

Covariate adjustment in multi-armed, possibly factorial experiments

Anqi Zhao and Peng Ding *

Abstract

Randomized experiments are the gold standard for causal inference, and justify simple comparisons across treatment groups. 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 multi-armed experiments. We start with the commonly used additive and fully interacted models for regression adjustment, and clarify the trade-offs between the resulting ordinary least-squares (OLS) estimators for estimating average treatment effects 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 average treatment effects 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-covariates interactions. Third, it can have better finite-sample performance than its unrestricted counterpart even if the restriction is moderately misspecified. It is thus our recommendation for covariate adjustment in multi-armed experiments 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 proposed theory of RLS also provides a powerful tool for studying OLS-based inference from general regression specifications. As an illustration, we demonstrate its unique value for studying OLS-based regression adjustment in factorial experiments via both theory and simulation. Importantly, although we analyze inferential procedures that are motivated by OLS, we do not invoke any assumptions required by the underlying linear models.

Keywords: Design-based inference; potential outcomes; restricted least squares; robust standard error

*Anqi Zhao, Department of Statistics and Data Science, National University of Singapore, 117546, Singapore (E-mail: staza@nus.edu.sg). Peng Ding, Department of Statistics, University of California, Berkeley, CA 94720 (E-mail: pengdingpku@berkeley.edu).

1. Introduction

1.1. Multi-armed experiment and covariate adjustment

Multi-armed 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 and Hamada 2009), and are becoming increasingly popular in social and biomedical sciences (see, e.g., Duflo et al. 2007; Karlan and McConnell 2014; Hainmueller et al. 2014; Egami and Imai 2019; Alsan et al. 2021; Torres et al. 2021; Blackwell and Pashley 2021). Linear models remain the dominant strategy for downstream analysis, delivering not only point estimates but also standard errors via one ordinary least-squares (OLS) fit. The flexibility of model specifications further provides a convenient way to incorporate covariate information for additional efficiency.

The additive and fully interacted regressions are two commonly used strategies for covariate adjustment by OLS. The theoretical superiority of the fully interacted regression under treatment-control experiments is well established in the literature (see, e.g., Freedman 2008a; Lin 2013; Li and Ding 2020). 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 under multi-armed 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. The results are highly 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 covariances for estimating the true sampling covariances under both the additive and fully interacted regressions. The result justifies the Wald-type inference based on OLS from the design-based perspective. This constitutes our first contribution on the design-based justification of regression-based covariate adjustment.

Despite being theoretically superior, however, 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 (Zhao and Ding 2021a). Simulation studies further suggest that the additive regression can have better finite-sample performance than both the unadjusted and fully interacted regressions under such circumstances so long as the treatment effects are not too different across treatment levels. The choice between the additive and fully interacted models is thus a trade-off between finite-sample performance and asymptotic efficiency. Of interest is the availability of alternative

strategies for achieving better middle ground.

To this end, we propose restricted least squares (RLS) as an alternative way to construct point estimates and standard errors from the fully interacted regression, and establish for the first time its properties for inferring average treatment effects 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-covariates interactions. Third, it can have better finite-sample performance than its unrestricted counterpart even if the restriction is moderately misspecified. It is thus our recommendation for covariate adjustment in multi-armed experiments when the OLS estimator from the fully interacted regression risks large finite-sample variability. We also propose 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, possibly unsaturated specifications. This constitutes our second contribution on the design-based theory of RLS for estimating average treatment effects. Importantly, the classical theory for 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 the factorial experiments (Box et al. 2005; Wu and Hamada 2009; Dasgupta et al. 2015; Zhao and Ding 2021a), and illustrate the value of the above theories for studying covariate adjustment under this special type of multi-armed 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 (Wu and Hamada 2009; Lu 2016a,b; Zhao and Ding 2021a), 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 so far focuses on *factor-saturated* regressions that include all possible interactions between the factor indicators (Lu 2016b; Zhao and Ding 2021b). 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, and clarify the design-based properties of the resulting OLS estimators for covariate adjustment under factorial experiments. 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 existing literature in the covariate-free setting (Zhao and Ding 2021a). The resulting theory includes the dominant specifications in practice as special cases, and offers practical guidelines on the causal interpretations of their results. This constitutes our third contribution on the factor-based inference of factorial experiments.

Importantly, although the regression-based covariate adjustment was originally motivated by a linear model, we do not invoke its underlying assumptions, but view the regression as a purely numeric procedure based on OLS or RLS. The sampling properties of the resulting point estimators and standard errors are then evaluated 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., Freedman 2008a,b; Schochet 2010; Lin 2013; Miratrix et al. 2013; Imbens and Rubin 2015; Bloniarz et al. 2016; Schochet 2018; Fogarty 2018; Liu and Yang 2020; Abadie et al. 2020; Guo and Basse 2021).

1.2. Notation and definitions

Let $Y_i \sim u_i$ denote the linear regression of Y_i on u_i free of any modeling assumptions. We focus on the numeric outputs of OLS and RLS, and evaluate their design-based properties.

Let 0_m and $0_{m \times n}$ be the $m \times 1$ vector and $m \times n$ matrix of zeros, respectively. Let 1_m and $1_{m \times n}$ be 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 $1(\cdot)$ be 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.

Following Li et al. (2020), we use the notion of peakedness (Sherman 1955) below to quantify the relative efficiency between different estimators.

Definition 1 (peakedness). For two symmetric random vectors A and B in \mathbb{R}^m , we say A is *more peaked* than B if $\text{pr}(A \in \mathcal{C}) \geq \text{pr}(B \in \mathcal{C})$ for every symmetric convex set $\mathcal{C} \in \mathbb{R}^m$, denoted by $A \succeq B$.

A more peaked random variable has smaller central quantile ranges. For A and B with finite second moments, $A \succeq B$ implies $\text{cov}(A) \leq \text{cov}(B)$. For A and B that are both Normally distributed with zero means, $A \succeq B$ is equivalent to $\text{cov}(A) \leq \text{cov}(B)$. Intuitively, $A \sim B$ if and only if $A \succeq B$ and $A \preceq B$.

For two sequences of random vectors $\{A_N\}_{N=1}^\infty$ and $\{B_N\}_{N=1}^\infty$ with $A_N \rightsquigarrow A$ and $B_N \rightsquigarrow B$ in \mathbb{R}^m , write $A_N \succeq_\infty B_N$ if $A \succeq B$, and write $A_N \dot{\sim} B_N$ if $A \sim B$.

Definition 2 (asymptotic relative efficiency). For two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ that are both consistent for some parameter $\theta \in \mathbb{R}^m$ as the sample size N tends to infinity, we say

- (i) $\hat{\theta}_1$ and $\hat{\theta}_2$ are asymptotically equally efficient if $\sqrt{N}(\hat{\theta}_1 - \theta) \dot{\sim} \sqrt{N}(\hat{\theta}_2 - \theta)$;
- (ii) $\hat{\theta}_1$ is asymptotically more efficient than $\hat{\theta}_2$ if $\sqrt{N}(\hat{\theta}_1 - \theta) \succeq_\infty \sqrt{N}(\hat{\theta}_2 - \theta)$.

For notational simplicity, we will abbreviate $\sqrt{N}(\hat{\theta}_1 - \theta) \dot{\sim} \sqrt{N}(\hat{\theta}_2 - \theta)$ as $\hat{\theta}_1 \dot{\sim} \hat{\theta}_2$, and $\sqrt{N}(\hat{\theta}_1 - \theta) \succeq_\infty \sqrt{N}(\hat{\theta}_2 - \theta)$ as $\hat{\theta}_1 \succeq_\infty \hat{\theta}_2$ or $\hat{\theta}_2 \preceq_\infty \hat{\theta}_1$, respectively, when the meaning of θ is clear from the context.

2. Causal inference with multi-armed 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 q (Neyman 1923). Let $\bar{Y}(q) = N^{-1} \sum_{i=1}^N Y_i(q)$ be the finite-population average, vectorized as $\bar{Y} = (\bar{Y}(1), \dots, \bar{Y}(Q))^T$. 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 $S_{qq'} = (N-1)^{-1} \sum_{i=1}^N \{Y_i(q) - \bar{Y}(q)\} \{Y_i(q') - \bar{Y}(q')\}$. The goal is to estimate the finite-population average treatment effect

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

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

The above setting is general and includes many common designs as special cases. We use the treatment-control experiment, the 2^2 factorial experiment, and the 2^K factorial experiment with general $K \geq 2$ as running examples to illustrate the main ideas. We give below their definitions for concreteness. The theory we are about to develop applies to general experiments with arbitrary structures of treatment levels; see, e.g., Karlan and List (2007); Sinclair et al. (2012); Alsan et al. (2021), and Zhao and Ding (2021a, Section A). We use customized indexes for treatment levels in Examples 1–3 to match the convention in the literature.

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

Example 2. The 2^2 factorial experiment (e.g., (Angrist et al. 2009)) has two factors of two levels, $A, B \in \{-1, +1\}$, and a total of $Q = 2^2 = 4$ treatment levels as their combinations: $q = (-1, -1)$, $(-1, +1)$, $(+1, -1)$, $(+1, +1)$; we abbreviate $(-1, -1)$ as $(--)$, etc. when no confusion would arise. Define $\bar{Y} = (\bar{Y}(--), \bar{Y}(-+), \bar{Y}(+-), \bar{Y}(++))^T$. 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). Intuitively, 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 between A and B compares the average potential outcomes when the two factors take the same and different levels, respectively.

Example 3. The 2^K factorial experiment (e.g., Pedulla 2020; Torres et al. 2021) has K binary factors, $k = 1, \dots, K$, and a total of $Q = 2^K$ treatment levels as their combinations: $q = (z_1, \dots, z_K)$, where $z_k \in \{-1, +1\}$ indicates the level of factor k . The treatment-control and 2^2 factorial experiments 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 and Hamada 2009; Dasgupta et al. 2015).

2.2. Assignment mechanism, estimators, and regression formulation

We focus on complete randomization defined below.

Definition 3 (complete randomization). The experimenter assigns completely at random N_q units to treatment level $q \in \mathcal{T}$, where the N_q 's are fixed integers with $\sum_{q \in \mathcal{T}} N_q = N$ and $e_q = N_q/N$.

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

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

for unit i . Let $\hat{Y}(q) = N_q^{-1} \sum_{i: Z_i = q} Y_i$ be the average observed outcome under treatment 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 . 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 $V_N = N \text{cov}(\hat{Y}_N) = \text{diag}(S_{qq}/e_q)_{q \in \mathcal{T}} - S$. Vectorize the treatment indicators for unit i as

$$t_i = (1(Z_i = 1), \dots, 1(Z_i = Q))^T.$$

Then \hat{Y}_N equals the coefficient vector of t_i from the OLS fit of

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

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

The presence of covariates promises the possibility of additional 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 centered $(x_i)_{i=1}^N$ 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

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

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

give two intuitive ways to adjust for covariates on the basis of (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. Denote by \hat{Y}_F and \hat{Y}_L the coefficient vectors of t_i from the OLS fits of (5) and (6), respectively, as two covariate-adjusted variants of \hat{Y}_N . 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 4 in Section 3 will clarify their respective connections with Neyman (1923), Fisher (1935), and Lin (2013). Replacing \hat{Y}_N with \hat{Y}_* in (3) yields

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

as two convenient 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 formulations in Section 2.3 below.

2.3. Derived linear model and target parameters

The *derived linear model* (Kempthorne 1952; Hinkelmann and Kempthorne 2008) provides the intuition for using the coefficients of t_i from (4)–(6) to estimate \bar{Y} .

To begin with, 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 the fitted model in (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 extension of the classical model 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 again a theoretical OLS fit, and yields the fitted model $Y_i(q) = \bar{Y}(q) + x_i^T \gamma_q + \epsilon_{L,i}(q)$. Plugging the fitted model in (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). With a slight abuse of terms, we call $\bar{Y}(q)$ and γ_q the *target parameters* of $1(Z_i = q)$ and $1(Z_i = q)x_i$ in (6), respectively, concatenated as $\theta_L = (\bar{Y}^T, \gamma^T)^T$.

Intuitively, the unadjusted specification (4) can be viewed as a *restricted variant* of (6), assuming that the coefficients of the $1(Z_i = q)x_i$ ’s are all zero. Likewise for the additive specification (5)

to be viewed as a restricted variant of (6), assuming that the coefficients 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 (4) and (5) relative to (6), but also the sufficient conditions for the asymptotic efficiency of $\hat{\tau}_*$ ($*$ = N, F), as we shall show shortly in Section 3.

Condition 1 (zero correlation). $\gamma_1 = \cdots = \gamma_Q = 0_J$.

Condition 2 (equal correlation). $\gamma_1 = \cdots = \gamma_Q$.

The variation in $\{\gamma_q : q \in \mathcal{T}\}$ measures heterogeneity in treatment effects (Ding et al. 2019). Condition 2 stipulates that the covariates have equal correlation with the potential outcomes across treatment levels, implying homogeneous treatment effects. Condition 1 is stronger than Condition 2, and stipulates uncorrelatedness within all levels. Condition 3 below 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, (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 notion of target parameters. Under the design-based framework, the derived linear model conditions on the potential outcomes, and attributes the randomness in Y_i solely to the randomness in the treatment indicators $1(Z_i = q)$'s. The covariances of $(\epsilon_{*,i})_{i=1}^N$ ($*$ = N, 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 of some hypothetical super-population. The covariance of the error terms is specified by model assumptions, with joint independence and homoskedasticity being two most commonly invoked ones. 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 covariances. We elaborate on the details in Section 3 below.

3. Regression adjustment via 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 design-based inference of τ . The result extends Fisher (1935), Freedman (2008a), and Lin (2013) to multi-armed experiments, and complements Freedman (2008b) on the asymptotics of

the fully interacted regression. Results under alternative super-population frameworks appeared in Tsiatis et al. (2008), Bugni et al. (2018), and Negi and Wooldridge (2021) for treatment-control experiments, and in Bugni et al. (2019) and Ye et al. (2021) for experiments with more than two treatment levels.

Condition 4 states the standard regularity conditions for design-based inference under complete randomization.

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; and (iii) there is a fixed $c < \infty$ independent of N such that $N^{-1} \sum_{i=1}^N Y_i^4(q) \leq c$, $N^{-1} \sum_{i=1}^N \|x_i\|_4^4 \leq c$, and $N^{-1} \sum_{i=1}^N \|x_i Y_i(q)\|_4^4 \leq c$.

Recall that γ_q denotes the coefficient vector of x_i from the OLS fit of $Y_i(q) \sim 1 + x_i$ over $i = 1, \dots, N$. Let $\gamma_F = \sum_{q \in \mathcal{T}} e_q \gamma_q$. 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 $\{Y_i(q) : q \in \mathcal{T}\}_{i=1}^N$, $\{Y_i(q; \gamma_F) : q \in \mathcal{T}\}_{i=1}^N$, and $\{Y_i(q; \gamma_q) : q \in \mathcal{T}\}_{i=1}^N$, respectively, with $S_N = S$. Let

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

Condition 4 ensures that \bar{Y} , τ , S_* , V_* , S_x^2 , γ_q , e_q , and γ_F all have finite limits. 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 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 below. The results clarify the consistency and asymptotic Normality of \hat{Y}_* for estimating \bar{Y} , and ensure 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) $\hat{Y}_L \succeq_\infty \hat{Y}_N$, \hat{Y}_F with $V_L \leq V_N, V_F$;
 - (iii) $\hat{Y}_F \sim \hat{Y}_L \succeq_\infty \hat{Y}_N$ under Condition 2 with $V_F = V_L \leq V_N$;
- $\hat{Y}_N \sim \hat{Y}_F \sim \hat{Y}_L$ under Condition 1 with $V_N = V_F = V_L$.

Lemma 1(i) justifies the Wald-type inference of τ based on $\hat{\tau}_* = C\hat{Y}_*$ and $C\hat{\Psi}_*C^T$ 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 $\hat{\tau}_F$, on the other hand, can be even less efficient than $\hat{\tau}_N$, 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, assume the finite population satisfies the equal correlation condition. Then $\hat{\tau}_F$ attains the same asymptotic

efficiency as $\hat{\tau}_L$ with a more parsimonious regression. Second, assume the finite population satisfies the zero correlation condition. Then $\hat{\tau}_N \sim \hat{\tau}_F \sim \hat{\tau}_L$, 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 4 below the results from Neyman (1923), Fisher (1935), and Lin (2013), and clarify their connections with (4)–(6).

Example 4. Inherit the notation from Example 1 with $\tau = \bar{Y}(1) - \bar{Y}(0)$, $Z_i \in \{0, 1\}$, and $\hat{\tau}_* = (-1, 1)\hat{Y}_*$, where \hat{Y}_* ($*$ = N, F, L) are the coefficient vectors of $t_i = (1(Z_i = 0), 1(Z_i = 1))^T = (1 - Z_i, Z_i)^T$ from the OLS fits of (4)–(6), respectively. Then $\hat{\tau}_N$ 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 . Fisher (1935) suggested to estimate τ by the coefficient of Z_i from the OLS fit of $Y_i \sim 1 + Z_i + x_i$. Freedman (2008a) showed that the resulting estimator does not necessarily improve the asymptotic efficiency over $\hat{\tau}_N$. Lin (2013) proposed an improved estimator as the coefficient of Z_i from the OLS fit of $Y_i \sim 1 + Z_i + x_i + Z_i x_i$, and showed that it is asymptotically at least as efficient as $\hat{\tau}_N$ and Fisher (1935)’s estimator.

The invariance of least squares to non-degenerate linear transformation of the regressor vector ensures that Fisher (1935)’s and Lin (2013)’s estimators equal $\hat{\tau}_F$ and $\hat{\tau}_L$, respectively. This clarifies the connections of (4)–(6) to Neyman (1923), Fisher (1935), and Lin (2013), respectively, and illustrates a way to compute $\hat{\tau}_*$ ($*$ = N, F, L) directly as OLS coefficients under the treatment-control experiment. Their design-based properties follow readily from Lemma 1.

3.2. Asymptotic efficiency of $\hat{\tau}_L$ over a class of linear estimators

We next extend Lemma 1 to a class of linear covariate-adjusted estimators that include \hat{Y}_* ($*$ = N, F, L) as special cases. Let

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

be a covariate-adjusted estimator of \bar{Y} with

$$\hat{Y}(q; b_q) = \hat{Y}(q) - \{\hat{x}(q)\}^T b_q, \quad \hat{x}(q) = N_q^{-1} \sum_{i: Z_i = q} x_i,$$

and $b = (b_1^T, \dots, b_Q^T)^T \in \mathbb{R}^{JQ}$ for some prespecified $b_q \in \mathbb{R}^J$ (Lin 2013; Li and Ding 2020). Intuitively, $\hat{Y}(q; b_q)$ equals the sample mean of the covariate-adjusted potential outcomes $Y_i(q; b_q) = Y_i(q) - x_i^T b_q$. We focus on estimators in

$$\mathcal{Y} = \{\hat{Y}\langle b \rangle : b \in \mathcal{B}\} \tag{9}$$

with

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

Recall that $\gamma = (\gamma_1^T, \dots, \gamma_Q^T)^T$ and $\gamma_F = \sum_{q \in \mathcal{T}} e_q \gamma_q$. Proposition S1 in the Supplementary Material shows that $\hat{Y}_* = \hat{Y}\langle b_* \rangle$ ($*$ = N, F, L) are all special cases of $\hat{Y}\langle b \rangle \in \mathcal{Y}$, with $b_N = 0_{JQ}$, $b_F = 1_Q \otimes \gamma_F + o_P(1)$, and $b_L = \gamma + o_P(1)$, respectively.

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

$$V_{b,\infty} = \text{diag}(S_{b,\infty,qq}/e_q)_{q \in \mathcal{T}} - S_{b,\infty}, \quad (10)$$

where $S_{b,\infty} = (S_{b,\infty,qq'})_{q,q' \in \mathcal{T}}$ denotes the finite-population covariance matrix of $\{Y_i(q; b_{q,\infty}) : q \in \mathcal{T}\}_{i=1}^N$. The V_* ($*$ = N, F, L) are special cases of $V_{b,\infty}$ for $b = b_*$.

Lemma 2. Assume complete randomization and Condition 4. For $\hat{Y}\langle b \rangle \in \mathcal{Y}$ with $\text{plim } b = b_\infty$, we have

$$\sqrt{N}\{\hat{Y}\langle b \rangle - \bar{Y}\} \rightsquigarrow \mathcal{N}(0_Q, V_{b,\infty})$$

with $V_{b,\infty} \geq V_L$ and $\hat{Y}\langle b \rangle \preceq_\infty \hat{Y}_L$. In particular, $\hat{Y}\langle b \rangle \sim \hat{Y}_L$ with $V_{b,\infty} = V_L$ if $b_\infty = \gamma$.

Lemma 2 includes Lemma 1 as a special case, and ensures the asymptotic efficiency of $\hat{\tau}_L$ over all estimators in $\mathcal{E} = \{C\hat{Y}\langle b \rangle : b \in \mathcal{B}\}$. This extends Li and Ding (2020, Example 9) to multi-armed experiments. Asymptotically, \hat{Y}_L is equivalent to the oracle estimator $\hat{Y}\langle \gamma \rangle = (\hat{Y}(1; \gamma_1), \dots, \hat{Y}(Q; \gamma_Q))^T$. The properties of OLS ensure full reduction of the variability due to the covariates, and thereby guarantee the efficiency of $\hat{Y}\langle \gamma \rangle$ over \mathcal{Y} . This provides the intuition for the asymptotic efficiency of $\hat{\tau}_L$. The equivalence of $\hat{\tau}_F$ and $\hat{\tau}_L$ under the equal correlation condition further ensures the asymptotic efficiency of $\hat{\tau}_F$ over \mathcal{E} when $\gamma_1 = \dots = \gamma_Q$.

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

Despite the gains in asymptotic efficiency, the fully interacted regression (6) involves $p = Q + JQ$ estimated coefficients, subjecting $\hat{\tau}_L$ to substantial finite-sample variability when the sample size N is moderate relative to Q and J . Heuristics under the Gauss–Markov model and simulation evidence, on the other hand, suggest 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 over \mathcal{E} 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, may not be a bad idea after all. 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 (6), assuming that the coefficients of $1(Z_i = q)x_i$ ’s are all equal. This motivates a more general approach to regression adjustment via restricted least squares (RLS), which estimates the coefficients

of (6) subject to some prespecified linear restrictions. We establish the design-based properties of the resulting inference in Section 4.

4. Regression adjustment via restricted least squares

4.1. Restricted least squares

Restricted least squares (RLS) 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 (Theil 1971; Rao 1973; Greene and Seaks 1991). The corresponding theory, however, is so far missing under the design-based framework, where the errors are intrinsically dependent and heteroskedastic from the derived linear model perspective (c.f. 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 powerful tool for studying OLS-based inference from general regression specifications. We focus on RLS-based inference for multi-armed experiments in this section, and demonstrate its unique value for studying OLS-based regression adjustment in factorial experiments in Section 5. Importantly, the classical theory for 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 $\chi_{L,i} = (t_i^T, t_i^T \otimes x_i^T)^T$ be the regressor vector of (6). The OLS fit of (6) estimates θ_L by

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

where \hat{Y}_L and $\hat{\beta}_L = (\hat{\beta}_{L,1}^T, \dots, \hat{\beta}_{L,Q}^T)^T$ denote the coefficient vectors of t_i and $t_i \otimes x_i$, respectively, with $\hat{\beta}_{L,q}$ 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 4 (restricted least squares). The RLS fit of (6) yields

$$\hat{\theta}_r = (\hat{Y}_r^T, \hat{\beta}_r^T)^T = \operatorname{argmin}_{\theta \in \mathbb{R}^p} \sum_{i=1}^N (Y_i - \chi_{L,i}^T \theta)^2 \quad \text{subject to } R\theta = r \quad (11)$$

for some prespecified restriction matrix R that has full row rank.

Let \hat{Y}_r and $\hat{\beta}_r = (\hat{\beta}_{r,1}^T, \dots, \hat{\beta}_{r,Q}^T)^T$ denote the RLS coefficient vectors of t_i and $t_i \otimes x_i$ in (11), respectively, with $\hat{\beta}_{r,q}$ corresponding to $1(Z_i = q)x_i$. This is as if we estimate θ_L subject to the

prior belief of

$$R\theta_L = r. \quad (12)$$

Of interest is the utility of the resulting $\hat{\tau}_r = C\hat{Y}_r$ for inferring τ .

We will use (12) to represent the restriction in (11) in terms of the target parameters, and refer to (11) as the RLS fit subject to (12). The *restriction* (12) as such is a purely numeric input for RLS that may or may not match the truth. We say (12) is *correctly specified* if it indeed matches the truth.

4.2. Examples

Examples 5 and 6 below illustrate both \hat{Y}_N and \hat{Y}_F as special cases of \hat{Y}_r under specific restrictions.

Example 5. $\hat{Y}_N = \hat{Y}_r$, where \hat{Y}_r is the coefficient vector of t_i from the RLS fit of (6) subject to the *zero correlation restriction* that $\gamma_q = 0_J$ for all q . A matrix representation of this restriction is $R\theta_L = 0_{JQ}$ with $R = (0_{JQ \times Q}, I_{JQ})$.

Example 6. Let $\hat{\beta}_F$ be the coefficient vector of x_i from the OLS fit of (5). Then $\hat{Y}_F = \hat{Y}_r$ and $\hat{\beta}_F = \hat{\beta}_{r,1} = \dots = \hat{\beta}_{r,Q}$, where \hat{Y}_r and $\hat{\beta}_{r,q}$ are the coefficients from the RLS fit of (6) subject to the *equal correlation restriction* that $\gamma_1 = \dots = \gamma_Q$. A matrix representation of this restriction is $R\theta_L = 0_{J(Q-1)}$ 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 (6), assuming the zero and equal correlation restrictions, respectively. Examples 5 and 6 illustrate the numeric correspondence between the RLS fit and the OLS fit of the corresponding restricted specification. Lemma S3 of the Supplementary Material states a more general result that includes Examples 5 and 6 as special cases. Importantly, the equal correlation restriction differs from the equal correlation condition in that we view the restriction as a purely numeric input for RLS that may or may not match the truth. The equal correlation condition, in contrast, represents our assumptions about the true data-generating process. The equal correlation restriction is as such 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 5 and 6 involve only the γ_q 's. Example 7 below instead imposes restrictions on both \bar{Y} and γ .

Example 7. Recall the setting of the 2^2 factorial experiment from Example 2. Define $\tau_{AB,i} = 2^{-1}\{Y_i(--)+Y_i(++)\} - 2^{-1}\{Y_i(-+)+Y_i(+-)\}$ as the interaction effect for unit i . Suppose we are willing to assume zero individual interaction effects that $\tau_{AB,i} = 0$ for all i . This implies

$$\tau_{AB} = 0, \quad \gamma_{--} - \gamma_{-+} - \gamma_{+-} + \gamma_{++} = 0_J, \quad (13)$$

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

4.3. Design-based theory of RLS

Write (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 (Theil 1971; Rao 1973). The design-based framework violates these assumptions, and leaves the sampling properties of \hat{Y}_r unclear. This is our focus for this subsection. The resulting theory includes \hat{Y}_* ($*$ = N, F) as special cases, and justifies the Wald-type inference of τ based on $\hat{\tau}_r = C\hat{Y}_r$.

To this end, we first review in Lemma 3 below the numeric expression of $\hat{\theta}_r$ free of any modeling assumptions. For simplicity, we assume that $\chi_L^T \chi_L$ is nonsingular; see Greene and Seaks (1991) for more general formulas.

Lemma 3. $\hat{\theta}_r = (I_p - M_r R) \hat{\theta}_L + M_r r$, with $M_r = (\chi_L^T \chi_L)^{-1} R^T \{R(\chi_L^T \chi_L)^{-1} R^T\}^{-1}$.

Recall the definition of \mathcal{B} from (9). Theorem S4 in the Supplementary Material is a direct consequence of Lemma 3, and ensures $\hat{\beta}_r \in \mathcal{B}$ regardless of whether the restriction $R\theta_L = r$ is correctly specified or not. Denote by $\text{plim } \hat{\beta}_r$ the probability limit of $\hat{\beta}_r$, and let V_r be the value of $V_{b,\infty}$ for $b = \hat{\beta}_r$ from (10).

4.3.1. Two types of restrictions

To simplify the presentation, we relegate the most general theory to the Supplementary Material, and 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.

Definition 5. A *separable* restriction restricts \bar{Y} separately from γ , with (12) taking the form of

$$\rho_Y \bar{Y} = r_Y, \quad \rho_\gamma \gamma = r_\gamma \tag{14}$$

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) .

A *correlation-only* restriction concerns only γ , with (14) reduced to

$$\rho_\gamma \gamma = r_\gamma. \tag{15}$$

The correlation-only restriction is a special type of the separable restriction with $\rho_Y = 0_J$. In Definition 5, $R = \text{diag}(\rho_Y, \rho_\gamma)$ 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} ; and $R = (0, \rho_\gamma)$ and $r = r_\gamma$ for a correlation-only restriction with non-empty restriction on γ . An empty restriction is correctly specified by definition.

Examples 5 and 6 are special cases of the correlation-only restriction, whereas Example 7 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 average treatment effects, namely $\rho_Y \bar{Y}$. The $\tau_{AB} = 0$ in Example 7 is an example with $\rho_Y = c_{AB}^T$. More generally, the assumptions of no higher-order interactions for analyzing factorial experiments all fall into this category; see Zhao and Ding (2021a). We give more examples under the 2^K factorial experiment in Section 5.

We focus on not only consistency but also asymptotic Normality and conservative covariance estimation for large-sample Wald-type inference. Sections 4.3.2 and 4.3.3 establish \hat{Y}_r as functions of $\hat{Y}\langle\hat{\beta}_r\rangle$ under both correlation-only and separable types of restrictions, with asymptotic sampling properties following immediately from Lemma 2. Section 4.3.4 then proposes a novel estimator of the true sampling covariance of \hat{Y}_r .

4.3.2. \hat{Y}_r under the correlation-only restriction

Proposition 1. Assume RLS subject to (15). Then $\hat{Y}_r = \hat{Y}\langle\hat{\beta}_r\rangle$, where $\hat{\beta}_r \in \mathcal{B}$ and satisfies $\text{plim } \hat{\beta}_r = \gamma$ if (15) is correctly specified.

Proposition 1 states the numeric equivalence between \hat{Y}_r and $\hat{Y}\langle\hat{\beta}_r\rangle$, and ensures that \hat{Y}_r belongs to $\mathcal{Y} = \{\hat{Y}\langle b \rangle : b \in \mathcal{B}\}$ with $b = \hat{\beta}_r$. Its design-based properties then follow immediately from Lemma 2.

Theorem 1. Assume complete randomization, Condition 4, and RLS subject to (15). Then

- (i) $\sqrt{N}(\hat{Y}_r - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, V_r)$ with $V_r \geq V_L$ and $\hat{Y}_r \preceq_\infty \hat{Y}_L$.
- (ii) Further assume that (15) is correctly specified. Then $\hat{Y}_r \sim \hat{Y}_L$ with $V_r = V_L$.

Recall that \hat{Y}_L ensures the maximum asymptotic efficiency over all estimators in \mathcal{Y} . 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 (15) is correctly specified or not. Theorem 1(ii) ensures that it has the same asymptotic efficiency as \hat{Y}_L over \mathcal{Y} when (15) 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 $\hat{\beta}_r$, namely $\rho_\gamma \hat{\beta}_r = r_\gamma$, reduces its variability relative to $\hat{\beta}_L$ from OLS in finite samples. Such reduction in variability, despite having no effect on the asymptotic efficiency of \hat{Y}_r when (15) is correctly specified, improves its precision in finite samples. This gives the third guarantee of \hat{Y}_r from the correlation-only restriction, with \hat{Y}_* ($*$ = N, F) as two special cases. 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.

4.3.3. \hat{Y}_r under the separable restriction with $\rho_Y \neq 0$

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

$$U = I_Q - \Pi^{-1} \rho_Y^\top (\rho_Y \Pi^{-1} \rho_Y^\top)^{-1} \rho_Y, \quad \mu_r = -\Pi^{-1} \rho_Y^\top (\rho_Y \Pi^{-1} \rho_Y^\top)^{-1} (\rho_Y \bar{Y} - r_Y) \quad (16)$$

for $\rho_Y \neq 0$.

Proposition 2. Assume RLS subject to (14) with $\rho_Y \neq 0$. Then

$$\hat{Y}_r - \bar{Y} = U \{ \hat{Y} \langle \hat{\beta}_r \rangle - \bar{Y} \} + \mu_r \quad (17)$$

with $\hat{\beta}_r \in \mathcal{B}$. In particular,

- (i) $\text{plim } \hat{\beta}_r = \gamma$ if $\rho_Y \gamma = r_Y$ is correctly specified;
- (ii) $\mu_r = 0$, and hence $\hat{Y}_r - \bar{Y} = U \{ \hat{Y} \langle \hat{\beta}_r \rangle - \bar{Y} \}$, if $\rho_Y \bar{Y} = r_Y$ is correctly specified.

Equation (17) is numeric and ensures that \hat{Y}_r is a linear function of $\hat{Y} \langle \hat{\beta}_r \rangle \in \mathcal{Y}$. Its sampling properties then follow immediately from those of $\hat{Y} \langle \hat{\beta}_r \rangle$ in Lemma 2. Specifically, Proposition 2(i) ensures that $\text{plim } \hat{\beta}_r = \gamma$ as long as the restriction on γ is correctly specified, and holds regardless of whether that on \bar{Y} is correct or not. This, together with Lemma 2, ensures that $\hat{Y} \langle \hat{\beta}_r \rangle \dot{\sim} \hat{Y}_L$ as long as $\rho_Y \gamma = r_Y$ is correctly specified. Proposition 2(ii), on the other hand, suggests the consistency of \hat{Y}_r when the restriction on \bar{Y} is correctly specified. The asymptotic bias and variance of \hat{Y}_r hence depend on distinct restrictions on \bar{Y} and γ , respectively. Theorem 2 below formalizes the intuition.

Theorem 2. Assume complete randomization, Condition 4, and RLS subject to (14) with $\rho_Y \neq 0$. Then

- (i) $\hat{Y}_r - \bar{Y} = \mu_r + o_P(1)$, where $\mu_r \neq 0$ in general unless $\rho_Y \bar{Y} = r_Y$ is correctly specified.
- (ii) Further assume that $\rho_Y \bar{Y} = r_Y$ is indeed correctly specified. Then

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

where $V_r \geq V_L$ and satisfies $V_r = V_L$ if $\rho_Y \gamma = r_Y$ is correctly specified.

Theorem 2(i) shows that the \hat{Y}_r from the separable restriction is in general not consistent 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 .

When the restriction on \bar{Y} is non-empty and indeed correctly specified, Theorem 2(ii) ensures that the resulting \hat{Y}_r is both consistent and asymptotically Normal. Moreover, it can secure additional efficiency over \hat{Y}_L under constant treatment effects.

Theorem 3. Assume complete randomization, Conditions 3–4, and RLS subject to (14) 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 $\hat{Y}_r - \bar{Y} = \mu_r + o_P(1)$ with \hat{Y}_r having smaller asymptotic covariance than \hat{Y}_L .
- (ii) Further assume that $\rho_Y \bar{Y} = r_Y$ is also correctly specified. Then \hat{Y}_r is the best linear consistent estimator of the form $L\hat{Y}\langle b \rangle + a$. That is, \hat{Y}_r is asymptotically Normal with mean \bar{Y} and $\hat{Y}_r \succeq_\infty L\hat{Y}\langle b \rangle + a$ for all linear consistent estimator in $\{L\hat{Y}\langle b \rangle + a : b \in \mathcal{B}\}$ that satisfies $L\hat{Y}\langle b \rangle + a = \bar{Y} + o_P(1)$. In particular, $\hat{Y}_r \succeq_\infty \hat{Y}_L$ given $\hat{Y}_L = \hat{Y}\langle b_L \rangle$ is a linear consistent estimator.

Theorem 3 extends the results in Zhao and Ding (2021a, Theorem A5) on unadjusted estimators to the covariate-adjusted settings, and illustrates the asymptotic bias-variance trade-off between \hat{Y}_r and \hat{Y}_L when the restriction on \bar{Y} is non-empty. Specifically, 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 non-diminishing bias. Further assume that the restriction on \bar{Y} is also correctly specified. Theorem 3(ii) states the design-based counterpart of the classical Gauss–Markov theorem for RLS, ensuring the asymptotic efficiency of \hat{Y}_r over all linear consistent estimators of the form $L\hat{Y}\langle b \rangle + a$, including \hat{Y}_* ($*$ = N, F, L).

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 non-diminishing bias when misspecified, yet ensures lower asymptotic variance 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-population performance of \hat{Y}_r . The choice of whether to restrict γ is thus a trade-off between finite-sample performance and asymptotic efficiency.

4.3.4. Robust covariance for RLS

Recall that $\hat{\theta}_r = (\hat{Y}_r^T, \hat{\beta}_r^T)^T = (I_p - M_r R)\hat{\theta}_L + M_r r$ from Lemma 3. Define

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

as a variant of the EHW covariance of $\hat{\theta}_L$, where we replace the OLS residuals with the RLS counterparts $\hat{\epsilon}_{r,i} = Y_i - t_i^T \hat{Y}_r - (t_i \otimes x_i)^T \hat{\beta}_r$ for $i = 1, \dots, N$. We propose to use the upper-left $Q \times Q$ submatrix of $(I_p - M_r R)\hat{\Sigma}_r(I_p - M_r R)^T$ to estimate the sampling covariance of \hat{Y}_r , denoted by $\hat{\Psi}_r$.

Let $\beta_{r,q} = \text{plim } \hat{\beta}_{r,q}$ ($q \in \mathcal{T}$) denote the probability limit of $\hat{\beta}_{r,q}$ under complete randomization and Condition 4. Let $S_r = (S_{r,qq'})_{q,q' \in \mathcal{T}}$ be the finite-population covariance matrix of $\{Y_i(q) - x_i^T \beta_{r,q} : q \in \mathcal{T}\}_{i=1}^N$, with $S_{r,qq'} = (N-1)^{-1} \sum_{i=1}^N \{Y_i(q) - x_i^T \beta_{r,q} - \bar{Y}(q)\} \{Y_i(q') - x_i^T \beta_{r,q'} - \bar{Y}(q')\}$. Then $V_r = \text{diag}(S_{r,qq}/e_q)_{q \in \mathcal{T}} - S_r$ by definition.

Theorem 4. Assume complete randomization and Condition 4.

- (i) Under RLS subject to (15), we have $N\hat{\Psi}_r - V_r = S_r + o_P(1)$ with $S_r \geq 0$.

(ii) Under RLS subject to (14) with $\rho_Y \neq 0$, we have

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

where $\mu_{r,q}$ denotes the q th element of μ_r , and $U\{S_r + \text{diag}(\mu_{r,q}^2/e_q)_{q \in \mathcal{T}}\}U^T \geq 0$. 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 (i) $\sqrt{N}(\hat{Y}_r - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, V_r)$ under RLS subject to (15) from Theorem 1, and (ii) $\sqrt{N}(\hat{Y}_r - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, UV_r U^T)$ under RLS subject to (14) 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 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 (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 $\hat{\Psi}_*$ ($*$ = N, F) from the OLS fits (c.f. Lemma 1). Let $\hat{\Psi}_{r,N}$ and $\hat{\Psi}_{r,F}$ 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 S3 in the Supplementary Material further gives a stronger result on their numeric equivalence, i.e., $\hat{\Psi}_* = \hat{\Psi}_{r,*}$ ($*$ = N, F), free of any distributional assumptions. This illustrates the equivalence between the EHW covariance from the RLS fit and that from the OLS fit of the corresponding restricted specification. See Lemma S3 in the Supplementary Material for a general result.

4.4. A concluding remark and an extension to rerandomization

This concludes our discussion on the various models for regression adjustment in multi-armed experiments. The RLS estimator $\hat{\tau}_r = C\hat{Y}_r$ includes $\hat{\tau}_*$ ($*$ = N, F) as special cases by Examples 5 and 6, and ensures a triple guarantee under complete randomization. First, it is asymptotically efficient when the restriction is correctly specified. Second, it is consistent as long as the restriction on \bar{Y} is correctly specified and separate from that on γ . Third, it can have better finite-sample performance than the unrestricted $\hat{\tau}_L$ as long as 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-based inference in finite samples.

Rerandomization, on the other hand, offers an alternative way to incorporate covariate information in the design stage of experiments (Cox 1982; Morgan and Rubin 2012). It is the experimental-design analogue of rejective sampling (Fuller 2009), and accepts a complete randomization if and only if it satisfies some prespecified covariate balance criterion. The resulting inference based on $\hat{\tau}_r$, as it turns out, inherits all guarantees from inference under complete randomization and, in addition, ensures less loss in asymptotic efficiency when the restriction is misspecified. It is therefore

our recommendation for covariate adjustment in multi-armed experiments when covariate data are available before units are exposed to treatments. We relegate the detailed theory to the Supplementary Material, and illustrate its finite-sample performance by simulation in Section 6. The results extend the existing literature on rerandomization and its unification with regression adjustment to multi-armed experiments (Morgan and Rubin 2012; Branson et al. 2016; Li et al. 2018, 2020; Li and Ding 2020; Zhao and Ding 2021c), and highlight the utility of rerandomization for providing additional protection against model misspecification.

5. Regression adjustment in factorial experiments

5.1. Overview of factorial experiments

Factorial experiments are a special type of multi-armed experiments, featuring treatments as combinations of two or more factors (Box et al. 2005; Wu and Hamada 2009; Dasgupta et al. 2015; Zhao and Ding 2021a; Pashley and Bind 2021). The treatment-based regressions hence provide a principled way of studying factorial experiments, enabling inference of arbitrary average treatment effects of interest by least-squares fits. Despite their generality and nice theoretical guarantees, however, they are not the dominant choice for analyzing 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 2016a,b; Zhao and Ding 2021b). The resulting regressions contain at least $2^K + J$ and $2^K(1 + J)$ estimated coefficients for a K -factor study under the additive and fully interacted specifications, respectively, subjecting subsequent inference to substantial 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 and standard factorial effects for notational simplicity. The results extend to general factorial effects in general factorial experiments with minimal modification (Zhao and Ding 2021a, 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 3. 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 $\mathcal{P}_K = \{\mathcal{K} : \emptyset \neq \mathcal{K} \subseteq [K]\}$ denote the set of the $2^K - 1$ non-empty 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 (Wu and Hamada 2009; Dasgupta et al. 2015). The $(\tau_A, \tau_B, \tau_{AB})$ in Example 2 are special cases with $\mathcal{P}_2 = \{\{A\}, \{B\}, \{A, B\}\}$. Vectorize the $2^K - 1$ standard factorial effects as

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

where $\tau_{\mathcal{K}} = c_{\mathcal{K}}^T \bar{Y}$ denotes the main effect or interaction corresponding to $\mathcal{K} \in \mathcal{P}_K$. Then $C_s = \{c_{\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\}$ is a $(Q - 1) \times Q$ contrast matrix with

$$2^{K-1} c_{\mathcal{K}} \in \{-1, +1\}^Q, \quad C_s 1_Q = 0_{Q-1}, \quad C_s C_s^T = 2^{-(K-2)} I_{Q-1}.$$

The standard factorial effects are hence *orthogonal* in terms of the contrast vectors that define them. Example 8 below 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 8. 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 2 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 . Let $Z_{i, \mathcal{K}} = \prod_{k \in \mathcal{K}} Z_{ik}$ for $\mathcal{K} \in \mathcal{P}_K$, representing the interaction between the factors in \mathcal{K} . The classical experimental design literature takes

$$Y_i \sim 1 + \sum_{k=1}^K Z_{ik} + \sum_{k \neq k'} Z_{ik} Z_{ik'} + \dots + \prod_{k=1}^K Z_{ik} \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i, \mathcal{K}} \quad (18)$$

as the standard specification for factor-based regression analysis, and estimates $\tau_{\mathcal{K}}$ by 2 times the OLS coefficient of $Z_{i, \mathcal{K}}$ (Wu and Hamada 2009; Lu 2016a). 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 (18) the *factor-saturated* unadjusted

regression, which includes all possible interactions between elements of $(Z_{ik})_{k=1}^K$. The presence of covariates further motivates

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

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} + x_i + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} x_i \quad (20)$$

as the additive and fully interacted variants, respectively.

Equations (18)–(20) 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 $\tilde{\tau}_{*,\mathcal{K}}$ ($*$ = N, F, L) be 2 times the coefficients of $Z_{i,\mathcal{K}}$ from the OLS fits of (18)–(20), respectively, vectorized as

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

Let $\tilde{\Omega}_*$ be the EHW covariance 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. Proposition 3 below follows from the invariance of OLS to non-degenerate linear transformation of the regressor vector, and justifies the Wald-type inference 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 and Ding 2021a), whereas that on $\tilde{\tau}_F$ is.

Proposition 3. $\tilde{\tau}_* = C_s \hat{Y}_*$ and $\tilde{\Omega}_* = C_s \hat{V}_* C_s^T$ for $*$ = N, F, L.

Proposition 3 is numeric, and ensures the equivalence between the factor-saturated specifications and their treatment-based counterparts for estimating τ_s . By Lemma 2, $\tilde{\tau}_L$ is asymptotically the most efficient among $\{C_s \hat{Y}(b) : b \in \mathcal{B}\}$, with $\tilde{\tau}_N$ and $\tilde{\tau}_F$ being special cases. In particular, $\tilde{\tau}_L \succeq_\infty \tilde{\tau}_*$ ($*$ = N, F), with $\tilde{\tau}_F \dot{\sim} \tilde{\tau}_L \succeq_\infty \tilde{\tau}_N$ under the equal correlation condition and $\tilde{\tau}_N \dot{\sim} \tilde{\tau}_F \dot{\sim} \tilde{\tau}_L$ under the zero correlation condition.

5.4. Factor-unsaturated regressions

Despite the conceptual straightforwardness and nice theoretical guarantees, the factor-saturated regressions (18)–(20) involve 2^K , $2^K + J$, and $2^K(1 + J)$ estimated coefficients, respectively, subjecting subsequent inference to substantial finite-sample variability even when K is moderate (Zhao and Ding 2021a). Oftentimes it is only the K main effects along with the $K(K - 1)/2$ two-way interactions that are of interest, with the higher-order effects believed to be small. A common practice is then 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 $Y_i \sim 1 + \sum_{k=1}^K Z_{ik} + \sum_{k \neq k'} Z_{ik} Z_{ik'}$.

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

base model	regression equation	OLS coefficients of $Z_{i,\mathcal{K}}$
$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}}$ (factor-saturated)	(18): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}}$	$\tilde{\tau}_{N,\mathcal{K}}$
	(19): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} + x_i$	$\tilde{\tau}_{F,\mathcal{K}}$
	(20): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} + x_i + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}} x_i$	$\tilde{\tau}_{L,\mathcal{K}}$
$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}}$ (factor-unsaturated)	(21): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}}$	$\tilde{\tau}_{N,r,\mathcal{K}}$
	(22): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} + x_i$	$\tilde{\tau}_{F,r,\mathcal{K}}$
	(23): $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} + x_i + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} x_i$	$\tilde{\tau}_{L,r,\mathcal{K}}$

More generally, for \mathcal{F}_+ as an arbitrary subset of \mathcal{P}_K with $\mathcal{F}_- = \mathcal{P}_K \setminus \mathcal{F}_+ \neq \emptyset$, define

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} \quad (21)$$

as a *factor-unsaturated* variant of (18), targeting specifically the effects associated with the factorial combinations in \mathcal{F}_+ . Throughout this section, we use $+$ and $-$ in the subscripts to indicate effects included in and excluded from the factor-unsaturated specifications, respectively. The first- and second-order specifications above are both special cases with $\mathcal{F}_+ = \{\{k\} : k \in [K]\}$ and $\mathcal{F}_+ = \{\{k\}, \{k, k'\} : k \neq k' \in [K]\}$. The additive and fully interacted variants are then

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} + x_i, \quad (22)$$

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} + x_i + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} x_i. \quad (23)$$

Equations (21)–(23) define three factor-unsaturated variants of (18)–(20), respectively, as arguably the most commonly used specifications in practice. The lower panel of Table 1 summarizes them. Of interest is the impact of such simplification on subsequent inference. We answer this question in Section 5.5 below.

5.5. Design-based properties of the factor-unsaturated regressions (21)–(23)

Define

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

as the *effects of interest* and *nuisance effects* corresponding to \mathcal{F}_+ and \mathcal{F}_- , respectively. Let $\tilde{\tau}_{*,r,\mathcal{K}}$ ($*$ = N, F, L) be 2 times the coefficients of $Z_{i,\mathcal{K}}$ from the OLS fits of (21)–(23), respectively,

vectorized as

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

We use the subscript “r” to signify that (21)–(23) are restricted variants of (18)–(20), respectively, assuming that the coefficients of $\{Z_{i,\mathcal{K}} : \mathcal{K} \in \mathcal{F}_-\}$ all equal zero. This slight abuse of notation is justified by the equivalence between restricted specifications and RLS; see Example S1 in the Supplementary Material.

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 (6) subject to

$$N : \quad \tau_{S,-} = 0, \quad \gamma = 0; \quad (24)$$

$$F : \quad \tau_{S,-} = 0, \quad \text{the } \gamma_q \text{'s are all equal}; \quad (25)$$

$$L : \quad \tau_{S,-} = 0, \quad (C_{S,-} \otimes I_J)\gamma = 0, \quad (26)$$

respectively. They are all special cases of \hat{Y}_r with separate restrictions on \bar{Y} and γ . Proposition 4 below states the numeric correspondence between $\tilde{\tau}_{*,r,+}$ and $\hat{Y}_{*,r}$.

Proposition 4. $\tilde{\tau}_{*,r,+} = C_{S,+}\hat{Y}_{*,r}$ for $* = N, F, L$.

The design-based properties of $\tilde{\tau}_{*,r,+}$ then follow from those of $\hat{Y}_{*,r}$ in Theorems 2 and 3. Recall the definition of V_r after Lemma 3. Let $V_{*,r}$ be the value of V_r associated with $\hat{Y}_{*,r}$ for $* = N, F, L$. Let U_S and $\mu_{r,s}$ be the values of U and μ_r at $\rho_Y = C_{S,-}$, respectively. Let

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

be the estimators of $\tau_{s,+}$ from the factor-saturated regressions (18)–(20). Let $\tilde{\Omega}_{*,r,+}$ be the EHW covariance of $\tilde{\tau}_{*,r,+}$ from the corresponding OLS fits of (21)–(23).

Corollary 1. Assume complete randomization, Condition 4, and $\tau_{S,-} = 0$. Then

(i) $\sqrt{N}(\tilde{\tau}_{*,r,+} - \tau_{s,+}) \rightsquigarrow \mathcal{N}(0, C_{S,+}U_S V_{*,r} U_S^T C_{S,+}^T)$ for $* = N, F, L$, with $\tilde{\Omega}_{*,r,+}$ being asymptotically conservative for estimating the true sampling covariance.

(ii) Further assume Condition 3. Then

$$\tilde{\tau}_{F,r,+} \overset{\cdot}{\sim} \tilde{\tau}_{L,r,+} \succeq_{\infty} \tilde{\tau}_{N,r,+}, \quad \tilde{\tau}_{F,r,+} \overset{\cdot}{\sim} \tilde{\tau}_{L,r,+} \succeq_{\infty} \tilde{\tau}_{F,+} \overset{\cdot}{\sim} \tilde{\tau}_{L,+} \succeq_{\infty} \tilde{\tau}_{N,+}.$$

Corollary 1(i) justifies the Wald-type inference based on the factor-unsaturated specifications when the nuisance effects excluded are indeed absent. Further assume Condition 3 with constant treatment effects. Corollary 1(ii) establishes the asymptotic efficiency of $\tilde{\tau}_{F,r,+} \overset{\cdot}{\sim} \tilde{\tau}_{L,r,+}$ among $\{\tilde{\tau}_{*,+}, \tilde{\tau}_{*,r,+} : * = N, F, L\}$ 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. The additional, correct knowledge on the nuisance effects

then secures extra precision on top of that over the otherwise optimal $\tilde{\tau}_{F,+} \sim \tilde{\tau}_{L,+}$. This illustrates the value of factor-unsaturated regressions in combination with covariate adjustment for improving efficiency.

One key limitation of the factor-unsaturated specifications is that the consistency of $\tilde{\tau}_{*,r,+}$ depends critically on the actual absence of the nuisance effects. This can never be verified exactly in reality, and subjects subsequent inference to possibly non-diminishing biases. This suggests the merit of the factor-saturated additive regression (19) as a trade-off 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 (20).

One surprising, and arguably extremely valuable, exception is when the design has equal treatment group sizes. The resulting $\tilde{\tau}_{*,r,+}$ from factor-unsaturated regression is consistent even if $\tau_{s,-} \neq 0$, thanks to the orthogonality of C_s in defining the standard factorial effects. We formalize the intuition in Condition 5 and Proposition 5 below.

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

Let $\hat{\beta}_{*,r}$ be the value of $\hat{\beta}_r$ associated with $\hat{Y}_{*,r}$, with $\hat{\beta}_{N,r} = 0_{JQ}$.

Proposition 5. Assume Condition 5. Then

$$\tilde{\tau}_{*,r,+} = C_{s,+} \hat{Y} \langle \hat{\beta}_{*,r} \rangle \quad (* = N, F, L),$$

with $\tilde{\tau}_{N,r,+} = C_{s,+} \hat{Y}_N = \tilde{\tau}_{N,+}$. Further assume complete randomization and Condition 4. Then

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

The identity $\tilde{\tau}_{*,r,+} = C_{s,+} \hat{Y} \langle \hat{\beta}_{*,r} \rangle$ is numeric, and ensures the consistency of $\tilde{\tau}_{*,r,+}$ regardless of whether $\tau_{s,-} = 0$ or not, as summarized by Proposition 5(i). This additional protection against model misspecification echos the emphasis of the classical experimental design literature on equal treatment group sizes and orthogonality of the factorial effects.

The asymptotic efficiency of $\tilde{\tau}_{*,r,+}$, nevertheless, still requires the associated restriction on γ in (24)–(26) to be correctly specified, and can no longer exceed that of $\tilde{\tau}_{L,+}$, as demonstrated by Proposition 5(ii). In particular, the condition $(C_{s,-} \otimes I_J)\gamma = 0$ implies that the restriction associated with $\hat{\beta}_{L,r}$ in (26) is correctly specified, and ensures that $\tilde{\tau}_{L,r,+}$ attains the same asymptotic efficiency as $\tilde{\tau}_{L,+}$. Condition 2 implies that the restrictions associated with $\hat{\beta}_{F,r}$ and $\hat{\beta}_{L,r}$ in (25) and (26) are both correctly specified, and ensures that $\tilde{\tau}_{*,r,+}$ ($* = F, L$) attain the same asymptotic efficiency as $\tilde{\tau}_{L,+}$. Condition 1 implies that the restrictions associated with $\hat{\beta}_{N,r}$, $\hat{\beta}_{F,r}$, and $\hat{\beta}_{L,r}$ in (24)–(26) are

all correctly specified, and ensures that $\tilde{\tau}_{*,r,+}$ ($*$ = 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 consistency at the cost of possible additional efficiency over $\tilde{\tau}_{L,+}$.

5.6. Extensions

We focused on specifications (18)–(23) because of their conceptual intuitiveness and prevalence in practice. The same results extend to general specifications like $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} + x_i + \sum_{\mathcal{K} \in \mathcal{F}'_+} Z_{i,\mathcal{K}} x_i$ for arbitrary $\mathcal{F}_+, \mathcal{F}'_+ \subseteq \mathcal{P}_K$ 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. Applications in social and biomedical sciences also 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 and Ding 2021a). The invariance of OLS to non-degenerate linear transformation of the regressor vector, 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 Supplementary Material.

6. Simulation

We now illustrate the finite-sample properties of the proposed method via simulation. Consider a 2^2 factorial experiment with $Q = 4$ treatment levels, $q \in \mathcal{T} = \{(--), (-+), (+-), (++)\}$, and a finite 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 2. 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 $Y_i(q) \sim \mathcal{N}(x_i^T \beta_q, 1)$ for $q = (-+), (+-), (++)$, 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 7, and ensures that restriction (13) is correctly specified with $\tau_{AB} = 0$ and $\gamma_{--} - \gamma_{-+} - \gamma_{+-} + \gamma_{++} = 2c_{AB}^T \gamma = 0_J$.

Fix $\{Y_i(q), x_i : q \in \mathcal{T}\}_{i=1}^N$ in the simulation. We consider inference under both complete randomization and rerandomization to illustrate the additional efficiency by covariate adjustment in the design stage. For $(N_{--}, N_{-+}, N_{+-}, N_{++}) = (22, 23, 24, 31)$, we draw a random permutation of N_q q 's to obtain the treatment assignments under complete randomization, and proceed with rerandomization based on contrasts of the covariate means. Specifically, let

$$\begin{aligned}\hat{\delta}_A &= 2^{-1}\{\hat{x}(+-) + \hat{x}(++)\} - 2^{-1}\{\hat{x}(--) + \hat{x}(-+)\}, \\ \hat{\delta}_B &= 2^{-1}\{\hat{x}(-+) + \hat{x}(++)\} - 2^{-1}\{\hat{x}(--) + \hat{x}(+-)\}\end{aligned}$$

be two linear contrasts of the $\hat{x}(q)$'s. They correspond to the c_A and c_B that define the standard main effects τ_A and τ_B , respectively, and provide two intuitive measures of the chance imbalances

Table 2: Six factor-based regressions and their respective restrictions relative to (20). We use “N”, “F”, and “L” to indicate the unadjusted, additive, and fully interacted adjustment schemes, respectively, with the suffix “_us” indicating the factor-unsaturated variants.

	regression equation	restriction relative to (20)	if correct
N	(18): $Y_i \sim 1 + A_i + B_i + A_i B_i$	$\gamma_q = 0_J$ for all $q \in \mathcal{T}$	no
F	(19): $Y_i \sim 1 + A_i + B_i + A_i B_i + x_i$	γ_q 's are all equal	no
L	(20): $Y_i \sim 1 + A_i + B_i + A_i B_i + x_i + A_i x_i + B_i x_i + A_i B_i x_i$		yes
N_us	(21): $Y_i \sim 1 + A_i + B_i$	$\tau_{AB} = 0$ and $\gamma_q = 0_J$ for all $q \in \mathcal{T}$	no
F_us	(22): $Y_i \sim 1 + A_i + B_i + x_i$	$\tau_{AB} = 0$ and γ_q 's are all equal	no
L_us	(23): $Y_i \sim 1 + A_i + B_i + x_i + A_i x_i + B_i x_i$	$\tau_{AB} = 0$ and $c_{AB}^T \gamma = 0_J$	yes

in the realized allocation. Building on Branson et al. (2016) and Li et al. (2020), we accept the initial complete randomization under rerandomization if and only if the Mahalanobis distance of $\hat{\delta} = (\hat{\delta}_A^T, \hat{\delta}_B^T)^T \in \mathbb{R}^{2J}$ is less or equal to the 0.01 quantile of the χ^2 distribution with $2J = 40$ degrees of freedom. The resulting procedure has an acceptance rate of approximately 0.01; see Section S1 in the Supplementary Material and Li et al. (2020).

Consider six models for estimating $\tau_S = (\tau_A, \tau_B, \tau_{AB})^T$ via OLS, summarized in Table 2. They are special cases of regressions (18)–(23) at $K = 2$ and $\mathcal{F}_+ = \{\{A\}, \{B\}\}$, respectively. 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 (23) is equivalent to the RLS fit of (20) subject to (13), which is correctly specified. The OLS fits of (18), (19), (21), and (22), on the other hand, are equivalent to fitting (20) subject to restrictions that are misspecified, summarized in the last column of Table 2.

Figure 1 shows the distributions of the differences between 2 times the OLS coefficients of $(A_i, B_i, A_i B_i)$ and the true values of $(\tau_A, \tau_B, \tau_{AB})$ over 100,000 independent initial complete randomizations. The results under rerandomization are summarized over the subsets of randomizations that satisfy the covariate balance criterion. Figure 1(a) corresponds to potential outcomes generated from $(\beta_{-+}, \beta_{+-}, \beta_{++}) = (1_J, 0_J, -1_J)$, suggesting considerable deviation from the equal correlation condition that justifies the additive regression (19) and its factor-unsaturated variant (22). Figure 1(b) corresponds to potential outcomes generated from $\beta_{-+} = \beta_{+-} = \beta_{++} = 1_J$, suggesting reasonable closeness to the equal correlation condition such that both (19) and (22) are approximately correctly specified.

The message is coherent across different values of β_q 's and in line with the theory. The correctly specified regression (23) (“L_us”) shows the smallest variability in Figure 1(a), with rerandomization bringing little extra improvement compared with complete randomization. The factor-unsaturated additive regression (22) (“F_us”) is only approximately correctly specified in Figure 1(b), but already delivers even better finite-sample performance than the correctly specified (23) thanks to the

more parsimonious model. The fully interacted regression (“L”), despite being correctly specified and asymptotically the most efficient among the three factor-saturated specifications, shows substantial variability in all cases without rerandomization. The four misspecified regressions, namely “N”, “F”, “N_us”, and “F_us”, on the other hand, show much stabler performance even in Figure 1(a), where the equal and zero correlation conditions are considerably violated. This illustrates the robustness of the proposed method to misspecification of the restriction. Rerandomization, on the other hand, brings visible improvements to the misspecified regressions in all cases.

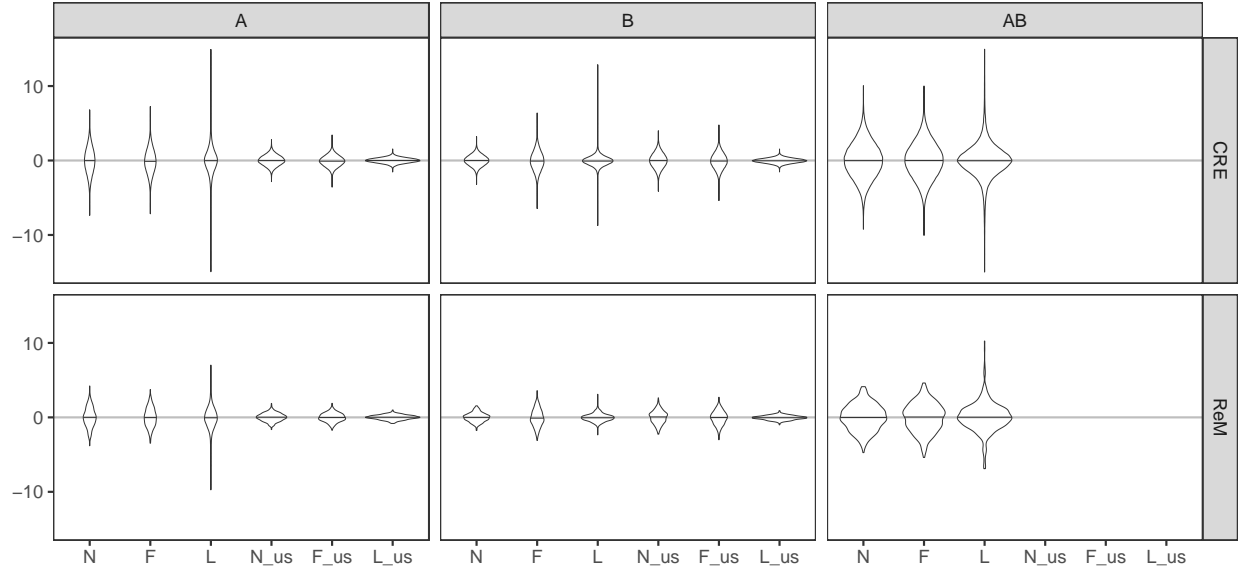
7. Discussion

Based on the asymptotic analysis and simulation, we recommend using restricted least squares on the fully interacted regression for covariate adjustment in multi-armed experiments when the sample size is moderate relative to the number of covariates or treatment levels. Assume 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 finite-population average treatment effects 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 correctly specified 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.

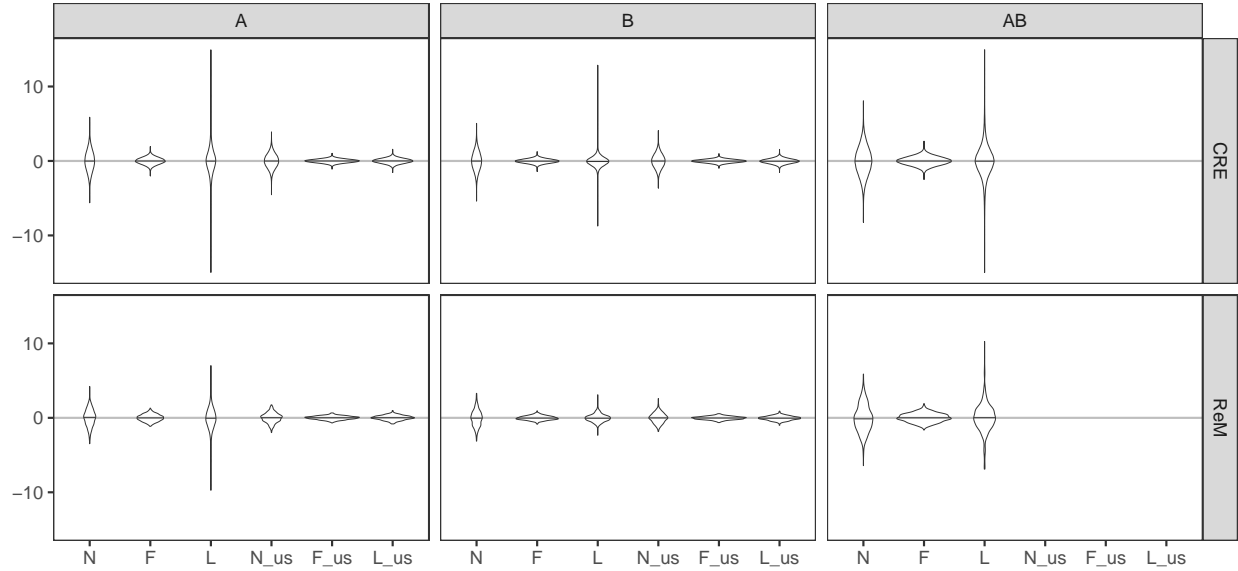
When prior knowledge on the average potential outcomes is less reliable, we recommend imposing restriction on only the correlations. The resulting estimator ensures consistency, yet can be at most as efficient as the OLS counterpart asymptotically. Importantly, all results are design-based, and hold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

References

- A. Abadie, S. Athey, G. W. Imbens, and J. M. Wooldridge. Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88:265–296, 2020.
- M. Alsan, F. C. Stanford, A. Banerjee, E. Breza, A. G. Chandrasekhar, S. Eichmeyer, P. Goldsmith-Pinkham, L. Ogbu-Nwobodo, B. A. Olken, C. Torres, A. Sankar, P.-L. Vautrey, and E. Duflo. Comparison of knowledge and information-seeking behavior after general COVID-19 public health messages and messages tailored for Black and Latinx communities. *Annals of Internal Medicine*, 174:484–492, 2021.
- J. Angrist, D. Lang, and P. Oropoulos. Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1:136–163, 2009.
- M. Blackwell and N. E. Pashley. Noncompliance and instrumental variables for 2^K factorial experiments. *Journal of the American Statistical Association*, in press, 2021.



(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.

Figure 1: Distributions of the differences between 2 times the OLS coefficients of $(A_i, B_i, A_i B_i)$ and the true values of $(\tau_A, \tau_B, \tau_{AB})$ over 100,000 independent initial complete randomizations. CRE and ReM stand for results under complete randomization and rerandomization, respectively.

- A. Bloniarz, H. Z. Liu, C. Zhang, J. Sekhon, and B. Yu. Lasso adjustments of treatment effect estimates in randomized experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 113:7383–7390, 2016.
- G. E. P. Box, J. S. Hunter, and G. W. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery*. New York: Wiley, 2nd edition, 2005.
- Z. Branson, T. Dasgupta, and D. B. Rubin. Improving covariate balance in 2^K factorial designs via rerandomization with an application to a New York City Department of Education High School Study. *The Annals of Applied Statistics*, 10:1958–1976, 2016.
- F. A. Bugni, I. A. Canay, and A. M. Shaikh. Inference under covariate-adaptive randomization. *Journal of the American Statistical Association*, 113:1784–1796, 2018.
- F. A. Bugni, I. A. Canay, and A. M. Shaikh. Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics*, 10:1747–1785, 2019.
- B. Chakraborty, L. M. Collins, V. J. Strecher, and S. A. Murphy. Developing multicomponent interventions using fractional factorial designs. *Statistics in Medicine*, 28:2687–2708, 2009.
- L. M. Collins, J. J. Dziak, and R. Z. Li. Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychological methods*, 14:202–224, 2009.
- 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.
- T. Dasgupta, N. Pillai, and D. B. Rubin. Causal inference from 2^K factorial designs by using potential outcomes. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 77:727–753, 2015.
- P. Ding, A. Feller, and L. Miratrix. Decomposing treatment effect variation. *Journal of the American Statistical Association*, 114:304–317, 2019.
- E. Duflo, R. Glennerster, and M. Kremer. Using randomization in development economics research: A toolkit. In T. P. Schultz and J. A. Strauss, editors, *Handbook of Development Economics*, volume 4, chapter 61, pages 3895–3962. Elsevier, 2007.
- N. Egami and K. Imai. Causal interaction in factorial experiments: Application to conjoint analysis. *Journal of the American Statistical Association*, 114:526–540, 2019.
- R. A. Fisher. *The Design of Experiments*. Edinburgh, London: Oliver and Boyd, 1st edition, 1935.
- C. B. Fogarty. Regression assisted inference for the average treatment effect in paired experiments. *Biometrika*, 105:994–1000, 2018.

- D. A. Freedman. On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40:180–193, 2008a.
- D. A. Freedman. On regression adjustments in experiments with several treatments. *Annals of Applied Statistics*, 2:176–196, 2008b.
- W. A. Fuller. Some design properties of a rejective sampling procedure. *Biometrika*, 96:933–944, 2009.
- W. H. Greene and T. G. Seaks. The restricted least squares estimator: A pedagogical note. *The Review of Economics and Statistics*, 73:563–567, 1991.
- K. Guo and G. Basse. The generalized Oaxaca–Blinder estimator. *Journal of the American Statistical Association*, page in press, 2021.
- J. Hainmueller, D. J. Hopkins, and T. Yamamoto. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22:1–30, 2014.
- K. Hinkelmann and O. Kempthorne. *Design and Analysis of Experiments: Introduction to Experimental Design*, volume 1. John Wiley & Sons, 2nd edition, 2008.
- G. W. Imbens and D. B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press, 2015.
- D. Karlan and J. A. List. Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *American Economic Review*, 97:1774–1793, 2007.
- D. Karlan and M. A. McConnell. Hey look at me: The effect of giving circles on giving. *Journal of Economic Behavior and Organization*, 106:402–412, 2014.
- O. Kempthorne. *The Design and Analysis of Experiments*. New York: John Wiley & Sons, 1952.
- X. Li and P. Ding. General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association*, 112:1759–1169, 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.

- H. Liu and Y. Yang. Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika*, 107:935–948, 2020.
- J. Lu. On randomization-based and regression-based inferences for 2^K factorial designs. *Statistics and Probability Letters*, 112:72–78, 2016a.
- J. Lu. Covariate adjustment in randomization-based causal inference for 2^K factorial designs. *Statistics and Probability Letters*, 119:11–20, 2016b.
- L. Miratrix, J. Sekhon, and B. Yu. Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 75:369–396, 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, T. Dasgupta, and D. B. Rubin. Using standard tools from finite population sampling to improve causal inference for complex experiments. *Journal of the American Statistical Association*, 113:868–881, 2018.
- A. Negi and J. M. Wooldridge. Revisiting regression adjustment in experiments with heterogeneous treatment effects. *Econometric Reviews*, 40:504–534, 2021.
- J. Neyman. On the application of probability theory to agricultural experiments. *Statistical Science*, 5:465–472, 1923.
- N. E. Pashley and M.-A. C. Bind. Causal inference for multiple treatments using fractional factorial designs, 2021.
- D. Pedulla. A field experiment of race, gender, and parental status discrimination in the United States. Technical report, Harvard Dataverse, <https://doi.org/10.7910/DVN/4TTCEY>, 2020.
- C. R. Rao. *Linear Statistical Inference and its Applications*. John Wiley & Sons, 2nd edition, 1973.
- P. Z. Schochet. Is regression adjustment supported by the Neyman model for causal inference? *Journal of Statistical Planning and Inference*, 140:246–59, 2010.
- P. Z. Schochet. Design-based estimators for average treatment effects for multi-armed RCTs. *Journal of Educational and Behavioral Statistics*, 43:568–593, 2018.
- S. Sherman. A theorem on convex sets with applications. *The Annals of Mathematical Statistics*, 26:763–767, 1955.
- B. Sinclair, M. A. McConnell, and D. P. Green. Detecting spillover effects: Design and analysis of multilevel experiments. *American Journal of Political Science*, 56:1055–1069, 2012.
- H. Theil. *Principles of Econometrics*. New York: John Wiley & Sons, 1971.

- C. Torres, L. Ogbu-Nwobodo, M. Alsan, F. C. Stanford, A. Banerjee, E. Breza, A. G. Chandrasekhar, S. Eichmeyer, M. Karnani, T. Loisel, P. Goldsmith-Pinkham, B. A. Olken, P.-L. Vautrey, E. Warner, E. Duflo, and COVID-19 Working Group. Effect of physician-delivered COVID-19 public health messages and messages acknowledging racial inequity on Black and White adults' knowledge, beliefs, and practices related to COVID-19: A randomized clinical trial. *JAMA Network Open*, 4:e2117115–e2117115, 2021.
- A. A. Tsiatis, M. Davidian, M. Zhang, and X. M. Lu. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine*, 27:4658–4677, 2008.
- C. F. J. Wu and M. Hamada. *Experiments: Planning, Analysis, and Optimization*. New York: John Wiley & Sons, 2009.
- T. Ye, J. Shao, Y. Y. Yi, and Q. Y. Zhao. Toward better practice of covariate adjustment in analyzing randomized clinical trials. *Technical report, arXiv*, 2021.
- A. Zhao and P. Ding. Regression-based causal inference with factorial experiments: estimands, model specifications, and design-based properties. *Biometrika*, accepted, 2021a.
- A. Zhao and P. Ding. Reconciling design-based and model-based causal inferences for split-plot experiments. *The Annals of Statistics*, accepted, 2021b.
- A. Zhao and P. Ding. Covariate-adjusted Fisher randomization tests for the average treatment effect. *Journal of Econometrics*, 225:278–294, 2021c.

Supplementary Material to “Covariate adjustment in multi-armed, possibly factorial experiments”

Sections S1–S4 give the additional results that complement the main text.

Section S5 gives the proofs of the results for $\hat{Y}\langle b \rangle$ under complete randomization and rerandomization.

Section S6 gives the proofs of the results on treatment-based regressions.

Section S7 gives the proofs of the results on factor-based regressions.

Notation

Assume centered covariates with $\bar{x} = 0_J$ throughout the Supplementary Material to simplify the presentation. For notational simplicity, we will use \hat{Y} and \hat{Y}_N interchangeably to denote the vector of $(\hat{Y}(1), \dots, \hat{Y}(Q))^T$ when no confusion would arise. Let $S_{xY(q)} = (N-1)^{-1} \sum_{i=1}^N x_i \{Y_i(q) - \bar{Y}(q)\}$ with $\gamma_q = (S_x^2)^{-1} S_{xY(q)}$.

For a sequence of random vectors $(A_N)_{N=1}^\infty$ that converges in distribution to 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 $\{B_N\}_{N=1}^\infty$, we say A_N and B_N are asymptotically equivalent if $\sqrt{N}(A_N - B_N) = o_P(1)$, denoted by $A_N \approx B_N$. For two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ that are both consistent for some parameter $\theta \in \mathbb{R}^m$, $\hat{\theta}_1 \approx \hat{\theta}_2$ implies that $\sqrt{N}(\hat{\theta}_1 - \theta) \sim \sqrt{N}(\hat{\theta}_2 - \theta)$ by Slutsky’s theorem.

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 $V_* = \text{diag}(S_{*,qq}/e_q)_{q \in \mathcal{T}} - S_*$ as special cases for $*$ = N, F, L. For random variables A and B , let

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

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

S1. Rerandomization in the design stage

S1.1. Overview

Rerandomization discards randomizations that do not satisfy a prespecified covariate balance criterion in the design stage of experiments (Cox 1982; Morgan and Rubin 2012), and provides an alternative way to incorporate covariate data for additional efficiency. A special type of rerandomization, ReM, uses the Mahalanobis distance of $\hat{\tau}_x = \hat{x}(1) - \hat{x}(0)$, denoted by $\|\hat{\tau}_x\|_{\mathcal{M}} = \hat{\tau}_x^T \{\text{cov}(\hat{\tau}_x)\}^{-1} \hat{\tau}_x$, as the balance criterion under the treatment-control experiment, and accepts a randomization if and only if $\|\hat{\tau}_x\|_{\mathcal{M}} \leq a$ for some prespecified threshold a (Morgan and Rubin 2012; Li et al. 2018; Li and Ding 2020; Zhao and Ding 2021c). Branson et al. (2016) and Li et al. (2020) extended their

discussion to 2^K factorial experiments, yet did not consider regression adjustment in the analysis stage.

We further the literature by providing 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 $\hat{\tau}_T$. 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 by $\hat{\tau}_T$. The combination of ReM and $\hat{\tau}_T$, in addition to delivering all guarantees as under complete randomization, further reduces the loss in asymptotic efficiency when the restriction is misspecified. It is thus our recommendation for covariate adjustment in multi-armed experiments when $\hat{\tau}_L$ is not practical.

S1.2. ReM under multi-armed experiments

ReM under the treatment-control experiment measures covariate balance by the Mahalanobis distance of $\hat{\tau}_x$. The presence of multiple treatment arms permits more flexible measures of covariate balance. Specifically, denote by $\hat{x} = (\hat{x}^T(1), \dots, \hat{x}^T(Q))^T$ the vectorization of covariate means $\hat{x}(q) = N_q^{-1} \sum_{i: Z_i=q} x_i$ over $q \in \mathcal{T}$. For a prespecified $Q \times 1$ contrast vector $g = (g_1, \dots, g_Q)^T$,

$$\hat{\delta}_g = \sum_{q \in \mathcal{T}} g_q \hat{x}(q) = (g^T \otimes I_J) \hat{x}$$

defines an intuitive measure of covariate balance across the Q treatment arms, and includes $\hat{\tau}_x$ as a special case with $Q = 2$ and $g = (-1, 1)^T$. Assume $H \geq 1$ of such measures are of interest, concatenated as

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

where $G = (g_1, \dots, g_H)^T$ for some prespecified, linearly independent g_h 's. We can conduct ReM based on the Mahalanobis distance of $\hat{\delta}$, with the formal definition given in Definition S1 below.

Definition S1 (ReM). Draw an initial treatment allocation by the complete randomization in Definition 3, 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 reality may choose G that is very different 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 practical relevance of developing the theory of ReM in terms of general G .

Recall the definition of $\hat{Y}\langle b \rangle \in \mathcal{Y}$ from (9), with $\hat{Y}\langle b \rangle \sim \hat{Y}_L$ if $\text{plim } b = \gamma$. Theorem S1 below clarifies 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 the comments after (9),
- (ii) $\hat{Y}_r = \hat{Y}\langle \hat{\beta}_r \rangle \in \mathcal{Y}$ under RLS subject to (15) by Proposition 1, and
- (iii) $\hat{Y}_r - \bar{Y} = U\{\hat{Y}\langle \hat{\beta}_r \rangle - \bar{Y}\}$ with $\hat{Y}\langle \hat{\beta}_r \rangle \in \mathcal{Y}$ under RLS subject to (14) when $\rho_Y \neq 0$ and $\rho_Y \bar{Y} = r_Y$ is correctly specified by Proposition 2.

The result of Theorem S1 hence implies the impact of ReM on $\hat{\tau}_*$ ($*$ = N, F, L) and $\hat{\tau}_r$ as direct consequences.

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 S1. Asymptotically, it is a convolution of two independent Normal and truncated Normal random vectors. The explicit form is not central to the discussion below and thus relegated to Theorem S3 in Section S2.2.

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

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

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 $\text{plim } \hat{\beta}_r = \gamma$ 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 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 necessarily be over-conservative when $\rho_\gamma \gamma = r_\gamma$ is misspecified, highlighting the need of ReM-specific inference for better calibration. We relegate the details to Section S2.3.

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 $E_\infty(\cdot \mid \mathcal{A})$ denote the asymptotic expectation under ReM.

Corollary S1. Assume the ReM in Definition S1 and Condition 4. Then

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

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

The $\hat{Y}\langle b \rangle$ in Corollary S1 includes \hat{Y}_* ($*$ = N, F, L) as special cases, and 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.

S2. Additional results on $\hat{Y}\langle b \rangle \in \mathcal{Y}$

We give in this section some additional results on the covariate-adjusted estimator $\hat{Y}\langle b \rangle \in \mathcal{Y}$ under complete randomization and the ReM in Definition S1, respectively. Assume throughout that $b = (b_1^T, \dots, b_Q^T)^T \in \mathcal{B}$ with $b_q \in \mathbb{R}^J$ and $\text{plim } b = b_\infty = (b_{1,\infty}^T, \dots, b_{Q,\infty}^T)^T$. Recall

$$\hat{Y}\langle \gamma \rangle = (\hat{Y}(1; \gamma_1), \dots, \hat{Y}(Q; \gamma_Q))^T$$

as the oracle estimator defined by $\gamma = (\gamma_1^T, \dots, \gamma_Q^T)^T$. Let $D_b = \text{diag}\{(b_q - \gamma_q)^T\}_{q \in \mathcal{T}}$ with

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

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

S2.1. Sampling properties under complete randomization

Lemma S1. 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 $V_x = N \text{cov}(\hat{x}) = (\Pi^{-1} - 1_{Q \times Q}) \otimes S_x^2$.

Lemma S1 and Lemma 2 together ensure that $\hat{Y}\langle \gamma \rangle \sim \hat{Y}_L \succeq_\infty \hat{Y}\langle b \rangle$ for all $\hat{Y}\langle b \rangle \in \mathcal{Y}$. This justifies $\hat{Y}\langle \gamma \rangle$ as the oracle estimator with fixed adjustment coefficients $b_q = \gamma_q$ ($q \in \mathcal{T}$), extending Li and Ding (2020, Example 9) to multi-armed experiments. Theorem S2 below states a stronger result than Lemma 2, characterizing the asymptotic distance of $\hat{Y}\langle b \rangle \in \mathcal{Y}$ from the oracle.

Theorem S2. Assume complete randomization and Condition 4. For $\hat{Y}\langle b \rangle \in \mathcal{Y}$ with $\text{plim } b = b_\infty$, we have

$$\hat{Y}\langle b \rangle \overset{\cdot}{\approx} \hat{Y}\langle \gamma \rangle - D_{b,\infty} \hat{x} \preceq_\infty \hat{Y}\langle \gamma \rangle$$

with $V_{b,\infty} = N\text{cov}_\infty\{\hat{Y}\langle b \rangle\} = V_L + D_{b,\infty} V_x D_{b,\infty}^\top$. In particular, $\hat{Y}\langle b \rangle \overset{\cdot}{\approx} \hat{Y}\langle \gamma \rangle$ if $D_{b,\infty} V_x D_{b,\infty}^\top = 0$. A sufficient condition for this is $b_\infty = \gamma$.

Theorem S2 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 $\hat{Y}\langle \gamma \rangle$. 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 $\hat{\beta}_{L,q}$, from the fully interacted regression (6), and by the OLS coefficient of x_i , namely $\hat{\beta}_F$, from the additive regression (5). Let $b_F = 1_Q \otimes \hat{\beta}_F$ and $b_L = (\hat{\beta}_{L,1}^\top, \dots, \hat{\beta}_{L,Q}^\top)^\top = \hat{\beta}_L$ be the resulting estimators of the full vector γ . Proposition S1 below establishes the numeric identity between \hat{Y}_* and $\hat{Y}\langle b_* \rangle$ for $* = F, L$.

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 $b_L = \hat{\beta}_L$. Further assume complete randomization and Condition 4. Then $\hat{\beta}_F = \gamma_F + o_P(1)$ and $\hat{\beta}_L = \gamma + o_P(1)$ with $\text{plim } b_F = 1_Q \otimes \gamma_F$ and $\text{plim } b_L = \gamma$.

As a result, b_L is consistent for γ , whereas b_F in general is not. By Theorem S2, the asymptotic efficiency of \hat{Y}_L over \mathcal{Y} in Lemma 2 is essentially the consequence of its asymptotic equivalence with $\hat{Y}\langle \gamma \rangle$ given $b_L = \gamma + o_P(1)$. The equal correlation condition, on the other hand, ensures $b_F = \gamma + o_P(1)$ such that $\hat{Y}_F \overset{\cdot}{\approx} \hat{Y}_L \overset{\cdot}{\approx} \hat{Y}\langle \gamma \rangle$. This gives the intuition behind Lemma 1. The asymptotic equivalence between $\hat{\beta}_{L,q}$ and γ_q is no coincidence but the consequence of the fully interacted regression being numerically equivalent to fitting Q treatment-specific regressions: $Y_i \sim 1 + x_i$ over $\{i : Z_i = q\}$. This ensures that $\hat{\beta}_{L,q}$ equals the coefficient vector of x_i in the q th regression, and converges in probability to γ_q as its population analog.

S2.2. Sampling properties under ReM

Let

$$V_{b,\infty}^\parallel = D_{b,\infty}(\Phi \otimes S_x^2)D_{b,\infty}^\top, \quad V_{b,\infty}^\perp = V_{b,\infty} - V_{b,\infty}^\parallel \quad (\text{S1})$$

with $\Phi = \Pi^{-1}G^\top(G\Pi^{-1}G^\top)^{-1}G\Pi^{-1}$. Following Li et al. (2020), let $(V_{b,\infty}^\parallel)^{1/2}_{JH}$ denote a $Q \times JH$ matrix that satisfies $(V_{b,\infty}^\parallel)^{1/2}_{JH}\{(V_{b,\infty}^\parallel)^{1/2}_{JH}\}^\top = V_{b,\infty}^\parallel$. 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 $\nu_{JH,a}I_{JH} < I_{JH}$. Theorem S3 below states the asymptotic distribution of $\hat{Y}\langle b \rangle$ under ReM.

Theorem S3. Assume the ReM in Definition S1 and Condition 4. For $\hat{Y}\langle b \rangle \in \mathcal{Y}$ with $\text{plim } b = b_\infty$, we have

$$\{\sqrt{N}(\hat{Y}\langle b \rangle - \bar{Y}) \mid \mathcal{A}\} \rightsquigarrow (V_{b,\infty}^\perp)^{1/2} \cdot \epsilon + (V_{b,\infty}^\parallel)_{JH}^{1/2} \cdot \mathcal{L} \quad (\text{S2})$$

with

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

In particular,

$$\{\sqrt{N}(\hat{Y}\langle b \rangle - \bar{Y}) \mid \mathcal{A}\} \rightsquigarrow V_L^{1/2} \cdot \epsilon \sim \mathcal{N}(0_Q, V_L)$$

if and only if $D_{b,\infty}(\Phi \otimes S_x^2)D_{b,\infty}^\top = 0$. A sufficient condition for this is $b_\infty = \gamma$.

Recall that $\sqrt{N}(\hat{Y}_L - \bar{Y}) \rightsquigarrow V_L^{1/2} \cdot \epsilon$ and $\sqrt{N}(\hat{Y}\langle b \rangle - \bar{Y}) \rightsquigarrow V_{b,\infty}^{1/2} \cdot \epsilon$ by Lemma 2. Theorem S3 implies Theorem S1 as a direct consequence, and clarifies the asymptotic relative efficiency between $(\hat{Y}\langle b \rangle \mid \mathcal{A})$, $\hat{Y}\langle b \rangle$, and \hat{Y}_L . The choice of $(V_{b,\infty}^\parallel)_{JH}^{1/2}$ is not unique, but the asymptotic distribution is.

In addition, recall that $\hat{Y}\langle b \rangle \approx \hat{Y}\langle \gamma \rangle - D_{b,\infty}\hat{x}$, with $\hat{Y}\langle \gamma \rangle$ being 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 affords the intuition for the asymptotic distribution of $\hat{Y}\langle b \rangle$ being a convolution of independent Normal and 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 $V_{b,\infty}^\perp - V_L = D_{b,\infty}\{(\Pi^{-1} - 1_{Q \times Q} - \Phi) \otimes S_x^2\}D_{b,\infty}^\top \geq 0$. 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 S6. ReM under the treatment-control experiment is a special case with $H = 1 = Q - 1$.

S2.3. ReM-specific inference based on plug-in distributions

Recall that ReM increases the peakedness of \hat{Y}_r when the restriction on γ is misspecified. The usual Normal approximation will therefore be over-conservative, deterring statistically significant findings. We consider below inference based on ReM-specific sampling distributions to avoid this issue.

Recall from (S2) 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 then conduct ReM-specific inference based on the distribution of

$$(\hat{V}_{b,\infty}^\perp)^{1/2} \cdot \epsilon + (\hat{V}_{b,\infty}^\parallel)_{JH}^{1/2} \cdot \mathcal{L}. \quad (\text{S3})$$

In particular, recall that $\hat{\beta}_{L,q}$ equals the coefficient vector of x_i from the OLS fit of $Y_i \sim 1 + x_i$ over $\{i : Z_i = q\}$, and hence gives an intuitive estimator of γ_q . Let

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

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

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

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

Proposition S2 below follows from Li et al. (2018, Lemma A10), and ensures the asymptotic validity of the inference based on (S3).

Proposition S2. Assume the ReM in Definition S1 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 fits, we can also estimate $V_{b,\infty} = V_*$ by $N\hat{\Psi}_*$, which is asymptotically equivalent to the $\hat{V}_{b,\infty}$ defined above.

S3. Additional results on RLS

S3.1. Properties under general restriction

Recall that $M_r = (\chi_L^\top \chi_L)^{-1} R^\top \{R(\chi_L^\top \chi_L)^{-1} R^\top\}^{-1}$ from Lemma 3. Let $M_{r,\infty} = \text{plim } M_r$ be the finite probability limit of M_r under complete randomization and Condition 4; see (S24) for the explicit form.

Let ξ_r be the first Q rows of $-M_r(R\theta_L - r)$. Let Σ_r be the upper-left $Q \times Q$ submatrix of $(I_p - M_{r,\infty}R)\{N\text{cov}_\infty(\hat{\theta}_L)\}(I_p - M_{r,\infty}R)^\top$; we give the explicit form of $N\text{cov}_\infty(\hat{\theta}_L)$ in Lemma S8.

Recall that $\Pi = \text{diag}(e_q)_{q \in \mathcal{T}}$ with $e_q = N_q/N$. Let

$$\Delta_0 = [R \text{diag}\{\Pi^{-1}, (\Pi \otimes S_x^2)^{-1}\} R^\top]^{-1}.$$

Write $R = (R_Y, R_\gamma)$, with R_Y and R_γ denoting the columns of R corresponding to \bar{Y} and γ , respectively. Theorem S4 below quantifies the asymptotic behaviors of \hat{Y}_r for general R , and suggests that \hat{Y}_r is in general not consistent for \bar{Y} unless $R_Y = 0$ or (12) is correctly specified.

Theorem S4. Assume RLS subject to (12). Then

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

where $\hat{\beta}_r \in \mathcal{B}$ with $\text{plim } \hat{\beta}_r = \gamma - (\Pi \otimes S_x^2)^{-1} R_\gamma^\top \Delta_0 (R\theta_L - r)$. Further assume complete randomization

and Condition 4. Then

$$\hat{Y}_r - \bar{Y} - \xi_r = o_P(1), \quad \sqrt{N}(\hat{Y}_r - \bar{Y} - \xi_r) \rightsquigarrow \mathcal{N}(0_Q, \Sigma_r)$$

with

- (i) $\xi_r = -\Pi^{-1}R_Y^T\Delta_0(R\theta_L - r) + o_P(1)$;
- (ii) $\xi_r = 0$, and hence $\sqrt{N}(\hat{Y}_r - \bar{Y}) \rightsquigarrow \mathcal{N}(0_Q, \Sigma_r)$, if (12) is correctly specified.

Theorem S4 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 S4 with Theorems 1 and 2 in the main text. The consistency of \hat{Y}_r in general requires the whole restriction (12) 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; likewise for the separable restriction to ensure consistency as long as the restriction on \bar{Y} is correctly specified.

S3.2. Asymptotic Normality under separable restriction

Theorem S5 below complements Theorem 2 in the main paper, and gives the asymptotic distribution of \hat{Y}_r when $\rho_Y\bar{Y} = r_Y$ is misspecified. Recall that $U = I_Q - \Pi^{-1}\rho_Y^T(\rho_Y\Pi^{-1}\rho_Y^T)^{-1}\rho_Y$ and $\mu_r = -\Pi^{-1}\rho_Y^T(\rho_Y\Pi^{-1}\rho_Y^T)^{-1}(\rho_Y\bar{Y} - r_Y)$ for $\rho_Y \neq 0$.

Theorem S5. Assume complete randomization, Condition 4, and RLS subject to (14) with $\rho_Y \neq 0$. Then

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

where

- (i) $\mu_r = 0$ if $\rho_Y\bar{Y} = r_Y$ is correctly specified;
- (ii) $V_r \geq V_L$, and satisfies $V_r = V_L$ if $\rho_Y\gamma = r_Y$ is correctly specified.

Echoing the comments after Theorem 3, the restrictions on \bar{Y} and γ affect the asymptotic bias and variance, respectively.

S3.3. Numeric properties of RLS

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 an $N \times 1$ vector Y and an $N \times p$ matrix X , 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 *robust covariance from RLS*.

The following lemma states the invariance of RLS to non-degenerate linear transformation of the regressor vector.

Lemma S2. Consider an $N \times 1$ vector Y and two $N \times p$ matrices, X and X' , that satisfy $X' = X\Gamma$ for some nonsingular $p \times p$ matrix Γ . The RLS fits of

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

yield $(\hat{\beta}, \hat{\epsilon}, \hat{V})$ and $(\hat{\beta}', \hat{\epsilon}', \hat{V}')$ as the coefficient vectors, residuals, and robust covariances by (S4), respectively. They satisfy

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

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. Importantly, whereas Lemma S2 allows for nonzero r , Lemma S3 requires that $r = 0$.

Lemma S3. Consider the RLS fit of

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

where $X \in \mathbb{R}^{N \times p}$, $R \in \mathbb{R}^{l \times p}$, and $\hat{\beta} \in \mathbb{R}^p$ 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

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

where $R_\perp \in \mathbb{R}^{(p-l) \times p}$ 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 $\hat{\beta}_{\text{OLS}} \in \mathbb{R}^{p-l}$ 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 of $R_\perp \hat{\beta}$ from the RLS fit of (S5) by (S4).

Let \hat{V}_{OLS} denote the EHW covariance 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 an $N \times 1$ vector Y , an $N \times k$ matrix X_1 , and an $N \times l$ matrix X_2 , 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 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 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 an $N \times 1$ vector Y , an $N \times k$ matrix X_0 , and $N \times l$ matrices X_1, \dots, X_Q , 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 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 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 $\hat{\Psi}_{\text{r},*}$ 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 $\hat{\Psi}_{\text{r},*}$ follows immediately from Examples S1 and S2.

Proposition S3. $\hat{\Psi}_* = \hat{\Psi}_{\text{r},*}$ for $*$ = N, F.

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

S4. Additional results on 2^K factorial experiments

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

For $\mathcal{F}_+ \subseteq \mathcal{P}_K$ and $\mathcal{F}'_+ \subseteq \mathcal{P}'_K = \{\emptyset\} \cup \mathcal{P}_K$, consider

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

where $Z_{i,\emptyset} x_i = x_i$ with $Z_{i,\emptyset} = 1$ if $\mathcal{K} = \emptyset \in \mathcal{F}'_+$. It is a general specification, and includes (18)–(23) 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 (18)–(23) as special cases.

Let

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

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 $\mathcal{F}'_- = \mathcal{P}'_K \setminus \mathcal{F}'_+$, and let $C'_{s,-}$ concatenate rows of $\{c_{s,\mathcal{K}} : \mathcal{K} \in \mathcal{F}'_-\}$, with $c_{s,\emptyset} = 2^{-(K-1)} 1_Q$ and $C'_{s,-} = 0_J^T$ if \mathcal{F}'_- is empty.

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

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

Let $\tilde{\Omega}_{r,+}$ be the EHW covariance 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

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

In case where $\mathcal{F}_- = \emptyset$ such that $C_{s,-} = 0_J^T$, then $C_{s,-}\bar{Y} = 0$ is correctly specified by definition, implying no restriction on \bar{Y} ; likewise for the case with $\mathcal{F}'_- = \emptyset$ to imply no restriction on γ . Let $\hat{\Psi}_{r,s}$ be the robust covariance of $\hat{Y}_{r,s}$ based on the definition in Section 4.3.4.

Proposition S4 below states the numeric correspondence between $(\tilde{\tau}_{r,+}, \tilde{\Omega}_{r,+})$ and $(\hat{Y}_{r,s}, \hat{\Psi}_{r,s})$.

Proposition S4. $\tilde{\tau}_{r,+} = C_{s,+}\hat{Y}_{r,s}$ and $\tilde{\Omega}_{r,+} = C_{s,+}\hat{\Psi}_{r,s}C_{s,+}^T$.

The one-to-one correspondence between $t_i = \{1(Z_i = q) : q \in \mathcal{T}\}$ and $\{1, Z_{i,\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\}$ ensures that $2^{-1}\tau_{\mathcal{K}}$, $2^{-1}(c_{s,\emptyset}^T \otimes I_J)\gamma$, and $2^{-1}(c_{\mathcal{K}}^T \otimes I_J)\gamma$ give the target parameters of $Z_{i,\mathcal{K}}$, x_i , and $Z_{i,\mathcal{K}}x_i$ for $\mathcal{K} \in \mathcal{P}_K$ in (20), respectively, from the derived linear model perspective; see the proof of Proposition S4 in Section S7. Regression (S7) can thus be seen as a restricted variant of (20), assuming (S8). This gives the intuition behind (S8) and the correspondence between $\tilde{\tau}_{r,+}$ and $\hat{Y}_{r,s}$.

The design-based properties of $\tilde{\tau}_{r,+}$ then follow from those of $\hat{Y}_{r,s}$ in Lemma 2 and Theorems 1–3. Let $\hat{\beta}_{r,s}$ and $V_{r,s}$ be the values of $\hat{\beta}_r$ and V_r associated with $\hat{Y}_{r,s}$. Then $\hat{\beta}_{r,s} \in \mathcal{B}$ with $\text{plim } \hat{\beta}_{r,s} = \gamma$ and $V_{r,s} = V_L$ if $(C'_{s,-} \otimes I_J)\gamma = 0$ by Propositions 1 and 2. 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, if $\mathcal{F}_- \neq \emptyset$. Recall $\tilde{\tau}_{*,+} = \{\tilde{\tau}_{*,\mathcal{K}} : \mathcal{K} \in \mathcal{F}_+\}$ as the estimators of $\tau_{s,+}$ from the factor-saturated regressions (18)–(20) with $\tilde{\tau}_{L,+} \succeq_\infty \tilde{\tau}_{N,+}, \tilde{\tau}_{F,+}$.

Corollary S2. (i) For (S7) that is factor-saturated with $\mathcal{F}_+ = \mathcal{P}_K$, we have

- $\tilde{\tau}_{r,+} = C_s\hat{Y}\langle\hat{\beta}_{r,s}\rangle$.
- $\sqrt{N}(\tilde{\tau}_{r,+} - \tau_s) \rightsquigarrow \mathcal{N}(0, C_sV_{r,s}C_s^T)$ under complete randomization and Condition 4, with $\tilde{\tau}_{r,+} \preceq_\infty \tilde{\tau}_L$ and $\tilde{\Omega}_{r,+}$ being asymptotically conservative for estimating the true sampling covariance. In particular, $\tilde{\tau}_{r,+} \dot{\sim} \tilde{\tau}_L$ if $(C'_{s,-} \otimes I_J)\gamma = 0$. A sufficient condition for this is Condition 2 and (S7) includes x_i .

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

- $\tilde{\tau}_{r,+} - \tau_{s,+} - C_{s,+}\mu_{r,s} = C_{s,+}U_s\{\hat{Y}\langle\hat{\beta}_{r,s}\rangle - \bar{Y}\}$, where $C_{s,+}\mu_{r,s} = 0$ if $\tau_{s,-} = 0$.
- $\sqrt{N}(\tilde{\tau}_{r,+} - \tau_{s,+} - C_{s,+}\mu_{r,s}) \rightsquigarrow \mathcal{N}(0, C_{s,+}U_sV_{r,s}U_s^TC_{s,+}^T)$ under complete randomization and Condition 4, with $\tilde{\Omega}_{r,+}$ being asymptotically conservative for estimating the true sampling covariance.

Further assume Condition 3. Then $V_{r,s} = V_L$ with $\tilde{\tau}_{r,+} \succeq_\infty \tilde{\tau}_{F,+} \dot{\sim} \tilde{\tau}_{L,+} \succeq_\infty \tilde{\tau}_{N,+}$ if (S7) includes x_i .

The implications of Corollary S2 are three-fold, generalizing the comment after Corollary 1. First, it justifies the Wald-type inference of τ_s based on factor-saturated specifications regardless

of the choice of \mathcal{F}'_+ . Second, it justifies the Wald-type inference of $\tau_{s,+}$ based on factor-unsaturated specifications when the nuisance effects excluded are indeed zero. Third, assume Condition 3 with constant treatment effects, it establishes the asymptotic efficiency of additive regressions like $Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}} + x_i$ when the nuisance effects excluded are indeed zero. Specifically, the resulting estimator is asymptotically as efficient as $\tilde{\tau}_L$ if the specification is factor-saturated with $\mathcal{F}_+ = \mathcal{P}_K$, and ensures additional efficiency, in the sense of $\tilde{\tau}_{r,+} \succeq_\infty \tilde{\tau}_{L,+}$, if the specification is factor-unsaturated with $\mathcal{F}_+ \subsetneq \mathcal{P}_K$. This illustrates the value of factor-unsaturated regressions in combination with covariate adjustment for improving efficiency.

Proposition S5 below generalizes Proposition 5 in the main paper, and ensures the consistency of $\tilde{\tau}_{r,+}$ under equal-sized designs even if $\tau_{s,-} \neq 0$. 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 S5. Assume Condition 5. Then $\tilde{\tau}_{r,+} = C_{s,+} \hat{Y} \langle \hat{\beta}_{r,s} \rangle$ with $\sqrt{N}(\tilde{\tau}_{r,+} - \tau_{s,+}) \rightsquigarrow \mathcal{N}(0, C_{s,+} V_{r,s} C_{s,+}^T)$ and $\tilde{\tau}_{r,+} \preceq_\infty \tilde{\tau}_{L,+}$ under complete randomization and Condition 4. In particular, $\tilde{\tau}_{r,+} \dot{\sim} \tilde{\tau}_{L,+}$ if $(C'_{s,-} \otimes I_J)\gamma = 0$.

Remark S1. Zhao and Ding (2021a) showed that weighted least squares can secure the same benefit as equal-sized designs, and ensures the consistency of $\tilde{\tau}_{N,r,+}$ 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 (20) yields

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}}^0 + x_i + \sum_{\mathcal{K} \in \mathcal{P}_K} Z_{i,\mathcal{K}}^0 x_i \quad (\text{S9})$$

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

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

with $c_0 = (1, 0_{Q-1}^T)^T$ 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 S4 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 $(Z_{i,\mathcal{K}}^0 x_i)_{\mathcal{K} \in \mathcal{P}_K}$ in (S9), respectively, from the derived linear model perspective; see the proof of Proposition S6 in Section S7. The elements of $\tau_0 = C_0 \bar{Y}$ define the analogs of the $Q - 1$ standard factorial effects under a different weighting scheme; see Remark S2 at the end of this section.

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

Consider

$$Y_i \sim 1 + \sum_{\mathcal{K} \in \mathcal{F}_+} Z_{i,\mathcal{K}}^0 + \sum_{\mathcal{K} \in \mathcal{F}'_+} Z_{i,\mathcal{K}}^0 x_i \quad (\text{S10})$$

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 $C_{0,-} = 0_J^T$ when \mathcal{F}_- is empty, and $C'_{0,-} = 0_J^T$ when \mathcal{F}'_- is empty, respectively.

Let $\tilde{\tau}_{r,+}^0$ be the coefficient vector of $\{Z_{i,\mathcal{K}} : \mathcal{K} \in \mathcal{P}_K\}$ from the OLS fit of (S10), with $\tilde{\Omega}_{r,+}^0$ as the associated EHW covariance. Let $\tilde{\tau}_{*,+}^0$ be the corresponding estimators of $\tau_{0,+}$ from the $\{0, 1\}$ -coded analogs of the factor-saturated regressions (18)–(20) for $*$ = N, F, L, analogous to $\tilde{\tau}_{*,+}$.

Let $(\hat{Y}_{r,0}, \hat{\beta}_{r,0})$ be the coefficient vectors of $(t_i, t_i \otimes x_i)$, respectively, from the RLS fit of (6) subject to

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

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 S6. Proposition S4 and Corollary S2 hold for inference of $\tau_{0,+}$ based on (S10) 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 S5 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 S2. 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 2021a). 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 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. Proof of the results on $\hat{Y}\langle b \rangle \in \mathcal{Y}$

Assume throughout this section that $b = (b_1^\top, \dots, b_Q^\top)^\top \in \mathcal{B}$, with $\text{plim } b = b_\infty = (b_{1,\infty}^\top, \dots, b_{Q,\infty}^\top)^\top$ 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 S1, respectively. Recall that $D_b = \text{diag}\{(b_q - \gamma_q)^\top\}_{q \in \mathcal{T}}$ and $D_{b,\infty} = \text{plim } D_b = \text{diag}\{(b_{q,\infty} - \gamma_q)^\top\}_{q \in \mathcal{T}}$.

S5.1. Lemmas

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

$$\text{cov}(\hat{\tau}) = \sum_{q \in \mathcal{T}} N_q^{-1} \Gamma_q S_{qq} \Gamma_q^\top - N^{-1} S_\tau^2,$$

where S_τ^2 is the finite-population covariance of $\{\tau_i = \sum_{q \in \mathcal{T}} \Gamma_q Y_i(q) : i = 1, \dots, N\}$. If for any $q, q' \in \mathcal{T}$, $S_{qq'}$ 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

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

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

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 conclusion 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 detailed calculation. \square

Recall that $\Phi = \Pi^{-1}G^T(G\Pi^{-1}G^T)^{-1}G\Pi^{-1}$, where G is a contrast matrix that has full row rank.

Lemma S6. $\Phi = \Pi^{-1} - 1_{Q \times Q}$ for all $(Q - 1) \times Q$ contrast matrices G that have full row rank.

Proof of Lemma S6. First, $G\Phi = G\Pi^{-1}$ such that $G(\Phi - \Pi^{-1}) = 0$. For contrast matrix G of rank $Q - 1$, this implies that the columns of $\Phi - \Pi^{-1}$ are all proportional to 1_Q , represented by $\Phi - \Pi^{-1} = 1_Q(a_1, \dots, a_Q)$ for some constants $a_q \in \mathbb{R}$ ($q = 1, \dots, Q$). The fact that $\Phi - \Pi^{-1}$ is symmetric further suggests that $a_1 = \dots = a_Q$. Let a denote the common value of a_q 's, with $\Phi - \Pi^{-1} = 1_Q(a1_Q^T) = a1_{Q \times Q}$. The result then follows from

$$Q - 1 = \text{tr}(I_{Q-1}) = \text{tr}\{\Pi^{-1}G^T(G\Pi^{-1}G^T)^{-1}G\} = \text{tr}(\Phi\Pi) = \text{tr}\{(\Pi^{-1} + a1_{Q \times Q})\Pi\} = Q + a$$

such that $a = -1$. \square

S5.2. Complete randomization

Proof of Lemma S1. Recall that $\hat{Y}\langle\gamma\rangle = \hat{Y} - \{\text{diag}(\gamma_q^T)_{q \in \mathcal{T}}\}\hat{x}$. The result follows from Lemma S5 with $\text{cov}\{\hat{Y}\langle\gamma\rangle, \hat{x}\} = \text{cov}(\hat{Y}, \hat{x}) - \{\text{diag}(\gamma_q^T)_{q \in \mathcal{T}}\}\text{cov}(\hat{x}) = 0$. \square

Proof of Lemma 2 and Theorem S2. The results follow from $\hat{Y}\langle b \rangle = \hat{Y}\langle\gamma\rangle - D_b\hat{x} \approx \hat{Y}\langle\gamma\rangle - D_{b,\infty}\hat{x}$ by Lemma S5 and Slutsky's theorem. \square

S5.3. ReM

We verify below Theorem S3, which implies Theorem S1 as a direct consequence. 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 $V_{\delta\delta} = N\text{cov}(\hat{\delta})$ and $V_{x\delta} = N\text{cov}(\hat{x}, \hat{\delta})$. A useful fact is

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

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

$$\text{cov}(\mu_x) = N^{-1}V_{x\delta}V_{\delta\delta}^{-1}V_{x\delta}^T = N^{-1}(\Phi \otimes S_x^2), \quad (\text{S11})$$

where $\Phi = \Pi^{-1}G^T(G\Pi^{-1}G^T)^{-1}G\Pi^{-1}$ from (S1).

Proof of Theorem S3. Let $A = \hat{Y}\langle\gamma\rangle - \bar{Y} - D_{b,\infty}r_x$ and $B = -D_{b,\infty}\mu_x$ with

$$A + B = \hat{Y}\langle\gamma\rangle - D_{b,\infty}\hat{x} - \bar{Y} \approx \hat{Y}\langle b\rangle - \bar{Y} \quad (\text{S12})$$

by Theorem S2. This ensures

$$\{\sqrt{N}(\hat{Y}\langle b\rangle - \bar{Y}) \mid \mathcal{A}\} \sim \{\sqrt{N}(A + B) \mid \mathcal{A}\}. \quad (\text{S13})$$

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

First, $(\hat{Y}\langle\gamma\rangle, r_x, \mu_x)$ are pairwise uncorrelated in finite samples, and asymptotically independent and jointly Normally distributed by Lemma S5. 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 $\text{cov}\{\hat{Y}\langle b_\infty\rangle\} = N^{-1}V_{b,\infty}$. Then $E(A) = E(B) = 0_Q$ with

$$\text{cov}(B) = D_{b,\infty}\text{cov}(\mu_x)D_{b,\infty}^T = N^{-1}V_{b,\infty}^\parallel, \quad \text{cov}(A) = \text{cov}\{\hat{Y}\langle b_\infty\rangle\} - \text{cov}(B) = N^{-1}V_{b,\infty}^\perp \quad (\text{S14})$$

by (S11) and the uncorrelatedness of A and B . This ensures

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

and hence $\{\sqrt{N}(A + B) \mid \mathcal{A}\} \sim (V_{b,\infty}^\perp)^{1/2} \cdot \epsilon + (V_{b,\infty}^\parallel)^{1/2}_{JH} \cdot \mathcal{L}$ 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 (S13)

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 (S14) to see that $V_{b,\infty}^\perp = V_L$ if and only if

$$D_{b,\infty}\text{cov}(r_x)D_{b,\infty}^T = 0. \quad (\text{S15})$$

Without restrictions on b , condition (S15) 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 (S11). The sufficiency of $H = Q - 1$ follows from Lemma S6. 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 S1. Direct algebra shows that

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

and likewise

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

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

S6. Proof of the results on treatment-based regressions

S6.1. Notation and useful facts

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

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

with design matrices

$$\chi_{\text{N}} = T, \quad \chi_{\text{F}} = (T, X), \quad \chi_{\text{L}} = (T, T_x),$$

OLS coefficients

$$\hat{\theta}_{\text{N}} = \hat{Y}_{\text{N}}, \quad \hat{\theta}_{\text{F}} = (\hat{Y}_{\text{F}}^\text{T}, \hat{\beta}_{\text{F}}^\text{T})^\text{T}, \quad \hat{\theta}_{\text{L}} = (\hat{Y}_{\text{L}}^\text{T}, \hat{\beta}_{\text{L}}^\text{T})^\text{T},$$

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

Let

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

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

$$\begin{cases} N^{-1} \chi_{\text{N}}^\text{T} \chi_{\text{N}} = N^{-1} T^\text{T} T = \Pi, \\ N^{-1} \chi_{\text{N}}^\text{T} Y = N^{-1} T^\text{T} Y = \Pi \hat{Y} = \Pi \bar{Y} + o_P(1); \end{cases} \quad (\text{S16})$$

$$\left\{ \begin{array}{l} N^{-1}\chi_F^T\chi_F = \begin{pmatrix} N^{-1}T^TT & N^{-1}T^TX \\ N^{-1}X^TT & N^{-1}X^TX \end{pmatrix} = \begin{pmatrix} \Pi & \Pi\hat{X} \\ \hat{X}^T\Pi & \kappa S_x^2 \end{pmatrix} = \begin{pmatrix} \Pi & \\ & S_x^2 \end{pmatrix} + o_P(1), \\ N^{-1}\chi_F^TY = \begin{pmatrix} N^{-1}T^TY \\ N^{-1}X^TY \end{pmatrix} = \begin{pmatrix} \Pi\hat{Y} \\ \sum_{q \in \mathcal{T}} e_q \hat{S}_{xY(q)} \end{pmatrix} = \begin{pmatrix} \Pi\bar{Y} \\ S_x^2\gamma_F \end{pmatrix} + o_P(1), \end{array} \right. \quad (\text{S17})$$

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

$$\left\{ \begin{array}{l} N^{-1}\chi_L^T\chi_L = \begin{pmatrix} N^{-1}T^TT & N^{-1}T^TT_x \\ N^{-1}T_x^TT & N^{-1}T_x^TT_x \end{pmatrix} \\ = \begin{pmatrix} \Pi & \Pi \text{diag}\{\hat{x}^T(q)\}_{q \in \mathcal{T}} \\ \text{diag}\{\hat{x}(q)\}_{q \in \mathcal{T}}\Pi & \text{diag}\{e_q \hat{S}_x^2(q)\}_{q \in \mathcal{T}} \end{pmatrix} = \begin{pmatrix} \Pi & \\ & \Pi \otimes S_x^2 \end{pmatrix} + o_P(1), \\ N^{-1}\chi_L^TY = \begin{pmatrix} N^{-1}T^TY \\ N^{-1}T_x^TY \end{pmatrix} = \begin{pmatrix} \Pi\hat{Y} \\ (\Pi \otimes I_J)\hat{S}_{xY} \end{pmatrix} = \begin{pmatrix} \Pi\bar{Y} \\ (\Pi \otimes I_J)S_{xY} \end{pmatrix} + o_P(1), \end{array} \right. \quad (\text{S18})$$

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

S6.2. Asymptotics of $\hat{\theta}_L$

We show below the asymptotic Normality of $\hat{\theta}_L$ under complete randomization. The result ensures the asymptotic Normality of \hat{Y}_r in Theorem S4.

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 $q \in \mathcal{T}$. Let

$$\hat{\theta} = \frac{\begin{pmatrix} N_1^{-1} \sum_{i: Z_i=1} Y_i(1; \gamma_1) \\ \vdots \\ N_Q^{-1} \sum_{i: Z_i=Q} Y_i(Q; \gamma_Q) \end{pmatrix}}{\begin{pmatrix} N_1^{-1} \sum_{i: Z_i=1} x_i \{Y_i(1; \gamma_1) - \bar{Y}(1)\} \\ \vdots \\ N_Q^{-1} \sum_{i: Z_i=Q} x_i \{Y_i(1; \gamma_Q) - \bar{Y}(Q)\} \end{pmatrix}} = \begin{pmatrix} \hat{Y}\langle\gamma\rangle \\ \hat{\psi} \end{pmatrix} \quad (\text{S19})$$

be the sample analog of

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

where

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

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 S7. 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

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

Proof of Lemma S7. 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 $(Y_i(q), x_i^\top)^\top$ from 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.

□

Lemma S8. Under complete randomization and Condition 4,

$$\sqrt{N}(\hat{\theta}_L - \theta_L) \rightsquigarrow \mathcal{N}(0_{Q+JQ}, A\Sigma_\theta A^T)$$

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

Proof of Lemma S8. Recall the definitions of $\hat{\theta}$ and θ from (S19). It follows from $\chi_L = (T, T_x)$ and $Y - \chi_L \theta_L = \{Y_i - \bar{Y}(Z_i) - x_i^T \gamma_{Z_i}\}_{i=1}^N$ that

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

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 (S18), ensures

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

by Slutsky's Theorem. The result then follows from Lemma S7.

□

S6.3. Results on outputs from OLS

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.

The unadjusted regression. The result follows from $\hat{\theta}_N = (\chi_N^T \chi_N)^{-1} (\chi_N^T Y) = \hat{Y}_N$ by (S16).

The additive regression. The first-order condition of OLS ensures

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

by (S17). Compare the first row to see $\Pi \hat{Y}_F + \Pi \hat{X} \hat{\beta}_F = \Pi \hat{Y}$, 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 $(\hat{Y}_F^T, \hat{\beta}_F^T)^T = (\chi_F^T \chi_F)^{-1} \chi_F^T Y$, where $(N^{-1} \chi_F^T \chi_F)^{-1} = \text{diag}\{\Pi^{-1}, (S_x^2)^{-1}\} + o_P(1)$ and $N^{-1} \chi_F^T Y = ((\Pi \bar{Y})^T, (S_x^2 \gamma_F)^T)^T + o_P(1)$ by (S17).

The fully interacted regression. The first-order condition of OLS ensures

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

by (S18). Compare the first row to see $\Pi \hat{Y}_L + \Pi \text{diag}\{\hat{x}^T(q)\}_{q \in \mathcal{T}} \hat{\beta}_L = \Pi \hat{Y}$, and hence $\hat{Y}_L = \hat{Y} \langle \hat{\beta}_L \rangle$.

The probability limit of $\hat{\beta}_L$ then follows from $(\hat{Y}_L^T, \hat{\beta}_L^T)^T = (\chi_L^T \chi_L)^{-1} \chi_L^T Y$ and (S18). \square

Proposition S1 ensures that

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

for $i = 1, \dots, N$. Denote by $\hat{\Sigma}_*$ ($*$ = N, F, L) the EHW covariance of $\hat{\theta}_*$ from the OLS fit of $Y = \chi_* \hat{\theta}_* + \hat{\epsilon}_*$, where $\hat{\epsilon}_* = (\hat{\epsilon}_{*,1}, \dots, \hat{\epsilon}_{*,N})^T$. Then $\hat{\Psi}_*$ is the upper-left $Q \times Q$ submatrix of $\hat{\Sigma}_*$, with

$$\hat{\Sigma}_* = (\chi_*^T \chi_*)^{-1} M_* (\chi_*^T \chi_*)^{-1},$$

where $M_* = \chi_*^T \text{diag}(\hat{\epsilon}_{*,1}^2, \dots, \hat{\epsilon}_{*,N}^2) \chi_*$.

Proof of Lemma 1. The asymptotic distributions of \hat{Y}_* follow from Proposition S1 and Lemma 2. 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

$$N^{-1} T^T \{\text{diag}(\hat{\epsilon}_{*,i}^2)_{i=1}^N\} T = \text{diag}(e_q S_{*,qq})_{q \in \mathcal{T}} + o_P(1) \quad (\text{S20})$$

by $T^T \{\text{diag}(\hat{\epsilon}_{*,i}^2)_{i=1}^N\} T = \text{diag}(\sum_{i:Z_i=q} \hat{\epsilon}_{*,i}^2)_{q \in \mathcal{T}}$ and $N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{*,i}^2 = S_{*,qq} + o_P(1)$.

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

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

The result then follows from (S20) .

Fully interacted regression (* = L). First,

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

where

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

It then follows from (S20) and analogous algebra that

$$N^{-1} M_1 = \text{diag}(e_q S_{L,qq})_{q \in \mathcal{T}} + o_P(1), \quad N^{-1} M_2 = O_P(1), \quad N^{-1} M_3 = O_P(1). \quad (\text{S21})$$

This, together with $(I_Q, 0_{Q \times JQ})(N^{-1} \chi_L^T \chi_L)^{-1} = (\Pi^{-1}, 0_{Q \times JQ}) + o_P(1)$ by (S18), ensures

$$\begin{aligned} N \hat{\Psi}_L &= (I_Q, 0_{Q \times JQ})(N \hat{\Sigma}_L)(I_Q, 0_{Q \times JQ})^T \\ &= (I_Q, 0_{Q \times JQ})(N^{-1} \chi_L^T \chi_L)^{-1} (N^{-1} M_L)(N^{-1} \chi_L^T \chi_L)^{-1} (I_Q, 0_{Q \times JQ})^T \\ &= \begin{pmatrix} \Pi^{-1}, & 0_{Q \times JQ} \end{pmatrix} \begin{pmatrix} N^{-1} M_1 & N^{-1} M_2 \\ N^{-1} M_2^T & N^{-1} M_3 \end{pmatrix} \begin{pmatrix} \Pi^{-1} \\ 0_{JQ \times Q} \end{pmatrix} + o_P(1) \\ &= \Pi^{-1} (N^{-1} M_1) \Pi^{-1} + o_P(1) \\ &= \text{diag}(S_{L,qq}/e_q)_{q \in \mathcal{T}} + o_P(1) \end{aligned} \quad (\text{S22})$$

from (S21). □

S6.4. Results on outputs from RLS

S6.4.1. RLS with general restriction

Proof of Theorem S4. 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 $(Y - \chi_L \theta)^T (Y - \chi_L \theta) -$

$2\lambda^\top(R\theta - r)$, 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 (S18), ensures

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

Extracting the first row of both sides yields $\hat{Y}_r + \text{diag}\{\hat{x}^\top(q)\}_{q \in \mathcal{T}} \hat{\beta}_r = \hat{Y} - \Pi^{-1} R_Y^\top (N^{-1} \lambda)$. The result then follows from $\hat{Y} - \text{diag}\{\hat{x}^\top(q)\}_{q \in \mathcal{T}} \hat{\beta}_r = \hat{Y} \langle \hat{\beta}_r \rangle$ such that

$$\hat{Y}_r = \hat{Y} \langle \hat{\beta}_r \rangle - \Pi^{-1} R_Y^\top (N^{-1} \lambda). \quad (\text{S23})$$

Asymptotic results. By (S18), we have $N(\chi_L^\top \chi_L)^{-1} = \text{diag}\{\Pi^{-1}, (\Pi \otimes S_x^2)^{-1}\} + o_P(1)$ and hence

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

This ensures

$$M_{r,\infty} = \text{plim } M_r = \begin{pmatrix} \Pi^{-1} & \\ & (\Pi \otimes S_x^2)^{-1} \end{pmatrix} R^\top \Delta_0 = \begin{pmatrix} \Pi^{-1} R_Y^\top \Delta_0 \\ (\Pi \otimes S_x^2)^{-1} R_\gamma^\top \Delta_0 \end{pmatrix} \quad (\text{S24})$$

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

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

by Lemma 3 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

$$\hat{\theta}_r - \theta_L + M_r(R\theta_L - r) = (I_p - M_r R)(\hat{\theta}_L - \theta_L) \overset{\cdot}{\approx} (I_p - M_{r,\infty} R)(\hat{\theta}_L - \theta_L). \quad (\text{S25})$$

The first equality in (S25) follows from Lemma 3. The asymptotic equivalence $\overset{\cdot}{\approx}$ follows from Slutsky's theorem and the asymptotic Normality of $\sqrt{N}(\hat{\theta}_L - \theta_L)$ by Lemma S8. \square

S6.4.2. RLS with correlation-only restriction

Proof of Proposition 1 and Theorem 1. Observe that $R\theta_L - r = \rho_\gamma \gamma - r_\gamma$ under the correlation-only restriction. Proposition 1 follows from Theorem S4, which ensures $\hat{Y}_r = \hat{Y} \langle \hat{\beta}_r \rangle$ with $\text{plim } \hat{\beta}_r = \gamma$ if (15) is correctly specified.

Theorem 1 then follows from Lemma 2, which ensures $\sqrt{N}\{\hat{Y} \langle \hat{\beta}_r \rangle - \bar{Y}\} \rightsquigarrow \mathcal{N}(0, V_r)$, with $V_r = V_L$

if $\text{plim } \hat{\beta}_r = \gamma$. □

S6.4.3. RLS with separable restriction

We next verify the results when the restriction satisfies (14) 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 2, Theorem 2, and Theorem S5. 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: $Y = T\theta_Y + T_x\theta_\gamma + \epsilon$. The Lagrangian for the restricted optimization problem equals

$$\begin{aligned} & (Y - T\theta_Y - T_x\theta_\gamma)^T (Y - T\theta_Y - T_x\theta_\gamma) - 2\lambda_Y^T (\rho_Y \theta_Y - r_Y) - 2\lambda_\gamma^T (\rho_\gamma \theta_\gamma - r_\gamma) \\ \text{or} \quad & (Y - T\theta_Y - T_x\theta_\gamma)^T (Y - T\theta_Y - T_x\theta_\gamma) - 2\lambda_Y^T (\rho_Y \theta_Y - r_Y) \end{aligned}$$

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

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

This, together with $(T^T T)^{-1} T^T Y = \hat{Y}$, $(T^T T)^{-1} T^T T_x = \text{diag}\{\hat{x}^T(q)\}_{q \in \mathcal{T}}$, and $(T^T T)^{-1} = N^{-1} \Pi^{-1}$ by (S18), ensures

$$\hat{Y}_r = \hat{Y} - [\text{diag}\{\hat{x}^T(q)\}_{q \in \mathcal{T}}] \hat{\beta}_r - N^{-1} \Pi^{-1} \rho_Y^T \lambda_Y = \hat{Y} \langle \hat{\beta}_r \rangle - N^{-1} \Pi^{-1} \rho_Y^T \lambda_Y. \quad (\text{S26})$$

The restriction $\rho_Y \hat{Y}_r = r_Y$ further suggests $\lambda_Y = N(\rho_Y \Pi^{-1} \rho_Y^T)^{-1} \{\rho_Y \hat{Y} \langle \hat{\beta}_r \rangle - r_Y\}$. Plugging this in (S26) verifies

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

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

The probability limit of $\hat{\beta}_r$ follows from $\text{plim } \hat{\beta}_r = \gamma - (\Pi \otimes S_x^2)^{-1} R_\gamma^T \Delta_0 (R\theta_L - r)$ by Theorem S4. Specially, 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 then follows from $\sqrt{N}\{\hat{Y} \langle \hat{\beta}_r \rangle - \bar{Y}\} \rightsquigarrow \mathcal{N}(0_Q, V_r)$ by Lemma 2. □

We next verify the asymptotic bias-variance trade-off and the design-based Gauss–Markov the-

orem under constant treatment effects.

Lemma S9. 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 S9. Let $U_Y = I_Q - U = \Pi^{-1}\rho_Y^T(\rho_Y\Pi^{-1}\rho_Y^T)^{-1}\rho_Y$. Then $L - U = U_Y + A\rho_Y = \{\Pi^{-1}\rho_Y^T(\rho_Y\Pi^{-1}\rho_Y^T)^{-1} + A\}\rho_Y$ such that

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

by $\rho_Y\Pi^{-1}U^T = 0$ and hence $\rho_Y(\Pi^{-1} - 1_{Q \times Q})U^T = 0$. This ensures

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

□

Proof of Theorem 3. Assume that the restriction on γ is correctly specified. Then $\text{plim } \hat{\beta}_r = \gamma$ by Proposition 2 with $\sqrt{N}(\hat{Y}_r - \bar{Y} - \mu_r) \rightsquigarrow \mathcal{N}(0_Q, UV_L U^T)$ by Theorem S5.

Condition 3 further ensures that

$$V_L = N\text{cov}_\infty(\hat{Y}_L) = s_0(\Pi^{-1} - 1_{Q \times Q}), \quad (\text{S27})$$

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 statement (i) and the design-based Gauss–Markov result in statement (ii), respectively.

The 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 (S27) and Lemma S9 when ρ_Y is a contrast matrix. .

The Gauss–Markov result when (14) is correctly specified. For an arbitrary $L\hat{Y}\langle b \rangle + a$ that is consistent, 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 special solution to the restriction with $L\bar{Y}_0 + a = \bar{Y}_0$. 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

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

with $LV_L L^\top \geq UV_L U^\top$ by (S27) and Lemma S9 when ρ_Y is a contrast matrix. □

S6.4.4. Robust covariance

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

Lemma S10. Assume complete randomization, Condition 4, and general R . Then

$$N\hat{\Sigma}_r = \begin{pmatrix} N(\hat{\Sigma}_r)_{[Q]} & O_P(1) \\ O_P(1) & O_P(1) \end{pmatrix}, \quad (\text{S28})$$

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 S10. Recall $M_L = \chi_L^\top \text{diag}(\hat{\epsilon}_{L,1}^2, \dots, \hat{\epsilon}_{L,N}^2) \chi_L$ as the “meat” matrix for computing the EHW covariance of \hat{Y}_L from the OLS fit. Define $M_{L,r} = \chi_L^\top \text{diag}(\hat{\epsilon}_{r,1}^2, \dots, \hat{\epsilon}_{r,N}^2) \chi_L$ as a variant of M_L based on the RLS residuals such that

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

The same algebra as in the proof of Lemma 1 ensures

$$M_{L,r} = \begin{pmatrix} M_1 & M_2 \\ M_2^\top & M_3 \end{pmatrix}, \quad (\text{S29})$$

where $M_1 = \text{diag}(\sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2)_{q \in \mathcal{T}}$, $M_2 = \text{diag}(\sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 x_i^\top)_{q \in \mathcal{T}}$, and $M_3 = \text{diag}(\sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 x_i x_i^\top)_{q \in \mathcal{T}}$.

Recall that $\beta_{r,q} = \text{plim } \hat{\beta}_{r,q}$, with $S_{r,qq} = (N-1)^{-1} \sum_{i=1}^N \{Y_i(q) - x_i^\top \beta_{r,q} - \bar{Y}(q)\}^2$. Then

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

for i with $Z_i = q$. This ensures

$$\begin{aligned} N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 &= N_q^{-1} \sum_{i:Z_i=q} \{Y_i(q) - x_i^\top \beta_{r,q} - \bar{Y}(q)\}^2 + \{\hat{Y}_r(q) - \bar{Y}(q)\}^2 + (\hat{\beta}_{r,q} - \beta_{r,q})^\top \hat{S}_x^2(q) (\hat{\beta}_{r,q} - \beta_{r,q}) \\ &\quad - 2\{\hat{Y}_r(q) - \bar{Y}(q)\} \{\hat{Y}_r(q) - \hat{x}^\top(q) \beta_{r,q} - \bar{Y}(q)\} - 2\{\hat{Y}_r(q) - \bar{Y}(q)\} (\hat{\beta}_{r,q} - \beta_{r,q})^\top \hat{x}(q) \end{aligned}$$

$$\begin{aligned}
& -2(\hat{\beta}_{r,q} - \beta_{r,q})^\top \{\hat{S}_{xY(q)} - \hat{S}_x^2(q)\beta_{r,q} - \hat{x}(q)\bar{Y}(q)\} \\
& = S_{r,qq} + \{\hat{Y}_r(q) - \bar{Y}(q)\}^2 + o_P(1);
\end{aligned}$$

the last equality follows from

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

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 $\hat{S}_{xY(q)} = S_{xY(q)} + o_P(1)$. Therefore, by (S29), we have

$$\begin{aligned}
N^{-1}M_1 &= \text{diag} \left(e_q N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 \right)_{q \in \mathcal{T}} \\
&= \text{diag}(e_q S_{r,qq})_{q \in \mathcal{T}} + \text{diag} \left[e_q \{\hat{Y}_r(q) - \bar{Y}(q)\} \right]_{q \in \mathcal{T}} + o_P(1).
\end{aligned} \tag{S30}$$

Similar algebra ensures that $N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 x_i = O_P(1)$ and $N_q^{-1} \sum_{i:Z_i=q} \hat{\epsilon}_{r,i}^2 x_i x_i^\top = O_P(1)$, 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 (S22) that

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

by (S30). □

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

Correlation-only restriction. When the restriction satisfies (15), (S24) simplifies to

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

such that

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

This, together with (S28), ensures

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

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 S10 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 (S24) that

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

This, together with (S28), ensures that

$$\begin{aligned}
& N(I_p - M_r R) \hat{\Sigma}_r (I_p - M_r R)^T \\
&= \begin{pmatrix} U & \\ & O(1) \end{pmatrix} \begin{pmatrix} N(\hat{\Sigma}_r)_{[Q]} & O_P(1) \\ O_P(1) & O_P(1) \end{pmatrix} \begin{pmatrix} U^T & \\ & O(1) \end{pmatrix} + o_P(1) \\
&= \begin{pmatrix} U \{N(\hat{\Sigma}_r)_{[Q]}\} U^T & O_P(1) \\ O_P(1) & O_P(1) \end{pmatrix} + o_P(1),
\end{aligned}$$

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 S10. 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 $R = (\rho_Y, 0)$, $\Delta_0 = \rho_Y^\top (\rho_Y \Pi^{-1} \rho_Y^\top)^{-1}$,

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

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

S6.4.5. Numeric properties

Proof of Lemma S3. Let $\Gamma = (R_\perp^\top, R^\top)^\top$ with

$$\Gamma \Gamma^\top = \begin{pmatrix} R_\perp & \\ & R \end{pmatrix} (R_\perp^\top, R^\top) = \begin{pmatrix} R_\perp R_\perp^\top & \\ & R R^\top \end{pmatrix}.$$

We have

$$\Gamma^{-1} = \Gamma^\top (\Gamma \Gamma^\top)^{-1} = \Gamma^\top \begin{pmatrix} (R_\perp R_\perp^\top)^{-1} & \\ & (R R^\top)^{-1} \end{pmatrix} = (R_\perp^\top (R_\perp R_\perp^\top)^{-1}, R^\top (R R^\top)^{-1})$$

such that $Y = (X \Gamma^{-1})(\Gamma \hat{\beta}) + \hat{\epsilon}$ by (S5) with

$$X \Gamma^{-1} = (X R_\perp^\top (R_\perp R_\perp^\top)^{-1}, X R^\top (R R^\top)^{-1}), \quad \Gamma \hat{\beta} = \begin{pmatrix} R_\perp \hat{\beta} \\ R \hat{\beta} \end{pmatrix} = \begin{pmatrix} R_\perp \hat{\beta} \\ 0 \end{pmatrix}.$$

This implies

$$Y = \{X R_\perp^\top (R_\perp R_\perp^\top)^{-1}\} (R_\perp \hat{\beta}) + \hat{\epsilon},$$

with

$$R_\perp \hat{\beta} = \text{argmin}_{\theta \in \mathbb{R}^{p-l}} \|Y - X R_\perp^\top (R_\perp R_\perp^\top)^{-1} \theta\|_2^2.$$

This justifies the form of (S6) with $\hat{\beta}_{\text{OLS}} = R_\perp \hat{\beta}$ and $\hat{\epsilon}_{\text{OLS}} = \hat{\epsilon}$.

The EHW covariance of $\hat{\beta}_{\text{OLS}}$ from the OLS fit of (S6) equals

$$\hat{V}_{\text{OLS}} = R_\perp R_\perp^\top (R_\perp X^\top X R_\perp^\top)^{-1} R_\perp X^\top \{\text{diag}(\hat{\epsilon}_i^2)_{i=1}^N\} X R_\perp^\top (R_\perp X^\top X R_\perp^\top)^{-1} R_\perp R_\perp^\top.$$

The robust covariance of $R_\perp \hat{\beta}$ from the RLS fit of (S5) by (S4) equals

$$\hat{V}_{\text{RLS}} = R_\perp (I_p - M_{\text{r}} R) (X^T X)^{-1} X^T \{\text{diag}(\hat{\epsilon}_i^2)_{i=1}^N\} X (X^T X)^{-1} (I_p - M_{\text{r}} R)^T R_\perp^T.$$

Juxtapose the expressions of \hat{V}_{OLS} and \hat{V}_{RLS} to see that a sufficient condition for $\hat{V}_{\text{OLS}} = \hat{V}_{\text{RLS}}$ is

$$\Delta = R_\perp^T (R_\perp X^T X R_\perp^T)^{-1} R_\perp - (I_p - M_{\text{r}} R) (X^T X)^{-1} = 0.$$

This is indeed correct. In particular,

$$\Delta (X^T X) R_\perp^T = R_\perp^T - R_\perp^T + M_{\text{r}} R R_\perp^T = 0, \quad \Delta R^T = 0 - (X^T X)^{-1} R^T + (X^T X)^{-1} R^T = 0$$

by $R R_\perp^T = 0$ and the definition of M_{r} . This ensures $\Delta = 0$ given $((X^T X) R_\perp^T, R^T)$ is nonsingular. \square

S7. Proof of the results on factor-based regressions

S7.1. Standard factorial effects

We verify below Proposition S4, Corollary S2, and Proposition S5 in Section S4. The results of Propositions 3–4, Corollary 1, and Proposition 5 then follow as special cases.

Proof of Proposition S4. Recall that $Z_{ik}^0 = 2^{-1}(Z_{ik} + 1)$ gives the $\{0, 1\}$ -counterpart of Z_{ik} . Then

$$t_i = \otimes_{k=1}^K (1 - Z_{ik}^0, Z_{ik}^0)^T \tag{S31}$$

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 (18)–(20). 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

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

where $c_{\text{s},\emptyset} = 2^{-(K-1)} 1_Q$ 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 (20) thus satisfy

$$\begin{pmatrix} t_i \\ t_i \otimes x_i \end{pmatrix} = \begin{pmatrix} \Gamma^T & \\ & \Gamma^T \otimes I_J \end{pmatrix} \begin{pmatrix} f_i \\ f_i \otimes x_i \end{pmatrix}. \quad (\text{S32})$$

Consider the RLS fit of (20) 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 robust covariance 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 (S32). The result then follows from $\tilde{\tau}_{r,+} = \tilde{\tau}'_{r,+}$ and $\tilde{\Omega}_{r,+} = \tilde{\Omega}'_{r,+}$ by Example S1.

In addition, (S32) 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 justify the forms of target parameters in the comments after Proposition S4. \square

Proof of Corollary S2. We verify below the results for $\mathcal{F}_+ = \mathcal{P}_K$ and $\mathcal{F}_+ \subsetneq \mathcal{P}_K$, respectively.

Factor-saturated regression with $\mathcal{F}_+ = \mathcal{P}_K$. The numeric result follows from $\tilde{\tau}_{r,+} = C_{s,+} \hat{Y}_{r,s} = C_{s,+} \hat{Y} \langle \hat{\beta}_{r,s} \rangle$ by Proposition S4 and Proposition 1. This ensures $\tilde{\tau}_{r,+} \preceq \tilde{\tau}_{L,+}$ by Lemma 2.

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 indeed the case, then $\text{plim } \hat{\beta}_{r,s} = \gamma$ with $\hat{Y} \langle \hat{\beta}_{r,s} \rangle \dot{\sim} \hat{Y}_L$ and $\tilde{\tau}_{r,+} \dot{\sim} \tilde{\tau}_{L,+}$ by Proposition 1 and Lemma 2.

Factor-unsaturated regression with $\mathcal{F}_+ \subsetneq \mathcal{P}_K$ and $\mathcal{F}_- \neq \emptyset$. The numeric result follows from

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

by Proposition S4 and Proposition 2. This ensures the asymptotic distribution by Lemma 2, and the asymptotic appropriateness of $\tilde{\Omega}_{r,+}$ by Proposition S4 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 . 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 S9. \square

Proof of Proposition S5. Recall that $\hat{Y}_{r,s}$ is the coefficient vector of t_i from the RLS fit of (6) subject to $C_{s,-} \bar{Y} = 0$ and $(C'_{s,-} \otimes I_J) \gamma = 0$. By Theorem S5, we have

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

where $\rho_Y = C_{s,-}$. The equal treatment sizes further ensure that $\Pi = Q^{-1}I_Q$ such that $C_{s,+}\Pi^{-1}\rho_Y^\top = QC_{s,+}C_{s,-}^\top = 0$. This, together with $\tilde{\tau}_{r,+} = C_{s,+}\hat{Y}_{r,s}$ by Proposition S4, ensures

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

The asymptotic results then follow from Lemma 2. \square

S7.2. 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 S6. Recall from (S31) that $t_i = \otimes_{k=1}^K (1 - Z_{ik}^0, Z_{ik}^0)^\top$. Let $f_i^0 = \otimes_{k=1}^K (1, Z_{ik}^0)^\top$ be the analog of $f_i = \otimes_{k=1}^K (1, Z_{ik})^\top$, vectorizing $\{1, Z_{i,\mathcal{K}}^0 : \mathcal{K} \in \mathcal{P}_K\}$ from (S9). Then $t_i = \Gamma_0^\top f_i^0$ by

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

The regressor vectors of (6) and (S9) thus satisfy

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

analogous to (S32). All results in Proposition S4 and Corollary S4 then follow from the same line of reasoning as that under the $\{-1, +1\}$ coding system. \square