}^T (C_{s,-} \Pi^{-1} C_{s,-}^T)^{-1} \tau_{s,-}$$

be the values of U and μ_r at $\rho_Y = C_{s,-}$, respectively, when $\mathcal{F}_- \neq \emptyset$. Recall $\tilde{\tau}_{*,+} = \{\tilde{\tau}_{*,K} : K \in \mathcal{F}_*\}$ as the estimators of $\tau_{s,+}$ from the factor-saturated regressions (17)–(19) with $\tilde{\tau}_{L,+} \succeq_\infty \tilde{\tau}_{N,+}, \tilde{\tau}_{F,+}$.

Corollary S2. For (S7) that is factor-unsaturated with $\mathcal{F}_+ \subsetneq \mathcal{P}_K$ and $\mathcal{F}_- \neq \emptyset$, we have

$$\tilde{\tau}_{r,+} - \tau_{s,+} - C_{s,+}\mu_{r,s} = C_{s,+}U_s(\hat{Y} \langle \hat{\gamma}_{r,s} \rangle - \bar{Y}),$$

where $C_{s,+}\mu_{r,s} = 0$ if $\tau_{s,-} = 0$. Further assume complete randomization and Condition 4. Then

- (i) $\sqrt{N}(\tilde{\tau}_{r,+} - \tau_{s,+} - C_{s,+}\mu_{r,s}) \rightsquigarrow \mathcal{N}(0, C_{s,+}U_s V_{r,s} U_s^T C_{s,+}^T)$;
- (ii) $\tilde{\Omega}_{r,+}$ is asymptotically conservative for estimating the true sampling covariance of $\tilde{\tau}_{r,+}$;
- (iii) $\tilde{\tau}_{r,+} \succeq_{\infty} \tilde{\tau}_{F,+} \sim \tilde{\tau}_{L,+} \succeq_{\infty} \tilde{\tau}_{N,+}$ if Condition 3 holds and (S7) includes x_i .

Recall that Condition 3 of constant treatment effects implies Condition 2 of equal correlations. Corollaries S1(iii) and S2(iii) together establish the asymptotic efficiency of additive regressions like $Y_i \sim 1 + \sum_{K \in \mathcal{F}_+} Z_{i,K} + x_i$ under constant treatment effects. Specifically, the resulting estimator is asymptotically as efficient as $\tilde{\tau}_L$ if the specification is factor-saturated with $\mathcal{F}_+ = \mathcal{P}_K$ by Corollary S1(iii), and ensures additional efficiency over $\tilde{\tau}_{L,+}$ if the specification is factor-unsaturated with $\mathcal{F}_+ \subsetneq \mathcal{P}_K$ by Corollary S2(iii). This illustrates the value of factor-unsaturated regressions in combination with covariate adjustment for improving efficiency.

Without constant treatment effects, a key limitation of the factor-unsaturated regressions with $\mathcal{F}_+ \subsetneq \mathcal{P}_K$ is that the consistency of $\tilde{\tau}_{r,+}$ depends critically on the actual absence of the nuisance effects. Proposition S6 below generalizes Proposition 3 and ensures the consistency of $\tilde{\tau}_{r,+}$ under equal-sized designs even when $\tau_{s,-} \neq 0$. The asymptotic efficiency of $\tilde{\tau}_{r,+}$, on the other hand, still requires the restriction on γ in (S8) to be correctly specified, and can no longer exceed that of $\tilde{\tau}_{L,+}$.

Proposition S6. Assume Condition 5. Then

$$\tilde{\tau}_{r,+} = C_{s,+}\hat{Y}\langle\hat{\gamma}_{r,s}\rangle. \quad (\text{S10})$$

Further assume complete randomization and Condition 4. Then

- (i) $\sqrt{N}(\tilde{\tau}_{r,+} - \tau_{s,+}) \rightsquigarrow \mathcal{N}(0, C_{s,+}V_{r,s}C_{s,+}^T)$ with $\tilde{\tau}_{r,+} \preceq_{\infty} \tilde{\tau}_{L,+}$;
- (ii) $\tilde{\tau}_{r,+} \sim \tilde{\tau}_{L,+}$ if $(C'_{s,-} \otimes I_J)\gamma = 0$.

Recall the definition of $\hat{Y}_{*,r}$ ($*$ = N, F, L) from Proposition 2. Let $\hat{\gamma}_{*,r}$ be the value of $\hat{\gamma}_r$ associated with $\hat{Y}_{*,r}$, with $\hat{\gamma}_{N,r} = 0_{JQ}$. Corollary S3 below is an immediate consequence of (S10) and underlies the asymptotic results in Proposition 3.

Corollary S3. Assume Condition 5. Then $\tilde{\tau}_{*,u,+} = C_{s,+}\hat{Y}\langle\hat{\gamma}_{*,r}\rangle$ for $* = N, F, L$ with $\tilde{\tau}_{N,u,+} = C_{s,+}\hat{Y}_N = \tilde{\tau}_{N,+}$.

Condition 5 depends on the design. Zhao and Ding (2022) showed that weighted least squares can secure the same benefit as equal-sized designs, and ensures the consistency of $\tilde{\tau}_{N,u,+}$ regardless of whether $\tau_{s,-}$ equals zero or not in covariate-free settings. The same result extends to the covariate-adjusted variant $\tilde{\tau}_{r,+}$ with minimal modification. We omit the details.

S4.2. Factorial effects under $\{0, 1\}$ -coded regressions

Define $Z_{ik}^0 = 2^{-1}(Z_{ik} + 1)$ as the counterpart of Z_{ik} under the $\{0, 1\}$ coding system.

Replacing $Z_{i,\mathcal{K}} = \prod_{k \in \mathcal{K}} Z_{ik}$ with $Z_{i,\mathcal{K}}^0 = \prod_{k \in \mathcal{K}} Z_{ik}^0$ in (19) yields

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}}^0 + x_i + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}}^0 \cdot x_i \quad (\text{S11})$$

as the fully interacted specification under the $\{0, 1\}$ coding system. Let

$$\Gamma_0 = \otimes_{k=1}^K \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} c_0^\top \\ C_0 \end{pmatrix}$$

with $c_0 = (1, 0_{Q-1}^\top)^\top$ and $C_0 1_Q = 0_{Q-1}$. Let $[-]$ denote the treatment level $q = (-1, \dots, -1) \in \mathcal{T}$ that has -1 in all dimensions. The comments after Proposition S5 extend here, and ensure that $\tau_0 = C_0 \bar{Y}$, $\gamma_{[-]}$, and $(C_0 \otimes I_J)\gamma$ give the target parameters of $(Z_{i,\mathcal{K}}^0)_{\mathcal{K} \in \mathcal{P}_K}$, x_i , and $(Z_{i,\mathcal{K}}^0 \cdot x_i)_{\mathcal{K} \in \mathcal{P}_K}$ in (S11), respectively; see the proof of Proposition S7 in Section S8. The elements of $\tau_0 = C_0 \bar{Y}$ define the analogs of the $2^K - 1$ standard factorial effects under a different weighting scheme; see Remark S1 at the end of this section. This is also known as *baseline parameterization* in the experimental design literature; see, e.g., Banerjee and Mukerjee (2008) and Mukerjee and Tang (2012).

Let $c_{0,\mathcal{K}}^\top$ be the row in C_0 that corresponds to $Z_{i,\mathcal{K}}^0$. Then $\tau_{0,\mathcal{K}} = c_{0,\mathcal{K}}^\top \bar{Y}$ and $(c_{0,\mathcal{K}}^\top \otimes I_J)\gamma$ give the target parameters of $Z_{i,\mathcal{K}}^0$ and $Z_{i,\mathcal{K}}^0 \cdot x_i$, respectively, for $\mathcal{K} \in \mathcal{P}_K$. Consider

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}}^0 + \sum_{\mathcal{K} \in \mathcal{F}'_+} Z_{i,\mathcal{K}}^0 \cdot x_i \quad (\text{S12})$$

as the $\{0, 1\}$ -coded analog of the general specification (S7). Let

$$\tau_{0,+} = \{\tau_{0,\mathcal{K}} : \mathcal{K} \in \mathcal{F}_+\} = C_{0,+} \bar{Y}, \quad \tau_{0,-} = \{\tau_{0,\mathcal{K}} : \mathcal{K} \in \mathcal{F}_-\} = C_{0,-} \bar{Y}$$

vectorize the effects of interest and nuisance effects corresponding to \mathcal{F}_+ and τ_0 , respectively, analogous to $\tau_{s,+}$ and $\tau_{s,-}$. Let $C'_{0,-}$ concatenate rows of $\{c_{0,\mathcal{K}} : \mathcal{K} \in \mathcal{F}'_-\}$, with $c_{0,\emptyset} = c_0$ if $\mathcal{K} = \emptyset \in \mathcal{F}'_-$. Following the convention in Section S4.1, define $C_{0,-} = 0_J^\top$ when \mathcal{F}_- is empty, and $C'_{0,-} = 0_J^\top$ when \mathcal{F}'_- is empty, respectively.

Let $\tilde{\tau}_{r,+}^0$ be the coefficient vector of $\{Z_{i,\mathcal{K}}^0 : \mathcal{K} \in \mathcal{F}_+\}$ from the OLS fit of (S12), with $\tilde{\Omega}_{r,+}^0$ as the associated EHW covariance estimator. Let $\tilde{\tau}_{*,+}^0$ be the corresponding estimators of $\tau_{0,+}$ from the $\{0, 1\}$ -coded analogs of the factor-saturated regressions (17)–(19) for $* = N, F, L$, analogous to $\tilde{\tau}_{*,+}$. Let $(\hat{Y}_{r,0}, \hat{\gamma}_{r,0})$ be the coefficient vectors of $(t_i, t_i \otimes x_i)$, respectively, from the RLS fit of (6) subject to the separable restriction

$$C_{0,-} \bar{Y} = 0, \quad (C'_{0,-} \otimes I_J) \otimes \gamma = 0.$$

Let $V_{r,0}$, U_0 , and $\mu_{r,0}$ be the corresponding values of V_r , U , and μ_r , respectively, paralleling $V_{r,s}$,

U_s , and $\mu_{r,s}$ from Section S4.1.

Proposition S7. Proposition S5 and Corollaries S1–S2 hold for inference of $\tau_{0,+}$ based on (S12) if we change (i) all $(\tilde{\tau}_{r,+}, \tilde{\Omega}_{r,+}, \tilde{\tau}_{*,+})$ to $(\tilde{\tau}_{r,+}^0, \tilde{\Omega}_{r,+}^0, \tilde{\tau}_{*,+}^0)$ for $* = N, F, L$, and (ii) all subscripts “s” to “0”.

A key distinction between the two coding systems is that the result in Proposition S6 under equal-sized designs no longer holds here, owing to the loss of orthogonality between the contrast vectors that define τ_0 . The consistency of $\tilde{\tau}_{r,+}^0$ thus in general requires the actual absence of the nuisance effects even under equal-sized designs.

Remark S1. We can show that $\tau_{0,\mathcal{K}}$ equals the effect of the factors in $\mathcal{K} \in \mathcal{P}_K$ when the rest of the factors are fixed at -1 (Zhao and Ding 2022). Denote by $-_k$ and $_k$ the -1 and $+1$ levels of factor k , respectively, when multiple factors are concerned. Let $(z_k, [-])$ denote the treatment combination with factor k at level $z_k \in \{-_k, +_k\}$ and the rest of the factors all at -1 . Then

$$\tau_{0,\{k\}} = \bar{Y}(+_k, [-]) - \bar{Y}(-_k, [-])$$

for $\mathcal{K} = \{k\}$, measuring the main effect of factor k when the rest of the factors are fixed at -1 . Let $(z_k, z_{k'}, [-])$ denote the treatment combination with factors k and k' ($\neq k$) at levels $z_k \in \{-_k, +_k\}$ and $z_{k'} \in \{-_{k'}, +_{k'}\}$, respectively, and the rest of the factors all at -1 . Then

$$\tau_{0,\{k,k'\}} = \bar{Y}(-_k, -_{k'}, [-]) + \bar{Y}(+_k, +_{k'}, [-]) - \bar{Y}(-_k, +_{k'}, [-]) - \bar{Y}(+_k, -_{k'}, [-])$$

for $\mathcal{K} = \{k, k'\}$, measuring the interaction effect of factors k and k' when the rest of the factors are fixed at -1 . The intuition extends to general $\mathcal{K} \in \mathcal{P}_K$ and elucidates the causal interpretation of $\tau_{0,\mathcal{K}}$.

S5. Rerandomization using the Mahalanobis distance

S5.1. Overview

Rerandomization discards treatment allocations that do not satisfy a prespecified covariate balance criterion in the design stage of experiments (Cox 1982; Morgan and Rubin 2012), enforcing covariate balance for additional efficiency. A special type of rerandomization, ReM, uses the Mahalanobis distance between covariate means by treatment group as the balance criterion under the treatment-control experiment, and accepts a randomization if and only if the distance does not exceed some prespecified threshold (Morgan and Rubin 2012; Li et al. 2018; Li and Ding 2020; Zhao and Ding 2021). Branson et al. (2016) and Li et al. (2020) extended the discussion to 2^K factorial experiments, yet did not consider regression adjustment in the analysis stage.

We provide in this section a unified theory for ReM and regression adjustment in multi-armed experiments. Specifically, we quantify the impact of ReM on the asymptotic efficiency of $\hat{\tau}_L$ and $\hat{\tau}_r$, respectively, with $\hat{\tau}_*$ ($*$ = N, F) being special cases of $\hat{\tau}_r$. The results are coherent with the existing

theory under the treatment-control experiment (Li and Ding 2020): ReM has no effect on $\hat{\tau}_L$ asymptotically, yet improves the asymptotic efficiency of $\hat{\tau}_r$ when the restriction on the correlations between potential outcomes and covariates is misspecified and separate from that on the average potential outcomes. The resulting estimator, though still not as efficient as $\hat{\tau}_L$ asymptotically, can have better finite sample performance when the sample size is moderate relative to the number of covariates or treatments. This illustrates the duality between ReM and regression adjustment for improving efficiency under multi-armed experiments, and further expands the theoretical guarantees of $\hat{\tau}_r$. The combination of ReM and $\hat{\tau}_r$, in addition to delivering all guarantees as under complete randomization, further reduces the loss in asymptotic efficiency when the restriction is misspecified.

Let

$$\|\delta\|_{\mathcal{M}} = \delta^T \{\text{cov}(\delta)\}^{-1} \delta$$

denote the Mahalanobis distance of a random vector δ from the origin.

S5.2. ReM under multi-armed experiments

Recall $\hat{x}(q) = N_q^{-1} \sum_{i:Z_i=q} x_i$ as the sample mean of covariates under treatment level $q \in \mathcal{T}$. ReM under the treatment-control experiment measures covariate balance by the Mahalanobis distance between $\hat{x}(0)$ and $\hat{x}(1)$, which is equivalent to the Mahalanobis distance of $\hat{\tau}_x = \hat{x}(1) - \hat{x}(0)$ from the origin. The presence of multiple treatment arms permits more flexible measures of covariate balance.

Specifically, denote by $\hat{x} = (\hat{x}^T(1), \dots, \hat{x}^T(Q))^T \in \mathbb{R}^{JQ}$ the vectorization of $\hat{x}(q)$'s over $q \in \mathcal{T}$. For a prespecified $Q \times 1$ contrast vector $g_1 = (g_{11}, \dots, g_{1Q})^T$ with $\sum_{q \in \mathcal{T}} g_{1q} = 0$, the contrast of $\hat{x}(q)$'s, denoted by

$$\hat{\delta}_{g_1} = \sum_{q \in \mathcal{T}} g_{1q} \hat{x}(q) = (g_1^T \otimes I_J) \hat{x} \in \mathbb{R}^J,$$

defines a measure of covariate balance across the Q treatment arms. Intuitively, a well-balanced allocation would have $\hat{\delta}_{g_1}$ that is close to 0_J . The $\hat{\tau}_x$ under the treatment-control experiment is a special case with $Q = 2$ and $g = (-1, 1)^T$.

Assume that $H \geq 1$ of such contrasts are of interest, denoted by $\hat{\delta}_{g_h} = (g_h^T \otimes I_J) \hat{x} \in \mathbb{R}^J$ for $h = 1, \dots, H$, with $(g_h)_{h=1}^H$ being H prespecified, linearly independent contrast vectors. We can vectorize them as

$$\hat{\delta} = (\hat{\delta}_{g_1}^T, \dots, \hat{\delta}_{g_H}^T)^T = (G \otimes I_J) \hat{x} \in \mathbb{R}^{JH}$$

with $G = (g_1, \dots, g_H)^T$, and conduct ReM based on the Mahalanobis distance between $\hat{\delta}$ and 0_{JH} . We formalize the intuition in Definition S3 below.

Definition S3 (ReM). Draw an initial treatment allocation by the complete randomization in Definition 1, and accept it if and only if the resulting $\hat{\delta}$ satisfies $\|\hat{\delta}\|_{\mathcal{M}} \leq a$ for some prespecified

threshold a .

Complete randomization is a special case with $a = \infty$. The linear independence of g_h 's limits the maximum number of contrast vectors to $H \leq Q - 1$. In the treatment-control experiment, this implies that there can only be one contrast vector, proportional to $g = (-1, 1)^T$. The resulting $\hat{\delta}$ is proportional to $\hat{\tau}_x$, illustrating the uniqueness of balance criterion when $Q = 2$. In the 2^K factorial experiment, Branson et al. (2016) and Li et al. (2020) considered $G = C_s$, with $H = Q - 1$ contrasts as those that define the $Q - 1$ standard factorial effects. Experimenters in practice may choose G that differs from C_s , with $H \ll Q$ when Q is large. An intuitive choice is $G = (c_{\{1\}}, \dots, c_{\{K\}})^T$, corresponding to the contrast vectors that define the K main effects. This suggests the need for theory of ReM with general G .

S5.3. Asymptotic properties

Recall the definition of $\hat{Y}\langle b \rangle$ and \mathcal{Y} from (S1), with $\hat{Y}\langle b \rangle \stackrel{\sim}{\rightarrow} \hat{Y}_L$ if $\text{plim } b = \gamma$ by Lemma S1. Proposition S8 and Theorem S4 below clarify the utility of ReM for improving the asymptotic efficiency of suboptimally adjusted $\hat{Y}\langle b \rangle$ with $\text{plim } b \neq \gamma$. Recall that

- (i) $\hat{Y}_* = \hat{Y}\langle b_* \rangle \in \mathcal{Y}$ for $* = N, F, L$ from Proposition S1;
- (ii) $\hat{Y}_r = \hat{Y}\langle \hat{\gamma}_r \rangle \in \mathcal{Y}$ under RLS subject to the correlation-only restriction (14) by Proposition S2;
- (iii) $\hat{Y}_r - \bar{Y} = U(\hat{Y}\langle \hat{\gamma}_r \rangle - \bar{Y})$ with $\hat{Y}\langle \hat{\gamma}_r \rangle \in \mathcal{Y}$ under RLS subject to the separable restriction (13) when $\rho_Y \neq 0$ and $\rho_Y \bar{Y} = r_Y$ is correctly specified by Proposition S3.

The results on $\hat{Y}\langle b \rangle$ hence imply the impact of ReM on $\hat{\tau}_*$ ($* = N, F, L$) and $\hat{\tau}_r$ as direct consequences.

To begin with, Proposition S8 below states the asymptotic distribution of $\hat{Y}\langle b \rangle$ under ReM.

Denote by $(\hat{Y}\langle b \rangle \mid \mathcal{A})$, where $\mathcal{A} = \{\|\hat{\delta}\|_{\mathcal{M}} \leq a\}$, the sampling distribution of $\hat{Y}\langle b \rangle$ under the ReM in Definition S3. Let

$$V_{b,\infty}^{\parallel} = D_{b,\infty}(\Phi \otimes S_x^2)D_{b,\infty}^T, \quad V_{b,\infty}^{\perp} = V_{b,\infty} - V_{b,\infty}^{\parallel} \quad (\text{S13})$$

with $\Phi = \Pi^{-1}G^T(G\Pi^{-1}G^T)^{-1}G\Pi^{-1}$. Following Li et al. (2020), let $(V_{b,\infty}^{\parallel})_{JH}^{1/2}$ denote a $Q \times JH$ matrix that satisfies $(V_{b,\infty}^{\parallel})_{JH}^{1/2}\{(V_{b,\infty}^{\parallel})_{JH}^{1/2}\}^T = V_{b,\infty}^{\parallel}$. Let $\epsilon \sim \mathcal{N}(0_Q, I_Q)$ and $\mathcal{L} \sim \epsilon' \mid (\|\epsilon'\|_2^2 \leq a)$ be two independent standard and truncated normal random vectors with $\epsilon' \sim \mathcal{N}(0_{JH}, I_{JH})$. Then $\mathcal{L} \succeq \epsilon'$ with mean 0_{JH} and covariance $\nu_{JH,a}I_{JH} < I_{JH}$.

Proposition S8. Assume the ReM in Definition S3 and Condition 4. For $\hat{Y}\langle b \rangle \in \mathcal{Y}$ with $\text{plim } b = b_{\infty}$, we have

- (i) $\{\sqrt{N}(\hat{Y}\langle b \rangle - \bar{Y}) \mid \mathcal{A}\} \rightsquigarrow (V_{b,\infty}^{\perp})^{1/2} \cdot \epsilon + (V_{b,\infty}^{\parallel})_{JH}^{1/2} \cdot \mathcal{L}$, where the limiting distribution satisfies

$$\mathcal{N}(0_Q, V_L) \succeq (V_{b,\infty}^{\perp})^{1/2} \cdot \epsilon + (V_{b,\infty}^{\parallel})_{JH}^{1/2} \cdot \mathcal{L} \succeq \mathcal{N}(0_Q, V_{b,\infty}). \quad (\text{S14})$$

(ii) $\{\sqrt{N}(\hat{Y}\langle b \rangle - \bar{Y}) \mid \mathcal{A}\} \rightsquigarrow \mathcal{N}(0_Q, V_L)$ if and only if

$$D_{b,\infty}(\Phi \otimes S_x^2)D_{b,\infty}^T = 0. \quad (\text{S15})$$

A sufficient condition for (S15) is $b_\infty = \gamma$.

Recall that $\sqrt{N}(\hat{Y}_L - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, V_L)$ and $\sqrt{N}(\hat{Y}\langle b \rangle - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, V_{b,\infty})$ under complete randomization by Lemma S1. The relationship in (S14) provides the basis for comparing the asymptotic relative efficiency between $(\hat{Y}\langle b \rangle \mid \mathcal{A})$, $\hat{Y}\langle b \rangle$, and \hat{Y}_L , which we formalized in Theorem S4 below. The choice of $(V_{b,\infty}^\parallel)_{JH}^{1/2}$ is not unique, but the asymptotic distribution is (Li et al. 2020).

In addition, recall that $\hat{Y}\langle b \rangle \approx \hat{Y}\langle \gamma \rangle - D_{b,\infty}\hat{x}$ from Theorem S1. Lemma S6 in Section S6.1 further ensures that $\hat{Y}\langle \gamma \rangle$ is asymptotically independent of \hat{x} . ReM based on $\hat{\delta} = (G \otimes I_J)\hat{x}$ thus affects only the $D_{b,\infty}\hat{x}$ part asymptotically, and increases its peakedness if $b_\infty \neq \gamma$. This gives the intuition for the asymptotic distribution of $\hat{Y}\langle b \rangle$ being the convolution of a normal and a truncated normal when $D_{b,\infty} \neq 0$, and ensures that the asymptotic sampling distribution of $\hat{Y}_L = \hat{Y}\langle b_L \rangle$ remains unchanged under ReM given $\text{plim } b_L = \gamma$.

A key distinction from the treatment-control experiment is that choosing a small a alone no longer suffices to ensure that $\hat{Y}\langle b \rangle$ is asymptotically almost as efficient as \hat{Y}_L when $b_\infty \neq \gamma$ (Li et al. 2018). In particular, a small a ensures $N\text{cov}_\infty\{\hat{Y}\langle b \rangle\} \approx V_{b,\infty}^\perp$, with

$$V_{b,\infty}^\perp - V_L = D_{b,\infty}\{(\Pi^{-1} - 1_{Q \times Q} - \Phi) \otimes S_x^2\}D_{b,\infty}^T \geq 0.$$

To have $V_{b,\infty}^\perp = V_L$ thus entails additional conditions. Without further assumptions on b_∞ , this requires $\Pi^{-1} - 1_{Q \times Q} - \Phi = 0$, with a sufficient condition given by $H = Q - 1$; see Lemma S7 in Section S6.1. ReM under the treatment-control experiment is a special case with $H = 1 = Q - 1$.

Theorem S4 builds on Proposition S8 and quantifies the asymptotic relative efficiency of $\hat{Y}\langle b \rangle$ under complete randomization and ReM.

Theorem S4. Assume the ReM in Definition S3 and Condition 4. For $\hat{Y}\langle b \rangle \in \mathcal{Y}$ with $\text{plim } b = b_\infty$, we have

$$\hat{Y}_L \stackrel{\sim}{\rightarrow} (\hat{Y}_L \mid \mathcal{A}) \succeq_\infty (\hat{Y}\langle b \rangle \mid \mathcal{A}) \succeq_\infty \hat{Y}\langle b \rangle$$

with $(\hat{Y}\langle b \rangle \mid \mathcal{A}) \sim \hat{Y}\langle b \rangle \sim \hat{Y}_L$ if $b_\infty = \gamma$.

Recall that $\text{plim } \hat{\gamma}_r = \gamma$ when the restriction on γ is correctly specified and separate from that on \bar{Y} . ReM thus has no effect on the asymptotic distribution of \hat{Y}_L , or \hat{Y}_r when the restriction on γ is correctly specified, but improves the asymptotic efficiency of \hat{Y}_r if $\rho_\gamma\gamma = r_\gamma$ is misspecified under the separable restriction. Inference based on \hat{Y}_L under ReM can therefore use the same normal approximation as under complete randomization; likewise for that based on \hat{Y}_r when the restriction on γ is correctly specified. The same normal approximation, however, will be overconservative when $\rho_\gamma\gamma = r_\gamma$ is misspecified, suggesting the need of more accurate ReM-specific inference. We relegate the details to Section S5.4.

In addition to the efficiency boost for the suboptimally adjusted $\hat{Y}\langle b \rangle$'s, ReM also improves the coherence between estimators based on different regression adjustments. Let $E_\infty(\cdot | \mathcal{A})$ denote the asymptotic expectation under ReM.

Corollary S4. Assume the ReM in Definition S3 and Condition 4. Then

$$E_\infty(\|\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle\|_2^2 | \mathcal{A}) = \nu_{JH,a} E_\infty(\|\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle\|_2^2) \leq E_\infty(\|\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle\|_2^2)$$

with $\nu_{JH,a} = P(\chi_{JH+2}^2 < a)/P(\chi_{JH}^2 < a) < 1$ for all $b \neq b' \in \mathcal{B}$.

Recall that \hat{Y}_* ($*$ = N, F, L) are all special cases of $\hat{Y}\langle b \rangle$ by Proposition S1. Corollary S4 ensures that the discrepancy between the unadjusted and adjusted estimators is smaller under ReM than under complete randomization. This is a desirable property in empirical research.

S5.4. ReM-specific inference

Recall that ReM increases the peakedness of \hat{Y}_r when the restriction on γ is misspecified. The usual normal approximation will therefore be overconservative. We consider below more accurate large-sample inference based on ReM-specific sampling distributions.

Recall from Proposition S8 the asymptotic distribution of $\hat{Y}\langle b \rangle$ under ReM. With ϵ and \mathcal{L} both following known distributions, the only parts that are unknown are $V_{b,\infty}^\perp$ and $V_{b,\infty}^\parallel$. We can estimate them using their respective sample analogs, denoted by $\hat{V}_{b,\infty}^\perp$ and $\hat{V}_{b,\infty}^\parallel$, respectively, and conduct ReM-specific inference based on the distribution of

$$(\hat{V}_{b,\infty}^\perp)^{1/2} \cdot \epsilon + (\hat{V}_{b,\infty}^\parallel)^{1/2} \cdot \mathcal{L}. \quad (\text{S16})$$

Proposition S9 below follows from Li et al. (2018, Lemma A10) and ensures the asymptotic validity of the inference based on (S16).

Recall that $\hat{\gamma}_q$ denotes the coefficient vector of x_i from the OLS fit of $Y_i \sim 1 + x_i$ over $\{i : Z_i = q\}$ and gives the sample analog of γ_q . Let

$$\hat{S}_{b,\infty,qq} = (N_q - 1)^{-1} \sum_{i:Z_i=q} \left[Y_i - b_q^T x_i - \{\hat{Y}(q) - b_q^T \hat{x}(q)\} \right]^2, \quad \hat{V}_{b,\infty} = \text{diag}(\hat{S}_{b,\infty,qq}/e_q)_{q \in \mathcal{T}}$$

be the sample analogs of $S_{b,\infty,qq}$ and $V_{b,\infty}$, respectively. Then

$$\hat{V}_{b,\infty}^\parallel = \hat{D}_{b,\infty}(\Phi \otimes S_x^2) \hat{D}_{b,\infty}^T, \quad \hat{V}_{b,\infty}^\perp = \hat{V}_{b,\infty} - \hat{V}_{b,\infty}^\parallel,$$

where $\hat{D}_{b,\infty} = \text{diag}\{(b_q - \hat{\gamma}_q)^T\}_{q \in \mathcal{T}}$.

Proposition S9. Assume the ReM in Definition S3 and Condition 4. Then

$$\hat{V}_{b,\infty}^\parallel = V_{b,\infty}^\parallel + o_P(1), \quad \hat{V}_{b,\infty}^\perp = V_{b,\infty}^\perp + S_{b,\infty} + o_P(1).$$

For $\hat{Y}_* = \hat{Y}\langle b_* \rangle$ ($*$ = N, F) as direct outputs from OLS, we can also estimate $V_{b,\infty} = V_*$ by $N\hat{\Psi}_*$, which is asymptotically equivalent to the $\hat{V}_{b,\infty}$ defined above.

S6. Proof of the results on $\hat{Y}\langle b \rangle \in \mathcal{Y}$

Assume throughout this section that $b = (b_1^T, \dots, b_Q^T)^T \in \mathcal{B}$, with $\text{plim } b = b_\infty = (b_{1,\infty}^T, \dots, b_{Q,\infty}^T)^T$ denoting its probability limit under complete randomization and Condition 4. We verify below the sampling properties of $\hat{Y}\langle b \rangle$ under complete randomization and the ReM in Definition S3, respectively. Recall that $D_b = \text{diag}\{(b_q - \gamma_q)^T\}_{q \in \mathcal{T}}$ and $D_{b,\infty} = \text{plim } D_b = \text{diag}\{(b_{q,\infty} - \gamma_q)^T\}_{q \in \mathcal{T}}$. Recall $\Pi = \text{diag}(e_q)_{q \in \mathcal{T}}$. Let \circ denote the Hadamard product of matrices. Then $V_{b,\infty} = (\Pi^{-1} - 1_{Q \times Q}) \circ S_{b,\infty}$ for $b \in \mathcal{B}$, with $V_* = \text{diag}(S_{*,qq}/e_q)_{q \in \mathcal{T}} - S_* = (\Pi^{-1} - 1_{Q \times Q}) \circ S_*$ as special cases for $* = \text{N, F, L}$.

S6.1. Lemmas

Lemma S4. (Li and Ding 2017, Theorems 3 and 5) Assume the complete randomization in Definition 1. Let $Y_i(q)$ be the $L \times 1$ potential outcome vector of unit i under treatment q . Let $\bar{Y}(q) = N^{-1} \sum_{i=1}^N Y_i(q)$ and $S_{qq'} = (N-1)^{-1} \sum_{i=1}^N \{Y_i(q) - \bar{Y}(q)\} \{Y_i(q') - \bar{Y}(q')\}^T$ be the finite population mean and covariance for $q, q' \in \mathcal{T}$, respectively. Let $\tau = \sum_{q \in \mathcal{T}} \Gamma_q \bar{Y}(q)$, where Γ_q is an arbitrary $K \times L$ coefficient matrix for $q \in \mathcal{T}$. Then

- (i) $\hat{\tau} = \sum_{q \in \mathcal{T}} \Gamma_q \hat{Y}(q)$ has mean τ and covariance $\text{cov}(\hat{\tau}) = \sum_{q \in \mathcal{T}} N_q^{-1} \Gamma_q S_{qq} \Gamma_q^T - N^{-1} S_\tau^2$, where S_τ^2 is the finite population covariance of $\{\tau_i = \sum_{q \in \mathcal{T}} \Gamma_q Y_i(q) : i = 1, \dots, N\}$;
- (ii) if for all $q, q' \in \mathcal{T}$, $S_{qq'}$ has a finite limit, N_q/N has a limit in $(0, 1)$, and $\max_{i=1, \dots, N} \|Y_i(q) - \bar{Y}(q)\|_2^2/N = o(1)$, then $N\text{cov}(\hat{\tau})$ has a limiting value, denoted by V , and

$$\sqrt{N}(\hat{\tau} - \tau) \rightsquigarrow \mathcal{N}(0, V).$$

Lemma S5. Assume complete randomization. Then

$$\begin{aligned} N\text{cov}(\hat{Y}_N) &= (\Pi^{-1} - 1_{Q \times Q}) \circ S, \\ V_x = N\text{cov}(\hat{x}) &= (\Pi^{-1} - 1_{Q \times Q}) \otimes S_x^2, \\ N\text{cov}(\hat{Y}_N, \hat{x}) &= \left(\begin{array}{cccc} (e_1^{-1} - 1)\gamma_1^T & -\gamma_1^T & \dots & -\gamma_1^T \\ -\gamma_2^T & (e_2^{-1} - 1)\gamma_2^T & \dots & -\gamma_2^T \\ \vdots & \vdots & & \vdots \\ -\gamma_Q^T & -\gamma_Q^T & \dots & (e_Q^{-1} - 1)\gamma_Q^T \end{array} \right) \otimes S_x^2. \end{aligned}$$

Further assume Condition 4. Then $\sqrt{N}((\hat{Y}_N - \bar{Y})^T, \hat{x}^T)^T \rightsquigarrow \mathcal{N}(0, V)$, where V is the finite limit of $N\text{cov}((\hat{Y}_N^T, \hat{x}^T)^T)$.

Proof of Lemma S5. See covariates as potential outcomes unaffected by the treatment. Define

$$\begin{pmatrix} Y_i(q) \\ x_i \end{pmatrix} \quad (i = 1, \dots, N)$$

as the pseudo potential outcome vectors under treatment q . The result follows from Lemma S4 with

$$\Gamma_q = \begin{pmatrix} a_{\cdot q} & \\ & a_{\cdot q} \otimes I_J \end{pmatrix}, \quad \text{where } a_{\cdot q} \text{ denotes the } q\text{th column of } I_Q.$$

We omit the algebraic details. \square

Lemma S6. Assume complete randomization. Then

$$E(\hat{Y}\langle\gamma\rangle) = \bar{Y}, \quad \text{cov}(\hat{Y}\langle\gamma\rangle) = N^{-1}V_L, \quad \text{cov}(\hat{Y}\langle\gamma\rangle, \hat{x}) = 0.$$

Further assume Condition 4. Then

$$\sqrt{N} \begin{pmatrix} \hat{Y}\langle\gamma\rangle - \bar{Y} \\ \hat{x} \end{pmatrix} \rightsquigarrow \mathcal{N} \left\{ 0, \begin{pmatrix} V_L & 0 \\ 0 & V_x \end{pmatrix} \right\},$$

where $V_x = N\text{cov}(\hat{x}) = (\Pi^{-1} - 1_{Q \times Q}) \otimes S_x^2$.

Proof of Lemma S6. Recall that $\hat{Y}\langle\gamma\rangle = \hat{Y}_N - \{\text{diag}(\gamma_q^T)_{q \in \mathcal{T}}\}\hat{x}$. The result follows from Lemma S5 with $\text{cov}(\hat{Y}\langle\gamma\rangle, \hat{x}) = \text{cov}(\hat{Y}_N, \hat{x}) - \{\text{diag}(\gamma_q^T)_{q \in \mathcal{T}}\}\text{cov}(\hat{x}) = 0$. \square

Lemma S7. Recall from (S13) that $\Phi = \Pi^{-1}G^T(G\Pi^{-1}G^T)^{-1}G\Pi^{-1}$, where G is an $H \times Q$ contrast matrix that has full row rank. We have $\Phi = \Pi^{-1} - 1_{Q \times Q}$ if $H = Q - 1$.

The proof of Lemma S7 follows from direct linear algebra and is omitted.

S6.2. Results under complete randomization

Proof of Lemma S1 and Theorem S1. The results follow from $\hat{Y}\langle b \rangle = \hat{Y}\langle\gamma\rangle - D_b\hat{x} \stackrel{\text{d}}{\approx} \hat{Y}\langle\gamma\rangle - D_{b,\infty}\hat{x}$ by Lemma S5 and Slutsky's theorem. \square

S6.3. Results under ReM

We verify below Proposition S8, which implies Theorem S4 as a direct consequence. For random variables A and B , let

$$\text{proj}(A \mid 1, B) = E(A) + \text{cov}(A, B)\{\text{cov}(B)\}^{-1}\{B - E(B)\}$$

denote the linear projection of A onto $(1, B)$, with residual $\text{res}(A \mid 1, B) = A - \text{proj}(A \mid 1, B)$.

Let

$$\mu_x = \text{proj}(\hat{x} \mid 1, \hat{\delta}) = V_{x\delta} V_{\delta\delta}^{-1} \hat{\delta}, \quad r_x = \text{res}(\hat{x} \mid 1, \hat{\delta}) = \hat{x} - \mu_x$$

with $V_{x\delta} = N\text{cov}(\hat{x}, \hat{\delta})$ and $V_{\delta\delta} = N\text{cov}(\hat{\delta})$. We have

$$\begin{aligned} V_{x\delta} &= N\text{cov}(\hat{x})(G^T \otimes I_J) = (\Pi^{-1} G^T) \otimes S_x^2, \\ V_{\delta\delta} &= (G \otimes I_J) \{N\text{cov}(\hat{x})\} (G^T \otimes I_J) = (G\Pi^{-1} G^T) \otimes S_x^2 \end{aligned}$$

by $\hat{\delta} = (G \otimes I_J)\hat{x}$, $N\text{cov}(\hat{x}) = (\Pi^{-1} - 1_{Q \times Q}) \otimes S_x^2$ from Lemma S5, and $1_{Q \times Q} G^T = 0$. This ensures

$$\text{cov}(\mu_x) = N^{-1} V_{x\delta} V_{\delta\delta}^{-1} V_{x\delta}^T = N^{-1} (\Phi \otimes S_x^2), \quad (\text{S17})$$

where $\Phi = \Pi^{-1} G^T (G\Pi^{-1} G^T)^{-1} G\Pi^{-1}$ as defined in (S13).

Proof of Proposition S8. Let $A = \hat{Y}\langle\gamma\rangle - \bar{Y} - D_{b,\infty}r_x$ and $B = -D_{b,\infty}\mu_x$ with

$$A + B = \hat{Y}\langle\gamma\rangle - D_{b,\infty}\hat{x} - \bar{Y} \stackrel{\cdot}{\approx} \hat{Y}\langle b \rangle - \bar{Y} \quad (\text{S18})$$

by Theorem S1. This ensures

$$\{\sqrt{N}(\hat{Y}\langle b \rangle - \bar{Y}) \mid \mathcal{A}\} \stackrel{\cdot}{\sim} \{\sqrt{N}(A + B) \mid \mathcal{A}\}. \quad (\text{S19})$$

We derive below the asymptotic distribution of $\{\sqrt{N}(A + B) \mid \mathcal{A}\}$.

First, it follows from Lemma S6 that $(\hat{Y}\langle\gamma\rangle, r_x, \mu_x)$ are pairwise uncorrelated in finite samples, and asymptotically independent and jointly normally distributed. This ensures that A and B are uncorrelated in finite samples, and asymptotically independent and jointly normally distributed.

Second, $A + B = \hat{Y}\langle b_\infty \rangle - \bar{Y}$ with $\text{cov}(\hat{Y}\langle b_\infty \rangle) = N^{-1}V_{b,\infty}$. We have

$$\begin{aligned} \text{cov}(B) &= D_{b,\infty} \text{cov}(\mu_x) D_{b,\infty}^T = N^{-1} V_{b,\infty}^\parallel, \\ \text{cov}(A) &= \text{cov}(\hat{Y}\langle b_\infty \rangle) - \text{cov}(B) = N^{-1} V_{b,\infty}^\perp \end{aligned} \quad (\text{S20})$$

by (S17) and the uncorrelatedness of A and B . This, together with $E(A) = E(B) = 0_Q$, ensures

$$(\sqrt{N}A \mid \mathcal{A}) \stackrel{\cdot}{\sim} (V_{b,\infty}^\perp)^{1/2} \cdot \epsilon, \quad (\sqrt{N}B \mid \mathcal{A}) \stackrel{\cdot}{\sim} (V_{b,\infty}^\parallel)_{JH}^{1/2} \cdot \mathcal{L},$$

and hence

$$\{\sqrt{N}(A + B) \mid \mathcal{A}\} \stackrel{\cdot}{\sim} (V_{b,\infty}^\perp)^{1/2} \cdot \epsilon + (V_{b,\infty}^\parallel)_{JH}^{1/2} \cdot \mathcal{L}$$

by the asymptotic independence between A and $(\hat{\delta}, B)$. The asymptotic distribution of $(\hat{Y}\langle b \rangle \mid \mathcal{A})$ then follows from (S19).

In addition, it follows from the definition of A and the uncorrelatedness of $\hat{Y}\langle\gamma\rangle$ and r_x that

$$\text{cov}(A) = \text{cov}(\hat{Y}\langle\gamma\rangle) + D_{b,\infty}\text{cov}(r_x)D_{b,\infty} = N^{-1}V_L + D_{b,\infty}\text{cov}(r_x)D_{b,\infty}.$$

Juxtapose this with (S20) to see that $V_{b,\infty}^\perp = V_L$ if and only if

$$D_{b,\infty}\text{cov}(r_x)D_{b,\infty}^T = 0. \quad (\text{S21})$$

Without restrictions on b , condition (S21) requires $\text{cov}(r_x) = 0$, which is equivalent to

$$\Phi = \Pi^{-1} - 1_{Q \times Q}$$

by $\text{cov}(r_x) = \text{cov}(\hat{x}) - \text{cov}(\mu_x) = (\Pi^{-1} - 1_{Q \times Q} - \Phi) \otimes S_x^2$ from (S17). The sufficiency of $H = Q - 1$ follows from Lemma S7. The necessity of $H = Q - 1$ follows from $\text{rank}(\Pi^{-1} - 1_{Q \times Q}) = Q - 1$ given $\text{rank}(\Pi^{-1} - 1_{Q \times Q}) + \text{rank}(1_{Q \times Q}) \geq \text{rank}(\Pi^{-1}) = Q$. \square

Proof of Corollary S4. Direct algebra shows that

$$\begin{aligned} E_\infty(\|\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle\|_2^2) &= E_\infty\left(\text{tr}\left[\left(\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle\right)\left(\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle\right)^T\right]\right) \\ &= \text{tr}\left[\text{cov}_\infty(\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle)\right] \end{aligned}$$

and likewise

$$E_\infty(\|\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle\|_2^2 | \mathcal{A}) = \text{tr}\left[\text{cov}_\infty(\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle | \mathcal{A})\right].$$

The result then follows from $\text{cov}_\infty(\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle | \mathcal{A}) = \nu_{JH,a} \cdot \text{cov}_\infty(\hat{Y}\langle b \rangle - \hat{Y}\langle b' \rangle)$. \square

S7. Proof of the results on treatment-based regressions

S7.1. Notation and useful facts

Let $Y = (Y_1, \dots, Y_N)^T$, $T = (t_1, \dots, t_N)^T$, $X = (x_1, \dots, x_N)^T$, and $T_x = (t_1 \otimes x_1, \dots, t_N \otimes x_N)^T$. We can unify the treatment-based regressions (4)–(6) in matrix form as

$$Y = \chi_* \hat{\theta}_* + \hat{\epsilon}_* \quad (* = N, F, L)$$

with design matrices

$$\chi_N = T, \quad \chi_F = (T, X), \quad \chi_L = (T, T_x),$$

OLS coefficients

$$\hat{\theta}_N = \hat{Y}_N, \quad \hat{\theta}_F = (\hat{Y}_F^T, \hat{\beta}_F^T)^T, \quad \hat{\theta}_L = (\hat{Y}_L^T, \hat{\gamma}_L^T)^T,$$

and residuals $\hat{\epsilon}_* = (\hat{\epsilon}_{*,1}, \dots, \hat{\epsilon}_{*,N})^T$. Denote by

$$\hat{\Sigma}_* = (\chi_*^T \chi_*)^{-1} M_* (\chi_*^T \chi_*)^{-1} \text{ with } M_* = \chi_*^T \text{diag}(\hat{\epsilon}_{*,1}^2, \dots, \hat{\epsilon}_{*,N}^2) \chi_* \quad (* = N, F, L) \quad (S22)$$

the EHW covariance estimator of $\hat{\theta}_*$ from the OLS fit of $Y = \chi_* \hat{\theta}_* + \hat{\epsilon}_*$. The EHW covariance estimator of \hat{Y}_* , namely $\hat{\Psi}_*$, is then the upper-left $Q \times Q$ submatrix of $\hat{\Sigma}_*$.

For $* = N, F, L$, let $\hat{Y}_*(q)$ denote the q th element of \hat{Y}_* , with $\hat{Y}_N(q) = \hat{Y}(q)$, $\hat{Y}_F(q) = \hat{Y}(q) - \hat{x}(q)^T \hat{\beta}_F$, and $\hat{Y}_L(q) = \hat{Y}(q) - \hat{x}(q)^T \hat{\gamma}_{L,q} = \hat{Y}(q) - \hat{x}(q)^T \hat{\gamma}_q$ by Proposition S1.

Let

$$\hat{S}_{xY(q)} = N_q^{-1} \sum_{i:Z_i=q} x_i Y_i, \quad \hat{S}_x^2(q) = N_q^{-1} \sum_{i:Z_i=q} x_i x_i^T \quad (q \in \mathcal{T})$$

be the sample means of $\{x_i Y_i(q)\}_{i=1}^N$ and $(x_i x_i^T)_{i=1}^N$, respectively, with $\hat{S}_{xY(q)} = S_{xY(q)} + o_P(1)$ and $\hat{S}_x^2(q) = S_x^2 + o_P(1)$ under complete randomization and Condition 4. Let $\hat{X} = (\hat{x}(1), \dots, \hat{x}(Q))^T$ be the $Q \times J$ matrix with $\hat{x}^T(q)$ as the q th row vector. We have

$$\begin{cases} N^{-1} \chi_N^T \chi_N = N^{-1} T^T T = \Pi, \\ N^{-1} \chi_N^T Y = N^{-1} T^T Y = \Pi \hat{Y}_N = \Pi \bar{Y} + o_P(1); \end{cases} \quad (S23)$$

$$\begin{cases} N^{-1} \chi_F^T \chi_F = \begin{pmatrix} N^{-1} T^T T & N^{-1} T^T X \\ N^{-1} X^T T & N^{-1} X^T X \end{pmatrix} = \begin{pmatrix} \Pi & \Pi \hat{X} \\ \hat{X}^T \Pi & \kappa S_x^2 \end{pmatrix} = \begin{pmatrix} \Pi & \\ & S_x^2 \end{pmatrix} + o_P(1), \\ N^{-1} \chi_F^T Y = \begin{pmatrix} N^{-1} T^T Y \\ N^{-1} X^T Y \end{pmatrix} = \begin{pmatrix} \Pi \hat{Y}_N \\ \sum_{q \in \mathcal{T}} e_q \hat{S}_{xY(q)} \end{pmatrix} = \begin{pmatrix} \Pi \bar{Y} \\ S_x^2 \bar{\gamma} \end{pmatrix} + o_P(1), \end{cases} \quad (S24)$$

where $\kappa = 1 - N^{-1}$; and

$$\begin{cases} N^{-1} \chi_L^T \chi_L = \begin{pmatrix} N^{-1} T^T T & N^{-1} T^T T_x \\ N^{-1} T_x^T T & N^{-1} T_x^T T_x \end{pmatrix} \\ = \begin{pmatrix} \Pi & \Pi \text{diag}\{\hat{x}^T(q)\}_{q \in \mathcal{T}} \\ \text{diag}\{\hat{x}(q)\}_{q \in \mathcal{T}} \Pi & \text{diag}\{e_q \hat{S}_x^2(q)\}_{q \in \mathcal{T}} \end{pmatrix} = \begin{pmatrix} \Pi & \\ & \Pi \otimes S_x^2 \end{pmatrix} + o_P(1), \\ N^{-1} \chi_L^T Y = \begin{pmatrix} N^{-1} T^T Y \\ N^{-1} T_x^T Y \end{pmatrix} = \begin{pmatrix} \Pi \hat{Y}_N \\ (\Pi \otimes I_J) \hat{S}_{xY} \end{pmatrix} = \begin{pmatrix} \Pi \bar{Y} \\ (\Pi \otimes I_J) S_{xY} \end{pmatrix} + o_P(1), \end{cases} \quad (S25)$$

where $\hat{S}_{xY} = (\hat{S}_{xY(1)}^T, \dots, \hat{S}_{xY(Q)}^T)^T$ and $S_{xY} = (S_{xY(1)}^T, \dots, S_{xY(Q)}^T)^T$.

S7.2. Asymptotics of $\hat{\theta}_L$

We show below the asymptotic normality of $\hat{\theta}_L$ under complete randomization. The result underlies the asymptotic normality of \hat{Y}_r in Theorem S3.

Recall that $\hat{Y}\langle\gamma\rangle$ vectorizes $\hat{Y}(q; \gamma_q) = \hat{Y}(q) - \hat{x}(q)^T \gamma_q = N_q^{-1} \sum_{i:Z_i=q} Y_i(q; \gamma_q)$ for $q \in \mathcal{T}$:

$$\hat{Y}(q; \gamma_q) = \begin{pmatrix} N_1^{-1} \sum_{i:Z_i=1} Y_i(1; \gamma_1) \\ \vdots \\ N_Q^{-1} \sum_{i:Z_i=Q} Y_i(Q; \gamma_Q) \end{pmatrix}.$$

Define

$$\hat{\psi} = \begin{pmatrix} N_1^{-1} \sum_{i:Z_i=1} x_i \{Y_i(1; \gamma_1) - \bar{Y}(1)\} \\ \vdots \\ N_Q^{-1} \sum_{i:Z_i=Q} x_i \{Y_i(Q; \gamma_Q) - \bar{Y}(Q)\} \end{pmatrix}.$$

With a slight abuse of notation, define

$$\hat{\theta} = \begin{pmatrix} \hat{Y}\langle\gamma\rangle \\ \hat{\psi} \end{pmatrix} \tag{S26}$$

as the sample analog of

$$\theta = \begin{pmatrix} N^{-1} \sum_{i=1}^N Y_i(1; \gamma_1) \\ \vdots \\ N^{-1} \sum_{i=1}^N Y_i(Q; \gamma_Q) \\ \hline N^{-1} \sum_{i=1}^N x_i \{Y_i(1; \gamma_1) - \bar{Y}(1)\} \\ \vdots \\ N^{-1} \sum_{i=1}^N x_i \{Y_i(1; \gamma_Q) - \bar{Y}(Q)\} \end{pmatrix} = \begin{pmatrix} \bar{Y} \\ 0_{JQ} \end{pmatrix}.$$

For $q, q' \in \mathcal{T}$, let $S_{Y,xY}(q, q'; \gamma)$ be the finite population covariance of $Y_i(q; \gamma_q)$ and $x_i \{Y_i(q'; \gamma_{q'}) - \bar{Y}(q')\}$, summarized in $S_{Y,xY}\langle\gamma\rangle = (S_{Y,xY}(q, q'; \gamma))_{q,q' \in \mathcal{T}}$. Let $S_{xY,xY}(q, q'; \gamma)$ be the finite population covariance of $x_i \{Y_i(q; \gamma_q) - \bar{Y}(q)\}$ and $x_i \{Y_i(q'; \gamma_{q'}) - \bar{Y}(q')\}$, summarized in $S_{xY,xY}\langle\gamma\rangle = (S_{xY,xY}(q, q'; \gamma))_{q,q' \in \mathcal{T}}$.

Lemma S8. Assume complete randomization. Then

$$\Sigma_\theta = N \text{cov}(\hat{\theta}) = \begin{pmatrix} V_\gamma & V_{Y,xY} \\ V_{Y,xY}^\top & V_{xY,xY} \end{pmatrix},$$

where $V_{Y,xY} = \text{diag}\{e_q^{-1}S_{Y,xY}(q, q; \gamma)\}_{q \in \mathcal{T}} - S_{Y,xY}\langle\gamma\rangle$ and $V_{xY,xY} = \text{diag}\{e_q^{-1}S_{xY,xY}(q, q; \gamma)\}_{q \in \mathcal{T}} - S_{xY,xY}\langle\gamma\rangle$. Further assume Condition 4. Then

$$\sqrt{N}(\hat{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, \Sigma_\theta).$$

Proof of Lemma S8. Let

$$\begin{pmatrix} Y_i(q; \gamma_q) \\ x_i\{Y_i(q; \gamma_q) - \bar{Y}(q)\} \end{pmatrix}$$

be a potential outcome vector analogous to the $(Y_i(q), x_i^\top)^T$ in the proof of Lemma S5. The result then follows from Lemma S4 with $\Gamma_q = \text{diag}(a_{\cdot q}, a_{\cdot q} \otimes I_J)$ for $q \in \mathcal{T}$ identical to those defined in the proof of Lemma S5. \square

Lemma S9. Under complete randomization and Condition 4,

$$\sqrt{N}(\hat{\theta}_L - \theta_L) \rightsquigarrow \mathcal{N}(0, \Lambda \Sigma_\theta \Lambda)$$

with $\Lambda = \text{diag}\{I_Q, I_Q \otimes (S_x^2)^{-1}\}$.

Proof of Lemma S9. Recall the definitions of $\hat{\theta}$ and θ from (S26). It follows from $\chi_L = (T, T_x)$ and $Y - \chi_L \theta_L = \{Y_i - \bar{Y}(Z_i) - x_i^\top \gamma_{Z_i}\}_{i=1}^N$ that

$$\begin{aligned} N^{-1} \chi_L^\top (Y - \chi_L \theta_L) &= N^{-1} \begin{pmatrix} T^\top (Y - \chi_L \theta_L) \\ T_x^\top (Y - \chi_L \theta_L) \end{pmatrix} = N^{-1} \begin{pmatrix} \sum_{i:Z_i=1} \{Y_i(1) - \bar{Y}(1) - x_i^\top \gamma_1\} \\ \vdots \\ \sum_{i:Z_i=Q} \{Y_i(Q) - \bar{Y}(Q) - x_i^\top \gamma_Q\} \\ \sum_{i:Z_i=1} x_i \{Y_i(1) - \bar{Y}(1) - x_i^\top \gamma_1\} \\ \vdots \\ \sum_{i:Z_i=Q} x_i \{Y_i(Q) - \bar{Y}(Q) - x_i^\top \gamma_Q\} \end{pmatrix} \\ &= \begin{pmatrix} \Pi & \\ & \Pi \otimes I_J \end{pmatrix} \begin{pmatrix} \hat{Y}\langle\gamma\rangle - \bar{Y} \\ \hat{\psi} \end{pmatrix} = \begin{pmatrix} \Pi & \\ & \Pi \otimes I_J \end{pmatrix} (\hat{\theta} - \theta). \end{aligned}$$

This, together with $(N^{-1}\chi_L^T\chi_L)^{-1} = \text{diag}\{\Pi^{-1}, \Pi^{-1} \otimes (S_x^2)^{-1}\} + o_P(1)$ from (S25), ensures

$$\begin{aligned}\hat{\theta}_L - \theta_L &= (N^{-1}\chi_L^T\chi_L)^{-1}\{N^{-1}\chi_L^T(Y - \chi_L\theta_L)\} \\ &= (N^{-1}\chi_L^T\chi_L)^{-1} \begin{pmatrix} \Pi & \\ & \Pi \otimes I_J \end{pmatrix} (\hat{\theta} - \theta) \\ &\stackrel{*}{\sim} \begin{pmatrix} \Pi^{-1} & \\ & \Pi^{-1} \otimes (S_x^2)^{-1} \end{pmatrix} \begin{pmatrix} \Pi & \\ & \Pi \otimes I_J \end{pmatrix} (\hat{\theta} - \theta) = \Lambda(\hat{\theta} - \theta)\end{aligned}$$

by Slutsky's Theorem. The result then follows from Lemma S8. \square

S7.3. Results on outputs from ordinary least squares

We next verify the results on $(\hat{Y}_*, \hat{\Psi}_*)$ ($*$ = N, F, L).

Proof of Proposition S1. We verify below the results on \hat{Y}_* for $*$ = N, F, L, respectively.

Unadjusted regression ($*$ = N). The result follows from $\hat{\theta}_N = (\chi_N^T\chi_N)^{-1}(\chi_N^TY) = \hat{Y}_N$ by (S23).

Additive regression ($*$ = F). By (S24), the first-order condition of OLS ensures

$$(\chi_F^T\chi_F)\hat{\theta}_F = \chi_F^TY \iff \begin{pmatrix} \Pi & \Pi\hat{X} \\ \hat{X}^T\Pi & \kappa S_x^2 \end{pmatrix} \begin{pmatrix} \hat{Y}_F \\ \hat{\beta}_F \end{pmatrix} = \begin{pmatrix} \Pi\hat{Y}_N \\ \sum_{q \in \mathcal{T}} e_q \hat{S}_{xY(q)} \end{pmatrix}.$$

Compare the first row to see that $\Pi\hat{Y}_F + \Pi\hat{X}\hat{\beta}_F = \Pi\hat{Y}_N$ and hence $\hat{Y}_F = \hat{Y}\langle 1_Q \otimes \hat{\beta}_F \rangle$.

The probability limit of $\hat{\beta}_F$ then follows from $(\hat{Y}_F^T, \hat{\beta}_F^T)^T = (\chi_F^T\chi_F)^{-1}\chi_F^TY$, where $(N^{-1}\chi_F^T\chi_F)^{-1} = \text{diag}\{\Pi^{-1}, (S_x^2)^{-1}\} + o_P(1)$ and $N^{-1}\chi_F^TY = ((\Pi\bar{Y})^T, (S_x^2\bar{\gamma})^T)^T + o_P(1)$ by (S24).

Fully interacted regression ($*$ = L). By (S25), the first-order condition of OLS ensures

$$(\chi_L^T\chi_L)\hat{\theta}_L = \chi_L^TY \iff \begin{pmatrix} \Pi & \Pi \text{diag}\{\hat{x}^T(q)\}_{q \in \mathcal{T}} \\ \text{diag}\{\hat{x}(q)\}_{q \in \mathcal{T}}\Pi & \text{diag}\{e_q \hat{S}_x^2(q)\}_{q \in \mathcal{T}} \end{pmatrix} \begin{pmatrix} \hat{Y}_L \\ \hat{\gamma}_L \end{pmatrix} = \begin{pmatrix} \Pi\hat{Y}_N \\ (\Pi \otimes I_J)\hat{S}_{xY} \end{pmatrix}.$$

Compare the first row to see that $\Pi\hat{Y}_L + \Pi \text{diag}\{\hat{x}^T(q)\}_{q \in \mathcal{T}}\hat{\gamma}_L = \Pi\hat{Y}_N$ and hence $\hat{Y}_L = \hat{Y}\langle \hat{\gamma}_L \rangle$.

The probability limit of $\hat{\gamma}_L$ then follows from $(\hat{Y}_L^T, \hat{\gamma}_L^T)^T = (\chi_L^T\chi_L)^{-1}\chi_L^TY$ and (S25). \square

Proof of Lemma 1. The asymptotic distributions of \hat{Y}_* follow from Proposition S1 and Lemma S1. We verify below the asymptotic conservativeness of $\hat{\Psi}_*$. Given $V_* = \text{diag}(S_{*,qq}/e_q)_{q \in \mathcal{T}} - S_*$, the results are equivalent to

$$N\hat{\Psi}_* - \text{diag}(S_{*,qq}/e_q)_{q \in \mathcal{T}} = o_P(1) \quad (* = N, F, L).$$

We verify below this for $* = N, L$. The proof for $* = F$ is almost identical to that for $* = L$ and thus omitted.

A useful fact is that

$$N^{-1}T^{\top}\{\text{diag}(\hat{\epsilon}_{*,i}^2)_{i=1}^N\}T = N^{-1}\text{diag}\left(\sum_{i:Z_i=q}\hat{\epsilon}_{*,i}^2\right)_{q \in \mathcal{T}} = \text{diag}(e_q S_{*,qq})_{q \in \mathcal{T}} + o_P(1) \quad (\text{S27})$$

for $* = N, F, L$. The second equality in (S27) follows from

$$\begin{aligned}\hat{\epsilon}_{N,i} &= Y_i - \hat{Y}_N(Z_i), \\ \hat{\epsilon}_{F,i} &= Y_i - \hat{Y}_F(Z_i) - x_i^{\top} \hat{\beta}_F, \\ \hat{\epsilon}_{L,i} &= Y_i - \hat{Y}_L(Z_i) - x_i^{\top} \hat{\gamma}_{L,Z_i}\end{aligned}$$

by Proposition S1, such that $N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{*,i}^2 = S_{*,qq} + o_P(1)$ by standard results.

Unadjusted regression ($* = N$). It follows from $\chi_N = T$ and $N^{-1}T^{\top}T = \Pi$ that

$$N\hat{\Psi}_N = N(T^{\top}T)^{-1}T^{\top}\{\text{diag}(\hat{\epsilon}_{N,i}^2)_{i=1}^N\}T(T^{\top}T)^{-1} = \Pi^{-1}[N^{-1}T^{\top}\{\text{diag}(\hat{\epsilon}_{N,i}^2)_{i=1}^N\}T]\Pi^{-1}.$$

The result then follows from (S27).

Fully interacted regression ($* = L$). First,

$$M_L = \begin{pmatrix} T^{\top} \\ T_x^{\top} \end{pmatrix} \{\text{diag}(\hat{\epsilon}_{L,i}^2)_{i=1}^N\}(T, T_x) = \begin{pmatrix} M_1 & M_2 \\ M_2^{\top} & M_3 \end{pmatrix},$$

where

$$M_1 = T^{\top}\{\text{diag}(\hat{\epsilon}_{L,i}^2)_{i=1}^N\}T, \quad M_2 = T^{\top}\{\text{diag}(\hat{\epsilon}_{L,i}^2)_{i=1}^N\}T_x, \quad M_3 = T_x^{\top}\{\text{diag}(\hat{\epsilon}_{L,i}^2)_{i=1}^N\}T_x.$$

Next, it follows from (S27) and analogous algebra that

$$N^{-1}M_1 = \text{diag}(e_q S_{L,qq})_{q \in \mathcal{T}} + o_P(1), \quad N^{-1}M_2 = O_P(1), \quad N^{-1}M_3 = O_P(1). \quad (\text{S28})$$

This, together with $(I_Q, 0_{Q \times JQ})(N^{-1}\chi_L^{\top}\chi_L)^{-1} = (\Pi^{-1}, 0_{Q \times JQ}) + o_P(1)$ by (S25), ensures

$$\begin{aligned}N\hat{\Psi}_L &= (I_Q, 0_{Q \times JQ})(N\hat{\Sigma}_L)(I_Q, 0_{Q \times JQ})^{\top} \\ &= (I_Q, 0_{Q \times JQ})(N^{-1}\chi_L^{\top}\chi_L)^{-1}(N^{-1}M_L)(N^{-1}\chi_L^{\top}\chi_L)^{-1}(I_Q, 0_{Q \times JQ})^{\top} \\ &= \left(\Pi^{-1}, 0_{Q \times JQ}\right) \begin{pmatrix} N^{-1}M_1 & N^{-1}M_2 \\ N^{-1}M_2^{\top} & N^{-1}M_3 \end{pmatrix} \begin{pmatrix} \Pi^{-1} \\ 0_{JQ \times Q} \end{pmatrix} + o_P(1) \\ &= \Pi^{-1}(N^{-1}M_1)\Pi^{-1} + o_P(1)\end{aligned}$$

$$= \text{diag}(S_{L,qq}/e_q)_{q \in \mathcal{T}} + o_P(1) \quad (\text{S29})$$

from (S28). \square

Recall that we propose to use $W = (R\hat{\theta}_L - r)^T(R\hat{\Sigma}_L R^T)^{-1}(R\hat{\theta}_L - r)$ as the test statistic for testing $H_0 : R\theta_L = r$, and compute a one-sided p -value by comparing W to χ_m^2 with $m = \text{rank}(R)$. Theorem S5 below follows from Lemma S9 and ensures that the resulting p -value preserves the nominal type one error rates asymptotically (Ding and Dasgupta 2018).

Theorem S5. Assume complete randomization and Condition 4. Under $H_0 : R\theta_L = r$, the asymptotic distribution of W is stochastically dominated by χ_m^2 .

Proof of Theorem S5. Recall $\hat{\Sigma}_L$ and M_L in (S22). Similar algebra as the proof of Lemma 1 ensures that $\hat{\Sigma}_L$ is asymptotically conservative for estimating the true sampling covariance of $\hat{\theta}_L$:

$$\text{plim } N \left\{ \hat{\Sigma}_L - \text{cov}(\hat{\theta}_L) \right\} = \Lambda \begin{pmatrix} S_L & S_{Y,xY} \langle \gamma \rangle \\ S_{Y,xY} \langle \gamma \rangle^T & S_{xY,xY} \langle \gamma \rangle \end{pmatrix} \Lambda \geq 0.$$

The result then follows from $\sqrt{N}(R\hat{\theta}_L - R\theta_L) \rightsquigarrow \mathcal{N}(0, R\Lambda\Sigma_\theta\Lambda^T R^T)$ by Lemma S9. \square

Remark S2. Let RSS_L and RSS_r denote the residual sum of squares from the OLS fit of (6) and the RLS fit of (6) subject to $R\theta_L = r$, respectively. The classical model-based framework tests $H_0 : R\theta_L = r$ by an F -test with

$$F = \frac{(\text{RSS}_r - \text{RSS}_L)/m}{\text{RSS}_L/(N - Q - JQ)} \quad (\text{S30})$$

as the test statistic, and compares it against the F -distribution with degrees of freedom m and $N - Q - JQ$, denoted by $F_{m,N-Q-JQ}$, to obtain the one-sided p -value.

We do not pursue this route because the validity of the above F -test depends critically on the classical linear model assumptions of uncorrelated and homoskedastic sampling errors. The design-based framework violates both assumptions and renders the sampling distribution of the F -statistic in (S30) no longer $F_{m,N-Q-JQ}$ under H_0 even asymptotically. The F -test as a result does not preserve the correct type one error rates even asymptotically. See Ding and Dasgupta (2018) and Wu and Ding (2021) for related discussions.

S7.4. Results on outputs from restricted least squares

S7.4.1. RLS subject to general restriction

Proof of Theorem S3. We verify below the numeric expression and asymptotic results, respectively.

Numeric expression. The numeric result follows from the method of Lagrange multipliers. In particular, the Lagrangian for the restricted optimization problem equals

$$(Y - \chi_L^\top \theta)^T (Y - \chi_L^\top \theta) - 2\lambda^T (R\theta - r) \quad (\text{S31})$$

and yields the first-order condition as

$$\chi_L^\top \chi_L \hat{\theta}_r = \chi_L^\top Y - R^\top \lambda,$$

where $\lambda = \{R(\chi_L^\top \chi_L)^{-1} R^\top\}^{-1} (R\hat{\theta}_L - r)$. This, together with (S25), ensures

$$\begin{pmatrix} \Pi & \Pi \operatorname{diag}\{\hat{x}^\top(q)\}_{q \in \mathcal{T}} \\ \operatorname{diag}\{\hat{x}(q)\}_{q \in \mathcal{T}} \Pi & \operatorname{diag}\{e_q \hat{S}_x^2(q)\}_{q \in \mathcal{T}} \end{pmatrix} \begin{pmatrix} \hat{Y}_r \\ \hat{\gamma}_r \end{pmatrix} = \begin{pmatrix} \Pi \hat{Y}_N \\ (\Pi \otimes I_J) \hat{S}_{xY} \end{pmatrix} - N^{-1} \begin{pmatrix} R_Y^\top \lambda \\ R_\gamma^\top \lambda \end{pmatrix}.$$

Extract the first row of both sides to see

$$\hat{Y}_r + \operatorname{diag}\{\hat{x}^\top(q)\}_{q \in \mathcal{T}} \hat{\gamma}_r = \hat{Y}_N - \Pi^{-1} R_Y^\top (N^{-1} \lambda).$$

The result then follows from $\hat{Y}_N - \operatorname{diag}\{\hat{x}^\top(q)\}_{q \in \mathcal{T}} \hat{\gamma}_r = \hat{Y} \langle \hat{\gamma}_r \rangle$ such that

$$\hat{Y}_r = \hat{Y} \langle \hat{\gamma}_r \rangle - \Pi^{-1} R_Y^\top (N^{-1} \lambda). \quad (\text{S32})$$

Asymptotic results. By (S25), we have $N(\chi_L^\top \chi_L)^{-1} = \operatorname{diag}\{\Pi^{-1}, (\Pi \otimes S_x^2)^{-1}\} + o_P(1)$ and hence

$$N\{R(\chi_L^\top \chi_L)^{-1} R^\top\}^{-1} = \Delta_0 + o_P(1).$$

This ensures

$$M_{r,\infty} = \operatorname{plim} M_r = \begin{pmatrix} \Pi^{-1} & \\ & (\Pi \otimes S_x^2)^{-1} \end{pmatrix} R^\top \Delta_0 = \begin{pmatrix} \Pi^{-1} R_Y^\top \Delta_0 \\ (\Pi \otimes S_x^2)^{-1} R_\gamma^\top \Delta_0 \end{pmatrix} \quad (\text{S33})$$

for arbitrary R . The probability limits of \hat{Y}_r and $\hat{\gamma}_r$ then follow from

$$\hat{\theta}_r - \theta_L = -M_{r,\infty}(R\theta_L - r) + o_P(1)$$

by Lemma 2 and the fact that $\hat{\theta}_L = \theta_L + o_P(1)$ by Proposition S1.

The asymptotic normality of $\hat{Y}_r - \xi_r$ follows from extracting the first Q rows of

$$\hat{\theta}_r - \theta_L + M_r(R\theta_L - r) = (I - M_r R)(\hat{\theta}_L - \theta_L) \approx (I - M_{r,\infty} R)(\hat{\theta}_L - \theta_L). \quad (\text{S34})$$

The first equality in (S34) follows from Lemma 2. The asymptotic equivalence \approx follows from

Slutsky's theorem and the asymptotic normality of $\sqrt{N}(\hat{\theta}_L - \theta_L)$ by Lemma S9. \square

S7.4.2. RLS subject to correlation-only restriction

Proof of Proposition S2 and Theorem 1. Observe that $R\theta_L - r = \rho_\gamma\gamma - r_\gamma$ under the correlation-only restriction. Proposition S2 follows from Theorem S3, which ensures $\hat{Y}_r = \hat{Y}\langle\hat{\gamma}_r\rangle$ when $R_Y = 0$, with $\text{plim } \hat{\gamma}_r = \gamma$ if (14) is correctly specified.

Theorem 1 then follows from Lemma S1, which ensures $\sqrt{N}(\hat{Y}\langle\hat{\gamma}_r\rangle - \bar{Y}) \rightsquigarrow \mathcal{N}(0, V_r)$, with $V_r = V_L$ if $\text{plim } \hat{\gamma}_r = \gamma$. \square

S7.4.3. RLS subject to separable restriction with $\rho_Y \neq 0$

Recall that

$$U = I_Q - \Pi^{-1}\rho_Y^T(\rho_Y\Pi^{-1}\rho_Y^T)^{-1}\rho_Y$$

with $\Pi = \text{diag}(e_q)_{q \in \mathcal{T}}$ for $\rho_Y \neq 0$.

Proof of Proposition S3 and Theorem 2. Recall that $\chi_L = (T, T_x)$, with $T = (t_1, \dots, t_N)^T$ and $T_x = (t_1 \otimes x_1, \dots, t_N \otimes x_N)^T$. Denote by $\theta = (\theta_Y^T, \theta_\gamma^T)^T$ the dummy for the model coefficients, with θ_Y and θ_γ corresponding to T and T_x , respectively: $Y = T\theta_Y + T_x\theta_\gamma + \epsilon$. Under the separable restriction with $\rho_Y \neq 0$ and $\rho_\gamma \neq 0$, $R\theta - r$ simplifies to the vectorization of $\rho_Y\theta_Y - r_Y$ and $\rho_\gamma\theta_\gamma - r_\gamma$, and we can write $\lambda = (\lambda_Y^T, \lambda_\gamma^T)^T$, with λ_Y and λ_γ corresponding to $\rho_Y\theta_Y - r_Y$ and $\rho_\gamma\theta_\gamma - r_\gamma$, respectively. Under the separable restriction with no restriction on θ_γ , $R\theta - r$ simplifies to $\rho_Y\theta_Y - r_Y$, and we write can $\lambda = \lambda_Y$. The Lagrangian in (S31) simplifies to

$$(Y - T\theta_Y - T_x\theta_\gamma)^T(Y - T\theta_Y - T_x\theta_\gamma) - 2\lambda_Y^T(\rho_Y\theta_Y - r_Y) - 2\lambda_\gamma^T(\rho_\gamma\theta_\gamma - r_\gamma)$$

or $(Y - T\theta_Y - T_x\theta_\gamma)^T(Y - T\theta_Y - T_x\theta_\gamma) - 2\lambda_Y^T(\rho_Y\theta_Y - r_Y)$

depending on if there is restriction on θ_γ or not. The first-order condition with regard to θ_Y equals

$$\begin{aligned} & T^T(Y - T\hat{Y}_r - T_x\hat{\gamma}_r) - \rho_Y^T\lambda_Y = 0 \\ \iff & \hat{Y}_r = (T^T T)^{-1} T^T Y - (T^T T)^{-1} T^T T_x \hat{\gamma}_r - (T^T T)^{-1} \rho_Y^T \lambda_Y. \end{aligned}$$

This, together with $(T^T T)^{-1} T^T Y = \hat{Y}_N$, $(T^T T)^{-1} T^T T_x = \text{diag}\{\hat{x}^T(q)\}_{q \in \mathcal{T}}$, and $(T^T T)^{-1} = N^{-1}\Pi^{-1}$ by (S25), ensures

$$\hat{Y}_r = \hat{Y}_N - [\text{diag}\{\hat{x}^T(q)\}_{q \in \mathcal{T}}] \hat{\gamma}_r - N^{-1}\Pi^{-1}\rho_Y^T\lambda_Y = \hat{Y}\langle\hat{\gamma}_r\rangle - N^{-1}\Pi^{-1}\rho_Y^T\lambda_Y. \quad (\text{S35})$$

The restriction $\rho_Y\hat{Y}_r = r_Y$ further suggests $\lambda_Y = N(\rho_Y\Pi^{-1}\rho_Y^T)^{-1}\{\rho_Y\hat{Y}\langle\hat{\gamma}_r\rangle - r_Y\}$. Plugging this in (S35) verifies

$$\hat{Y}_r = \hat{Y}\langle\hat{\gamma}_r\rangle - \Pi^{-1}\rho_Y^T(\rho_Y\Pi^{-1}\rho_Y^T)^{-1}\{\rho_Y\hat{Y}\langle\hat{\gamma}_r\rangle - r_Y\} = U\hat{Y}\langle\hat{\gamma}_r\rangle + \Pi^{-1}\rho_Y^T(\rho_Y\Pi^{-1}\rho_Y^T)^{-1}r_Y \quad (\text{S36})$$

with $\hat{Y}_r - \bar{Y} = U(\hat{Y}\langle\hat{\gamma}_r\rangle - \bar{Y}) + \mu_r$.

The probability limit of $\hat{\gamma}_r$ follows from $\text{plim } \hat{\gamma}_r = \gamma - (\Pi \otimes S_x^2)^{-1} R_\gamma^T \Delta_0 (R\theta_L - r)$ by Theorem S3. Specifically, we have (i) $R_\gamma = 0$ if $R = (\rho_Y, 0)$ with no restriction on γ , and (ii) $R_\gamma = (0, \rho_\gamma^T)^T$ and thus

$$R_\gamma^T \Delta_0 (R\theta_L - r) = \rho_\gamma^T \{ \rho_\gamma (\Pi \otimes S_x^2)^{-1} \rho_\gamma^T \}^{-1} (\rho_\gamma \gamma - r_\gamma)$$

if $R = \text{diag}(\rho_Y, \rho_\gamma)$ with non-empty restrictions on both \bar{Y} and γ .

The asymptotic normality follows from $\sqrt{N}(\hat{Y}\langle\hat{\gamma}_r\rangle - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, V_r)$ by Lemma S1. \square

We next verify the asymptotic bias-variance trade-off and the design-based Gauss–Markov theorem under constant treatment effects.

Lemma S10. Assume that ρ_Y is a contrast matrix that has full row rank. For $U = I_Q - \Pi^{-1} \rho_Y^T (\rho_Y \Pi^{-1} \rho_Y^T)^{-1} \rho_Y$ and $L = I_Q + A \rho_Y$, where A is an arbitrary matrix, we have

$$U(\Pi^{-1} - 1_{Q \times Q})U^T \leq L(\Pi^{-1} - 1_{Q \times Q})L^T.$$

Proof of Lemma S10. Let $U_Y = I_Q - U = \Pi^{-1} \rho_Y^T (\rho_Y \Pi^{-1} \rho_Y^T)^{-1} \rho_Y$. Then $L - U = U_Y + A \rho_Y = \{\Pi^{-1} \rho_Y^T (\rho_Y \Pi^{-1} \rho_Y^T)^{-1} + A\} \rho_Y$ with

$$\begin{aligned} (L - U)(\Pi^{-1} - 1_{Q \times Q})U^T &= \{\Pi^{-1} \rho_Y^T (\rho_Y \Pi^{-1} \rho_Y^T)^{-1} + A\} \rho_Y (\Pi^{-1} - 1_{Q \times Q})U^T \\ &= 0; \end{aligned}$$

the last equality follows from $\rho_Y \Pi^{-1} U^T = 0$ and hence $\rho_Y (\Pi^{-1} - 1_{Q \times Q})U^T = 0$. This ensures

$$\begin{aligned} L(\Pi^{-1} - 1_{Q \times Q})L^T &= (L - U + U)(\Pi^{-1} - 1_{Q \times Q})(L - U + U)^T \\ &= (L - U)(\Pi^{-1} - 1_{Q \times Q})(L - U)^T + U(\Pi^{-1} - 1_{Q \times Q})(L - U)^T \\ &\quad + (L - U)(\Pi^{-1} - 1_{Q \times Q})U^T + U(\Pi^{-1} - 1_{Q \times Q})U^T \\ &= (L - U)(\Pi^{-1} - 1_{Q \times Q})(L - U)^T + U(\Pi^{-1} - 1_{Q \times Q})U^T \\ &\geq U(\Pi^{-1} - 1_{Q \times Q})U^T. \end{aligned}$$

\square

Proof of Theorems 3 and S2. Assume that the restriction on γ is correctly specified. Then $\text{plim } \hat{\gamma}_r = \gamma$ by Proposition S3 with $\sqrt{N}(\hat{Y}_r - \bar{Y} - \mu_r) \rightsquigarrow \mathcal{N}(0_Q, UV_L U^T)$ by Theorem 2.

Condition 3 further ensures that

$$V_L = N \text{cov}_\infty(\hat{Y}_L) = s_0(\Pi^{-1} - 1_{Q \times Q}), \quad (\text{S37})$$

where s_0 denotes the common value of $S_{L,qq'}$ for all $q, q' \in \mathcal{T}$ under Condition 3. We verify below the asymptotic bias-variance trade-off result in Theorem 3(i) and the design-based Gauss–Markov result in Theorem S2, respectively.

Asymptotic bias-variance trade-off. The result follows from

$$N\text{cov}_\infty(\hat{Y}_r - \mu_r) = UV_L U^T \leq V_L = N\text{cov}_\infty(\hat{Y}_L)$$

by (S37) and Lemma S10.

Gauss–Markov theorem when (13) is correctly specified. For an arbitrary $L\bar{Y}\langle b \rangle + a \in \mathcal{Y}'$ that is consistent for \bar{Y} , the definition of consistency implies that $L\bar{Y} + a = \bar{Y}$ for all \bar{Y} that satisfies $\rho_Y \bar{Y} = r_Y$. Let \bar{Y}_0 be a solution to $\rho_Y \bar{Y} = r_Y$ with $\rho_Y \bar{Y}_0 = r_Y$ and hence $L\bar{Y}_0 + a = \bar{Y}_0$. Then $(L - I_Q)(\bar{Y} - \bar{Y}_0) = 0$ for all \bar{Y} that satisfy $\rho_Y \bar{Y} = r_Y$. This ensures $L - I_Q = A\rho_Y$ for some matrix A .

The result then follows from

$$N\text{cov}_\infty(L\bar{Y}\langle b \rangle + a) = LV_{b,\infty} L^T \geq LV_L L^T, \quad N\text{cov}_\infty(\hat{Y}_r) = UV_L U^T,$$

with $LV_L L^T \geq UV_L U^T$ by (S37) and Lemma S10. □

S7.4.4. Robust covariance estimator

Recall that $\hat{\Sigma}_r = (\chi_L^T \chi_L)^{-1} \{ \chi_L^T \text{diag}(\hat{\epsilon}_{r,1}^2, \dots, \hat{\epsilon}_{r,N}^2) \chi_L \} (\chi_L^T \chi_L)^{-1}$, where $(\hat{\epsilon}_{r,i})_{i=1}^N$ are the residuals from the RLS fit of (6). Let $(\hat{\Sigma}_r)_{[Q]}$ denote the upper-left $Q \times Q$ submatrix of $\hat{\Sigma}_r$. Let $\hat{Y}_r(q)$ denote the q th element of \hat{Y}_r .

Lemma S11. Assume complete randomization, Condition 4, and general R . Then

$$N\hat{\Sigma}_r = \begin{pmatrix} N(\hat{\Sigma}_r)_{[Q]} & O_P(1) \\ O_P(1) & O_P(1) \end{pmatrix}, \quad (S38)$$

where $N(\hat{\Sigma}_r)_{[Q]} = \text{diag}(S_{r,qq}/e_q)_{q \in \mathcal{T}} + \text{diag}[\{\hat{Y}_r(q) - \bar{Y}(q)\}^2/e_q]_{q \in \mathcal{T}} + o_P(1)$.

Proof of Lemma S11. Recall (S22), where M_L is the matrix in the middle for computing the EHW covariance estimator of \hat{Y}_L from the OLS fit. Define $M_{L,r} = \chi_L^T \text{diag}(\hat{\epsilon}_{r,1}^2, \dots, \hat{\epsilon}_{r,N}^2) \chi_L$ as a variant of M_L based on the RLS residuals such that

$$\hat{\Sigma}_r = (\chi_L^T \chi_L)^{-1} M_{L,r} (\chi_L^T \chi_L)^{-1}.$$

The same algebra as in the proof of Lemma 1 ensures

$$M_{L,r} = \begin{pmatrix} M_1 & M_2 \\ M_2^T & M_3 \end{pmatrix}, \quad (S39)$$

where $M_1 = \text{diag}(\sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2)_{q \in \mathcal{T}}$, $M_2 = \text{diag}(\sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 x_i^T)_{q \in \mathcal{T}}$, and $M_3 = \text{diag}(\sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 x_i x_i^T)_{q \in \mathcal{T}}$. Recall that $\gamma_{r,q} = \text{plim } \hat{\gamma}_{r,q}$, with $S_{r,qq} = (N-1)^{-1} \sum_{i=1}^N \{Y_i(q) - x_i^T \gamma_{r,q} - \bar{Y}(q)\}^2$. Then

$$\begin{aligned}\hat{\epsilon}_{r,i} &= Y_i(q) - \hat{Y}_r(q) - x_i^T \hat{\gamma}_{r,q} \\ &= \{Y_i(q) - x_i^T \gamma_{r,q} - \bar{Y}(q)\} - \{\hat{Y}_r(q) - \bar{Y}(q)\} - x_i^T (\hat{\gamma}_{r,q} - \gamma_{r,q})\end{aligned}$$

for i with $Z_i = q$. This ensures

$$\begin{aligned}N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 &= N_q^{-1} \sum_{i:Z_i=q} \{Y_i(q) - x_i^T \gamma_{r,q} - \bar{Y}(q)\}^2 + \{\hat{Y}_r(q) - \bar{Y}(q)\}^2 \\ &\quad + (\hat{\gamma}_{r,q} - \gamma_{r,q})^T \hat{S}_x^2(q) (\hat{\gamma}_{r,q} - \gamma_{r,q}) \\ &\quad - 2\{\hat{Y}_r(q) - \bar{Y}(q)\} \{\hat{Y}_r(q) - \hat{x}^T(q) \gamma_{r,q} - \bar{Y}(q)\} \\ &\quad - 2\{\hat{Y}_r(q) - \bar{Y}(q)\} (\hat{\gamma}_{r,q} - \gamma_{r,q})^T \hat{x}(q) \\ &\quad - 2(\hat{\gamma}_{r,q} - \gamma_{r,q})^T \{\hat{S}_{xY(q)} - \hat{S}_x^2(q) \gamma_{r,q} - \hat{x}(q) \bar{Y}(q)\} \\ &= S_{r,qq} + \{\hat{Y}_r(q) - \bar{Y}(q)\}^2 + o_P(1),\end{aligned}\tag{S40}$$

where the last equality follows from

$$N_q^{-1} \sum_{i:Z_i=q} \{Y_i(q) - x_i^T \gamma_{r,q} - \bar{Y}(q)\}^2 = S_{r,qq} + o_P(1), \quad \hat{\gamma}_{r,q} = \gamma_{r,q} + o_P(1), \quad \hat{Y}_r(q) - \bar{Y}(q) = O_P(1)$$

in addition to $\hat{Y}(q) = \bar{Y}(q) + o_P(1)$, $\hat{x}(q) = o_P(1)$, $\hat{S}_x^2(q) = S_x^2 + o_P(1)$, and $\hat{S}_{xY(q)} = S_{xY(q)} + o_P(1)$. Plugging (S40) in the definition of M_1 ensures

$$\begin{aligned}N^{-1} M_1 &= \text{diag} \left(e_q N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 \right)_{q \in \mathcal{T}} \\ &= \text{diag}(e_q S_{r,qq})_{q \in \mathcal{T}} + \text{diag} \left[e_q \{\hat{Y}_r(q) - \bar{Y}(q)\}^2 \right]_{q \in \mathcal{T}} + o_P(1).\end{aligned}\tag{S41}$$

Similar algebra ensures that $N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 x_i = O_P(1)$ and $N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 x_i x_i^T = O_P(1)$, such that $N^{-1} M_2 = O_P(1)$ and $N^{-1} M_3 = O_P(1)$. It then follows from the same reasoning as in (S29) that

$$\begin{aligned}N(\hat{\Sigma}_r)_{[Q]} &= (I_Q, 0_{Q \times JQ})(N\hat{\Sigma}_r)(I_Q, 0_{Q \times JQ})^T \\ &= \Pi^{-1}(N^{-1} M_1) \Pi^{-1} + o_P(1) \\ &= \text{diag}(S_{r,qq}/e_q)_{q \in \mathcal{T}} + \text{diag} \left[\{\hat{Y}_r(q) - \bar{Y}(q)\}^2 / e_q \right]_{q \in \mathcal{T}} + o_P(1)\end{aligned}$$

by (S41). \square

Proof of Theorem 4. We verify below the result for the correlation-only and separable restrictions, respectively.

Correlation-only restriction. When the restriction satisfies (14), (S33) simplifies to

$$M_{r,\infty} = \begin{pmatrix} 0 \\ (\Pi \otimes S_x^2)^{-1} R_\gamma^T \Delta_0 \end{pmatrix}$$

such that

$$I - M_{r,\infty} R = I - \begin{pmatrix} 0 \\ (\Pi \otimes S_x^2)^{-1} R_\gamma^T \Delta_0 \end{pmatrix} (0, R_\gamma) = \begin{pmatrix} I_Q \\ I_{JQ} - (\Pi \otimes S_x^2)^{-1} R_\gamma^T \Delta_0 R_\gamma \end{pmatrix}.$$

This, together with (S38), ensures

$$\begin{aligned} & N(I - M_r R) \hat{\Sigma}_r (I - M_r R)^T \\ &= \begin{pmatrix} I_Q \\ O(1) \end{pmatrix} \begin{pmatrix} N(\hat{\Sigma}_r)_{[Q]} & O_p(1) \\ O_p(1) & O_p(1) \end{pmatrix} \begin{pmatrix} I_Q \\ O(1) \end{pmatrix} + o_p(1) \\ &= \begin{pmatrix} N(\hat{\Sigma}_r)_{[Q]} & O_p(1) \\ O_p(1) & O_p(1) \end{pmatrix} + o_p(1), \end{aligned}$$

and hence $N\hat{\Psi}_r = N(\hat{\Sigma}_r)_{[Q]} + o_p(1)$.

Theorem 1 further ensures that $\hat{Y}_r(q) = \bar{Y}(q) + o_p(1)$ under the correlation-only restriction. This ensures $N(\hat{\Sigma}_r)_{[Q]} = \text{diag}(S_{r,qq}/e_q)_{q \in \mathcal{T}} + o_p(1) = V_r + S_r + o_p(1)$ by Lemma S11 and the definition of V_r .

Separable restriction with $\rho_Y \neq 0$. Consider first the case with $\rho_Y \neq 0$ and $\rho_\gamma \neq 0$. Then $R = \text{diag}(\rho_Y, \rho_\gamma)$, and it follows from

$$M_{r,\infty} = \begin{pmatrix} \Pi^{-1} \\ (\Pi \otimes S_x^2)^{-1} \end{pmatrix} R^T \Delta_0 = \begin{pmatrix} \Pi^{-1} \rho_Y^T (\rho_Y \Pi^{-1} \rho_Y^T)^{-1} \\ (\Pi \otimes S_x^2)^{-1} \rho_\gamma^T \{ \rho_\gamma (\Pi \otimes S_x^2)^{-1} \rho_\gamma^T \}^{-1} \end{pmatrix}$$

by (S33) that

$$I - M_{r,\infty} R = \text{diag}\{U, O(1)\}.$$

This, together with (S38), ensures that

$$N(I - M_r R) \hat{\Sigma}_r (I - M_r R)^T$$

$$\begin{aligned}
&= \begin{pmatrix} U \\ O(1) \end{pmatrix} \begin{pmatrix} N(\hat{\Sigma}_r)_{[Q]} & O_p(1) \\ O_p(1) & O_p(1) \end{pmatrix} \begin{pmatrix} U^T \\ O(1) \end{pmatrix} + o_p(1) \\
&= \begin{pmatrix} U\{N(\hat{\Sigma}_r)_{[Q]}\}U^T & O_p(1) \\ O_p(1) & O_p(1) \end{pmatrix} + o_p(1),
\end{aligned}$$

and hence $N\hat{\Psi}_r = U\{N(\hat{\Sigma}_r)_{[Q]}\}U^T + o_p(1)$ with

$$N(\hat{\Sigma}_r)_{[Q]} - V_r = S_r + \text{diag}[\{\hat{Y}_r(q) - \bar{Y}(q)\}^2/e_q]_{q \in \mathcal{T}} + o_p(1)$$

by Lemma S11. Theorem 2 further ensures that $\hat{Y}_r(q) = \bar{Y}(q) + \mu_{r,q} + o_p(1)$ under the separable restriction when $\rho_Y \neq 0$. This verifies the result for $R = \text{diag}(\rho_Y, \rho_\gamma)$.

The proof for the case with $\rho_\gamma = 0$ is analogous with $R = (\rho_Y, 0)$, $\Delta_0 = \rho_Y^T(\rho_Y\Pi^{-1}\rho_Y^T)^{-1}$,

$$M_{r,\infty} = \begin{pmatrix} \Pi^{-1} \\ (\Pi \otimes S_x^2)^{-1} \end{pmatrix} R^T \Delta_0 = \begin{pmatrix} \Pi^{-1}\rho_Y^T(\rho_Y\Pi^{-1}\rho_Y^T)^{-1} \\ 0 \end{pmatrix},$$

and $I - M_{r,\infty}R = \text{diag}(U, 0_{JQ \times JQ})$. We omit the details. \square

S8. Proof of the results on factor-based regressions

S8.1. Standard factorial effects

We verify below Proposition S5, Corollaries S1–S2, and Proposition S6 in Section S4. The results of Propositions 1–2, Corollary 2, and Proposition 3 then follow as special cases.

Proof of Proposition S5. Recall that $Z_{ik}^0 = 2^{-1}(Z_{ik} + 1)$ gives the $\{0, 1\}$ -counterpart of Z_{ik} . Then

$$t_i = \otimes_{k=1}^K (1 - Z_{ik}^0, Z_{ik}^0)^T \quad (\text{S42})$$

gives the vector of $\{1(Z_i = q) : q \in \mathcal{T}\}$ from the treatment-based regressions (4)–(6). Let

$$f_i = \otimes_{k=1}^K (1, Z_{ik})^T$$

vectorize $\{1, Z_{i,\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\}$ from the factor-based regressions (17)–(19). It follows from

$$\begin{pmatrix} 1 - Z_{ik}^0 \\ Z_{ik}^0 \end{pmatrix} = 2^{-1} \begin{pmatrix} 1 - Z_{ik} \\ 1 + Z_{ik} \end{pmatrix} = 2^{-1} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ik} \end{pmatrix} \quad (k = 1, \dots, K)$$

that $t_i = \Gamma^T f_i$ for nonsingular

$$\Gamma = 2^{-K} \left\{ \otimes_{k=1}^K \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \right\} = 2^{-K} \left\{ 2^{K-1} \begin{pmatrix} c_{S,\emptyset}^T \\ C_S \end{pmatrix} \right\} = 2^{-1} \begin{pmatrix} c_{S,\emptyset}^T \\ C_S \end{pmatrix},$$

where $c_{S,\emptyset} = 2^{-(K-1)} \mathbf{1}_Q$ and C_S is the contrast matrix corresponding to τ_S . The regressor vectors of the treatment-based and factor-based fully interacted regressions (6) and (19) thus satisfy

$$\begin{pmatrix} t_i \\ t_i \otimes x_i \end{pmatrix} = \begin{pmatrix} \Gamma^T & \\ & \Gamma^T \otimes I_J \end{pmatrix} \begin{pmatrix} f_i \\ f_i \otimes x_i \end{pmatrix}. \quad (\text{S43})$$

Consider the RLS fit of (19) subject to (S8). Let $\tilde{\tau}'_{r,+}$ denote 2 times the coefficient vector of $\{Z_{i,\mathcal{K}} : \mathcal{K} \in \mathcal{F}_+\}$, with $\tilde{\Omega}'_{r,+}$ as the associated double-decker-taco robust covariance estimator by (S4). Lemma S2 ensures that

$$\tilde{\tau}'_{r,+} = C_{S,+} \hat{Y}_{r,S}, \quad \tilde{\Omega}'_{r,+} = C_{S,+} \hat{\Psi}_r C_{S,+}^T$$

by (S43). The result then follows from $\tilde{\tau}_{r,+} = \tilde{\tau}'_{r,+}$ and $\tilde{\Omega}_{r,+} = \tilde{\Omega}'_{r,+}$ by Example S1.

In addition, (S43) ensures that

$$t_i^T \bar{Y} + (t_i \otimes x_i)^T \gamma = f_i^T (\Gamma \bar{Y}) + (f_i \otimes x_i)^T \{(\Gamma \otimes I_J) \gamma\}$$

in (8). This justifies the forms of target parameters in the comments after Proposition S5. □

Proof of Corollary S1. The numeric result follows from $\tilde{\tau}_{r,+} = C_{S,+} \hat{Y}_{r,S} = C_{S,+} \hat{Y} \langle \hat{\gamma}_{r,S} \rangle$ by Proposition S5 and Proposition S2. This ensures the asymptotic distribution by Lemma S1, and the asymptotic conservativeness of $\tilde{\Omega}_{r,+}$ by Proposition S5 and Theorem 4.

Further assume Condition 2. Then $(C'_{S,-} \otimes I_J) \gamma = 0$ is correctly specified as long as $C'_{S,-}$ is a contrast matrix, which is equivalent to $\mathcal{K} = \emptyset \notin \mathcal{F}'_-$ with (S7) including x_i . When this is true, we have $\text{plim } \hat{\gamma}_{r,S} = \gamma$ with $\hat{Y} \langle \hat{\gamma}_{r,S} \rangle \dot{\sim} \hat{Y}_L$ and $\tilde{\tau}_{r,+} \dot{\sim} \tilde{\tau}_{L,+}$ by Proposition S2 and Lemma S1. □

Proof of Corollary S2. The numeric result follows from

$$\tilde{\tau}_{r,+} - \tau_{S,+} - C_{S,+} \mu_{r,S} = C_{S,+} (\hat{Y}_{r,S} - \bar{Y} - \mu_{r,S}) = C_{S,+} U_S \{ \hat{Y} \langle \hat{\gamma}_{r,S} \rangle - \bar{Y} \}$$

by Proposition S5 and Proposition S3. This ensures the asymptotic distribution by Lemma S1, and the asymptotic conservativeness of $\tilde{\Omega}_{r,+}$ by Proposition S5 and Theorem 4.

Further assume Condition 3. The same reasoning as above ensures that $V_{r,S} = V_L$ as long as (S7) includes x_i . When this is true, the result on $\tilde{\tau}_{r,+} \succeq_\infty \tilde{\tau}_{F,+} \dot{\sim} \tilde{\tau}_{L,+} \succeq_\infty \tilde{\tau}_{N,+}$ then follows from

$$N\text{cov}_\infty(\tilde{\tau}_{r,+}) = C_{S,+} U_S V_{r,S} U_S^T C_{S,+}^T, \quad N\text{cov}_\infty(\tilde{\tau}_{*,+}) = C_{S,+} V_* C_{S,+}^T \quad (* = N, F, L)$$

with $U_S V_{r,S} U_S^T = U_S V_L U_S^T \leq V_L = V_F \leq V_N$ by Lemma S10.

□

Proof of Proposition S6. Recall $\hat{Y}_{r,S}$ as the coefficient vector of t_i from the RLS fit of (6) subject to (S8). By (S36), we have

$$\hat{Y}_{r,S} = \hat{Y} \langle \hat{\gamma}_{r,S} \rangle - \Pi^{-1} \rho_Y^T (\rho_Y \Pi^{-1} \rho_Y^T)^{-1} \rho_Y \hat{Y} \langle \hat{\gamma}_{r,S} \rangle,$$

where $\rho_Y = C_{S,-}$. The equal treatment sizes further ensure that $\Pi = Q^{-1} I_Q$ such that $C_{S,+} \Pi^{-1} \rho_Y^T = Q C_{S,+} C_{S,-}^T = 0$. This, together with $\tilde{\tau}_{r,+} = C_{S,+} \hat{Y}_{r,S}$ by Proposition S5, ensures

$$\tilde{\tau}_{r,+} = C_{S,+} \hat{Y} \langle \hat{\gamma}_{r,S} \rangle - C_{S,+} \Pi^{-1} \rho_Y^T (\rho_Y \Pi^{-1} \rho_Y^T)^{-1} \rho_Y \hat{Y} \langle \hat{\gamma}_{r,S} \rangle = C_{S,+} \hat{Y} \langle \hat{\gamma}_{r,S} \rangle.$$

The asymptotic results then follow from Lemma S1. □

S8.2. Factorial effects under $\{0, 1\}$ -coded regressions

First, it follows from

$$\Gamma_0 1_Q = \left\{ \otimes_{k=1}^K \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right\} \left\{ \otimes_{k=1}^K \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} = \otimes_{k=1}^K \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \otimes_{k=1}^K \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0_{Q-1} \end{pmatrix}$$

that C_0 is indeed a contrast matrix.

Proof of Proposition S7. Recall from (S42) that $t_i = \otimes_{k=1}^K (1 - Z_{ik}^0, Z_{ik}^0)^T$. Let $f_i^0 = \otimes_{k=1}^K (1, Z_{ik}^0)^T$ be the analog of $f_i = \otimes_{k=1}^K (1, Z_{ik})^T$, vectorizing $\{1, Z_{i,\mathcal{K}}^0 : \mathcal{K} \in \mathcal{P}_K\}$ from (S11). It follows from

$$\begin{pmatrix} 1 - Z_{ik}^0 \\ Z_{ik}^0 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ Z_{ik}^0 \end{pmatrix} \quad (k = 1, \dots, K)$$

that $t_i = \Gamma_0^T f_i^0$. The regressor vectors of (6) and (S11) thus satisfy

$$\begin{pmatrix} t_i \\ t_i \otimes x_i \end{pmatrix} = \begin{pmatrix} \Gamma_0^T & \\ & \Gamma_0^T \otimes I_J \end{pmatrix} \begin{pmatrix} f_i^0 \\ f_i^0 \otimes x_i \end{pmatrix},$$

analogous to (S43). All results in Proposition S5 and Corollary S5 then follow from the same reasoning as that under the $\{-1, +1\}$ coding system. □

References

- T. Banerjee and R. Mukerjee. Optimal factorial designs for cdna microarray experiments. *The Annals of Applied Statistics*, 2:366–385, 2008.
- 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.
- D. R. Cox. Randomization and concomitant variables in the design of experiments. In G. Kallianpur, P. R. Krishnaiah, and J. K. Ghosh, editors, *Statistics and Probability: Essays in Honor of C. R. Rao*, pages 197–202. North-Holland, Amsterdam, 1982.
- P. Ding and T. Dasgupta. A randomization-based perspective of analysis of variance: a test statistic robust to treatment effect heterogeneity. *Biometrika*, 105:45–56, 2018.
- X. Li and P. Ding. General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112:1759–1169, 2017.
- X. Li and P. Ding. Rerandomization and regression adjustment. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 82:241–268, 2020.
- X. Li, P. Ding, and D. B. Rubin. Asymptotic theory of rerandomization in treatment-control experiments. *Proceedings of the National Academy of Sciences*, 115:9157–9162, 2018.
- X. Li, P. Ding, and D. B. Rubin. Rerandomization in 2^K factorial experiments. *The Annals of Statistics*, 48:43–63, 2020.
- W. Lin. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique. *The Annals of Applied Statistics*, 7:295–318, 2013.
- K. L. Morgan and D. B. Rubin. Rerandomization to improve covariate balance in experiments. *The Annals of Statistics*, 40:1263–1282, 2012.
- R. Mukerjee and B. Tang. Optimal fractions of two-level factorials under a baseline parameterization. *Biometrika*, 99:71–84, 2012.
- N. E. Pashley and M.-A. C. Bind. Causal inference for multiple treatments using fractional factorial designs. *The Canadian Journal of Statistics*, in press, 2022.
- S. Sherman. A theorem on convex sets with applications. *The Annals of Mathematical Statistics*, 26:763–767, 1955.
- C. F. J. Wu and M. Hamada. *Experiments: Planning, Analysis, and Optimization*. New York: John Wiley & Sons, 2009.

- J. Wu and P. Ding. Randomization tests for weak null hypotheses in randomized experiments. *Journal of the American Statistical Association*, 116:1898–1913, 2021.
- A. Zhao and P. Ding. Covariate-adjusted Fisher randomization tests for the average treatment effect. *Journal of Econometrics*, 225:278–294, 2021.
- A. Zhao and P. Ding. Regression-based causal inference with factorial experiments: estimands, model specifications, and design-based properties. *Biometrika*, 109:799–815, 2022.