Reconciling design-based and model-based causal inferences for split-plot experiments

Anqi Zhao and Peng Ding *

Abstract

The split-plot design arose from the agricultural science with experimental units, also known as the sub-plots, nested within groups known as the whole-plots. It assigns different interventions at the whole-plot and sub-plot levels, respectively, providing a convenient way to accommodate hard-to-change factors. By design, sub-plots within the same whole-plot receive the same level of the whole-plot intervention, and thereby induce a group structure on the final treatment assignments. A common strategy is to use the ordinary-least-squares (ols) fit of the outcome on the treatment indicators coupled with the robust standard errors clustered at the whole-plot level. It does not give consistent estimators for the treatment effects of interest when the whole-plot sizes vary. Another common strategy is to fit a linear mixed-effects model of the outcome with Normal random effects and errors. It is a purely model-based approach and can be sensitive to violations of parametric assumptions. In contrast, the design-based inference assumes no outcome models and relies solely on the controllable randomization mechanism determined by the physical experiment. We first extend the existing design-based inference based on the Horvitz–Thompson estimator to the Hajek estimator, and establish the finite-population central limit theorem for both under split-plot randomization. We then reconcile the results with those under the model-based approach, and propose two regression strategies, namely (i) the weighted-least-squares (wls) fit of the unit-level data based on the inverse probability weighting and (ii) the ols fit of the aggregate data based on whole-plot total outcomes, to reproduce the Hajek and Horvitz–Thompson estimators from least squares, respectively. This, together with the asymptotic conservativeness of the corresponding cluster-robust covariances for estimating the true design-based covariances as we establish in the process, justifies the validity of the regression estimators for design-based inference. In light of the flexibility of regression formulation with covariate adjustment, we further extend the theory to the case with covariates and demonstrate the efficiency gain by regression-based covariate adjustment via both asymptotic theory and simulation. Importantly, all design-based properties of the regression estimators hold regardless of whether the regression equations are correctly specified or not.

Keywords: Cluster randomization; cluster-robust standard error; covariate adjustment; inverse probability weighting; potential outcome; randomization inference

*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). Peng Ding was partially funded by the U.S. National Science Foundation (grant # 1945136). We thank the Associate Editor, two referees, Zhichao Jiang, and Rahul Mukerjee for most insightful comments.
1. Introduction

The split-plot design originated from agricultural experiments (Yates 1935, 1937) and affords a convenient way to accommodate hard-to-change factors. It remains one of the most popular designs in industrial and engineering applications (Jones and Nachtsheim 2009), and is gaining increasing popularity in social sciences (e.g., Olken 2007; Chong et al. 2020; Breza et al. 2021). It also has deep connections with causal inference with interference (e.g., Hudgens and Halloran 2008; Basse and Feller 2018; Imai et al. 2021).

Split-plot randomization subjects all units within the same group to the same level of the whole-plot intervention and, consequently, poses challenges to subsequent inference for the treatment effects. Model-based analyses often require strong assumptions on the functional forms of the outcome models, for example, Liang and Zeger (1986)'s marginal model, and sometimes impose additional assumptions on the distributions of the error terms and random effects, for example, the mixed-effects model (e.g., Kempthorne 1952; Cox and Reid 2000; Wu and Hamada 2009). The design-based inference, on the other hand, assumes no outcome models and draws its justification solely from the randomization mechanism. Kempthorne (1952) and Hinkelmann and Kempthorne (2008) initiated the discussion on the design-based inference for split-plot designs under the assumption of additive treatment effects. Due to the complexity of the randomization distributions, they invoked additional parametric assumptions for statistical inference. Under the pure design-based inference framework, Zhao et al. (2018) discussed causal inference for split-plot designs without assuming additive treatment effects, and developed the finite-sample exact theory for uniform split-plot designs where all groups are of the same size and have an equal number of units under each level of the sub-plot intervention. Mukerjee and Dasgupta (2019) extended the discussion to possibly non-uniform variants, and considered the Horvitz–Thompson estimator that guarantees unbiased inference. Both works, however, focused on the finite-sample properties of the proposed estimators and left their asymptotic distributions, as the theoretical basis for statistical inference, an open question. To fill this gap, we extend the discussion to the Hajek estimator under possibly non-uniform split-plot randomization, and derive the asymptotic distributions of both the Horvitz–Thompson and Hajek estimators thereunder based on a martingale central limit theorem (Ohlsson 1989). The result includes the sample-mean estimator under uniform designs as a special case, and justifies the design-based inference of possibly non-uniform split-plot experiments based on the Horvitz–Thompson and Hajek estimators, respectively. This constitutes our first contribution.

In addition, Zhao et al. (2018) made a heuristic link between the design-based inference and the regression-based inference in the context of uniform split-plot designs, motivating with the idea of the derived linear model (Kempthorne 1952; Hinkelmann and Kempthorne 2008). We extend their discussion to possibly non-uniform split-plot designs, and propose two regression formulations to reproduce the Hajek and Horvitz–Thompson estimators from least squares, respectively. In particular, we demonstrate that the Hajek estimator is numerically identical to the coefficient from the weighted-least-squares (WLS) fit with unit data based on the inverse probability of treatment, and
the Horvitz–Thompson estimator is numerically identical to the coefficient from the ordinary-least-squares (OLS) fit with aggregate data based on the whole-plot totals. More interestingly, we show that the associated cluster-robust covariances (Liang and Zeger 1986) are asymptotically conservative for the true design-based sampling covariances of the Hajek and Horvitz–Thompson estimators, respectively. These results justify the corresponding regression-based inferences for split-plot data from the design-based perspective. Although the regression procedures were originally motivated by some outcome modeling assumptions, their design-based properties hold independent of those assumptions as long as the data arise from the split-plot design. The analysis as such is justified by the design of the experiment rather than the modeling assumptions. This constitutes our second contribution on the unification of the model-based and design-based inferences.

Last but not least, the regression formulation offers a flexible way to incorporate covariate information, and promises the opportunity to improve asymptotic efficiency under complete randomization (Fisher 1935; Lin 2013). We extend the discussion to possibly non-uniform split-plot randomization, and establish the design-based properties of the additive and fully-interacted formulations for covariate adjustment under the unit and aggregate regressions, respectively. The OLS estimator based on the fully-interacted aggregate model, as it turns out, ensures the highest asymptotic efficiency when (i) the covariates are relatively homogeneous within whole-plots and (ii) we include the whole-plot size factor as an additional covariate. The additive formulation, on the other hand, affords an alternative when the number of whole plots is small. This constitutes our third contribution on the design-based justification of regression-based covariate adjustment.

We start with the $2^2$ split-plot design to lay down the main ideas, and then extend the results to general designs with multiple factors of multiple levels. Our paper furthers the growing literature on design-based causal inference with various types of experimental data (e.g., Neyman 1923, 1935; Kempthorne 1952; Box and Andersen 1955; Wu 1981; Rosenbaum 2002; Hudgens and Halloran 2008; Imai et al. 2009; Schochet 2010; Lin 2013; Miratrix et al. 2013; Sabbaghi and Rubin 2013; Dagsupta et al. 2015; Middleton and Aronow 2015; Imbens and Rubin 2015; Li et al. 2017; Liu and Ding 2017; Fogarty 2018a,b; Basse and Feller 2018; Mukerjee et al. 2018; Liu and Yang 2020; Abadie et al. 2021; Pashley and Miratrix 2021; Schochet et al. 2021; Su and Ding 2021).

We use the following notation for convenience. 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 $I_m$ be the $m \times m$ identity matrix. We suppress the dimensions when they are clear from the context. Let $\otimes$ and $\circ$ denote the Kronecker and Hadamard products of matrices, respectively. Let $1(\cdot)$ be the indicator function. Let var$_\infty$ and cov$_\infty$ denote the asymptotic variance and covariance, respectively. We use $Y_i \sim x_i$ to denote the least-squares regression of $Y_i$ on $x_i$ and focus on the associated cluster-robust covariance for inference motivated by Abadie et al. (2017), Basse and Feller (2018), Imai et al. (2021), and Su and Ding (2021). The terms “regression” and “cluster-robust covariance” refer to the numerical outputs of the least-squares fit free of any modeling assumptions; we evaluate their sampling properties under the design-based framework.
2. Setup

2.1. Motivating examples

Consider a study with multiple factors of interest and a study population nested in different groups. The split-plot design assigns subsets of the factors at the group and unit levels, respectively, providing a convenient way to accommodate hard-to-change factors. We give below two examples from neuroscience and economics to add intuition.

Example 1. Mouse models are widely used in neuroscience research. Fricano et al. (2014) conducted a two-stage randomized experiment to study the effects of fatty acid and Pten knockdown on neuronal hypertrophy. The first stage took place at the mouse level and randomly assigned the mice to three levels of the first intervention, coded by “rapamycin”, “fatty acid”, and “vehicle”, respectively. The second stage took place at the neuron level and randomly infected neurons within each mouse with an shRNA against Pten (“Pten knockdown”) or an mCherry control (“control”) as two levels of the second intervention. The outcome of interest was measured by the soma size of the neurons following the delivery of the treatments. The number of neurons extracted from each mouse varied depending on the level of infection of each virus. This defines a non-uniform split-plot experiment with the “rapamycin”/“fatty acid”/“vehicle” intervention and the “Pten knockdown”/“control” intervention constituting the whole-plot and sub-plot factors, respectively.

Example 2. Olken (2007) conducted a randomized experiment on 608 villages in two of Indonesia’s most populous provinces to study the effects of three interventions on reducing corruption: increasing the probability of external audits (“audits”), increasing participation in accountability meetings (“invitations”), and providing an anonymous comment form to villages (“comment forms”). The villages are nested in subdistricts, which typically contain between 10 and 20 villages. The randomization for audits was clustered by subdistrict such that all study villages in a subdistrict received audits or none did to circumvent interference. The randomization for invitations and comment forms, on the other hand, was done village by village due to less of such concerns. This defines a non-uniform split-plot experiment with the audit and the two participation interventions constituting the whole-plot and sub-plot factors, respectively.

2.2. $2^2$ split-plot randomization, potential outcomes, and causal estimands

To simplify the presentation, we start with the $2^2$ split-plot design with two binary factors of interest, $A, B \in \{0, 1\}$. Consider a study population of $N$ units, $S = \{ws : w = 1, \ldots, W; s = 1, \ldots, M_w\}$, nested in $W$ groups of possibly different sizes, $M_w (w = 1, \ldots, W; \sum_{w=1}^W M_w = N)$. A $2^2$ split-plot design compounds a cluster randomization with a stratified randomization and assigns the treatments to units in two steps:

(I) the first step features a cluster randomization and assigns completely at random $W_a$ groups to receive level $a \in \{0, 1\}$ of factor $A$ with $W_0 + W_1 = W$;
(II) the second step then runs a stratified randomization and assigns completely at random \( M_{wb} \)
units in group \( w \) to receive level \( b \in \{0, 1\} \) of factor B with \( M_{w0} + M_{w1} = M_w \) for \( w = 1, \ldots, W \).

Refer to the first and second steps as the stage (I) and stage (II) randomizations, respectively, with
\( \{W_a, M_{wb} : a, b = 0, 1; w = 1, \ldots, W\} \) being some prespecified, fixed positive integers that satisfy
\( W_a \geq 2 \) and \( M_{wb} \geq 2 \). The final treatment received by a unit is the combination of the level of
factor A its group receives in stage (I) and the level of factor B the unit itself receives in stage (II),
indexed by \( (a, b) \in T = \{(0, 0), (0, 1), (1, 0), (1, 1)\} \); we abbreviate \( (a, b) \) as \( (ab) \) when no confusion
would arise. Refer to each unit as a sub-plot and each group as a whole-plot by convention of
the literature on agricultural experiments. Factors A and B become the whole-plot and sub-plot
factors, respectively.

Let \( A_{ws} = A_w \) and \( B_{ws} \) indicate the levels of factors A and B received by sub-plot \( ws \), respectives, with
\( \mathbb{P}(A_w = a) = W_a/W = p_a \) and \( \mathbb{P}(B_{ws} = b) = M_{wb}/M_w = q_{wb} \) for \( a, b = 0, 1 \). We
suppress the subscript \( s \) in \( A_{ws} \) to highlight its identicalness over all sub-plots within the same
whole-plot. Let \( Z_{ws} = (A_w, B_{ws}) \in T \) indicate the final treatment for sub-plot \( ws \) with
\[
p_{ws}(z) = \mathbb{P}(Z_{ws} = z) = \mathbb{P}(A_w = a) \cdot \mathbb{P}(B_{ws} = b) = p_a q_{wb} \quad \text{for } z = (ab) \in T.
\]

Refer to \( p_{ws}(z) \) as the inclusion probability of sub-plot \( ws \) to receive treatment \( z = (ab) \). It is
identical for units in the same whole-plot yet varies across different whole-plots unless \( q_{wb} = q_{w'b} \)
for \( w \neq w' \).

Let \( \bar{M} = N/W \) be the average whole-plot size, and let \( \alpha_w = M_w/\bar{M} \) be the whole-plot size
factor with \( \bar{\alpha} = W^{-1} \sum_{w=1}^{W} \alpha_w = 1 \). Intuitively, the variance of \( (\alpha_w)^W \)
measures the variability in the whole-plot sizes. The sample size under treatment \( z = (ab) \) equals
\( N_z = \sum_{w:A_w=a} M_{wb} = \bar{M} \sum_{w:A_w=a} \alpha_w q_{wb} \) and is in general stochastic unless \( \alpha_w q_{wb} \) is identical across all \( w \). We call a
split-plot design uniform if \( M_w \) and \( \{M_{wb} : b = 0, 1\} \) are identical across all \( w = 1, \ldots, W \).

**Condition 1.** \( M_w = M \) and \( M_{wb} = M_b \) for all \( w = 1, \ldots, W \) and \( b = 0, 1 \).

Condition 1 ensures that \( p_{ws}(z) = p_a q_b = N_z/N \) is identical for all \( ws \in S \) with \( q_b = M_b/M \)
and \( N_z = N p_a q_b \). A uniform split-plot design is balanced, in the sense of \( N_z \) being identical for
all \( z \in T \), if and only if \( W_0 = W_1 \) and \( M_0 = M_1 \). Zhao et al. (2018) focused on the uniform
\( 2^2 \) split-plot design, and established the unbiasedness of the sample-mean estimator. Most real-
world experiments in social and biomedical sciences, however, are not uniform (see, for example,
Examples 1 and 2).

Let \( Y_{ws}(z) \) be the potential outcome of sub-plot \( ws \) if assigned to treatment \( z \). Let \( \bar{Y}(z) =
N^{-1} \sum_{ws \in S} Y_{ws}(z) \) be the finite-population average, vectorized as \( \bar{Y} = (\bar{Y}(00), \bar{Y}(01), \bar{Y}(10), \bar{Y}(11))^T \).

Contrasts
\[
\tau_a = 2^{-1}\{\bar{Y}(11) + \bar{Y}(10)\} - 2^{-1}\{\bar{Y}(01) + \bar{Y}(00)\},
\]
\[
\tau_b = 2^{-1}\{\bar{Y}(11) + \bar{Y}(01)\} - 2^{-1}\{\bar{Y}(10) + \bar{Y}(00)\},
\]

(1)
\[ \tau_{AB} = \bar{Y}(11) - \bar{Y}(10) - \bar{Y}(01) + \bar{Y}(00) \]

define the standard main effects and interaction under \(2^2\) factorial designs. Sometimes the interaction is also defined as \(\tau_{AB}/2\) (Dasgupta et al. 2015); the difference causes no essential change to our discussion. We will discuss inference of the general estimand \(\tau = G\bar{Y}\)

for arbitrary \(F \times 4\) coefficient matrix \(G\). The standard main effects and interaction in (1) correspond to a special \(G = G_0 = (g_A, g_B, g_{AB})^T\) with \(g_A = 2^{-1}(-1, -1, 1, 1)^T\), \(g_B = 2^{-1}(-1, -1, 1, 1)^T\), and \(g_{AB} = (1, -1, -1, 1)^T\).

We consider the design-based inference of \(Y\), conditioning on the potential outcomes and viewing the treatment assignment as the sole source of randomness.

The observed outcome equals \(Y_{ws} = \sum_{z \in T} 1(Z_{ws} = z) \cdot Y_{ws}(z)\) for sub-plot \(ws\). We focus on estimators of the form \(\hat{\tau} = G\bar{Y}\), where \(\bar{Y}\) is some estimator of \(\bar{Y}\) based on \((Y_{ws})_{ws \in S}, (Z_{ws})_{ws \in S}\), and possibly some pre-treatment covariates. Without loss of generality, we start with inference of \(\bar{Y}\) and then translate the result back to that of \(\tau\). Assume \(a, b \in \{0, 1\}\) index the levels of factors \(A\) and \(B\) in treatment combination \(z \in T\) throughout unless specified otherwise. Let \((Y_{ws}, A_w, B_{ws}, Z_{ws})_{ws \in S}\) be a set of split-plot type data with \(A_w, B_{ws} \in \{0, 1\}\) and \(Z_{ws} = (A_w, B_{ws}) \in T\).

### 3. Horvitz–Thompson and Hajek estimators

We review in this section three design-based estimators for estimating \(\bar{Y}\). Let \(S(z) = \{ws : Z_{ws} = z, ws \in S\}\) be the set of sub-plots under treatment \(z\). The restriction in the stage (I) randomization ensures that there are only two treatment levels, namely \(z = (A_w, 0)\) and \(z = (A_w, 1)\), observed in whole-plot \(w\). For \(z \in T\), let \(W(z)\) denote the set of whole-plots that contain at least one observation under treatment \(z\). By definition, \(W(z) = \{w : A_w = a\}\) with \(|W(z)| = W_a\) for \(z \in \{(a0), (a1)\}\) with level \(a\) of factor \(A\).

First, the sample-mean estimator of \(\bar{Y}(z)\) equals

\[ \hat{Y}_{sm}(z) = |S(z)|^{-1} \sum_{ws \in S(z)} Y_{ws} = |S(z)|^{-1} \sum_{ws \in S} 1(Z_{ws} = z) \cdot Y_{ws}, \]

averaging over all units under treatment \(z\). It is neither unbiased nor consistent in general.

Second, the Horvitz–Thompson estimator is unbiased for \(\bar{Y}(z)\):}

\[ \hat{Y}_{ht}(z) = N^{-1} \sum_{ws \in S(z)} p_{ws}^{-1}(z) \cdot Y_{ws} = N^{-1} \sum_{ws \in S} \frac{1(Z_{ws} = z)}{p_{ws}(z)} \cdot Y_{ws}(z). \] (2)
Split-plot randomization ensures \( p_{ws}(z) = p_a q_{wb} \) and simplifies (2) to
\[
\hat{Y}_{ht}(z) = W_a^{-1} \sum_{w \in W(z)} \alpha_w \hat{Y}_w(z),
\]
where \( \hat{Y}_w(z) = M_w^{-1} \sum_{s:Z_{ws}=z} Y_{ws} \) is the whole-plot sample mean under treatment \( z \). Let \( \bar{Y}_w(z) = M_w^{-1} \sum_{s=1}^{M_w} Y_{ws}(z) \) be the whole-plot average potential outcome, as the population analog of \( \hat{Y}_w(z) \). Let \( U_w(z) = M^{-1} \sum_{s=1}^{M_w} Y_{ws}(z) = \alpha_w \bar{Y}_w(z) \) be the scaled whole-plot total potential outcome with sample analog \( \hat{U}_w(z) = \alpha_w \hat{Y}_w(z) \). We have
\[
\bar{Y}(z) = W^{-1} \sum_{w=1}^{W} \alpha_w \bar{Y}_w(z) = W^{-1} \sum_{w=1}^{W} U_w(z), \quad \hat{Y}_{ht}(z) = W_a^{-1} \sum_{w \in W(z)} \hat{U}_w(z). \tag{3}
\]
This illustrates \( \hat{Y}_{ht}(z) \) as a two-stage sample-mean estimator of \( \bar{Y}(z) \) by first using \( \hat{U}(z) = W_a^{-1} \sum_{w \in W(z)} U_w(z) \) to estimate \( \bar{Y}(z) = W^{-1} \sum_{w=1}^{W} U_w(z) \) and then using \( \hat{U}_w(z) \) to estimate \( U_w(z) \) in \( \hat{U}(z) \) for \( w \in W(z) \). Standard results ensure that the two steps are unbiased with regard to the whole-plot and sub-plot randomizations, respectively.

A main criticism of the Horvitz–Thompson estimator is that it is not invariant to location shifts in general (e.g., Fuller 2009; Middleton and Aronow 2015; Su and Ding 2021). This is intuitively because the sum of the individual weights involved in constructing \( \hat{Y}_{ht}(z) \),
\[
\hat{1}_{ht}(z) = N^{-1} \sum_{w \in S(z)} p_{ws}^{-1}(z),
\]
is stochastic and does not equal 1 in general. In contrast, the Hajek estimator
\[
\hat{Y}_{haj}(z) = \frac{\hat{Y}_{ht}(z)}{\hat{1}_{ht}(z)}
\]
normalizes the Horvitz–Thompson estimator with \( \hat{1}_{ht}(z) \) and ensures location invariance by construction. We can view \( \hat{1}_{ht}(z) \) as the Horvitz–Thompson estimator of constant 1 when all potential outcomes equal 1, and thus the Hajek estimator as a ratio estimator for \( \bar{Y}(z) = \bar{Y}(z)/1 \) with the numerator and denominator estimated by \( \hat{Y}_{ht}(z) \) and \( \hat{1}_{ht}(z) \), respectively.

This gives us three estimators of \( \bar{Y} \), denoted by \( \hat{Y}_* = (\hat{Y}_s(00), \hat{Y}_s(01), \hat{Y}_s(10), \hat{Y}_s(11))^T \) for \( * = sm, ht, haj \). They differ in general but coincide under uniform split-plot designs.

**Proposition 1.** Under Condition 1, we have \( \hat{Y}_{sm} = \hat{Y}_{ht} = \hat{Y}_{haj} \) with
\[
\hat{Y}_{sm}(z) = \hat{Y}_{ht}(z) = \hat{Y}_{haj}(z) = W_a^{-1} \sum_{w \in W(z)} \hat{Y}_w(z) \quad \text{for} \quad z = (ab) \in T.
\]

We study the design-based properties of \( \hat{Y}_* \) \( * = sm, ht, haj \) under split-plot randomization in the next section.
4. Design-based properties under split-plot randomization

4.1. Finite-sample results for the Horvitz–Thompson estimator

Define the scaled between and within whole-plot covariances of \( \{Y_{ws}(z), Y_{ws}(z')\}_{ws \in S} \) as

\[
S(z, z') = (W - 1)^{-1} \sum_{w=1}^{W} \left\{ \alpha_w \bar{Y}_w(z) - \bar{Y}(z) \right\} \left\{ \alpha_w \bar{Y}_w(z') - \bar{Y}(z') \right\},
\]

\[
S_w(z, z') = (M_w - 1)^{-1} \alpha_w^2 \sum_{s=1}^{M_w} \left\{ Y_{ws}(z) - \bar{Y}_w(z) \right\} \left\{ Y_{ws}(z') - \bar{Y}_w(z') \right\},
\]

summarized in \( S = (S(z, z'))_{4 \times 4} \) and \( S_w = (S_w(z, z'))_{4 \times 4} \), respectively (Mukerjee and Dasgupta 2019). They measure the between and within whole-plot heterogeneity in potential outcomes after adjusting for the whole-plot sizes. Let

\[
H = \text{diag}(p_0^{-1}, p_1^{-1}) \otimes 1_{2 \times 2} - 1_{4 \times 4}, \quad H_w = \text{diag}(p_0^{-1}, p_1^{-1}) \otimes \{\text{diag}(q_0^{-1}, q_1^{-1}) - 1_{2 \times 2}\}
\]

be two symmetric \( 4 \times 4 \) matrices defined by the design parameters. Lemma 1 quantifies the sampling covariance of \( \hat{Y}_{ht} \) in finite samples.

**Lemma 1.** Under the \( 2^2 \) split-plot randomization, we have

\[
E(\hat{Y}_{ht}) = \bar{Y}, \quad \text{cov}(\hat{Y}_{ht}) = W^{-1} (H \circ S_{ht} + \Psi)
\]

with \( S_{ht} = S \) and \( \Psi = W^{-1} \sum_{w=1}^{W} M_w^{-1} (H_w \circ S_w) \).

Consider \( \Psi \) as a summary of \( (S_w)_{w=1}^{W} \) after adjusting for the whole-plot sizes. Lemma 1 decomposes the variability in \( \hat{Y}_{ht} \) into that due to the stage (I) randomization, namely \( W^{-1} (H \circ S_{ht}) \), and that due to the stage (II) randomization, namely \( W^{-1} \Psi \). Lemma S3 in the Supplementary Material quantifies this statement rigorously. A direct implication is \( \text{var}(g^T \hat{Y}_{ht}) = W^{-1} g^T (H \circ S_{ht} + \Psi) g \) for arbitrary \( g \in \mathbb{R}^4 \). This gives a more compact matrix form of Mukerjee and Dasgupta (2019, Theorem 1).

Quantification of the Hajek estimator is, on the other hand, hard in finite samples in general. We thus cast the discussion under an asymptotic framework, and establish the asymptotic Normality of \( \hat{Y}_{ht} \) and \( \hat{Y}_{haj} \) under split-plot randomization in Section 4.2.

4.2. Asymptotic Normality of the Horvitz–Thompson and Hajek estimators

To facilitate the discussion, we introduce an intermediate quantity

\[
\hat{Y}'_{ht}(z) = N^{-1} \sum_{ws \in S(z)} p_{ws}^{-1}(z) \cdot Y'_{ws}(z), \quad \text{where} \quad Y'_{ws}(z) = Y_{ws}(z) - \bar{Y}(z),
\]
as the Horvitz-Thompson estimator defined on the centered potential outcomes \( Y_{ws}'(z) \). The difference between the Hajek estimator and the true finite-population average equals

\[
\hat{Y}_{haj}(z) - \hat{Y}(z) = \frac{\hat{Y}_{ht}(z) - \hat{I}_{ht}(z)\hat{Y}(z)}{\hat{I}_{ht}(z)} = \frac{\hat{Y}'_{ht}(z)}{I_{ht}(z)}.
\]

Let \( S_{haj} = (S_{haj}(z, z'))_{k \times k} \), where

\[
S_{haj}(z, z') = (W - 1)^{-1} \sum_{w=1}^{W} \alpha_w^2 (\hat{Y}_w(z) - \hat{Y}(z)) \{\hat{Y}_w(z') - \hat{Y}(z')\},
\]

be the scaled between whole-plot covariance matrix defined on \( \{Y_{ws}'(z) : ws \in S, \ z \in T\} \). Let \( \bar{\alpha} = W^{-1} \sum_{w=1}^{W} \alpha_w^k \) be the \( k \)th moment of \( \alpha_w^k \) for \( k = 1, 2, 4 \) with \( \bar{\alpha} = W^{-1} \sum_{w=1}^{W} \alpha_w = 1 \). Let \( \bar{Y}_{ws}'(z) = M_w^{-1} \sum_{a=1}^{M_w} Y_{ws}^a(z) \) be the uncentered fourth moment of \( Y_{ws}(z) \) in whole-plot \( w \). We state in Condition 2 the regularity conditions for asymptotics under split-plot randomization.

**Condition 2.** As \( W \) goes to infinity, for \( a, b = 0, 1 \) and \( z \in T \),

(i) \( \bar{\alpha}^2 = O(1) \); \( \bar{\alpha}^4 = o(W) \);

(ii) \( p_a \) has a limit in \( (0, 1) \); \( \epsilon \leq \min_{w=1,...,W} q_{wb} \leq \max_{w=1,...,W} q_{wb} \leq 1 - \epsilon \) for some \( \epsilon \in (0, 1/2] \) independent of \( W \);

(iii) \( \bar{Y}, S_{ht}, S_{haj}, \) and \( \Psi \) have finite limits;

(iv) \( \max_{w=1,...,W} |\alpha_w \bar{Y}_w(z) - \bar{Y}(z)|^2/W = o(1) \);

(v) \( W^{-1} \sum_{w=1}^{W} \alpha_w^2 \bar{Y}_w^4(z) = O(1) \); \( W^{-2} \sum_{w=1}^{W} \alpha_w^4 \bar{Y}_w^4(z) = o(1) \).

Without introducing new symbols, we will also use \( p_a, \bar{Y}, S_{ht}, S_{haj}, \) and \( \Psi \) to denote their limiting values when no confusion would arise. The exact meaning should be clear from the context.

With \( \bar{\alpha} = 1 \), Condition 2(ii) requires the finite-population variance of \( \alpha_w \) to be uniformly bounded and thereby protects against the possibility of superlarge whole-plots. It also allows for diverging fourth moment yet stipulates the growth rate to be slower than \( W \).

Condition 2(iii)–(v), on the other hand, ensure that \( \text{cov}(\bar{Y}_{ht}) \) decays at the rate of \( W^{-1} \) and thereby guarantee the consistency of \( \bar{Y}_{ht} \) for estimating \( \bar{Y} \). We do not need \( q_{wb} \) to converge but only be uniformly bounded as long as \( \Psi \) has a finite limit; see Lemma 3.4 in the Supplementary Material for more details. In the neuroscience experiment in Example 2, for example, this imposes bounds on the number of neurons affected by each level of the “Pten knockdown”/“control” intervention. Condition 2(ii) stipulates the bounded fourth moment condition peculiar to the split-plot randomization. Provided Condition 2(ii), it is satisfied as long as the \( \bar{Y}_w^4(z) \)'s are uniformly bounded for all \( w \).

Further, Condition 2 requires only \( W \) goes to infinity and includes both of the following asymptotic regimes as special cases:
(i) $(M_w)_{w=1}^W$ go to infinity as $W$ goes to infinity;

(ii) $(M_w)_{w=1}^W$ are uniformly bounded as $W$ goes to infinity.

Recall from Lemma \( \text{II} \) that $H \circ S_{ht}$ and $\Psi = W^{-1} \sum_{w=1}^W M_w^{-1}(H_w \circ S_w)$ characterize the variability in $\tilde{Y}_{ht}$ due to the stage (I) and stage (II) randomizations, respectively. Regime (i) ensures that $\Psi = o(1)$, and thus the variability from the stage (I) randomization dominates that from stage (II), as long as $(S_w)_{w=1}^W$ are uniformly bounded. Regime (ii), on the other hand, requires $(S_w)_{w=1}^W$ to have a stable mean to ensure $\Psi = O(1)$. A third asymptotic regime is to have $(M_w)_{w=1}^W$ go to infinity while keeping $W$ fixed (Liu and Hudgens 2014). Asymptotic Normality is lost under this regime, and we omit it from the ensuing discussion. Overall, it is crucial to have large $W$ to have reliable asymptotic approximations under our framework. This requires a large number of mice in Example \( \text{II} \), and a large number of subdistricts in Example \( \text{I} \).

**Remark 1.** Although our theory does not impose any stochastic assumptions on the potential outcomes, we can invoke a working model to gain intuition for the requirements on $S_{ht} = O(1)$ and $S_{haj} = O(1)$ in Condition \( \text{II} \)\( \text{III} \). Consider $N = WM$ units in $W$ equal-sized groups, \{w,s : w = 1, \ldots, W; s = 1, \ldots, M\}, with $Y_{ws}(z) \sim [\mu_w, \sigma^2_w]$, where $\mu_w \sim [\mu_0, \sigma^2]$. This defines a classical model for characterizing data nested in clusters. Denote by $P'$ the probability measure induced by the potential outcomes generating process. Standard result shows that $S_{ht} = O_{P'}(1)$ as $W$ and $M$ go to infinity, and degenerates to $S_{ht} = \sigma_{P'}(1)$ if $\sigma_0 = 0$ and $Y_{ws}(z) \sim [\mu_0, \sigma^2]$. We state in Theorem \( \text{III} \) the asymptotic Normality of $\tilde{Y}_{ht}$ and $\tilde{Y}_{haj}$.

**Theorem 1.** Let $\Sigma_s = H \circ S_s + \Psi$ for $* = ht, haj$ with $\Sigma_{ht} = W \text{cov}(\tilde{Y}_{ht})$ and $\Sigma_{haj} = W \text{cov}(\tilde{Y}_{haj})$ in finite samples. Under the $2^2$ split-plot randomization and Condition \( \text{II} \), we have

$$\sqrt{W(\tilde{Y}_s - \bar{Y})} \sim N(0, \Sigma_s) \quad \text{for } * = ht, haj.$$

Theorem \( \text{II} \) ensures the consistency of $\tilde{Y}_{ht}$ and $\tilde{Y}_{haj}$ for estimating $\bar{Y}$, and establishes $\Sigma_{haj}$ as the asymptotic covariance of $\sqrt{W}(\tilde{Y}_{haj} - \bar{Y})$. The large-sample relative efficiency between $\tilde{Y}_{ht}$ and $\tilde{Y}_{haj}$ then follows from the comparison of $\Sigma_{ht}$ and $\Sigma_{haj}$.

**Corollary 1.** Under the $2^2$ split-plot randomization and Condition \( \text{II} \), we have

$$W \left[ \text{var}_{\infty} \{\tilde{Y}_{haj}(z)\} - \text{var}_{\infty} \{\tilde{Y}_{ht}(z)\} \right] = (p^{-1}_a - 1) \{S_{haj}(z, z) - S_{ht}(z, z)\} \quad \text{for } z = (ab) \in T$$

with

- $S_{haj}(z, z) = S_{ht}(z, z)$ if (i) $\tilde{Y}(z) = 0$ or (ii) $\alpha_w = 1$ for all $w$;
- $0 = S_{haj}(z, z) \leq S_{ht}(z, z)$ if $Y_w(z)$ is constant over all $w$;
- $0 = S_{ht}(z, z) \leq S_{haj}(z, z)$ if $U_w(z) = \alpha_w \tilde{Y}_w(z)$ is constant over all $w$. 

10
Intuitively, \( \hat{Y}_{\text{haj}}(z) \) is asymptotically more efficient than \( \hat{Y}_{\text{ht}}(z) \) if the whole-plots have similar average potential outcomes; vice versa if the whole-plots have similar total potential outcomes. The whole-plot averages are often more homogeneous than the whole-plot totals in realistic data generating processes. This affords another angle for perceiving the advantage of \( \hat{Y}_{\text{haj}}(z) \) over \( \hat{Y}_{\text{ht}}(z) \).

4.3. **Estimation of the sampling covariances**

The expressions of \( \Sigma_* \) \((* = \text{ht, haj})\) involve unobserved potential outcomes. We need to estimate them for the Wald-type inference. Let

\[
\hat{S}_{\text{ht}}(z, z') = (W_a - 1)^{-1} \sum_{w:A_0 = a} \{\alpha_w \hat{Y}_w(z) - \hat{Y}_{\text{ht}}(z)\} \{\alpha_w \hat{Y}_w(z') - \hat{Y}_{\text{ht}}(z')\},
\]

\[
\hat{S}_{\text{haj}}(z, z') = (W_a - 1)^{-1} \sum_{w:A_0 = a} \alpha_w^2 \{\hat{Y}_w(z) - \hat{Y}_{\text{haj}}(z)\} \{\hat{Y}_w(z') - \hat{Y}_{\text{haj}}(z')\}
\]

be the sample analogs of \( S_{\text{ht}}(z, z') \) and \( S_{\text{haj}}(z, z') \) for \( z = (ab) \) and \( z' = (ab') \) that share the same level of factor \( A \). Split-plot randomization assigns all sub-plots within the same whole-plot to receive the same level of factor \( A \), and thus deifies the definition of \( \hat{S}_{\text{ht}}(z, z') \) and \( \hat{S}_{\text{haj}}(z, z') \) for \( z = (ab) \) and \( z' = (a'b') \) with \( a \neq a' \). We use

\[
\hat{V}_s = \begin{pmatrix}
W_0^{-1} \begin{pmatrix}
\hat{S}_s(00, 00) & \hat{S}_s(00, 01) \\
\hat{S}_s(00, 01) & \hat{S}_s(01, 01)
\end{pmatrix} & 0_{2 \times 2} \\
0_{2 \times 2} & W_1^{-1} \begin{pmatrix}
\hat{S}_s(10, 10) & \hat{S}_s(10, 11) \\
\hat{S}_s(10, 11) & \hat{S}_s(11, 11)
\end{pmatrix}
\end{pmatrix}
\]

(5)

to estimate the covariance matrix of \( \hat{Y}_s \) for \(* = \text{ht, haj}\), respectively.

**Theorem 2.** Under the \( 2^2 \) split-plot randomization and Condition \( \Box \) we have

\[
W \hat{V}_s - \Sigma_* = S_* + \sigma_p(1) \quad \text{for } * = \text{ht, haj}.
\]

*Mukerjee and Dasgupta (2019, Theorem 2)* implied \( E(\hat{V}_{\text{ht}}) - \text{cov}(\hat{Y}_{\text{ht}}) = W^{-1} S_{\text{ht}} \geq 0 \) such that \( \hat{V}_{\text{ht}} \) is a conservative estimator of \( \text{cov}(\hat{Y}_{\text{ht}}) \) in finite samples, extending *Zhao et al. (2018)* to possibly non-uniform \( 2^2 \) split-plot randomization. Theorem 2 extends the discussion to the finite-population asymptotics, and establishes the asymptotic conservativeness of \( \hat{V}_s \) for estimating the asymptotic covariance of \( \hat{Y}_s \). This, together with Theorem 1, justifies the Wald-type inference of \( \tau = G \hat{Y} \) based on \( \hat{\tau} = G \hat{Y}_s \) with estimated sampling covariance \( G \hat{V}_s G^T \) for \(* = \text{ht, haj}\).
5. Reconciliation with model-based inference

5.1. Overview

Despite the nice properties of the design-based estimators, their reception among practitioners is at best lukewarm due to the dominance of the more convenient model-based counterparts. Can these convenient model-based estimators match their design-based counterparts and deliver inferences that are valid from the design-based perspective? The answer is affirmative with the aid of appropriate weighting schemes and cluster-robust covariances.

Consider

\[ Y_{ws} \sim 1(Z_{ws} = 00) + 1(Z_{ws} = 01) + 1(Z_{ws} = 10) + 1(Z_{ws} = 11) \]  

(6)

as a standard, treatment-based regression formulation that regresses the outcome on the treatment indicators. We propose two general strategies, namely the inverse probability weighting and aggregate model, to recover the Hajek and Horvitz-Thompson estimators of \( \hat{Y} \) directly as coefficients from least-squares fits, and establish the appropriateness of the corresponding cluster-robust covariances for estimating the true sampling covariances. The result reconciles the regression estimators with their design-based counterparts free of any modeling assumptions, and ensures the validity of the resulting inferences regardless of how well the model represents the true outcome generating process.

5.2. Least-squares estimators from treatment-based regressions

We introduce in this subsection three fitting schemes, denoted by “ols”, “wls”, and “ag”, respectively, for estimating \( \hat{Y} \) from least-squares fits, and establish their respective design-based properties under split-plot randomization.

First, the “ols” fitting scheme represents the dominant choice, and takes the ols coefficients from (6) to estimate \( \hat{Y} \). Let \( \hat{\beta}_{ols} \) denote the resulting estimator.

Next, inspired by the use of inverse probability weighting in constructing \( \hat{Y}_{ht}(z) \) and \( \hat{Y}_{haj}(z) \), the “wls” fitting scheme weights \( Y_{ws} \) by the inverse of its realized inclusion probability, \( p_{ws}(Z_{ws}) \), in the least-squares fit of (6), and estimates \( \hat{Y} \) by the resulting wls coefficients. Let \( \hat{\beta}_{wls} \) denote the resulting estimator.

Finally, recall \( \hat{U}_w(z) = \alpha_w \hat{Y}_w(z) \) as an intuitive estimator of the scaled whole-plot total potential outcome, \( U_w(z) \), for \( w \in W(z) \). The restriction in the stage (I) randomization ensures that there are only two treatment levels, namely \( z = (A_w0) \) and \( z = (A_w1) \), observed in whole-plot \( w \), resulting in a total of \( 2W \) whole-plot level observations: \( \{ \hat{U}_w(A_wb) : w = 1, \ldots, W; \ b = 0,1 \} \). We propose to fit

\[ \hat{U}_w(A_wb) \sim 1(A_wb = 00) + 1(A_wb = 01) + 1(A_wb = 10) + 1(A_wb = 11) \]  

(7)
Table 1: Regression estimators of \( \tilde{Y} \) under the “ols”, “wls”, and “ag” fitting schemes, respectively, along with their design-based equivalents. The design-based properties in the last two columns are with regard to general split-plot designs. All six estimators coincide under uniform split-plot designs.

| fitting scheme | model     | weight               | regression estimator | design-based equivalent | unbiased | consistent |
|---------------|-----------|----------------------|----------------------|-------------------------|----------|------------|
| ols           | \( \hat{Y} \) | 1                    | \( \tilde{\beta}_{\text{ols}} \) | \( \hat{Y}_{\text{sm}} \) | no       | no         |
| wls           | \( \{p_w(Z_{ws})\}^{-1} \) | 1                    | \( \tilde{\beta}_{\text{wls}} \) | \( \hat{Y}_{\text{haj}} \) | no       | yes        |
| ag            | \( \tilde{\beta}_{\text{ag}} \) | 1                    | \( \tilde{\beta}_{\text{ag}} \) | \( \hat{Y}_{\text{ht}} \) | yes      | yes        |

over \( \{(w, b) : w = 1, \ldots, W; b = 0, 1\} \) for these 2W observations as an aggregate analog of \( \hat{Y} \). The “ag” fitting scheme takes the resulting OLS coefficients to estimate \( \tilde{Y} \). Let \( \tilde{\beta}_{\text{ag}} \) denote the resulting estimator. The idea of regression based on aggregate data appeared before: Basse and Feller (2018) discussed it in a two-stage experiment for estimating treatment effects in the presence of interference; Su and Ding (2021) recommended it for analyzing the one-stage cluster-randomized experiment.

This gives us three regression estimators, \( \{\tilde{\beta} : \dag = \text{ols, wls, ag}\} \), summarized in Table 1. As a convention, we use the tilde symbol “\( \tilde{\} \)” to denote outputs from least-squares fits. Proposition 2 states their numerical equivalence with the design-based \( \hat{Y}_{\text{sm}}, \hat{Y}_{\text{haj}}, \) and \( \hat{Y}_{\text{ht}} \), respectively.

**Proposition 2.** \( \tilde{\beta}_{\text{ols}} = \hat{Y}_{\text{sm}}, \tilde{\beta}_{\text{wls}} = \hat{Y}_{\text{haj}}, \) and \( \tilde{\beta}_{\text{ag}} = \hat{Y}_{\text{ht}}. \)

Proposition 2 is numeric and shows the utility of inverse probability weighting and aggregate model in reproducing the Hajek and Horvitz–Thompson estimators from least squares, respectively. The correspondence between the three fitting schemes, \( \{\text{ols, wls, ag}\} \), and the three design-based estimation schemes, \( \{\text{sm, haj, ht}\} \), turns out a recurring theme at the heart of the rest of the discussion.

**Remark 2.** Alternatively, the least-squares fit of \( \alpha_wY_{ws} \sim 1(Z_{ws} = 00) + 1(Z_{ws} = 01) + 1(Z_{ws} = 10) + 1(Z_{ws} = 11) \) with weights \( \alpha_w^{-1}\{p_w(Z_{ws})\}^{-1} \) recovers the Horvitz–Thompson estimator from scaled unit-level outcomes. We exclude it from the discussion due to the unnaturalness in both its weighting and outcome transformation schemes. Miratrix et al. (2021) reviewed an alternative WLS in the context of stratified experiments, which corresponds to the least-squares fit of \( \tilde{Y} \) with weights \( N_{Z_{ws}}/p_w(Z_{ws}) \). Lemma S12 in the Supplementary Material ensures that this slightly different weighting scheme leads to identical regression coefficients and cluster-robust covariance as those under the “wls” fitting scheme.

A key virtue of the regression-based approach is its ability to deliver also estimators of the standard errors via the same least-squares fit. Of interest is how these convenient covariance estimators approximate the true sampling covariances from the design-based perspective. Denote
by $\hat{V}_{ht}$ the classic cluster-robust covariance for $\hat{\beta}_\dagger$ ($\dagger = \text{wls}, \text{ag}$) from the same least-squares fit. Theorem 3 shows their asymptotic equivalence with the design-based $\hat{V}_{ht}$ and $\hat{V}_{haj}$, respectively.

**Theorem 3.** Define $\hat{I}_{ht} = \text{diag}\{\hat{I}_{ht}(z)\}_{z \in \mathcal{T}}$. We have

$$
\hat{V}_{\text{wls}} = \hat{I}_{ht}^{-1} \text{diag} \left( \frac{W_0 - 1}{W_0} I_2, \frac{W_1 - 1}{W_1} I_2 \right) \cdot \hat{V}_{\text{haj}} \hat{I}_{ht}^{-1},
$$

$$
\hat{V}_{\text{ag}} = \text{diag} \left( \frac{W_0 - 1}{W_0} I_2, \frac{W_1 - 1}{W_1} I_2 \right) \cdot \hat{V}_{ht}.
$$

Further assume Condition 2. We have $\hat{I}_{ht}(z) = 1 + o_P(1)$ and thus

$$
W(\hat{V}_{\text{wls}} - \hat{V}_{\text{haj}}) = o_P(1), \quad W(\hat{V}_{\text{ag}} - \hat{V}_{ht}) = o_P(1).
$$

With $\hat{I}_{ht}(z) = 1 + o_P(1)$, the asymptotic equivalence between the cluster-robust covariances and their design-based counterparts follows directly from the numerical correspondence, and ensures the asymptotic conservativeness of $\hat{V}_{\text{ag}}$ for estimating the true sampling covariance of $\hat{\gamma}$ for $\dagger = \text{wls}, \text{ag}$. This, together with Proposition 2, justifies the Wald-type inference of $\gamma = G \hat{Y}$ based on point estimator $\hat{\gamma} = G \hat{Y}$ and estimated covariance $G \hat{V}_{ht} G^T$ for $\dagger = \text{wls}, \text{ag}$. Importantly, the cluster-robust covariance is necessary for valid regression-based inferences because the heteroskedasticity-robust covariance can be asymptotically anti-conservative.

**Remark 3.** A number of other options exist for constructing cluster-robust covariances from linear models (Liang and Zeger 1986; Cameron and Miller 2015). In particular, the HC2 variant of $\hat{V}_{ag}$ recovers $\hat{V}_{ht}$ exactly in finite samples (Bell and McCaffrey 2002; Basse and Feller 2018; Imai et al. 2021). The difference between the classic and HC2 estimators vanishes as the sample size goes to infinity. We relegate the details to the Supplementary Material. With small $W$, Bell and McCaffrey (2002), Cameron and Miller (2015), Pustejovsky and Tipton (2018), and MacKinnon et al. (2021) proposed various confidence intervals to achieve better finite-sample coverage properties. They are likely to improve the simple cluster-robust covariance and its HC2 variant under the design-based framework as well. We leave this to future research. Lastly, $\hat{V}_{ht}$ is unbiased for estimating $\text{cov}(\hat{Y}_{ht})$ if $\alpha_w \hat{Y}_w(z)$ is identical over $w = 1, \ldots, W$ for all $z \in \mathcal{T}$. Modification to $\hat{V}_{ht}$ is proposed by Mukerjee and Dasgupta (2019) that ensures unbiased estimation of $\text{cov}(\hat{Y}_{ht})$ under a different additivity assumption. Alternative, likely less common, model specification is needed to recover this variant via least-squares fit.

### 6. Regression-based covariate adjustment

#### 6.1. Covariate adjustment under complete randomization

The regression formulation offers a natural way to incorporate covariates to further improve the estimation efficiency. We briefly review the theory of covariate adjustment under complete randomization to motivate our extension to split-plot randomization.
Consider a treatment-control experiment with two levels of intervention, $T = \{0, 1\}$, and a study population of $N$ units with potential outcomes $\{Y_i(0), Y_i(1) : i = 1, \ldots, N\}$. The finite-population average treatment effect equals $\tau = \bar{Y}(1) - \bar{Y}(0)$, where $\bar{Y}(z) = \frac{1}{N} \sum_{i=1}^{N} Y_i(z)$.

Denote by $Z_i$ the treatment indicator of unit $i$ under complete randomization. The difference-in-means estimator is unbiased for $\tau$, and equals the coefficient of $Z_i$ from the OLS fit of $Y_i \sim 1 + Z_i$. Given covariates $x_i = (x_{i1}, \ldots, x_{ij})^T$ for unit $i$ $(i = 1, \ldots, N)$, Fisher (1935) proposed to use the coefficient of $Z_i$ from the OLS fit of $Y_i \sim 1 + Z_i + x_i$ to estimate $\tau$; Freedman (2008) criticized its potential efficiency loss compared to the difference-in-means estimator; Lin (2013) proposed to use the coefficient of $Z_i$ from the OLS fit of $Y_i \sim 1 + Z_i + (x_i - \bar{x}) + Z_i(x_i - \bar{x})$ with centered covariates and treatment-covariates interactions to estimate $\tau$, and proved that it is at least as efficient as the difference-in-means and Fisher (1935)’s estimators asymptotically. We call Fisher (1935)’s regression the additive specification and Lin (2013)’s regression the fully-interacted specification hence when no confusion would arise.

In the following two subsections, we extend their results to split-plot randomization. We will focus on four covariate-adjusted regressions depending on whether we use the unit or aggregate data to form the model and whether we use the additive or fully-interacted specification for covariate adjustment. We will study their design-based properties and compare their efficiency gains over the unadjusted counterparts. Given $\hat{\tau}_1$ and $\hat{\tau}_2$ as two consistent and asymptotically Normally distributed estimators for $\tau$, we say $\hat{\tau}_1$ is asymptotically more efficient than $\hat{\tau}_2$, or equivalently, $\hat{\tau}_1$ guarantees gains in asymptotic efficiency over $\hat{\tau}_2$, if $N\text{cov}_{\infty}(\hat{\tau}_1) \leq N\text{cov}_{\infty}(\hat{\tau}_2)$ for all possible values of $\{Y_{ws}(z) : ws \in S, z \in T\}$, and the strict inequality holds for at least one set of $Y_{ws}(z)$’s.

### 6.2. Additive treatment-based regressions

Let $x_{ws} = (x_{ws[1]}, \ldots, x_{ws[J^T]}$ be the $J \times 1$ covariate vector for sub-plot $ws$. Adding $x_{ws}$ to (3) yields

$$Y_{ws} \sim \sum \mathbb{1}(Z_{ws} = z) + x_{ws} \sim d_{ws} + x_{ws}$$

as the additive unit regression over $ws \in S$ with $d_{ws} = (1(Z_{ws} = 00), 1(Z_{ws} = 01), 1(Z_{ws} = 10), 1(Z_{ws} = 11))^T$. Let $\hat{\beta}_{ols,f}$ and $\hat{\beta}_{wls,f}$ denote the coefficients of $d_{ws}$ from the OLS and WLS fits of (8), respectively; we use the subscript “f” to indicate Fisher (1935).

Let $\hat{v}_{ws}(z) = \alpha_w \hat{x}_{ws}(z)$ be the covariate analog of $\bar{U}_w(z) = \alpha_w \bar{X}_w(z)$ with $\hat{x}_{ws}(z) = M_{ws}^{-1} \sum_{s:Z_{ws}=z} x_{ws}$. Adding $\hat{v}_{ws}(A_{wb})$ to (3) defines

$$\hat{U}_w(A_{wb}) \sim \sum \mathbb{1}(A_{wb} = z) + \hat{v}_{ws}(A_{wb}) \sim d_w(A_{wb}) + \hat{v}_{ws}(A_{wb})$$

as the additive aggregate regression over $\{ (w, b) : w = 1, \ldots, W; b = 0, 1 \}$ with $d_w(z) = (1(z = 00), 1(z = 01), 1(z = 10), 1(z = 11))^T$ for $z = (A_w, 0), (A_w, 1)$. Let $\hat{\beta}_{ag,f}$ denote the coefficient of $d_w(A_{wb})$ from the OLS fit of (9). This, together with the above $\hat{\beta}_{ols,f}$ and $\hat{\beta}_{wls,f}$, defines three
covariate-adjusted regression estimators of $\bar{Y}$.

Oftentimes we may want to include some whole-plot level attributes to both the unit and aggregate regressions; examples include the weight of a mouse in Example \textcircled{1} and the population of a subdistrict in Example \textcircled{2}. The definition of $x_{ws} = (x_{ws}[1], \ldots, x_{ws}[j])^T$ is flexible enough to accommodate both unit and whole-plot level covariates. In particular, given $c_w$ as a whole-plot level attribute that is prognostic to the unit outcome, $Y_{ws}$, we can simply let $x_{ws[j]} = c_w$ for $s = 1, \ldots, M_w$ to include it as the $j$th covariate in the unit regression. The definition of $\hat{v}_w(A_w b)$ then ensures that it enters into the aggregate regression as $\alpha_w c_w$ for all $w$ and $b$.

We now derive the design-based properties of the above three covariate-adjusted regression estimators, $\hat{\beta}_{1,Y}$ ($\dagger = \text{ols, wls, ag}$). To simplify the presentation, we center the covariates to have $\bar{x} = N^{-1} \sum_{ws \in S} x_{ws} = 0_J$. Let $\hat{x}_s = (\hat{x}_s(00), \hat{x}_s(01), \hat{x}_s(10), \hat{x}_s(11))^T$ be the $4 \times J$ matrix concatenating $\{\hat{x}_s(z)\}_{z \in T}$ for $* = \text{sm, ht, haj}$. Let $\hat{\gamma}_\text{ols}$, $\hat{\gamma}_\text{wls}$, and $\hat{\gamma}_\text{ag}$ be the coefficients of $x_{ws}$ or $\hat{v}_w(A_w b)$ from the respective least-squares fits of \textcircled{3} and \textcircled{4}. Proposition \textcircled{5} parallels Proposition \textcircled{2} and states the numerical correspondence between \textcircled{5} and \textcircled{4}. Proposition \textcircled{5} links the covariate-adjusted $\hat{\beta}_{1,Y}$’s back to the unadjusted $\hat{\beta}_Y$’s from \textcircled{5} and \textcircled{4}, respectively, and establishes $\hat{\beta}_{\text{ols},Y}$, $\hat{\beta}_{\text{wls},Y}$, and $\hat{\beta}_{\text{ag},Y}$ as the sample-mean, Hajek, and Horvitz–Thompson estimators based on the covariate-adjusted outcomes $Y_{ws} - x_{ws}^T \hat{\beta}_Y$ for $\dagger = \text{ols, wls, ag}$. The correspondence between the fitting schemes \{ols, wls, ag\} and the design-based estimation schemes \{sm, haj, ht\} is preserved in these covariate-adjusted fits as well.

The unbiasedness of the Horvitz–Thompson estimator is in general lost after covariate adjustment due to the correlation between $\hat{x}_{ht}$ and $\hat{\gamma}_{ag}$ in $\hat{\beta}_{\text{ag},Y} = \bar{Y}_{ht} - \bar{x}_{ht} \hat{\gamma}_{ag}$. The consistency of $\bar{Y}_s$ and $\hat{x}_s(z)$, where $* = \text{ht, haj}$, for estimating $\bar{Y}$ and $\bar{x}$ under mild conditions, on the other hand, ensures the consistency of $\hat{\beta}_{\text{wls},Y}$ and $\hat{\beta}_{\text{ag},Y}$ so long as $\hat{\gamma}_{wls}$ and $\hat{\gamma}_{ag}$ have finite probability limits. Theorem \textcircled{11} below formalizes the intuition. The sample-mean analog $\hat{\beta}_{\text{ols},Y}$, on the other hand, is in general neither unbiased nor consistent unless the design is uniform. We thus deprioritize it in the following discussion but outline its theoretical guarantees under uniform designs in Section \textcircled{8} in the Supplementary Material.

To simplify the presentation, let $\bar{x}_w = M_w^{-1} \sum_{s=1}^{M_w} x_{ws}$ and $\|x_{ws}\|_4 = M_w^{-1} \sum_{s=1}^{M_w} \|x_{ws}\|_4$ be the whole-plot average covariates and uncentered fourth norms, respectively. Let $S_{xx}$, $S_{xx,w}$, $S_{xY(z)}$, and $S_{xY(z),w}$ be the scaled between and within whole-plot covariances of $(x_{ws})_{ws \in S}$ and $(x_{ws}, Y_{ws})_{ws \in S}$, respectively. Let $S_{xY'(z)}$ be the scaled between whole-plot covariance of $(x_{ws})_{ws \in S}$ with the centered potential outcomes $\{Y_{ws}'(z)\}_{ws \in S}$; the scaled within whole-plot covariance of $\{x_{ws}, Y_{ws}'(z)\}_{ws \in S}$ coincides with $S_{xY(z),w}$. To avoid too many formulas in the main paper, we relegate the explicit forms of $S_{xx}$, $S_{xx,w}$, $S_{xY(z)}$, $S_{xY(z),w}$, and $S_{xY'(z)}$ to Section \textcircled{8} in the Supplementary Material.

Recall $\Psi(z, z) = W^{-1} \sum_{w=1}^{W} M_w^{-1} H_w(z, z) S_w(z, z)$ as the $(z, z)$th element of $\Psi$ from Lemma \textcircled{12}. Let $\Psi_{xx}(z, z)$ and $\Psi_{xY}(z, z)$ be the analogs after replacing $S_w(z, z)$ with $S_{xx,w}$ and $S_{xY(z),w}$,
respectively. Let \( Q_{xx} = (N-1)^{-1} \sum_{w \in S} x_{ws} x_{ws}^T \) and \( Q_{xY(z)} = (N-1)^{-1} \sum_{w \in S} x_{ws} Y_{ws}(z) \) be the standard, unscaled finite-population covariances of \( (x_{ws})_{w \in S} \) and \( \{x_{ws}, Y_{ws}(z)\}_{w \in S} \), respectively.

**Condition 3.** As \( W \) goes to infinity,

(i) \( S_{xx}, \Psi_{xx}(z, z), S_{xY(z)}, S_{xY'(z)}, \Psi_{xY}(z, z), Q_{xx}, \) and \( Q_{xY(z)} \) have finite limits for all \( z \in \mathcal{T} \);

(ii) \( \max_{w=1, \ldots, W} \|a_w \bar{x}_w\|_2^2 / W = o(1) \);

(iii) \( W^{-1} \sum_{w=1}^W \alpha_w^2 \|x_w\|_2^4 = O(1) \); \( W^{-2} \sum_{w=1}^W \alpha_w^4 \|x_w\|_4^4 = o(1) \).

Condition 3 is the analog of Condition (11)-(13) for the covariates as potential outcomes unaffected by the treatment. The additional requirement on \( Q_{xx} \) and \( Q_{xY(z)} \) ensures the convergence of \( \gamma_{wls} \) under split-plot randomization.

Let \( \gamma_{\dagger} \) be the finite probability limit of \( \gamma_{\dagger} \) for \( \dagger = \text{wls}, \text{ag} \) under split-plot randomization and Conditions 2 & 3; we relegate the proof of their existence and explicit forms to Lemma S11 in the Supplementary Material. Let \( S_{wls, f} \) and \( \Sigma_{wls, f} \) be the analogs of \( S_{haj} \) and \( \Sigma_{haj} \) defined on the adjusted potential outcomes \( Y_{ws}(z; \gamma_{wls}) = Y_{ws}(z) - x_{ws}^T \gamma_{wls} \). Let \( S_{ag, f} \) and \( \Sigma_{ag, f} \) be the analogs of \( S_{ht} \) and \( \Sigma_{ht} \) defined on the adjusted potential outcomes \( Y_{ws}(z; \gamma_{ag}) = Y_{ws}(z) - x_{ws}^T \gamma_{ag} \). Let \( \tilde{V}_{\dagger, f} \) be the cluster-robust covariance corresponding to \( \tilde{\beta}_{\dagger, f} \) for \( \dagger = \text{wls}, \text{ag} \).

**Theorem 4.** Under the \( 2^2 \) split-plot randomization and Conditions 2 & 3, we have

\[
\sqrt{W} (\tilde{\beta}_{\dagger, f} - \bar{Y}) \sim N(0, \Sigma_{\dagger, f}), \quad \sqrt{W} \tilde{V}_{\dagger, f} - \Sigma_{\dagger, f} = S_{\dagger, f} + o_P(1)
\]

for \( \dagger = \text{wls}, \text{ag} \), with \( S_{\dagger, f} \geq 0 \).

Theorem 4 states the asymptotic Normality of \( \tilde{\beta}_{\dagger, f} \) (\( \dagger = \text{wls}, \text{ag} \)) under split-plot randomization, and ensures the asymptotic conservativeness of \( \tilde{V}_{\dagger, f} \) for estimating the true sampling covariance. It justifies the regression-based inference of \( \tau = G \bar{Y} \) from the additive models with point estimator \( G \tilde{\beta}_{\dagger, f} \) and estimated sampling covariance \( G \tilde{V}_{\dagger, f} G^T \) for \( \dagger = \text{wls}, \text{ag} \). Neither \( \tilde{\beta}_{wls, f} \) nor \( \tilde{\beta}_{ag, f} \) always guarantees efficiency gain over their unadjusted counterparts, namely \( \tilde{\beta}_{\dagger} \) for \( \dagger = \text{wls}, \text{ag} \), even asymptotically. Similar discussion by Freedman (2008) and Liu (2013) under complete randomization suggests including the interactions between the treatment indicators and covariates in the regression could be one possible remedy. We discuss the utility of such modification in the next two subsections, and give a sufficient condition for guaranteed gains in asymptotic efficiency.

### 6.3. Fully-interacted treatment-based regressions

Modify (S) and (4) with full interactions between the treatment indicators and covariates:

\[
Y_{ws} \sim d_{ws} + \sum_{z \in \mathcal{T}} 1(Z_{ws} = z) \cdot x_{ws} \tag{10}
\]

\[
\tilde{U}_w(A_w b) \sim d_w(A_w b) + \sum_{z \in \mathcal{T}} 1(A_w b = z) \cdot \tilde{v}_w(A_w b). \tag{11}
\]
Let $\tilde{\beta}_{\text{ols},L}$ and $\tilde{\beta}_{\text{wls},L}$ denote the coefficients of $d_{ws}$ from the OLS and WLS fits of the unit regression (III), respectively. Let $\tilde{\beta}_{ag,L}$ denote the coefficient of $d_w(A_{wb})$ from the OLS fit of the aggregate regression (IV). We use the subscript “L” to indicate \textit{Model 2013}. This gives us three more estimators of $\bar{Y}$, \{ $\tilde{\beta}_{t,l} : t = \text{ols, wls, ag}\}$, summarized in Table 4. We now derive their respective design-based properties.

Let $\hat{\gamma}_{t,z}$ be the coefficient of $1(Z_{ws} = z) \cdot x_{ws}$ or $1(A_{wb} = z) \cdot \hat{v}_w(A_{wb})$ from the corresponding regression for $z \in T$ and $t \in \{\text{ols, wls, ag}\}$. Let $\tilde{\beta}_{t,l}(z)$ be the element in $\tilde{\beta}_{t,L}$ that corresponds to treatment $z$.

Proposition 4. $\tilde{\beta}_{\text{ols},L}(z) = \hat{Y}_{\text{sm}}(z) - \hat{x}_{\text{sm}}^T(z)\hat{\gamma}_{\text{ols},z}$, $\tilde{\beta}_{\text{wls},L}(z) = \hat{Y}_{\text{haj}}(z) - \hat{x}_{\text{haj}}^T(z)\hat{\gamma}_{\text{wls},z}$, and $\tilde{\beta}_{ag,L}(z) = \hat{Y}_{\text{ht}}(z) - \hat{x}_{\text{ht}}^T(z)\hat{\gamma}_{\text{ag},z}$ for $z \in T$.

Parallel to the discussion under the additive models, Proposition 4 links the covariate-adjusted $\tilde{\beta}_{t,l}$’s back to the unadjusted $\hat{\beta}_t$’s from (I) and (II), respectively, and establishes $\tilde{\beta}_{\text{ols},L}(z)$, $\tilde{\beta}_{\text{wls},L}(z)$, and $\tilde{\beta}_{ag,L}(z)$ as the sample-mean, Hajek, and Horvitz–Thompson estimators based on the covariate-adjusted outcomes $Y_{ws} - x_{ws}^T\hat{\gamma}_{t,z}$ for $t = \text{ols, wls, ag}$, respectively. A key distinction is that the adjustment is now based on treatment-specific coefficients, $\hat{\gamma}_{t,z}$ for $z \in T$.

The correlation between $\hat{x}_{\text{ht}}(z)$ and $\hat{\gamma}_{\text{ag},z}$ likewise leaves the covariate-adjusted Horvitz–Thompson estimator, $\tilde{\beta}_{ag,L}$, biased in finite samples. The consistency of $\tilde{\beta}_{\text{wls},L}$ and $\tilde{\beta}_{ag,L}$ is retained as long as the $\hat{\gamma}_{t,z}$’s have finite probability limits for $t = \text{wls, ag}$. This is guaranteed by Conditions 2, 3 such that $\hat{\gamma}_{t,z}$ has a finite probability limit, denoted by $\gamma_{t,z}$, under split-plot randomization for $t = \text{wls, ag}$; see Section 3.4 in the Supplementary Material for analogous guarantees by $\tilde{\beta}_{\text{wls},L}$ under uniform designs.

Let $S_{\text{wls},L}$ and $\Sigma_{\text{wls},L}$ be the analogs of $S_{\text{haj}}$ and $\Sigma_{\text{haj}}$ defined on the adjusted potential outcomes $Y_{ws}(z; \gamma_{\text{wls},z}) = Y_{ws}(z) - x_{ws}^T\gamma_{\text{wls},z}$. Let $S_{\text{ag},L}$ and $\Sigma_{\text{ag},L}$ be the analogs of $S_{\text{ht}}$ and $\Sigma_{\text{ht}}$ defined on the adjusted potential outcomes $Y_{ws}(z; \gamma_{\text{ag},z}) = Y_{ws}(z) - x_{ws}^T\gamma_{\text{ag},z}$. Let $\tilde{V}_{\text{L}}$ be the cluster-robust covariance corresponding to $\tilde{\beta}_{t,L}$ for $t = \text{wls, ag}$.

Theorem 5. Under the $2^2$ split-plot randomization and Conditions 2–3, we have

\[
\sqrt{W}(\tilde{\beta}_{t,L} - \bar{Y}) \sim N(0, \Sigma_{t,L}), \quad W\tilde{V}_{t,L} - \Sigma_{t,L} = S_{t,L} + o_p(1)
\]
for \( \hat{\tau} = \text{wls, ag} \), with \( S_{1,1} \geq 0 \).

Theorem 3 establishes the asymptotic Normality of \( \tilde{\beta}_{1,1} \) under split-plot randomization, and ensures the asymptotic conservativeness of \( \tilde{V}_{1,1} \) for estimating the true sampling covariance. It justifies the regression-based inference of \( \tau = G\bar{Y} \) from the fully-interacted models with point estimator \( G\tilde{\beta}_{1,1} \) and estimated sampling covariance \( G\tilde{V}_{1,1}G^T \) for \( \hat{\tau} = \text{wls, ag} \).

6.4. Guaranteed gains in asymptotic efficiency

A natural next question is if the inclusion of the interactions is not just as good but delivers extra gains in asymptotic efficiency. The answer is affirmative when the right covariates are used in combination with the aggregate model.

Inspired by the utility of cluster-size adjustment in improving asymptotic efficiency under one-stage cluster randomization (Middleton and Aronow 2015; Su and Ding 2021), one simple extension to the aggregate regressions in (9) and (11) is to also include the centered whole-plot size factor, \((\alpha_w - 1)\), as an additional whole-plot level covariate in addition to \( \hat{v}_w(A_wb) \). Intuitively, \( \alpha_w \) reflects the size of the whole-plot and is thus prognostic to \( \bar{U}_w(z) \) as the outcome of the aggregate regressions. Let \( \tilde{\beta}_{ag,p}(\alpha, v) \) and \( \tilde{\beta}_{ag,l}(\alpha, v) \) be the resulting OLS coefficients of \( d_w(A_wb) \) under the additive and fully-interacted specifications, respectively; we use the suffix \( \text{"(\alpha, v)"} \) to emphasize the components of the corresponding augmented covariates. Proposition 5 demonstrates the asymptotic efficiency of \( \tilde{\beta}_{ag,l}(\alpha, v) \).

Proposition 5. Under

\[
\Psi_{xx}(z, z) = o(1) \quad \text{for all } z \in \mathcal{T}
\]

and Conditions 2 3, \( \tilde{\beta}_{ag,l}(\alpha, v) \) has the smallest asymptotic covariance matrix among

\[
B = \{ \tilde{\beta}_{\text{wls}}, \tilde{\beta}_{\text{wls},o}; \tilde{\beta}_{ag}, \tilde{\beta}_{ag,o}, \tilde{\beta}_{ag,o}(\alpha, v); \diamond = \text{F, L} \}.
\]

Proposition 5 establishes the optimality of \( \tilde{\beta}_{ag,l}(\alpha, v) \) among the eight consistent regression estimators in \( B \), highlighting the utility of including \((\alpha_w - 1)\) as an additional covariate in the aggregate regression for ensuring additional asymptotic efficiency. The asymptotic efficiency over the unadjusted design-based estimators then follows from the numerical identities in Proposition 4. Condition (12) holds if (i) \( x_{ws} = \bar{x}_w \) or (ii) \( S_{xx, w} \) is uniformly bounded while \( M_w \) goes to infinity for all \( w \). We thus recommend using \( \tilde{\beta}_{ag,l}(\alpha, v) \) when the covariates are relatively homogeneous within whole-plots or when it is reasonable to consider whole-plot level covariates only. The latter ensures gains in asymptotic efficiency over the unadjusted case even if the unit-level covariates show great heterogeneity within each whole-plot. The discussion becomes more complicated when heterogeneous unit-level covariates enter the picture. We leave the more general theory to future work.
This guaranteed minimum asymptotic covariance, however, should not be the basis for dismissing the additive models completely. In particular, the fully-interacted model \((11)\) entails \(|T| \times (1+J)\) parameters compared with the \((|T| + J)\) parameters in the additive model \((9)\), subjecting \(\hat{\beta}_{ag,t}\) to possibly substantial finite-sample variability when \(J\) is large. We thus recommend keeping both strategies in the toolkit and making decisions on a case by case basis contingent on the nature of the design and the abundance of data.

7. Extensions

7.1. Factor-based regressions: practical implementations

Despite the generality of the treatment-based formulations and their theoretical guarantees, they are nevertheless not the dominant choice in practice when the goal is to estimate the standard factorial effects as those defined in \((1)\). Factor-based regressions, as the more popular practice, estimate the factorial effects directly by the regression coefficients.

With the treatment combinations of interest exhibiting a \(2^2\) factorial structure, the factor-based approach regresses the outcome on the factors themselves via specifications like

\[ Y_{ws} \sim 1 + A_w + B_{ws} + A_w B_{ws}, \tag{13} \]

and interprets the coefficients of the non-intercept terms as the main effects and interaction, respectively. Let \(\hat{\tau}'_{A,0}, \hat{\tau}'_{B,0},\) and \(\hat{\tau}'_{AB,0}\) be the OLS coefficients of \(A_{ws}, B_{ws},\) and \(A_{ws}B_{ws}\) from \((13)\), respectively. Standard least-squares theory suggests

\[
\hat{\tau}'_{A,0} = \hat{Y}_sm(10) - \hat{Y}_sm(00), \\
\hat{\tau}'_{B,0} = \hat{Y}_sm(01) - \hat{Y}_sm(00), \\
\hat{\tau}'_{AB,0} = \hat{Y}_sm(11) - \hat{Y}_sm(10) - \hat{Y}_sm(01) + \hat{Y}_sm(00),
\]

equaling the sample-mean estimators of \(\tau_{A,0} = \bar{Y}(10) - \bar{Y}(00), \tau_{B,0} = \bar{Y}(01) - \bar{Y}(00),\) and \(\tau_{AB}\) from \((11)\), respectively. When the goal is to estimate the standard main effects \((\tau_A, \tau_B)\) as defined in \((1)\), an algebraic trick is to center the factor indicators by \(1/2\) and form the regression as

\[ Y_{ws} \sim 1 + (A_w - 1/2) + (B_{ws} - 1/2) + (A_w - 1/2)(B_{ws} - 1/2) \tag{14} \]

over \(ws \in S\). The resulting OLS coefficients of the three non-intercept terms equal the sample-mean estimators of \(\tau_A, \tau_B,\) and \(\tau_{AB}\), respectively \((Zhao and Ding 2021b)\).

Following the intuition from the OLS fit, let \(\hat{\tau}'_{wls}\) \(\hat{\Omega}'_{wls}\) be the WLS coefficients and the corresponding cluster-robust covariance of the non-intercept terms from \((14)\) under fitting scheme “wls”. Let

\[ \hat{U}_w(A_{w}b) \sim 1 + (A_w - 1/2) + (b - 1/2) + (A_w - 1/2)(b - 1/2) \tag{15} \]
be the aggregate analog of (14) over \( \{ (w, b) : w = 1, \ldots, W; b = 0, 1 \} \), with \( \tau_{ag}^i \) and \( \Omega_{ag}^i \) as the corresponding OLS coefficients and cluster-robust covariance of the non-intercept terms under fitting scheme “ag”. As a convention, we use the combination of tilde and prime to denote outputs from factor-based models like (14) and (15).

Proposition 6 follows from the invariance of least squares to non-degenerate linear transformation, and ensures the validity of \( \tilde{\tau}_{i}^T, \tilde{\Omega}_{i}^T \) for the Wald-type inference of the standard factorial effects \( (\tau_A, \tau_B, \tau_{AB})^T = G_0 \bar{Y} \).

**Proposition 6.** \( \tilde{\tau}_{i}^T = G_0 \tilde{\tau}_i \) and \( \tilde{\Omega}_{i}^T = G_0 \tilde{\Omega}_i G_0^T \) for \( \dagger = \text{wls, ag} \).

Specifications (14)–(15) thus deliver the Hajek and Horvitz–Thompson estimators of the standard factorial effects, namely \( \hat{\tau}_{haj} = G_0 \hat{Y}_{haj} \) and \( \hat{\tau}_{ht} = G_0 \hat{Y}_{ht} \), directly as regression coefficients.

We thus recommend using specifications (14)–(15) if the goal is the standard factorial effects and switching to specifications (6)–(7) if otherwise.

The results on covariate adjustment are almost identical to those under the treatment-based regressions. The covariate-adjusted estimator from the fully-interacted aggregate regression after adjusting for the whole-plot sizes ensures asymptotic efficiency under Conditions 2–3 and (12), and is thus our recommendation for estimating the standard factorial effects under \( 2^2 \) split-plot design. We relegate the details to Section S4.3 in the Supplementary Material.

### 7.2. Split-plot designs with multiple factors of multiple levels

All discussion so far concerns the \( 2^2 \) split-plot design with one whole-plot factor and one sub-plot factor, each of two levels. We now extend the result to general split-plot designs with multiple factors of multiple levels.

Consider a set of \( K \geq 2 \) interventions of interest among which there are \( K_A < K \) that we wish to randomize at the whole-plot level. Let \( T_A \) be the set of all possible combinations of these \( K_A \) whole-plot factors, indexed by \( a = 0, \ldots, T_A - 1 \) with \( T_A = |T_A| \). Let \( T_B \) be the set of all possible combinations of the rest \( K_B = K - K_A \) factors to be randomized within each whole-plot, indexed by \( b = 0, \ldots, T_B - 1 \) with \( T_B = |T_B| \). A general \( T_A \times T_B \) split-plot randomization first runs a cluster randomization at the whole-plot level and assigns completely at random \( W_a \) of the \( W \) whole-plots to receive level \( a \in T_A \) of the whole-plot factors with \( \sum_{a \in T_A} W_a = W \). It then conducts an independent randomization within each whole-plot and assigns completely at random \( M_{wb} \) sub-plots in whole-plot \( w \) to receive level \( b \in T_B \) of the sub-plot factors with \( \sum_{b \in T_B} M_{wb} = M_w \) for \( w = 1, \ldots, W \). Refer to the cluster and stratified randomizations as stage (I) and stage (II) of the assignment, respectively. The final treatment of sub-plot \( ws \) is the combination of its whole-plot factors assignment in stage (I) and its sub-plot factors assignment in stage (II), taking values from \( T = \{ (ab) : a \in T_A, b \in T_B \} \). Example 1 defines a \( 3 \times 2 \) split-plot experiment, and Example 2 defines a \( 2 \times 3 \) split-plot experiment.

All notations and results from the \( 2^2 \) case extend to the current setting with minimal modification. We relegate the details on the design-based inference to Section S4.4 in the Supplementary Material.
Material and focus below on the regression-based inference from the factor-based linear models.

Renew \( \{Y_{ws}(z) : ws \in S, \ z \in T\} \) as the potential outcomes under the \( T_A \times T_B \) design with finite-population averages \( \{\bar{Y}(z)\}_{z \in T} \), vectorized as \( \bar{Y} \). Assume 0 as the baseline levels for both \( T_A \) and \( T_B \). We define

\[
\tau_{ab} = T_B^{-1} \sum_{b \in T_B} \{\bar{Y}(ab) - \bar{Y}(0b)\}, \quad \tau_{ab} = T_A^{-1} \sum_{a \in T_A} \{\bar{Y}(ab) - \bar{Y}(a0)\}, \\
\tau_{ab,bb} = \bar{Y}(ab) - \bar{Y}(0b) - \bar{Y}(a0) + \bar{Y}(00),
\]

where \( a = 1, \ldots, T_A - 1 \) and \( b = 1, \ldots, T_B - 1 \), as the standard main effects and interactions under the \( T_A \times T_B \) split-plot design, vectorized as \( \tau = G_0 \bar{Y} \). The definitions reduce to \( \tau_A, \tau_B, \) and \( \tau_{AB} \) from (12) when \( T_A = T_B = 2 \).

Inspired by the utility of centered factors in recovering \( \tau_A \) and \( \tau_B \) directly as least-squares coefficients from (13) and (14), consider

\[
Y_{ws} \sim 1 + \sum_{a=1}^{T_A-1} 1_c(A_w = a) + \sum_{b=1}^{T_B-1} 1_c(B_{ws} = b) + \sum_{a=1}^{T_A-1} \sum_{b=1}^{T_B-1} 1_c(A_w = a)1_c(B_{ws} = b), \quad (17)
\]
\[
\hat{Y}_w(A_wb) \sim 1 + \sum_{a=1}^{T_A-1} 1_c(A_w = a) + \sum_{b'=1}^{T_B-1} 1_c(b = b') + \sum_{a=1}^{T_A-1} \sum_{b'=1}^{T_B-1} 1_c(A_w = a)1_c(b = b'), \quad (18)
\]
as two generalizations under the \( T_A \times T_B \) design with \( 1_c(A_w = a) = 1(A_w = a) - T_A^{-1}, \ 1_c(B_{ws} = b) = 1(B_{ws} = b) - T_B^{-1}, \) and \( 1_c(b = b') = 1(b = b') - T_B^{-1} \).

Renew \( \bar{Y}_w \) (\( \dagger = \text{ols}, \text{wls}, \text{ag} \)) as the coefficients of the non-intercept terms from (12) and (13) under fitting schemes “ols”, “wls”, and “ag”, respectively. Renew \( \hat{Y}_w \) as the sample-mean, Horvitz–Thompson, and Hajek estimators of \( \bar{Y} \) for \* = sm, ht, haj. Proposition 1 states their numerical correspondence paralleling Proposition 3.

**Proposition 7.** \( \bar{\tau}'_{\text{ols}} = G_0 \bar{Y}_w, \bar{\tau}'_{\text{wls}} = G_0 \bar{Y}_w, \) and \( \bar{\tau}'_{\text{ag}} = G_0 \bar{Y}_w \).

The unbiasedness and consistency results then follow from the properties of \( \hat{Y}_w \) for \* = sm, ht, haj such that \( \hat{\tau}'_{\text{wls}} \) and \( \hat{\tau}'_{\text{ag}} \) are both consistent for estimating \( \tau \). The results on the cluster-robust covariances are similar and thus omitted. This justifies the validity of regression-based inferences from (12)–(15) under fitting schemes \( \dagger = \text{wls}, \text{ag} \).

The results on covariate adjustment are almost identical to those under the \( 2^2 \) case and thus omitted to avoid repetition. The covariate-adjusted estimator from the fully-interacted aggregate regression after adjusting for the whole-plot sizes ensures asymptotic efficiency under the generalized version of Conditions 2–4 and (12). It is thus our recommendation for estimating the standard factorial effects under general \( T_A \times T_B \) split-plot design.
7.3. Fisher randomization test

The Fisher randomization test targets the strong null hypothesis of no treatment effect on any unit in its original form, and delivers finite-sample exact p-values regardless of the choice of test statistic. Moreover, the theory under complete randomization demonstrates that the Fisher randomization test with robustly-studentized test statistics is finite-sample exact under the strong null hypothesis, asymptotically valid under the weak null hypothesis of zero average treatment effect, and allows for flexible covariate adjustment to secure additional power (Wu and Ding 2020; Zhao and Ding 2021a). The theory extends naturally to split-plot randomization.

Renew \( B = \{ \tilde{\beta}_{\text{wls}}, \tilde{\beta}_{\text{wls} \circ}, \tilde{\beta}_{\text{ag}}, \tilde{\beta}_{\text{ag} \circ}, (\alpha, v) : \diamond = \text{f, l} \} \) as the collection of regression estimators of \( \tilde{Y} \) that are consistent under general \( T_a \times T_b \) split-plot design. We propose to use the Fisher randomization test with statistic \( t^2(\tilde{\beta}) = (G\tilde{\beta})^T(G\tilde{V}G^T)^{-1}G\tilde{\beta} \) for \( \tilde{\beta} \in B \) and \( \tilde{V} \) as the corresponding cluster-robust covariance. The resulting test is finite-sample exact for testing the strong null hypothesis and asymptotically valid for testing the weak null hypothesis under split-plot randomization for all \( \tilde{\beta} \in B \). Under (12), the test based on \( t^2(\tilde{\beta}_{\text{ag} \circ}, (\alpha, v)) \) has the highest power asymptotically. By duality, we can also construct confidence regions for factorial effects by inverting a sequence of Fisher randomization tests. We relegate the details to Section S4.5 in the Supplementary Material.

8. Numerical example

Due to the space limit, we relegate the simulation studies to Section S5 in the Supplementary Material. We now apply the proposed methods to estimate the standard factorial effects in the 3 \( \times \) 2 neuroscience experiment from Example 11.

We use the data set from Moen et al. (2016) to illustrate our theory and methods. The data set consists of \( N = 1,143 \) neuron level observations nested within \( W = 14 \) mice. The treatment sizes by mouse are summarized in the matrix below:

| Mouse \((w)\) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| \( A_w \) | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| \( M_w \) | 74 | 126 | 51 | 13 | 132 | 94 | 88 | 133 | 152 | 49 | 56 | 27 | 87 | 60 |
| \( M_{w0} \) | 30 | 58 | 18 | 2 | 56 | 39 | 33 | 58 | 60 | 15 | 27 | 7 | 34 | 22 |
| \( M_{w1} \) | 44 | 68 | 33 | 11 | 76 | 55 | 55 | 75 | 92 | 34 | 29 | 20 | 53 | 38 |

Denote by “fa1” and “fa2” the standard main effects of the whole-plot intervention, by “pten” the standard main effect of the sub-plot intervention, and by “fa1:pten” and “fa2:pten” their interactions, with definitions given by (14). We apply four regression schemes as the combinations of two fitting schemes, namely “wls” for unit model and “ag” for aggregate model, and the presence or absence of covariate adjustment. Table 4 shows the point estimators, cluster-robust standard errors, p-values based on large-sample approximations of the t-statistics, and p-values based on...
the Fisher randomization tests, respectively. We use $w^* = M_w / \bar{M}$ as the covariate for neurons in mouse $w$, and conduct covariate adjustment using the additive models due to the small number of whole-plots at $W = 14$. The covariate adjustment reduces the standard errors of the estimators of “fa1” and “fa2” under both the unit and aggregate models, yet has no effect on the estimators of “pten” and the two interactions. The identicalness of the unadjusted and additive models for estimating “pten” and the two interactions is no coincidence but due to the use of whole-plot level covariate, namely $x_{ws} = \alpha_w$, for covariate adjustment. Intuitively, this causes the covariate matrix to be orthogonal to the centered regressors for the subplot factor and interactions after accounting for the least-squares weights, leaving their estimation unaffected by the inclusion of covariates. Proposition $S3$ in the Supplementary Material offers a rigorous statement.

The $p$-values from large-sample approximations and Fisher randomization tests concur in most cases at significance level 0.05. The two exceptions are the tests for “fa2” under the adjusted unit model and those for “fa1” under the unadjusted aggregate model. This is likely due to the relatively small number of whole-plots that leaves the asymptotic approximation dubious. Based on the theory, the $p$-values based on the Fisher randomization tests should be trusted more given their additional guarantee of finite-sample exactness under the strong null hypothesis. This becomes especially important in the current case given its relatively small $W$.

Recall that the regression estimators under the “wls” and “ag” fitting schemes correspond to the Hajek and Horvitz–Thompson estimators, respectively. The estimators and $p$-values under the two models concur in most cases with the two exceptions being the $p$-values from large-sample approximations for “fa1” under the unadjusted models and those for “fa2” under the adjusted models. This is likely due to the relatively small $W$.

Overall the four regression schemes and two types of $p$-values lead to coherent conclusions: the Pten knockdown increased the soma sizes whereas the effect of fatty acid delivery, along with its interaction with Pten, is statistically insignificant.

9. Discussion

Based on the asymptotic analysis, we recommend using the OLS outputs from the fully-interacted aggregate regression after adjusting for the whole-plot sizes if the sample size permits, and switching to the additive model otherwise. The OLS coefficient is consistent for estimating the finite-population average treatment effect, with the corresponding cluster-robust covariance being an asymptotically conservative estimator of the true sampling covariance. The resulting regression-based inference is valid from the design-based perspective regardless of how well the model represents the true relationship between the outcome, treatments, and covariates.
Table 3: Re-analyzing the data from [Moen et al. (2016)]. Suffix “.x” indicates results from the covariate-adjusted models. “p.normal” and “p.frt” indicate the $p$-values from large-sample approximations and Fisher randomization tests, respectively.

### (a) regression based on unit data

|      | est  | se   | p.normal | p.frt | est.x | se.x | p.normal.x | p.frt.x |
|------|------|------|----------|-------|-------|------|------------|---------|
| fa1  | 8.30 | 4.57 | 0.07     | 0.14  | 6.16  | 3.60 | 0.09       | 0.22    |
| fa2  | 5.29 | 5.31 | 0.32     | 0.39  | 10.36 | 4.86 | 0.03       | 0.11    |
| pten | 14.00| 1.81 | 0.00     | 0.00  | 14.00 | 1.81 | 0.00       | 0.00    |
| fa1:pten | 8.64 | 5.24 | 0.10     | 0.23  | 8.64  | 5.24 | 0.10       | 0.23    |
| fa2:pten | −1.33| 2.71 | 0.62     | 0.68  | −1.33 | 2.71 | 0.62       | 0.68    |

### (b) regression based on aggregate data

|      | est  | se   | p.normal | p.frt | est.x | se.x | p.normal.x | p.frt.x |
|------|------|------|----------|-------|-------|------|------------|---------|
| fa1  | 9.84 | 4.22 | 0.03     | 0.08  | 5.66  | 3.69 | 0.14       | 0.21    |
| fa2  | 5.64 | 6.23 | 0.38     | 0.42  | 8.24  | 5.53 | 0.15       | 0.23    |
| pten | 13.15| 1.95 | 0.00     | 0.00  | 13.15 | 1.95 | 0.00       | 0.00    |
| fa1:pten | 7.13 | 5.55 | 0.21     | 0.28  | 7.13  | 5.55 | 0.21       | 0.28    |
| fa2:pten | −2.51| 3.05 | 0.42     | 0.49  | −2.51 | 3.05 | 0.42       | 0.49    |

### References

A. Abadie, S. Athey, G. W. Imbens, and J. M. Wooldridge. When should you adjust standard errors for clustering? *National Bureau of Economic Research*, 2017.

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.

G. Basse and A. Feller. Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association*, 113:41–55, 2018.

R. M. Bell and D. F. McCaffrey. Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, 28:169–181, 2002.

G. E. P. Box and S. L. Andersen. Permutation theory in the derivation of robust criteria and the study of departures from assumption. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 17:1–34, 1955.

E. Breza, F. C. Stanford, M. Alsan, B. Alsan, A. Banerjee, A. G. Chandrasekhar, S. Eichmeyer, T. Glushko, P. Goldsmith-Pinkham, K. Holland, E. Hoppe, M. Karnani, S. Liegl, T. Loisel, L. Ogbu-Nwobodo, B. A. Olken, C. Torres, P.-L. Vautrey, E. T. Warner, S. Wootton, and E. Duñó. Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: A cluster randomized controlled trial. *Nature Medicine*, 27:1622–1628, 2021.

A. Cameron and D. Miller. A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50:317–372, 2015.
A. Chong, M. Gonzalez-Navarro, D. Karlan, and M. Valdivia. Do information technologies improve teenagers’ sexual education? Evidence from a randomized evaluation in Colombia. *World Bank Economic Review*, 34:371–392, 2020.

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

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.

R. A. Fisher. *The Design of Experiments*. Edinburgh, London: Oliver and Boyd, 1st edition, 1935.

C. B. Fogarty. On mitigating the analytical limitations of finely stratified experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80:1035–1056, 2018a.

C. B. Fogarty. Regression-assisted inference for the average treatment effect in paired experiments. *Biometrika*, 105:994–1000, 2018b.

D. A. Freedman. On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40:180–193, 2008.

C. J. Fricano, T. DeSpenza Jr., P. W. Frazel, M. Li, J. A. O’Malley, G. L. Westbrook, and B. W. Luikart. Fatty acids increase neuronal hypertrophy of Pten knockdown neurons. *Frontiers in Molecular Neuroscience*, 7:30, 2014.

W. A. Fuller. *Sampling Statistics*. New York: Wiley, 2009.

K. Hinkelmann and O. Kempthorne. *Design and Analysis of Experiments*, volume 1. New York: John Wiley and Sons, 2nd edition, 2008.

M. G. Hudgens and M. E. Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103:832–842, 2008.

K. Imai, G. King, and C. Nall. The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. *Statistical Science*, 24:29–53, 2009.

K. Imai, Z. Jiang, and A. Malani. Causal inference with interference and noncompliance in two-stage randomized experiments. *Journal of the American Statistical Association*, 116:632–644, 2021.

G. W. Imbens and D. B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press, 2015.
X. Ji, G. Fink, P. J. Robyn, and D. S. Small. Randomization inference for stepped-wedge cluster-randomized trials: An application to community-based health insurance. *Annals of Applied Statistics*, 11:1–20, 2017.

B. Jones and C. J. Nachtsheim. Split-plot designs: What, why, and how. *Journal of Quality Technology*, 41:340–361, 2009.

O. Kempthorne. *The Design and Analysis of Experiments*. New York: John Wiley and 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.

K. Liang and S. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22, 1986.

W. Lin. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique. *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.

L. Liu and M. G. Hudgans. Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association*, 109:288–301, 2014.

J. G. MacKinnon, M. Nielsen, and M. D. Webb. Cluster-robust inference: A guide to empirical practice. Working Paper 1456, Economics Department, Queen’s University, 2021.

J. A. Middleton and P. M. Aronow. Unbiased estimation of the average treatment effect in cluster-randomized experiments. *Statistics, Politics and Policy*, 6:39–75, 2015.

L. W. 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)*, 75:369–96, 2013.

L. W. Miratrix, M. J. Weiss, and B. Henderson. An applied researcher’s guide to estimating effects from multisite individually randomized trials: Estimands, estimators, and estimates. *Journal of Research on Educational Effectiveness*, 14:270–308, 2021.

E. L. Moen, C. J. Fricano-Kugler, B. W. Luikart, and A. J. O’Malley. Analyzing clustered data: Why and how to account for multiple observations nested within a study participant. *PLoS ONE*, 11:e0146721–e0146721, 2016.

R. Mukerjee and T. Dasgupta. Causal inference from possibly unbalanced split-plot designs: A randomization-based perspective. *ArXiv*, 2019. https://arxiv.org/abs/1906.08420.
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.

J. Neyman. On the application of probability theory to agricultural experiments. *Statistical Science*, 5:465–472, 1923. translated by D. M. Dabrowska and T. P. Speed.

J. Neyman. Statistical problems in agricultural experimentation (with discussion). *Supplement to the Journal of the Royal Statistical Society*, 2:107–180, 1935.

E. Ohlsson. Asymptotic normality for two-stage sampling from a finite population. *Probability Theory and Related Fields*, 81:341–352, 1989.

B. A. Olken. Monitoring corruption: Evidence from a field experiment in Indonesia. *Journal of Political Economy*, 115:200–249, 2007.

N. E. Pashley and L. W. Miratrix. Insights on variance estimation for blocked and matched pairs designs. *Journal of Educational and Behavioral Statistics*, 46:271–296, 2021.

J. E. Pustejovsky and E. Tipton. Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, 36:672–683, 2018.

P. R. Rosenbaum. *Observational Studies*. New York: Springer, 2nd edition, 2002.

A. Sabbaghi and D. B. Rubin. Comments on the Neyman–Fisher controversy and its consequences. *Statistical Science*, 29:267–284, 2014.

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, N. E. Pashley, L. W. Miratrix, and T. Kautz. Design-based ratio estimators and central limit theorems for clustered, blocked RCTs. *Journal of the American Statistical Association*, in press, 2021.

G. P. H. Styan. Hadamard products and multivariate statistical analysis. *Linear Algebra and its Applications*, 6:217–240, 1973.

F. Su and P. Ding. Model-assisted analyses of cluster-randomized experiments. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, in press, 2021.

C. F. J. Wu. On the robustness and efficiency of some randomized designs. *Annals of Statistics*, 9:1168–1177, 1981.

C. F. J. Wu and M. Hamada. *Experiments: Planning, Analysis, and Optimization*. New York: John Wiley & Sons, 2009.

28
J. Wu and P. Ding. Randomization tests for weak null hypotheses in randomized experiments. *Journal of the American Statistical Association*, in press, 2020.

F. Yates. Complex experiments. *Supplement to the Journal of the Royal Statistical Society*, 2: 181–247, 1935.

F. Yates. *The Design and Analysis of Factorial Experiments*. Harpenden: Imperial Bureau of Soil Science London, 1937.

A. Zhao and P. Ding. Covariate-adjusted Fisher randomization tests for the average treatment effect. *Journal of Econometrics*, in press, 2021a.

A. Zhao and P. Ding. Regression-based causal inference with factorial experiments: estimands, model specifications, and design-based properties. *ArXiv*, 2021b. https://arxiv.org/abs/2101.02400.

A. Zhao, P. Ding, R. Mukerjee, and T. Dasgupta. Randomization-based causal inference from split-plot designs. *Annals of Statistics*, 46:1876–1903, 2018.
Supplementary Material

Section S1 gives the proofs for the design-based inference under the $2^2$ split-plot design.
Section S2 gives the proofs for the regression-based inference without covariate adjustment.
Section S3 gives the proofs for the regression-based covariate adjustment.
Section S4 gives the results for the special case of uniform designs and extensions to the HC2 correction for the cluster-robust covariance estimators, covariate adjustment via factor-based regression, the general $T_A \times T_B$ design, and the Fisher randomization test.

Section S5 gives the results of the simulation studies to assess the finite-sample properties of the estimators, complementing the asymptotic theory in the main paper.

S1. Design-based inference for the $2^2$ split-plot design

S1.1. Notation and useful facts

We review in this subsection the key notation and algebraic facts for verifying the results under the $2^2$ split-plot design. The results are stated in terms of the general $T_A \times T_B$ design to facilitate generalization. Let $a \in T_A = \{0, 1, \ldots, T_A - 1\}$ and $b \in T_B = \{0, 1, \ldots, T_B - 1\}$ indicate the levels of factors A and B in treatment combination $z \in T = T_A \times T_B$ throughout unless specified otherwise. The $T_A$ and $T_B$ reduce to $\{0, 1\}$ under the $2^2$ split-plot design. Assume lexicographical order of $z$ for all vectors and matrices when applicable.

Let $M = N/W$ be the average whole-plot size, and let $\alpha_w = M_w/M$ be the whole-plot size factor for $w = 1, \ldots, W$. Let $Z_{ws} = (A_w, B_{ws}) \in T$ indicate the treatment received by sub-plot $ws$, with $\mathbb{P}(A_w = a) = W_a/W = p_a$, $\mathbb{P}(B_{ws} = b) = M_{wb}/M_w = q_{wb}$, and $p_{ws}(z) = \mathbb{P}(Z_{ws} = z) = p_a q_{wb}$. For $z = (ab)$, let $W(z) = \{w : A_w = a\}$ be the set of whole-plots that contain at least one observation under treatment $z$.

Let $\bar{Y}_w(z) = M_w^{-1} \sum_{s=1}^{M_w} Y_{ws}(z)$ and $U_w(z) = \bar{M}^{-1} \sum_{w=1}^{W} Y_{ws}(z)$ be the whole-plot average potential outcome and the scaled whole-plot total potential outcome for whole-plot $w$ under treatment $z$, respectively, with $\bar{Y}(z) = W^{-1} \sum_{w=1}^{W} \alpha_w \bar{Y}_w(z) = W^{-1} \sum_{w=1}^{W} U_w(z)$. Let $Y'_w(z) = Y_{ws}(z) - \bar{Y}(z)$ be the centered potential outcome with $\bar{Y}'_w(z) = \bar{Y}_w(z) - \bar{Y}(z)$ and $\bar{Y}'(z) = 0$.

Let $S_{ht} = S = (S(z, z'))_{z, z' \in T}$ and $S_{haj} = (S_{haj}(z, z'))_{z, z' \in T}$ be the $|T| \times |T|$ scaled between whole-plot covariance matrices of $\{Y_{ws}(z), Y_{ws}(z')\}_{ws \in S}$ and $\{Y'_{ws}(z), Y'_{ws}(z')\}_{ws \in S}$, respectively, with

$$S_{ht}(z, z') = S(z, z') = (W - 1)^{-1} \sum_{w=1}^{W} \{\alpha_w Y_w(z) - \bar{Y}(z)\}\{\alpha_w Y_w(z') - \bar{Y}(z')\},$$

$$S_{haj}(z, z') = (W - 1)^{-1} \sum_{w=1}^{W} \{\alpha_w Y_w(z) - \alpha_w \bar{Y}(z)\}\{\alpha_w Y_w(z') - \alpha_w \bar{Y}(z')\}. $$

A key observation is that $S(z, z')$ equals the finite-population covariance of $\{U_w(z), U_w(z')\}_{w=1}^{W}$. A
useful fact is

\[
\lambda_W \{S_{haj}(z, z') - S_{ht}(z, z')\} = \bar{Y}(z)\bar{Y}(z') \cdot (\alpha^2 + 1) - \bar{Y}(z') \cdot \alpha\bar{U}(z')
\]  

(S1)

for \( z, z' \in \mathcal{T} \), where \( \lambda_W = 1 - W^{-1} \), \( \alpha^2 = W^{-1} \sum_{w=1}^{W} \alpha_w^2 \), and \( \alpha\bar{U}(z) = W^{-1} \sum_{w=1}^{W} \alpha_w U_w(z) \).

Let \( \hat{Y}_w(z) = M_w^{-1} \sum_{s:Z_{ws}=z} Y_{ws} \) and \( \hat{U}_w(z) = \alpha_w \hat{Y}_w(z) \) be the sample analogs of \( \bar{Y}_w(z) \) and \( U_w(z) \), respectively, with \( \hat{Y}_w(z) = \bar{U}_w(z) = 0 \) for \( w \not\in \mathcal{W}(z) \). A key observation is that

\[
\hat{Y}_{ht}(z) = W_a^{-1} \sum_{w \in \mathcal{W}(z)} \hat{U}_w(z)
\]

\[
= W_a^{-1} \sum_{w \in \mathcal{W}(z)} U_w(z) + W_a^{-1} \sum_{w \in \mathcal{W}(z)} \{\hat{U}_w(z) - U_w(z)\}
\]

\[
= \mu(z) + \sum_{w=1}^{W} \delta_w(z) \quad \text{for} \quad z = (ab) \in \mathcal{T},
\]

where

\[
\mu(z) = W_a^{-1} \sum_{w \in \mathcal{W}(z)} U_w(z), \quad \delta_w(z) = 1(A_w = a) \cdot W_a^{-1}\{\hat{U}_w(z) - U_w(z)\};
\]

the randomness in \( \mu(z) \) comes solely from the stage (I) randomization. Let \( \mu = (\mu(z))_{z \in \mathcal{T}} \) and \( \delta_w = (\delta_w(z))_{z \in \mathcal{T}} \) be the vectorizations of \( \mu(z) \) and \( \delta_w(z) \), respectively, to write

\[
\hat{Y}_{ht} = \mu + \delta, \quad \text{where} \quad \delta = \sum_{w=1}^{W} \delta_w.
\]  

(S2)

Let \( \mathcal{A} = \sigma(A_1, \ldots, A_W) \) be the \( \sigma \)-algebra generated by \( (A_w)_{w=1}^{W} \). The independence between the stage (I) and stage (II) randomizations ensures that \( (\delta_w)_{w=1}^{W} \) are jointly independent conditioning on \( \mathcal{A} \). We have \( E(\delta_w \mid \mathcal{A}) = 0 \), \( E(\delta \mid \mathcal{A}) = 0 \), \( \text{cov}(\delta \mid \mathcal{A}) = \sum_{w=1}^{W} \text{cov}(\delta_w \mid A_w) \), and

\[
E(\delta_w) = 0; \quad E(\delta) = 0; \quad \text{cov}(\delta) = \sum_{w=1}^{W} \text{cov}(\delta_w) \quad \text{with} \quad \text{cov}(\delta_w) = E\{\text{cov}(\delta_w \mid \mathcal{A})\};
\]

\[
\text{cov}(\mu, \delta) = E\{\text{cov}(\mu, \delta \mid \mathcal{A})\} + E\{\text{cov}(\mu \mid \mathcal{A}), E(\delta \mid \mathcal{A})\} = 0.
\]  

(S3)

Further let \( U_{ws}(z) = \alpha_w Y_{ws}(z) \) be the scaled potential outcome at the unit level. We have \( U_w(z) = M_w^{-1} \sum_{w=1}^{W} U_{ws}(z) \) and \( S_w(z, z') = (M_w - 1)^{-1} \sum_{w=1}^{W} \{U_{ws}(z) - U_w(z)\}\{U_{ws}(z') - U_w(z')\} \) equal the finite-population mean and covariance of \( \{U_{ws}(z)\}_{w=1}^{W} \) in whole-plot \( w \), respectively, with \( \hat{U}_w(z) = M_w^{-1} \sum_{w, Z_{ws}=z} U_{ws}(z) \) as the corresponding sample mean under treatment \( z \). Let \( H_w(z, z') \) denote the \((z, z')\)th element of matrix \( H_w \). Standard result ensures

\[
\text{cov}(\hat{U}_w(z), \hat{U}_w(z') \mid A_w = a) = M_w^{-1}q_{w1}(z = z') - 1\}S_w(z, z')
\]

\[
= p_a M_w^{-1}H_w(z, z')S_w(z, z')
\]  

(S4)
with $H_w(z, z') = p^{-1}_a \{ q_{ab}^{-1}(z = z') - 1 \}$ for $z = (ab)$ and $z' = (ab')$ that share the same level of factor A. Let $\bar{U}_w^4(z) = M_w^{-1} \sum_{s=1}^{M_w} U_w^4(z) = \alpha_w^4 \bar{Y}_w^4(z)$ be the uncentered fourth moment of $U_{w,s}(z)$ in whole-plot $w$.

Last but not least, for arbitrary $L \times 1$ vectors $u = (u_1, \ldots, u_L)^t$ and $v = (v_1, \ldots, v_L)^t$, we have

$$(u^tv)^4 \leq \|u\|^2_2 \|v\|^2_2; \quad \bar{u}^4 \leq (\bar{u}^2)^2 \leq \bar{u}^4,$$

where $\bar{u}^k = L^{-1} \sum_{i=1}^L u_i^k$, (S5)

due to the Cauchy–Schwarz inequality. Setting $L = 2$ in $\bar{u}^4 \leq \bar{u}^4$ ensures $(u_1 + u_2)^4 \leq 8u_1^4 + 8u_2^4$.

S1.2. Established lemmas

Lemma S1. ([Li and Ding 2017], Theorems 3 and 5) In a completely-randomized experiment with $N$ units and $Q$ treatment groups of sizes $N_q$ ($q = 1, \ldots, Q$), let $Y_i(q)$ be the $L \times 1$ vector potential outcome of unit $i$ under treatment $q$, and let $S_{qq'} = (N-1)^{-1} \sum_{i=1}^N \{ Y_i(q) - \bar{Y}(q) \} \{ Y_i(q') - \bar{Y}(q') \}^t$ be the finite-population covariance for $1 \leq q, q' \leq Q$. Let $\tau = \sum_{q=1}^Q G_q \bar{Y}(q)$ be the finite-population average treatment effect of interest, and let $\hat{\tau} = \sum_{q=1}^Q G_q \bar{Y}(q)$ be the corresponding moment estimator with $\bar{Y}(q) = N_q^{-1} \sum_{i,Z_i=q} Y_i$, where $Z_i$ and $Y_i$ are the treatment indicator and observed outcome for unit $i$, respectively. We have

$$\text{cov}(\hat{\tau}) = \sum_{q=1}^Q N_q^{-1} G_q S_{qq} G_q^t - N^{-1} S_\tau^2,$$

where $S_\tau^2$ is the finite-population covariance of $\tau_i = \sum_{q=1}^Q G_q Y_i(q)$ for $i = 1, \ldots, N$. Further assume that for all $1 \leq q, q' \leq Q$, (i) $S_{qq'}$ has a finite limit, (ii) $r_q = N_q/N$ has a limit in $(0,1)$, and (iii) $\max_{1 \leq i \leq N} |Y_i(q) - \bar{Y}(q)|^2/N = o(1)$. We have $\sqrt{N}(\hat{\tau} - \tau) \rightsquigarrow N(0, V)$ with $V$ denoting the limiting value of $N\text{cov}(\hat{\tau})$.

Lemma S2. ([Ohlsson 1989], Theorem A.1) For $W = 1, 2, 3, \ldots$, let $\{\xi_{W,w} : w = 1, \ldots, W\}$ be a martingale difference sequence relative to the filtration $\{\mathcal{F}_{W,w} : w = 0, \ldots, W\}$, and let $X_W$ be an $\mathcal{F}_{W,0}$-measurable random variable. Set $\xi_W = \sum_{w=1}^W \xi_{W,w}$. Suppose that the following three conditions are fulfilled as $W \to \infty$.

(i) $\sum_{w=1}^W E(\xi_{W,w}^4) = o(1)$.

(ii) For some sequence of non-negative real numbers $\{\beta_W : W = 1, 2, 3, \ldots\}$ with $\sup_W \beta_W < \infty$, we have $E[\{ \sum_{w=1}^W E(\xi_{W,w}^4 | \mathcal{F}_{W,w-1}) - \beta_W^2 \}] = o(1)$.

(iii) $\mathcal{L}(X_W) \ast \mathcal{N}(0, \beta_W^2) \rightsquigarrow \mathcal{L}_0$ for some probability distribution $\mathcal{L}_0$, where $\ast$ denotes convolution.

Then we have $\mathcal{L}(X_W + \xi_W) \rightsquigarrow \mathcal{L}_0$ as $W \to \infty$. 

S3
S1.3. New lemmas

We give in this subsection the key lemmas for verifying the results under the $2^2$ split-plot design. The lemmas and their proofs extend to the general $T_\lambda \times T_\beta$ design with minimal modification.

**Decomposition of $\text{cov}(\hat{Y}_{ht})$ in finite samples.** Lemma S3 separates the parts in $\text{cov}(\hat{Y}_{ht})$ that are due to $\mu$ and $\delta$ from (S20), respectively. The decomposition furnishes an alternative proof of Lemma 1 relative to [Mukerjee and Dasgupta (2019)].

**Lemma S3.** Under the $2^2$ split-plot randomization, we have

$$\text{cov}(\hat{Y}_{ht}) = \text{cov}(\mu) + \text{cov}(\delta)$$

with $\text{cov}(\mu) = W^{-1}(H \circ S_{ht})$ and $\text{cov}(\delta) = \sum_{w=1}^{W} \text{cov}(\delta_w) = W^{-1}\Psi$, where $\text{cov}(\delta_w) = W^{-2}M_w^{-1}(H_w \circ S_w)$.

**Proof of Lemma S3.** Identities $\text{cov}(\hat{Y}_{ht}) = \text{cov}(\mu) + \text{cov}(\delta)$ and $\text{cov}(\delta) = \sum_{w=1}^{W} \text{cov}(\delta_w)$ follow from (S3). We verify below the analytic forms of $\text{cov}(\mu)$ and $\text{cov}(\delta_w)$, respectively.

For the analytic form of $\text{cov}(\mu)$, define $U_w(a) = (U_w(a0), U_w(a1))^T$ as the vector potential outcomes of whole-plot $w$ under $A_w = a \in \{0,1\}$ with means $\bar{U}(a) = W^{-1}\sum_{w=1}^{W} U_w(a) = (\bar{Y}(a0), \bar{Y}(a1))^T$ and covariances

$$S_{ht}(a) = (W - 1)^{-1} \sum_{w=1}^{W} \{U_w(a) - \bar{U}(a)\}^2 = \begin{pmatrix} S_{ht}(a0,a0) & S_{ht}(a0,a1) \\ S_{ht}(a1,a0) & S_{ht}(a1,a1) \end{pmatrix}.$$

Direct comparison shows that $\mu$ equals the sample analog of $\bar{Y} = (\bar{U}(0)^{T}, \bar{U}(1)^{T})^{T} = (I_2, 0_{2 \times 2})^{T} \bar{U}(0) + (0_{2 \times 2}, I_2)^{T} \bar{U}(1)$ with regard to the stage (I) randomization. It then follows from Lemma S1 that

$$\text{cov}(\mu) = W_0^{-1} \begin{pmatrix} I_2 \\ 0_{2 \times 2} \end{pmatrix} S_{ht}(0) (I_2, 0_{2 \times 2}) + W_1^{-1} \begin{pmatrix} 0_{2 \times 2} \\ I_2 \end{pmatrix} S_{ht}(1) (0_{2 \times 2}, I_2) - W^{-1}S_{ht}$$

$$= W^{-1}(H \circ S_{ht}).$$

For the analytic form of $\text{cov}(\delta_w)$, recall $\hat{U}_w(z)$ as the sample mean of the scaled potential outcomes $\{U_{ws}(z) : Z_{ws} = z\}$ in whole-plot $w$. It follows from the definition of $\delta_w$ and (S4) that

$$\text{cov}(\delta_w | A_w = a) = W_a^{-2} p_a M_w^{-1} \{H_w(a) \circ S_w\} = p_a^{-1}W^{-2}M_w^{-1} \{H_w(a) \circ S_w\}$$

with $H_w(0) = \text{diag}(p_0^{-1}, 0) \otimes \{\text{diag}(q_{w0}^{-1}, q_{w1}^{-1}) - I_{2 \times 2}\}$ and $H_w(1) = \text{diag}(0, p_1^{-1}) \otimes \{\text{diag}(q_{w0}^{-1}, q_{w1}^{-1}) - I_{2 \times 2}\}$ “extracting” the upper-left and lower-right $2 \times 2$ block matrices of $H_w$, respectively. It then follows from (S3) and $H_w(0) + H_w(1) = H_w$ that $\text{cov}(\delta_w) = E\{\text{cov}(\delta_w | A_w)\} = \sum_{a=0,1} P(A_w = a) \cdot \text{cov}(\delta_w | A_w = a) = W^{-2}M_w^{-1}(H_w \circ S_w).$
The weak law of large numbers. We give in this part the weak law of large numbers for quantifying the probability limits of the Horvitz-Thompson estimator and the sample covariances $\hat{V}_s$ ($*=ht, haj$) under some weaker conditions than Condition 4, summarized in Condition S1.

**Condition S1.** As $W$ goes to infinity, for all $a, b = 0, 1$, and $z \in \mathcal{T}$,

(i) $p_a$ has a limit in $(0, 1)$; $\varepsilon \leq \min_{w=1,\ldots,W} q_{wb} \leq \max_{w=1,\ldots,W} q_{wb} \leq 1 - \varepsilon$ for some $\varepsilon \in (0, 1/2]$ independent of $W$;

(ii) $\bar{Y}$ has a finite limit; $S = O(1)$ and $\Psi = O(1)$;

(iii) $W^{-2} \sum_{w=1}^{W} \alpha^2_w Y^2_w(z) = o(1)$.

Condition 4 ensures that $\{Y_{ws}(z)\}_{ws \in S}$, $\{Y'_{ws}(z)\}_{ws \in S}$, and the finite population of all ones, namely $\{Y_{ws}(z)=1 : ws \in S, z \in \mathcal{T}\}$, all satisfy Condition S1(i)-(iii).

**Lemma S4.** Assume split-plot randomization. We have

(i) $\hat{Y}_{ht} - \bar{Y} = o_{P}(1)$ provided Condition S1(i)-(ii);

(ii) $\hat{I}_{ht} \equiv \text{diag}\{\hat{I}_{ht}(z)\}_{z \in \mathcal{T}} = I_{|\mathcal{T}|} + o_{P}(1)$ provided Condition S1(i) and $\alpha^2 = O(1)$.

**Proof of Lemma S4.** Standard result ensures $E(\hat{Y}_{ht}) = \bar{Y}$. Lemma II ensures $\text{cov}(\hat{Y}_{ht}) = W^{-1}(H \circ S_{ht} + \Psi) = o(1)$ under Condition S1(iii). The result for $\hat{Y}_{ht}$ then follows from Markov’s inequality. The result for $\hat{I}_{ht}$ follows from applying statement (ii) to the finite population of all ones.

**Lemma S5.** Assume split-plot randomization and Condition S1(i) and (iii). We have

(i) $W^{-2} \sum_{w=1}^{W} E(\hat{U}^2_w(z)\hat{U}^2_w(z') \mid A_w = a) = o(1)$ for $z = (ab)$ and $z' = (ab')$ with the same level of factor $A$;

(ii) $W^2 \sum_{w=1}^{W} E(\|\delta_w\|_2^4 \mid A_w = a) = o(1)$; $W^2 \sum_{w=1}^{W} E(\|\delta_w\|_2^4) = o(1)$.

**Proof of Lemma S4.** For statement (i), it follows from $\hat{U}_w(z) = M^{-1}_{wb} \sum_{s:Z_{ws}=z} U_{ws}(z)$ and (S7) that

$$\hat{U}^4_w(z) \leq M^{-1}_{wb} \sum_{s:Z_{ws}=z} U^4_{ws}(z) \leq M^{-1}_{wb} \sum_{s=1}^{M_w} U^4_{ws}(z) = q_{wb}^{-1} \cdot \bar{U}^4_w(z),$$

recalling $\bar{U}^4_w(z) = M^{-1}_{w} \sum_{s=1}^{M_w} U^4_{ws}(z)$. This ensures $E\{\hat{U}^4_w(z) \mid A_w = a\} \leq \varepsilon^{-1} \cdot \bar{U}^4_w(z)$ and thus $W^{-2} \sum_{w=1}^{W} E\{\hat{U}^4_w(z) \mid A_w = a\} \leq W^{-2} \varepsilon^{-1} \sum_{w=1}^{W} \bar{U}^4_w(z) = o(1)$ by Condition S1(i) and (iii). The result follows from $\hat{U}^2_w(z)\hat{U}^2_w(z') \leq 2^{-1} \{\hat{U}^2_w(z) + \hat{U}^2_w(z')\}$.

For statement (ii), it suffices to verify the first equality, namely

$$W^2 \sum_{w=1}^{W} E(\|\delta_w\|_2^4 \mid A_w = a) = o(1).$$

(S7)
The second equality \( W^2 \sum_{w=1}^{W} E(\| \delta_w \|_2^2) = o(1) \) then follows from \( E(\| \delta_w \|_2^2) = p_0 E(\| \delta_w \|_2^2 | A_w = 0) + p_1 E(\| \delta_w \|_2^2 | A_w = 1) \) by the law of total expectation.

To verify (S7), note that \( \| \delta_w \|_2^2 = W^{-1} \left\{ \tilde{U}_w(a1) - U_w(a1) \right\}^2 + \{ \tilde{U}_w(a0) - U_w(a0) \}^2 \) for \( w \) with \( A_w = a \) such that
\[
\| \delta_w \|_2^2 = (\| \delta_w \|_2^2)^2 \leq 2W_a^{-4} \left\{ \tilde{U}_w(a1) - U_w(a1) \right\}^4 + 2W_a^{-4} \left\{ \tilde{U}_w(a0) - U_w(a0) \right\}^4
\]
by the Cauchy–Schwarz inequality. A sufficient condition for (S7) is thus
\[
\sum_{w=1}^{W} E\left[ \{ \tilde{U}_w(z) - U_w(z) \}^4 | A_w = a \right] = o(W^2) \quad \text{for } z = (ab).
\]

With \( \sum_{w=1}^{W} E[\{ \tilde{U}_w(z) - U_w(z) \}^4 | A_w = a] \leq 8 \sum_{w=1}^{W} E\{ \tilde{U}_w^4(z) | A_w = a \} + 8 \sum_{w=1}^{W} U_w^4(z) \) by \( \{ \tilde{U}_w(z) - U_w(z) \}^4 \leq \tilde{U}_w^4(z) + 8U_w^4(z) \), (S8) is guaranteed by statement (S9) and \( W^{-2} \sum_{w=1}^{W} U_w^4(z) \leq W^{-2} \sum_{w=1}^{W} \tilde{U}_w^4(z) = o(1) \) by (S8) and Condition S1(S1).

Let
\[
\hat{T}_{z,z'} = W_a^{-1} \sum_{w:A_w=a} \tilde{U}_w(z)\tilde{U}_w(z')
\]
be the sample analog of \( U(z)U(z') = W^{-1} \sum_{w=1}^{W} U_w(z)U_w(z') \) for \( z = (ab) \) and \( z' = (ab') \) with the same level of factor A. Let \( \Psi(z, z') = W^{-1} \sum_{w=1}^{W} M_w^{-1} H_w(z, z')S_w(z, z') \) be the \((z, z')\)th element of \( \Psi \). Lemma S6 gives the explicit form of \( E(\hat{T}_{z,z'}) \) in terms of \( U(z)U(z') \) and \( \Psi(z, z') \), affording the basis for computing the probability limits of \( \hat{V}_{ht} \) and \( \hat{V}_{haj} \).

**Lemma S6.** Assume split-plot randomization and Condition S1. We have
\[
\hat{T}_{z,z'} - E(\hat{T}_{z,z'}) = o_2(1) \quad \text{for } z = (ab), z' = (ab')
\]
with \( E(\hat{T}_{z,z'}) = U(z)U(z') + p_a \Psi(z, z') = (1 - W^{-1})S_{ht}(z, z') + \hat{Y}(z)\hat{Y}(z') + p_a \Psi(z, z') \).

**Proof of Lemma S6.** Let \( X_w = 1(A_w = a) \cdot \tilde{U}_w(z)\tilde{U}_w(z') \) to write \( \hat{T}_{z,z'} = W_a^{-1} \sum_{w=1}^{W} X_w \). Markov’s inequality ensures that it suffices to verify the explicit form of \( E(\hat{T}_{z,z'}) \) and \( \text{cov}(\hat{T}_{z,z'}) = o(1) \).

For the explicit form of \( E(\hat{T}_{z,z'}) \), it follows from (S8) that
\[
\mu_w \equiv E(X_w | A_w = a) = E(\tilde{U}_w(z)\tilde{U}_w(z') | A_w = a) = \text{cov}(\tilde{U}_w(z), \tilde{U}_w(z') | A_w = a) + E(\tilde{U}_w(z) | A_w = a) \cdot E(\tilde{U}_w(z') | A_w = a) = p_a M_w^{-1} H_w(z, z')S_w(z, z') + U_w(z)U_w(z'),
\]
\[
E(X_w) = E\{E(X_w | A_w)\} = p_a E(X_w | A_w = a) = p_a \mu_w.
\]
This verifies

\[E(T_{z,z'}) = W_a^{-1} \sum_{w=1}^W E(X_w) = W_a^{-1} \sum_{w=1}^W \mu_w = W_a^{-1} \sum_{w=1}^W U_w(z) U_w(z') + p_a \Psi(z, z') \]

\[= (1 - W^{-1}) S_{ht}(z, z') + \tilde{Y}(z) \tilde{Y}(z') + p_a \Psi(z, z'); \]

the last equality follows from \(W_a^{-1} \sum_{w=1}^W U_w(z) U_w(z') = (1 - W^{-1}) S_{ht}(z, z') + \tilde{Y}(z) \tilde{Y}(z'). \)

For the limit of \( \text{cov}(\hat{T}_{z,z'}) = W_a^{-2} \sum_{w=1}^W \text{cov}(X_w) + \sum_{w \neq k} \text{cov}(X_w, X_k) \}

\[
E\{\text{cov}(X_w \mid A_w)\} = p_a \text{cov}(X_w \mid A_w = a) = p_a E(X_w^2 \mid A_w = a) - p_a \mu_w^2, \\
\text{cov}\{E(X_w \mid A_w)\} = E[\{E(X_w \mid A_w)\}^2] - \{E(X_w)\}^2 \\
= p_a \{E(X_w \mid A_w = a)\}^2 - p_a \mu_w^2 = p_a \mu_w^2 - p_a \mu_w^2, \\
E\{E(X_w \mid A_w) \cdot E(X_k \mid A_k)\} = \mathbb{P}(A_w = A_k = a) \cdot E(X_w \mid A_w = a) \cdot E(X_k \mid A_k = a) \\
= \frac{W_a - 1}{W - 1} \mu_w \mu_k, \\
E(X_w)E(X_k) = \frac{p_a^2}{W_a} \mu_w \mu_k
\]

by (S9) that

\[
\text{cov}(X_w) = E\{\text{cov}(X_w \mid A_w)\} + \text{cov}\{E(X_w \mid A_w)\} \\
= p_a E(X_w^2 \mid A_w = a) - p_a \mu_w^2, \\
\text{cov}(X_w, X_k) = \text{cov}\{E(X_w \mid A_w), E(X_k \mid A_k)\} + E\{\text{cov}(X_w, X_k \mid A_w, A_k)\} \\
= \text{cov}\{E(X_w \mid A_w), E(X_k \mid A_k)\} \\
= E\{E(X_w \mid A_w) \cdot E(X_k \mid A_k)\} - E(X_w)E(X_k) \\
= -p_0 p_1 (W - 1)^{-1} \mu_w \mu_k.
\]

This ensures

\[
W_a^2 \text{cov}(\hat{T}_{z,z'}) = \sum_{w=1}^W \text{cov}(X_w) + \sum_{w \neq k} \text{cov}(X_w, X_k) \\
= p_a \sum_{w=1}^W E(X_w^2 \mid A_w = a) - p_a \sum_{w=1}^W \mu_w^2 p_0 p_1 \cdot \frac{1}{W - 1} \sum_{w \neq k} \mu_w \mu_k \\
= p_a \sum_{w=1}^W E(X_w^2 \mid A_w = a) - p_a \sum_{w=1}^W \mu_w^2 + p_0 p_1 \cdot \frac{1}{W - 1} \sum_{w=1}^W \mu_w^2 - p_0 p_1 \cdot \frac{1}{W - 1} \sum_{w,k} \mu_w \mu_k \\
\leq p_a \sum_{w=1}^W E(X_w^2 \mid A_w = a) - \left( p_a^2 - p_0 p_1 \cdot \frac{1}{W - 1} \right) \sum_{w=1}^W \mu_w^2 \\
\leq \infty p_a \sum_{w=1}^W E(X_w^2 \mid A_w = a) = p_a \sum_{w=1}^W E\{\hat{U}_w^2(z) \hat{U}_w^2(z') \mid A_w = a\},
\]

S7
where \( \leq_{\infty} \) indicates less or equal to as \( W \to \infty \). The result then follows from Lemma S2(3). \( \square \)

S1.4. Proof of the main results

**Proof of Theorem 1.** We first prove the result for \( \hat{Y}_{ht} \), and then use it to verify that for \( \hat{Y}_{haj} \).

**Asymptotic Normality of \( \hat{Y}_{ht} \).** Recall that \( \hat{Y}_{ht} = \mu + \delta \) from (S2). The Cramer–Wold device ensures that \( \sqrt{W}(\hat{Y}_{ht} - \bar{Y}) \sim \mathcal{N}(0, \Sigma_{ht}) \) as long as

\[
\eta^T \sqrt{W}(\hat{Y}_{ht} - \bar{Y}) = \eta^T \sqrt{W}(\mu - \bar{Y} + \delta) \sim \mathcal{N}(0, \eta^T \Sigma_{ht} \eta)
\]

(S10)

for arbitrary non-random \( 4 \times 1 \) unit vector \( \eta \).

Let \( X = \eta^T \sqrt{W}(\mu - \bar{Y}) \) and \( \xi_w = \eta^T \sqrt{W} \delta_w \) to write the right-hand side of (S10) as \( X + \xi \), where \( \xi = \sum_{w=1}^{W} \xi_w = \sqrt{W} \eta^T \delta \). Let \( F_{W,0} = A = \sigma(A_1, \ldots, A_W) \) be the \( \sigma \)-algebra generated by \( (A_w)_w=1^W \), and let \( F_{W,w} \) be the \( \sigma \)-algebra generated by \( (A_w)_w=1^W \) and \( \{ (B_{vw})_{s=1}^{M_v} : v = 1, \ldots, w \} \) for \( w = 1, \ldots, W \). We verify the sufficient condition (S11) by checking that \( (\xi_w)_w=1^W \) and \( X \) satisfy the conditions of Lemma S2 with \( \beta_w^2 = \eta^T \Psi \eta \) with regard to filtration \( \{ F_{W,w} : w = 0, \ldots, W \} \).

Technically, \( X = X_W, \xi_w = \xi_{W,w}, \) and \( \Psi = \Psi_W \) all depend on \( W \); we suppress the \( W \) in the subscripts when no confusion would arise.

First, \( F_{W,0} \subset F_{W,1} \subset \cdots \subset F_{W,W} \) such that \( \{ F_{W,w} : w = 0, \ldots, W \} \) is indeed a filtration. Intuitively, \( F_{W,0} \) contains all the information on the stage (I) cluster randomization, whereas \( F_{W,w} \) contains all the information on the stage (I) cluster randomization plus the subset of stage (II) stratified randomization in the first \( w \) whole-plots, \( v = 1, \ldots, w \).

For Lemma S2 condition (ii), (S11) ensures

\[
\xi_w^4 = W^2(\eta^T \delta_w)^4 \leq W^2 \eta^2 \| \eta \|_2^4 \cdot \| \delta_w \|_2^4 = W^2 \| \delta_w \|_2^4.
\]

(S11)

The result follows from \( \sum_{w=1}^{W} E(\xi_w^4) \leq W^2 \sum_{w=1}^{W} E(\| \delta_w \|_2^4) = o(1) \) by Lemma S2(3).

For Lemma S2 condition (i), it follows from (S3) and Lemma S2 that \( E(\xi | F_{W,0}) = 0 \) and \( \text{var}(\xi) = \eta^T \Psi \eta = \beta_w^2 \). This, together with \( E(\xi_w^2 | F_{W,w-1}) = E(\xi_w^2 | F_{W,0}) = \text{var}(\xi_w | F_{W,0}) \), ensures

\[
\sum_{w=1}^{W} E(\xi_w^2 | F_{W,w-1}) = \sum_{w=1}^{W} \text{var}(\xi_w | F_{W,0}) = \text{var}(\xi | F_{W,0}) \equiv \sigma, \text{ and}
\]

\[
\beta_w^2 = \text{var}(\xi) = E\{ \text{var}(\xi | F_{W,0}) \} + \text{var}\{ E(\xi | F_{W,0}) \} = E(\sigma)
\]

such that \( E\left[ \sum_{w=1}^{W} E(\xi_{w}^2 | F_{W,w-1}) - \beta_w^2 \right]^2 = E\left[ (\sigma - E(\sigma))^2 \right] = \text{var}(\sigma) \). Lemma S2 condition (i) is thus equivalent to

\[
\text{var}(\sigma) = o(1).
\]

(S12)

To verify (S12), let

\[
\sigma_w = \text{var}(\xi_w | F_{W,0}) = \text{var}(\xi_w | A_w) = W \eta^T \text{cov}(\delta_w | A_w) \eta
\]
to write $\sigma = \sum_{w=1}^W \sigma_w$. Treat $\sigma_w$ as the observed outcome from

$$
\sigma_w(a) = E(\xi_w^2 \mid A_w = a) = W \eta^T \text{cov} (\delta_w \mid A_w = a) \eta
$$

(S13)

with mean $\bar{\sigma}(a) = W^{-1} \sum_{w=1}^W \sigma_w(a)$, variance $S_{\bar{\sigma}(a)}^2 = (W - 1)^{-1} \sum_{w=1}^W (\sigma_w(a) - \bar{\sigma}(a))^2$, and sample mean $\tilde{\sigma}(a) = W_a^{-1} \sum_{w : A_w = a} \sigma_w(a)$ for $a = 0, 1$. Standard result ensures $\text{var}\{\tilde{\sigma}(a)\} = W^{-1} p_a^{-1} (1 - p_a) S_{\bar{\sigma}(a)}^2$ such that, with $\sigma = W_0 \tilde{\sigma}_w(0) + W_1 \tilde{\sigma}_w(1)$, we have

$$
\text{var}(\sigma) = \text{var}\{W_0 \tilde{\sigma}_w(0) + W_1 \tilde{\sigma}_w(1)\}
\leq 2 \text{var}\{W_0 \tilde{\sigma}_w(0)\} + 2 \text{var}\{W_1 \tilde{\sigma}_w(1)\} = 2 W p_0 p_1 (S_{\sigma(0)}^2 + S_{\sigma(1)}^2).
$$

(S14)

Sufficient condition (S12) thus holds as long as $S_{\sigma(1)}^2 = o(W^{-1})$.

Indeed, with

$$
\{\sigma_w(a)\}^2 = \{E(\xi_w^2 \mid A_w = a)\}^2 \leq E(\xi_w^4 \mid A_w = a) \leq W^2 E(\|\delta_w\|_2^4 \mid A_w = a)
$$

by Jensen’s inequality and (S11), we have

$$
(W - 1)^{-1} \sum_{w=1}^W \{\sigma_w(a)\}^2 \leq (W - 1)^{-1} W^2 \sum_{w=1}^W E(\|\delta_w\|_2^4 \mid A_w = a) = o(W^{-1})
$$

by Lemma S5(ii). This, together with

$$
\tilde{\sigma}(a) = \eta^T \sum_{w=1}^W \text{cov}(\delta_w \mid A_w = a) \eta = \eta^T \left\{ p_a^{-1} W^{-2} \sum_{w=1}^W M_a^{-1} H_w(\alpha) \circ S_w \right\} \eta = o(W^{-1})
$$

by (S13) and (S10), ensures that $S_{\tilde{\sigma}(a)}^2 = (W - 1)^{-1} \sum_{w=1}^W \{\sigma_w(a)\}^2 - \lambda_a^{-1} \{\tilde{\sigma}(a)\}^2 = o(W^{-1})$. Plugging this in (S11) verifies (S12) and hence Lemma S2 condition (ii).

Lemma S2 condition (iii) then follows from Lemmas S1 and S2 which ensure $\sqrt{W}(\mu - \hat{Y}) \Rightarrow \mathcal{N}(0, H \circ \hat{S}_T)$ under Condition $\mathfrak{A}$. The convolution of $\mathcal{L}(X)$ with $\mathcal{N}(0, \eta^T \Psi \eta)$ thus converges in distribution to $\mathcal{N}(0, \eta^T \Sigma_{\mathfrak{A}} \eta)$.

This verifies that $(\xi_{wa})_{w=1}^W$ and $X$ satisfy the conditions of Lemma S2. The sufficient condition (S11) then follows from Lemma S2 and ensures the result for $\hat{Y}_{\mathfrak{A}}$.

**Asymptotic Normality of $\hat{Y}_{\mathfrak{A}}$.** We have $\hat{Y}_{\mathfrak{A}} - Y = \hat{Y}_{\mathfrak{A}} - Y'_{\mathfrak{A}}$ with $\hat{Y}_{\mathfrak{A}} = \text{diag}\{\hat{I}_{\mathfrak{A}}(z)\}_{z \in \mathcal{T}}$ and $Y'_{\mathfrak{A}}$ as the vectorization of $\{Y'_{\mathfrak{A}}(z)\}_{z \in \mathcal{T}}$. The result for $\hat{Y}_{\mathfrak{A}}$ then follows from the asymptotic Normality of $Y'_{\mathfrak{A}}$, $\hat{I}_{\mathfrak{A}} = I_{\mathcal{T}} + o_p(1)$, and Slutsky’s theorem.

**Proof of Corollary $\mathfrak{A}$.** By (S11), we have

$$
(W - 1)\{S_{\mathfrak{A}}(z, z) - S_T(z, z)\} = \{\tilde{Y}(z)\}^2 \left(\sum_{w=1}^W \alpha_w^2 + W\right) - 2 \tilde{Y}(z) \sum_{w=1}^W \alpha_w^2 \tilde{Y}_w(z).
$$
When $\tilde{Y}_w(z) = c$ for all $w = 1, \ldots, W$, we have $\tilde{Y}(z) = W^{-1} \sum_{w=1}^{W} \alpha_w \tilde{Y}_w(z) = c$ such that $(W - 1) \{S_{\text{haj}}(z, z) - S_{\text{ht}}(z, z)\} = c^2(W - \sum_{w=1}^{W} \alpha_w^2) \leq 0$; the equality holds if and only if $\alpha_w = 1$ for all $w$ or $c = 0$.

When $U_w(z) = c$ for all $w = 1, \ldots, W$, we have $\tilde{Y}(z) = W^{-1} \sum_{w=1}^{W} U_w(z) = c$ such that $(W - 1) \{S_{\text{haj}}(z, z) - S_{\text{ht}}(z, z)\} = c^2 \sum_{w=1}^{W} (\alpha_w - 1)^2 \geq 0$; the equality holds if and only if $\alpha_w = 1$ for all $w$ or $c = 0$.

Proof of Theorem \ref{thm:main}. Let $\hat{V}_s(z, z')$ be the $(z, z')$th element of $\hat{V}_s$ for $* = \text{ht, haj}$.

For $\hat{V}_{\text{ht}}$, standard algebra shows that

$$\hat{V}_{\text{ht}}(z, z') = W_a^{-1} \hat{S}_{\text{ht}}(z, z') = (W_a - 1)^{-1} \{(\hat{T}_{z,z'}) - \hat{Y}_\text{ht}(z) \hat{Y}_\text{ht}(z')\}$$

for $z = (ab)$ and $z' = (ab')$ with the same level of factor A. It then follows from Lemmas \ref{lem:ht-1} and \ref{lem:ht-2} that

$$W \hat{V}_{\text{ht}}(z, z') = p_a^{-1} \{E(\hat{T}_{z,z'}) - \hat{Y}(z) \hat{Y}(z')\} + o_P(1)$$

$$= p_a^{-1} \{S_{\text{ht}}(z, z') + p_a \Psi(z, z')\} + o_P(1)$$

$$= \Sigma_{\text{ht}}(z, z') + S_{\text{ht}}(z, z') + o_P(1),$$

where $\Sigma_{\text{ht}}(z, z') = (p_a^{-1} - 1)S_{\text{ht}}(z, z') + \Psi(z, z')$ is the $(z, z')$th element of $\Sigma_{\text{ht}}$. This verifies the result for $\hat{V}_{\text{ht}}$.

For $\hat{V}_{\text{haj}}$, standard algebra shows that

$$(W_a - 1) \hat{V}_{\text{haj}}(z, z') = \hat{T}_{z,z'} + \hat{Y}_{\text{haj}}(z) \hat{Y}_{\text{haj}}(z') \left( W_a^{-1} \sum_{w:A_w = a} \alpha_w^2 \right)$$

$$- \hat{Y}_{\text{haj}}(z) \left\{ W_a^{-1} \sum_{w:A_w = a} \alpha_w \hat{U}_w(z') \right\}$$

$$- \hat{Y}_{\text{haj}}(z') \left\{ W_a^{-1} \sum_{w:A_w = a} \alpha_w \hat{U}_w(z) \right\}$$

(S15)

for $z = (ab)$ and $z' = (ab')$. We proceed to compute the probability limit of (S15).

First, Condition \ref{cond:finite} ensures that the finite population of $\{Y_{ws}(z), Y_{ws}(z')\}_{ws \in S}$ with $Y_{ws}(z') = 1$ for all $ws \in S$ satisfies Condition \ref{cond:finite}. Let $Y_{ws}(z') = 1$ for all $ws \in S$ in Lemma \ref{lem:finite} to see

$$W_a^{-1} \sum_{w:A_w = a} \alpha_w \hat{U}_w(z) - \bar{o}(z) = o_P(1).$$

This, together with $\bar{o}(z) = O(1)$ by $|\bar{o}(z)| = W^{-1} |\sum_{w=1}^{W} \alpha_w \{U_w(z) - \bar{Y}(z)\}| \leq \bar{z}^2$. 

S10
\( S_{ht}(z, z) \), ensures

\[
\hat{Y}_{ht}(z') \left\{ W_a^{-1} \sum_{w: A_w = a} \alpha_w \hat{U}_w(z) \right\} - \bar{Y}(z') \cdot \overline{\alpha U(z)} = o_p(1). \tag{S16}
\]

Likewise for \( W_a^{-1} \sum_{w: A_w = a} \alpha_w^2 - \overline{\alpha^2} = o_p(1) \) by letting \( Y_{ws}(z) = Y_{ws}(z') = 1 \) for all \( ws \in S \) in Lemma S6. This, together with \( \overline{\alpha^2} = O(1) \) by Condition (i), ensures

\[
\hat{Y}_{ht}(z) \hat{Y}_{ht}(z') \left( W_a^{-1} \sum_{w: A_w = a} \alpha_w^2 \right) - \bar{Y}(z) \bar{Y}(z') \cdot \overline{\alpha^2} = o_p(1). \tag{S17}
\]

Plug (S16)–(S17) and the probability limit of \( \hat{T}_{z,z'} \) from Lemma S6 in (S13) to see

\[
(W_a - 1)\hat{V}_{ht}(z, z') = \{ S_{ht}(z, z') + \bar{Y}(z) \bar{Y}(z') + p_a \Psi(z, z') \} + (1 - W^{-1}) \{ S_{ht}(z, z') - S_{ht}(z, z') \} + o_p(1)
\]

by (S11). This ensures \( W_a \hat{V}_{ht}(z, z') = p_a^{-1} S_{ht}(z, z') + \Psi(z, z') + o_p(1) \) such that the result follows from \( p_a^{-1} S_{ht}(z, z') = H(z, z') S_{ht}(z, z') + S_{ht}(z, z') \).

\[ \square \]

S2. Reconciliation with model-based inference

S2.1. Notation and useful facts

Assume \( \pi_{ws} = N^{-1} \{ p_{ws}(Z_{ws}) \}^{-1} \) as the weight for \( Y_{ws} \) under fitting scheme “wls”. It differs from the original weight, \( \{ p_{ws}(Z_{ws}) \}^{-1} \), by a constant factor of \( N^{-1} \) and thus does not affect the result of the least-squares fit. The Horvitz–Thompson estimators satisfy

\[
\hat{Y}_{ht}(z) = \sum_{ws \in S(z)} \pi_{ws} Y_{ws}, \quad \hat{1}_{ht}(z) = \sum_{ws \in S(z)} \pi_{ws}
\]

with

\[
\pi_{ws} = N^{-1} p_a^{-1} q_w^{-1} = \alpha_w W_a^{-1} M_w^{-1} \quad \text{for } ws \in S(z) \tag{S18}
\]

under split-plot randomization.

Let \( Y \) and \( U \) be the vectorizations of \( \{ Y_{ws} : ws \in S \} \) and \( \{ \hat{U}_w(A_w b) : w = 1, \ldots, W; b = 0, 1 \} \), respectively, as the vectors of outcomes in matrix forms of (iT) and (iQ). Assume lexicographical orders of \( ws \) and \( (w, b) \) throughout unless specified otherwise.

Recall \( d_{ws} = (1(Z_{ws} = 00), 1(Z_{ws} = 01), 1(Z_{ws} = 10), 1(Z_{ws} = 11))^T \) and \( d_w(A_w b) = (1(A_w b = 00), 1(A_w b = 01), 1(A_w b = 10), 1(A_w b = 11))^T \) as the vectors of regressors in (iT) and (iQ). Let \( D \) and \( D_{ag} \) be the corresponding design matrices, concatenating \( \{ d_{ws} : ws \in S \} \) and \( \{ d_w(A_w b) : \)}
where \( w = 1, \ldots, W; \ b = 0, 1 \) in the same orders of \( ws \) and \( (w, b) \) as in \( Y \) and \( U \), respectively. Let 
\[
\Pi = \text{diag}(\pi_{ws})_{ws \in S}
\]
be the weighting matrix with diagonal elements in the same order of \( ws \) as in \( Y \).

Let \( Y_w = (Y_{w1}, \ldots, Y_{wM_w})^T \), \( D_w = (d_{w1}, \ldots, d_{wM_w})^T \), \( \Pi_w = \text{diag}(\pi_{w1}, \ldots, \pi_{wM_w}) \), \( U_w = (\hat{U}_w(A_w0), \hat{U}_w(A_w1))^T \), and \( D_{ag,w} = (d_w(A_w0), d_w(A_w1))^T \) be the sub-vectors or sub-matrices of \( Y, D, \Pi, U, \) and \( D_{ag} \) corresponding to whole-plot \( w \), respectively. The cluster-robust covariances equal
\[
\hat{V}_{wls} = (D^T \Pi D)^{-1} \left( \sum_{w=1}^{W} D_w^T \Pi_w e_{wls,w} e_{wls,w}^T \Pi_w D_w \right) (D^T \Pi D)^{-1},
\]
与
\[
\hat{V}_{ag} = (D^T_{ag} D_{ag})^{-1} \left( \sum_{w=1}^{W} D_{ag,w}^T e_{ag,w} e_{ag,w}^T D_{ag,w} \right) (D^T_{ag} D_{ag})^{-1},
\]
where \( e_{wls,w} = (e_{wls,w1}, \ldots, e_{wls,wM_w})^T \) and \( e_{ag,w} = (e_{ag,w}(A_w0), e_{ag,w}(A_w1))^T \) are the residuals in whole-plot \( w \) from \( \textbf{(1)} \) and \( \textbf{(2)} \), respectively.

Let \( \tilde{\beta}_1(z) \) be the \( z \)th element in \( \tilde{\beta}_1 \) that gives the regression estimator of \( \tilde{Y}(z) \) for \( \dag = \text{ols}, \text{wls}, \text{ag} \).

Let \( P = \text{diag}(p_0, p_1) \otimes I_2 \) and \( Q_w = I_2 \otimes \text{diag}(q_{w0}, q_{w1}) \) with \( \text{diag}\{p_{ws}(z)\}_{z \in T} = PQ_w \) for all \( ws \) in whole-plot \( w \). Let \( R = \text{diag}(r_z)_{z \in T} \), where \( r_z = N_z/N \). Equation \( \textbf{(S21)} \) lists some useful algebraic facts for split-plot type data; the proof follows from direct algebra and is thus omitted:

\[
N^{-1} D^T D = R, \quad D^T \Pi D = \hat{1}_{ht}, \quad W^{-1} D^T_{ag} D_{ag} = P,
\]
\[
N^{-1} D^T Y = R \hat{Y}_{sm}, \quad D^T \Pi Y = \hat{Y}_{ht}, \quad W^{-1} D^T_{ag} U = P \hat{Y}_{ht};
\]
\[
D^T_w Y_w = M_w Q_w \hat{Y}_w, \quad D^T_w \Pi Y_w = W^{-1} P^{-1} \alpha_w \hat{Y}_w, \quad D^T_{ag,w} U_w = \hat{U}_w,
\]
where \( \hat{Y}_w \) and \( \hat{U}_w \) are the 4 \times 1 vectors of \( \{\hat{Y}_w(z)\}_{z \in T} \) and \( \{\hat{U}_w(z)\}_{z \in T} \) in lexicographical order of \( z \), respectively, with \( \hat{Y}_w(z) = \hat{U}_w(z) = 0 \) for \( z \notin \{(A_w0), (A_w1)\} \) by definition.

**S2.2. Proof of the main results**

*Proof of Proposition \( \Box \).* The result follows from \( \tilde{\beta}_{wls} = (D^T D)^{-1} D^T Y, \) \( \tilde{\beta}_{wls} = (D^T \Pi D)^{-1} D^T \Pi Y, \)
\( \tilde{\beta}_{ag} = (D^T_{ag} D_{ag})^{-1} D^T_{ag} U, \) and \( \textbf{(S21)}. \)

*Proof of Theorem \( \Box \).* We verify below the algebraic results of \( \hat{V}_{wls} \) and \( \hat{V}_{ag} \) in finite samples. The asymptotic equivalence then follows from Lemma \( \textbf{S3}. \)

**Algebraic result for \( \hat{V}_{wls} \).** Suppress the subscript “wls” in \( e_{wls,w} \) and \( e_{wls,ws} \) in \( \textbf{(S14)} \) when verifying the result for \( \hat{V}_{wls} \). Proposition \( \Box \) ensures \( e_{ws} = Y_{ws} - \beta_{wls}(Z_{ws}) = Y_{ws} - \hat{Y}_{waj}(Z_{ws}) \).

Let \( \hat{e}_w(z) = M_{wb}^{-1} \sum_{ws:Z_{ws}=z} e_{ws} \) be the sample mean of \( e_{ws} \) under treatment \( z \) with \( \hat{e}_w(z) = \hat{Y}_w(z) - \hat{Y}_{waj}(z) \) for \( z \in \{(A_w0), (A_w1)\} \) and \( \hat{e}_w(z) = 0 \) for \( z \notin \{(A_w0), (A_w1)\} \). Setting \( Y_w = e_w \) in \( \textbf{(S21)} \) to see \( D^T_w \Pi_w e_w = W^{-1} P^{-1} \alpha_w \hat{e}_w \) with \( \hat{e}_w = (\hat{e}_w(00), \hat{e}_w(01), \hat{e}_w(10), \hat{e}_w(11))^T \) the middle part of

S12
\[ (S19) \text{ thus equals} \]
\[
\sum_{w=1}^{W} D_{w}^{T} \Pi_{w} e_{w} e_{w}^{T} \Pi_{w} D_{w} = W^{-2} P^{-1} \left( \sum_{w=1}^{W} \alpha_{w}^{2} \hat{e}_{w} \hat{e}_{w}^{T} \right) P^{-1} = (\Omega_{\text{wls}}(z, z'))_{z, z' \in T},
\]

where
\[
\Omega_{\text{wls}}(z, z') = W_{a}^{-1} W_{a'}^{-1} \sum_{w=1}^{W} \alpha_{w}^{2} \hat{e}_{w}(z) \hat{e}_{w}(z') = \begin{cases} 
W_{a}^{-2}(W_{a} - 1) \hat{S}_{ha}(z, z') & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a = a', \\
0 & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a \neq a'.
\end{cases}
\]

The result for \( \hat{V}_{\text{wls}} \) then follows from \( D^{T} \Pi D = \hat{I}_{ht} \) by \((S24)\).

**Algebraic result for \( \hat{V}_{\text{ag}} \).** Proposition 2 ensures \( e_{ag, w}(z) = \hat{U}_{w}(z) - \hat{\beta}_{ag}(z) = \hat{U}_{w}(z) - \hat{Y}_{\text{ht}}(z) \) for \( z \in \{(A_{w}0), (A_{w}1)\} \). Define \( e_{ag, w}(z) = 0 \) for \( z \not\in \{(A_{w}0), (A_{w}1)\} \) for notational simplicity. Setting \( U_{w} = e_{ag, w} \) in \((S20)\) to see \( D_{ag, w}^{T} e_{ag, w} = \hat{e}_{ag, w} \) with \( \hat{e}_{ag, w} = (e_{ag, w}(00), e_{ag, w}(01), e_{ag, w}(10), e_{ag, w}(11))^{T} \). The middle part of \((S20)\) thus equals
\[
\sum_{w=1}^{W} D_{ag, w}^{T} (e_{ag, w} e_{ag, w}^{T}) D_{ag, w} = \sum_{w=1}^{W} \hat{e}_{ag, w} e_{ag, w}^{T} = (\Omega_{ag}(z, z'))_{z, z' \in T}, \tag{S22}
\]

where
\[
\Omega_{ag}(z, z') = \sum_{w=1}^{W} e_{ag, w}(z) \cdot e_{ag, w}(z') = \begin{cases} 
(W_{a} - 1) \hat{S}_{ht}(z, z') & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a = a', \\
0 & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a \neq a'.
\end{cases}
\]

The result for \( \hat{V}_{ag} \) then follows from \( (D_{ag}^{T} D_{ag})^{-1} = \text{diag}(W_{0}^{-1}, W_{1}^{-1}) \otimes I_{2} \) by \((S21)\). \( \square \)

### S3. Regression-based covariate adjustment

#### S3.1. Notation and lemmas

Inherit all notations from Section S2.1. In addition, recall \( \bar{x}_{w} = M_{w}^{-1} \sum_{s=1}^{M_{w}} x_{w,s} \) as the whole-plot average covariates with \( W^{-1} \sum_{w=1}^{W} \alpha_{w} \bar{x}_{w} = \bar{x} = 0_{J} \). Let \( v_{w,s} = \alpha_{w} x_{w,s}, w_{v} = \alpha_{w} \bar{x}_{w}, \) and \( U_{w}^{t}(z) = \alpha_{w} \bar{Y}_{w}(z) = \alpha_{w} \bar{Y}_{w}(z) - \alpha_{w} \bar{Y}(z) \) be the analogs of \( U_{w}(z) = \alpha_{w} Y_{w}(z) \) and \( U_{w}(z) = \bar{Y}_{w}(z) \) defined on the covariates and centered potential outcomes, respectively. We have
\[
S_{xx} = (W - 1)^{-1} \sum_{w=1}^{W} \alpha_{w}^{2} \bar{x}_{w} \bar{x}_{w}^{T} = (W - 1)^{-1} \sum_{w=1}^{W} v_{w} v_{w}^{T},
\]
\[
S_{xx, w} = (M_{w} - 1)^{-1} \alpha_{w}^{2} \sum_{s=1}^{M_{w}} (x_{w,s} - \bar{x}_{w})(x_{w,s} - \bar{x}_{w})^{T} = (M_{w} - 1)^{-1} \sum_{s=1}^{M_{w}} (v_{w,s} - v_{w})(v_{w,s} - v_{w})^{T},
\]

\[ S13 \]
\[ S_{XY}(z) = S_{Y(z)x}^T = (W - 1)^{-1} \sum_{u=1}^{W} \alpha_w x_w \{ \alpha_w Y_w(z) - \bar{Y}(z) \} = (W - 1)^{-1} \sum_{u=1}^{W} v_w U_w(z), \]

\[ S_{Y'(z)x} = S_{Y'(z)x}^T = (W - 1)^{-1} \sum_{u=1}^{W} \alpha_w x_w \{ \alpha_y Y'_w(z) \} = (W - 1)^{-1} \sum_{u=1}^{W} v_w U'_w(z), \]

\[ S_{xY'(z),uw} = S_{xY'(z),uw}^T = (M_w - 1)^{-1} \alpha^2 w \sum_{s=1}^{M_w} (x_{us} - \bar{x}_u) \{ Y_{ws}(z) - \bar{Y}(z) \} \]

\[ = (M_w - 1)^{-1} \sum_{s=1}^{M_w} (v_{us} - v_u) \{ U_{ws}(z) - U_w(z) \} = S_{xY'(z),uw} = S_{Y'(z)x,sw}, \]

Recall \( \Psi(z, z) = W^{-1} \sum_{w=1}^{W} M_w^{-1} H_w(z, z) S_w(z, z) \) as the \((z, z)\)th element of \( \Psi \) from Lemma 1 with \( H_w(z, z) = p_a^{-1}(q_{wb} - 1) \). We have

\[ \Psi_{xx}(z, z) = W^{-1} \sum_{w=1}^{W} M_w^{-1} H_w(z, z) S_{xx, uw}, \quad \Psi_{xY}(z, z) = W^{-1} \sum_{w=1}^{W} M_w^{-1} H_w(z, z) S_{xY, uw} \]

as the analogs by replacing \( S_u(z, z) \) with \( S_{xx, uw} \) and \( S_{xY, uw} \), respectively. Recall \( Q_{xx} = (N - 1)^{-1} \sum_{w \in S} x_{ws} x_{w}^T \) and \( Q_{xY}(z) = (N - 1)^{-1} \sum_{w \in S} x_{ws} Y_{ws}(z) \) as the standard, unscaled finite-population covariances. Let

\[ \hat{Q}_{xx}(z) = \lambda_N^{-1} \sum_{w \in S(x)} \pi_{ws} x_{ws} x_{w}^T, \quad \hat{Q}_{xY}(z) = \lambda_N^{-1} \sum_{w \in S(z)} \pi_{ws} x_{ws} Y_{ws}, \]

where \( \lambda_N = 1 - N^{-1} \), be their Horvitz–Thompson estimators based on sub-plots under treatment \( z \). Let

\[ \hat{S}_{xx}(z) = W_a^{-1} \sum_{w \in W(z)} \hat{v}_u(z) v_w^T, \quad \hat{S}_{xY}(z) = W_a^{-1} \sum_{w \in W(z)} \hat{v}_u(z) \hat{U}_w(z) \]

be the sample analogs of \( S_{xx} \) and \( S_{xY}(z) \) based on sub-plots under treatment \( z = (ab) \). Let

\[ T_{xx}(z) = S_{xx} + p_a \Psi_{xx}(z, z), \quad T_{xY}(z) = S_{xY}(z) + p_a \Psi_{xY}(z, z). \]

**Lemma S7.** Under the \( 2^2 \) split-plot randomization and Conditions 2–3, we have

\[ \hat{Q}_{xx}(z) - Q_{xx} = o_P(1), \quad \hat{Q}_{xY}(z) - Q_{xY}(z) = o_P(1), \]

\[ \hat{S}_{xx}(z) - T_{xx}(z) = o_P(1), \quad \hat{S}_{xY}(z) - T_{xY}(z) = o_P(1). \]

**Proof of Lemma S7.** We verify below the result for scalar covariate \( x_{ws} \in \mathbb{R} \) to simplify the presentation.

For the result on \( \hat{Q}_{xx}(z) \), let \( \sigma_{ws} = x_{ws}^2 \) to write \( Q_{xx} = (N - 1)^{-1} \sum_{w \in S} \sigma_{ws} \). Lemma S7 ensures that \( \hat{Q}_{xx}(z) - Q_{xx} = o_P(1) \) as long as Condition S14 holds for the finite population of
$(\sigma_{ws})_{w \in S}$. To verify this, let $\bar{\sigma}_w = M_w^{-1} \sum_{s=1}^{M_w} \sigma_{ws}$ with $\bar{\sigma}_w^2 \leq M_w^{-1} \sum_{s=1}^{M_w} \sigma_{ws}^2 = M_w^{-1} \sum_{s=1}^{M_w} x_{ws}^4$ by (S3). This ensures
$$W^{-1} \sum_{w=1}^{W} \alpha_w^2 \sigma_w^2 \leq W^{-1} \sum_{w=1}^{W} \alpha_w^2 \left( M_w^{-1} \sum_{s=1}^{M_w} x_{ws}^4 \right) = O(1) \quad (S23)$$
by Condition 3(II). Condition S1(II) is thus satisfied with
- $\bar{\sigma} = N^{-1} \sum_{w \in S} \sigma_{ws} = \lambda_N Q_{xx}$ having a finite limit by Condition 3(I);
- $S_{w\sigma} = (W - 1)^{-1} \sum_{w=1}^{W} \left( \alpha_w \sigma_w - \bar{\sigma} \right)^2 \leq (W - 1)^{-1} \sum_{w=1}^{W} \alpha_w^2 \sigma_w^2 = O(1)$ by (S24);
- $\Psi_{w\sigma} = W^{-1} \sum_{w=1}^{W} M_w^{-1} \{ H_w \circ (S_{w\sigma} \circ \omega \circ x) \} = W^{-1} \sum_{w=1}^{W} M_w^{-1} S_{w\sigma} H_w = O(1)$ given $S_{w\sigma} = (M_w - 1)^{-1} \alpha_w^2 \left( \sum_{s=1}^{M_w} \sigma_{ws}^2 - M_w \bar{\sigma}_w^2 \right) \leq (M_w - 1)^{-1} \alpha_w^2 \sum_{s=1}^{M_w} \sigma_{ws}^2$ and (S24).

This verifies $\hat{Q}_{xx}(z) = Q_{xx} = o_p(1)$.

The proof for $\hat{Q}_{xx}(z) = Q_{xx} = o_p(1)$ is almost identical by verifying Condition S1(II) for $\{x_{ws} Y_{ws}(z) : w \in S\}$ and thus omitted.

The result on $\hat{S}_{xx}(z)$ follows by applying Lemma S8 to the finite population with $Y_{ws}(z') = Y_{ws}(z') = x_{ws}$; the corresponding Condition S1 is ensured by Condition 3. Likewise for the result on $\hat{S}_{xx}(z)$ to follow from letting $Y_{ws}(z') = x_{ws}$ in Lemma S8.

For a set of $J \times 1$ vectors $\gamma = (\gamma_z)_{z \in T}$, let $Y_{ws}(z; \gamma_z) = Y_{ws}(z) - x_{ws}^T \gamma_z$ be the adjusted potential outcome based on $\gamma_z$, and let $S_{s,\gamma}, \Sigma_{s,\gamma}, S_{w,\gamma}, \Psi_{\gamma}, \hat{Y}_{s,\gamma},$ and $\hat{V}_{s,\gamma}$ be the covariate-adjusted analogs of $S_s, \Sigma_s, S_w, \Psi, \hat{Y}_s,$ and $\hat{V}_s$ defined on $\{Y_{ws}(z; \gamma_z) : w \in S, z \in T\}$ for $* = ht, haj$, respectively. We have $\Psi_s = W^{-1} \sum_{w=1}^{W} M_w^{-1} H_w \circ S_{w,\gamma}$ and $S_{s,\gamma} = H \circ S_{s,\gamma} + \Psi_{\gamma}$. Lemma S8 states the asymptotic Normality and conservativeness of $\hat{Y}_{s,\gamma}$ and $\hat{V}_{s,\gamma}$, respectively.

**Lemma S8.** Assume fixed $\gamma = (\gamma_z)_{z \in T}$ and split-plot randomization. We have $E(\hat{Y}_{ht,\gamma}) = \tilde{\gamma}$, $\text{cov}(\hat{Y}_{ht,\gamma}) = W^{-1} \{ H \circ S_{ht,\gamma} + \Psi_s \}$, and $E(\hat{V}_{ht,\gamma}) - \text{cov}(\hat{Y}_{ht,\gamma}) = W^{-1} S_{ht,\gamma} \geq 0$. Further assume Conditions 2, 3 and $\hat{\gamma} = (\hat{\gamma}_z)_{z \in T}$ with $\hat{\gamma}_z = \gamma_z + o_p(1)$ for $* = ht, haj$, we have

(i) $\sqrt{W}(\hat{Y}_{s,\gamma} - \bar{Y}) \sim N(0, \Sigma_{s,\gamma})$, $W \hat{V}_{s,\gamma} - \Sigma_{s,\gamma} = S_{s,\gamma} + o_p(1)$;

(ii) $\sqrt{W}(\hat{Y}_{s,\gamma} - \hat{\gamma}_{s,\gamma}) = o_p(1)$, $W(\hat{V}_{s,\gamma} - \hat{\gamma}_{s,\gamma}) = o_p(1)$, $W \hat{V}_{s,\gamma} - \Sigma_{s,\gamma} = S_{s,\gamma} + o_p(1)$.

**Proof of Lemma S8.** The $(z, z')$th elements of $S_{ht,\gamma}, S_{w,\gamma}$, and $S_{haj,\gamma}$ equal

$$S_{ht}(z, z'; \gamma) = S_{ht}(z, z') - \gamma_z^T S_{h Y}(z') - S_{Y}(z) \gamma_{z'} + \gamma_z^T S_{xx} \gamma_{z'},$$
$$S_{w}(z, z'; \gamma) = S_{w}(z, z') - \gamma_z^T S_{h Y}(z')_{w} - S_{Y}(z)_{w} \gamma_{z'} + \gamma_z^T S_{xx, w} \gamma_{z'},$$
$$S_{haj}(z, z'; \gamma) = S_{haj}(z, z') - \gamma_z^T S_{h Y}(z') - S_{Y}(z) \gamma_{z'} + \gamma_z^T S_{xx} \gamma_{z'},$$
respectively, by standard algebra. Conditions 2, 3 together imply that Condition 2 holds for the adjusted potential outcomes $\{Y_{ws}(z; \gamma_z)\}_{w \in S, z \in T}$ for fixed $\gamma = (\gamma_z)_{z \in T}$. The finite-sample result and statement (ii) follow from Lemma III and Theorems III, IV.
We verify below statement (ii) for the Horvitz–Thompson estimator. The proof for * = haj is almost identical and thus omitted.

The first equality $\sqrt{W}(\hat{Y}_{ht\gamma} - \hat{Y}_{ht\gamma}) = o_P(1)$ follows from

$$\sqrt{W}\{\hat{Y}_{ht}(z; \hat{\gamma}_z) - \hat{Y}_{ht}(z; \gamma_z)\} = -(\hat{\gamma}_z - \gamma_z)^{T}\sqrt{W}\hat{x}_{ht}(z) = o_P(1)$$  \hspace{1cm} (S24)$$

by Slutsky’s theorem with $\hat{\gamma}_z = \gamma_z + o_P(1)$ and $\sqrt{W}\hat{x}_{ht}(z)$ being asymptotically Normal by Theorem II.

For the second equality $W(\hat{V}_{ht\gamma} - \hat{V}_{ht\gamma}) = o_P(1)$, component-wise comparison ensures that it suffices to verify $\hat{S}_{ht}(z, z'; \hat{\gamma}) - \hat{S}_{ht}(z, z'; \gamma) = o_P(1)$, where

$$\hat{S}_{ht}(z, z'; \gamma) = (W_a - 1)^{-1} \left\{ \sum_{w:A_w = a} \hat{U}_w(z; \hat{\gamma}_z)\hat{U}_w(z'; \hat{\gamma}_z) - W_a \cdot \hat{Y}_{ht}(z; \hat{\gamma}_z)\hat{Y}_{ht}(z'; \hat{\gamma}_z) \right\},$$

$$\hat{S}_{ht}(z, z'; \gamma) = (W_a - 1)^{-1} \left\{ \sum_{w:A_w = a} \hat{U}_w(z; \gamma_z)\hat{U}_w(z'; \gamma_z) - W_a \cdot \hat{Y}_{ht}(z; \gamma_z)\hat{Y}_{ht}(z'; \gamma_z) \right\},$$

for all $z = (ab)$ and $z' = (ab')$ with the same level of factor A. Given $\hat{Y}_{ht}(z; \hat{\gamma}_z) - \hat{Y}_{ht}(z; \gamma_z) = o_P(1)$ by (S24), it suffices to verify

$$\Delta = W_a^{-1} \left\{ \sum_{w:A_w = a} \hat{U}_w(z; \hat{\gamma}_z)\hat{U}_w(z'; \hat{\gamma}_z) - \sum_{w:A_w = a} \hat{U}_w(z; \gamma_z)\hat{U}_w(z'; \gamma_z) \right\} = o_P(1).$$

To this end, let $\Delta_w(z) = (\hat{\gamma}_z - \gamma_z)^{T}\hat{v}_w(z)$ to write $\hat{U}_w(z; \hat{\gamma}_z) = \hat{U}_w(z; \gamma_z) - \Delta_w(z)$ and

$$\hat{U}_w(z; \hat{\gamma}_z)\hat{U}_w(z'; \hat{\gamma}_z) - \hat{U}_w(z; \gamma_z)\hat{U}_w(z'; \gamma_z) = \Delta_w(z)\Delta_w(z') - \Delta_w(z)\hat{U}_w(z'; \gamma_z) - \Delta_w(z')\hat{U}_w(z; \gamma_z).$$

We have

$$\Delta = W_a^{-1} \sum_{w:A_w = a} \left\{ \Delta_w(z)\Delta_w(z') - \Delta_w(z)\hat{U}_w(z'; \gamma_z) - \Delta_w(z')\hat{U}_w(z; \gamma_z) \right\}$$

$$= (\hat{\gamma}_z - \gamma_z)^{T}\left\{ W_a^{-1} \sum_{w:A_w = a} \hat{v}_w(z)\hat{v}_w^{T}(z') \right\} (\hat{\gamma}_z - \gamma_z)$$

$$- (\hat{\gamma}_z - \gamma_z)^{T}\cdot W_a^{-1} \sum_{w:A_w = a} \hat{v}_w(z)\hat{U}_w(z'; \gamma_z) - (\hat{\gamma}_z - \gamma_z)^{T}\cdot W_a^{-1} \sum_{w:A_w = a} \hat{v}_w(z')\hat{U}_w(z; \gamma_z)$$

$$= o_P(1);$$

the last equality follows from $\hat{\gamma}_z - \gamma_z = o_P(1)$ and the fact that $W_a^{-1} \sum_{w:A_w = a} \hat{v}_w(z)\hat{v}_w^{T}(z')$ and

$$W_a^{-1} \sum_{w:A_w = a} \hat{v}_w(z)\hat{U}_w(z'; \gamma_z) = W_a^{-1} \sum_{w:A_w = a} \hat{v}_w(z)\hat{U}_w(z') - \left\{ W_a^{-1} \sum_{w:A_w = a} \hat{v}_w(z)\hat{v}_w^{T}(z') \right\} \gamma_z.$$
are both $O_p(1)$ by Lemmas S9 and S7. This ensures $W(\hat{V}_{ht,\gamma} - \hat{V}_{ht,\gamma}) = o_p(1)$.

The third equality $W\hat{V}_{ht,\gamma} - \Sigma_{ht,\gamma} = S_{*,\gamma} + o_p(1)$ then follows from statement (\textbullet).

Lemma S9. Styan (1973, Lemma 3.4) If $G_1$ is positive definite and $G_2$ is positive semi-definite with $r$ positive diagonal elements then $\text{rank}(G_1 \circ G_2) = r$.

S3.2. Regression estimators from the fully-interacted models

We verify in this part the main results under the fully-interacted models (\textbullet\textbullet\textbullet) and (\textbullet\textbullet). The proofs provide important building blocks for verifying the results under the additive models (S) and (\textbullet\textbullet\textbullet).

Lemma S10. Under the $2^2$ split-plot randomization and Conditions S2-S3, we have $\gamma_{wls,z} = Q_{xx}^{-1}Q_{xY}(z)$ and $\gamma_{ag,z} = T_{xx}^{-1}(z)T_{xY}(z)$.

Proof of Proposition S4 and Lemma S7. For $\dagger = \text{ols}, \text{wls}$ and $z \in \mathcal{T}$, the inclusion of full interactions ensures that $\tilde{\beta}_{\dagger,1}(z)$ and $\tilde{\gamma}_{\dagger,z}$ from (\textbullet\textbullet\textbullet) equal the coefficients of 1 and $x_{ws}$ from the treatment-specific regression

$$Y_{ws} \sim 1 + x_{ws} \quad \text{over } ws \in S(z), \quad (S25)$$

respectively, under their respective fitting schemes. Likewise for $\tilde{\beta}_{\text{ag},1}(z)$ and $\tilde{\gamma}_{\text{ag},z}$ from (\textbullet\textbullet) to equal the OLS coefficients of 1 and $\hat{v}_{w}(z)$ from the treatment-specific regression

$$\hat{U}_{w}(z) \sim 1 + \hat{v}_{w}(z) \quad \text{over } w \in W(z) \quad (S26)$$

for $z \in \mathcal{T}$.

Let $Y_z$ and $X_z$ be the concatenation of $Y_{ws}$ and $x_{ws}$ over $ws$ under treatment $z$. Specification (S25) can be expressed in matrix form as $Y_z \sim 1_{N_z} + X_z$ for $z \in \mathcal{T}$. Let $\Pi_z = \text{diag}(\pi_{ws})_{ws \in S(z)}$ be the weighting matrix under fitting scheme “wls”. The first-order condition ensures

$$\{(1_{N_z}, X_z)^T \Pi_z (1_{N_z}, X_z) \}(\tilde{\beta}_{wls,l}(z), \tilde{\gamma}_{wls,z}^T)^T = (1_{N_z}, X_z)^T \Pi_z Y_z, \quad (S27)$$

where

$$(1_{N_z}, X_z)^T \Pi_z (1_{N_z}, X_z) = \begin{pmatrix} 1_{N_z}^T \Pi_z 1_{N_z} & 1_{N_z}^T \Pi_z X_z \\ X_z^T \Pi_z 1_{N_z} & X_z^T \Pi_z X_z \end{pmatrix} = \begin{pmatrix} \hat{1}_{ht}(z) & x_{ht}^T(z) \\ x_{ht}(z) & \lambda_N \hat{Q}_{xx}(z) \end{pmatrix} = \begin{pmatrix} 1 & x_{ht}^T(z) \\ 0 & \lambda_N \hat{Q}_{xx}(z) \end{pmatrix} + o_p(1),$$

$$(1_{N_z}, X_z)^T \Pi_z Y_z = \begin{pmatrix} 1_{N_z}^T \Pi_z Y_z \\ X_z^T \Pi_z Y_z \end{pmatrix} = \begin{pmatrix} \hat{Y}_{ht}(z) \\ \lambda_N \hat{Q}_{xY}(z) \end{pmatrix} = \begin{pmatrix} \tilde{Y}(z) \\ Q_{xY}(z) \end{pmatrix} + o_p(1)$$

by the definition of $\pi_{ws}$ and Lemmas S3 and S7. Compare the first row of (S27) to see $\hat{1}_{ht}(z)\tilde{\beta}_{wls,l}(z) + x_{ht}^T(z)\tilde{\gamma}_{wls,z} = \hat{Y}_{ht}(z)$ and hence the numerical result for $\tilde{\beta}_{wls,l}(z)$. The probability limits of $\tilde{\beta}_{wls,l}$ and $\tilde{\gamma}_{wls,z}$ follow from $(\tilde{\beta}_{wls,l}(z), \tilde{\gamma}_{wls,z}^T)^T = \{(1_{N_z}, X_z)^T \Pi_z (1_{N_z}, X_z)\}^{-1}\{(1_{N_z}, X_z)^T \Pi_z Y_z\}$. 

S17
The numerical result for $\tilde{\beta}_{\text{obs},l}(z)$ follows from identical reasoning with $\Pi_z$ replaced by the identity matrix.

For the aggregate regression (14), let $U_z$ be the vectorization of the $W_a$ observations, namely \{\tilde{U}_w(z) : w \in W_z\}, under treatment $z$, and let $\Lambda_z$ be the concatenation of the corresponding \{\tilde{v}_w(z) : w \in W_z\}. Specification (S28) can be expressed in matrix form as $U_z \sim 1_{W_a} + \Lambda_z$. The first-order condition ensures

\[ \{(1_{W_a}, \Lambda_z)^T(1_{W_a}, \Lambda_z)\}(\tilde{\beta}_{\text{ag},l}(z), \tilde{\gamma}_{ag,z})^T = (1_{W_a}, \Lambda_z)^T U_z, \quad (S28) \]

where

\[ W^{-1}(1_{W_a}, \Lambda_z)^T (1_{W_a}, \Lambda_z) = W^{-1} \begin{pmatrix} 1_{W_a}^T & 1_{W_a} \\ 1_{W_a} & \Lambda_z \end{pmatrix} \]

\[ = p_a \begin{pmatrix} 1 & \tilde{\beta}_{\text{ht}}(z) \\ \tilde{\beta}_{\text{ht}}(z) & \tilde{\gamma}_{ag,z} \end{pmatrix} = p_a \begin{pmatrix} 1 & 0_f^T \\ 0_f & T_{xx}(z) \end{pmatrix} + o_P(1), \]

\[ W^{-1}(1_{W_a}, \Lambda_z)^T U_z = W^{-1} \begin{pmatrix} 1_{W_a}^T U_z \\ \Lambda_z U_z \end{pmatrix} = p_a \begin{pmatrix} \tilde{Y}_{\text{ht}}(z) \\ \tilde{S}_{xY}(z) \end{pmatrix} = p_a \begin{pmatrix} \tilde{Y}(z) \\ T_x Y(z) \end{pmatrix} \]

by Lemmas S3 and S7. Compare the first row of (S28) to see $\tilde{\beta}_{\text{ag},l}(z) + \tilde{\beta}_{\text{ht}}(z) \tilde{\gamma}_{ag,z} = \tilde{Y}_{\text{ht}}(z)$ and hence the numerical result for $\tilde{\beta}_{\text{ag},l}(z)$. The probability limits of $\tilde{\beta}_{\text{ag},l}$ and $\tilde{\gamma}_{ag,z}$ follow from $(\tilde{\beta}_{\text{ag},l}(z), \tilde{\gamma}_{ag,z})^T = \{(1, \Lambda_z)^T(1, \Lambda_z)\}^{-1}(1, \Lambda_z)^T U_z$.

\[ \square \]

\textbf{Proof of Theorem 4.} Let $\gamma_{\text{ht},l} = (\gamma_{\text{ht},z})_{z \in T}$ for $\dagger = \text{wls, ag}$. With a slight repetition of notation, let $\tilde{Y}_{\text{haj}}(\gamma_{\text{wls},l})$ and $\tilde{Y}_{\text{ht}}(\gamma_{\text{ag},l})$ be the vectorizations of $\tilde{Y}_{\text{haj}}(z; \gamma_{\text{wls},z}) = \tilde{Y}_{\text{haj}}(z) - \{\tilde{x}_{\text{haj}}(z)\}^T \gamma_{\text{wls},z}$, and $\tilde{Y}_{\text{ht}}(z; \gamma_{\text{ag},z}) = \tilde{Y}_{\text{ht}}(z) - \{\tilde{x}_{\text{ht}}(z)\}^T \gamma_{\text{ag},z}$, respectively. Lemma S3 ensures

\[ \sqrt{W} \{\tilde{\beta}_{\text{wls},l} - \tilde{Y}_{\text{haj}}(\gamma_{\text{wls},l})\} = o_P(1), \quad \sqrt{W} \{\tilde{Y}_{\text{haj}}(\gamma_{\text{wls},l}) - \tilde{Y}\} \rightsquigarrow \mathcal{N}(0, \Sigma_{\text{wls},l}), \]

\[ \sqrt{W} \{\tilde{\beta}_{\text{ag},l} - \tilde{Y}_{\text{ht}}(\gamma_{\text{ag},l})\} = o_P(1), \quad \sqrt{W} \{\tilde{Y}_{\text{ht}}(\gamma_{\text{ag},l}) - \tilde{Y}\} \rightsquigarrow \mathcal{N}(0, \Sigma_{\text{ag},l}). \]

The asymptotic Normality of $\tilde{\beta}_{\dagger,l}$ follows from Slutsky’s theorem.

We verify below the result on the cluster-robust covariances for $\tilde{V}_{\text{wls},l}$ from (14). The proof for $\tilde{V}_{\text{ag},l}$ follows from identical reasoning and is thus omitted.

Let $\chi$ be the concatenation of $\chi_{ws} = d_{ws} \otimes x_{ws}$ over all $ws \in S$. The design matrix of (14) equals $C_L = (D, \chi)$. Let $C_{L, w} = (D_w, \chi_w)$ be the sub-matrix of $C_L$ corresponding to whole-plot $w$. Let $\epsilon_w = (\epsilon_{w1}, \ldots, \epsilon_{wM_w})^T$, where $\epsilon_{ws} = Y_{ws} - \tilde{\beta}_{\text{wls},l}(Z_{ws}) - x_{ws}^T \tilde{\gamma}_{\text{wls},Z_{ws}}$ is the least-squares residual from the wls fit of (14). The covariance $W V_{\text{wls},l}$ is by definition the upper-left $4 \times 4$ matrix of

\[ (C_L^T \Pi C_L)^{-1} \left( W \sum_{w=1}^W C_{L,w}^T \Pi_w \epsilon_w \epsilon_w^T \Pi_w C_{L,w} \right) (C_L^T \Pi C_L)^{-1}. \quad (S29) \]
Direct algebra shows that the middle part of (S29) equals

\[
W \sum_{w=1}^{W} C_{L,w}^{T} \Pi_{w} e_{w} e_{w}^{T} \Pi_{w} C_{L,w} = W \sum_{w=1}^{W} \left( \frac{D_{w}^{T}}{\chi_{w}^{T}} \right) \Pi_{w} e_{w} e_{w}^{T} \Pi_{w} (D_{w}, \chi_{w})
\]

\[
= W \left( \sum_{w=1}^{W} \frac{D_{w}^{T}}{\chi_{w}^{T}} \Pi_{w} e_{w} e_{w}^{T} \Pi_{w} D_{w} \right) = \left( \frac{D^{T} \Pi D}{G_{1}^{T}} \right) \left( \frac{D^{T} \Pi D}{G_{2}^{T}} \right),
\]

where \( G_{1} = W \sum_{w=1}^{W} \frac{D_{w}^{T}}{\chi_{w}^{T}} \Pi_{w} e_{w} e_{w}^{T} \Pi_{w} \chi_{w} \), and

\[
\hat{\Omega}_{wls,l} = (D^{T} \Pi D)^{-1} \left( \sum_{w=1}^{W} \frac{D_{w}^{T}}{\chi_{w}^{T}} \Pi_{w} e_{w} e_{w}^{T} \Pi_{w} D_{w} \right) (D^{T} \Pi D)^{-1}.
\]

The result on \( \check{\nu}_{wls,l} \) holds as long as

\[
W \hat{\Omega}_{wls,l} - \Sigma_{wls,l} = \omega_{wls,l} + \omega_{l}(1) \quad \text{and} \quad W (\check{\nu}_{wls.l} - \hat{\Omega}_{wls,l}) = \omega_{l}(1). \quad (S31)
\]

We proceed to verify these two conditions one by one.

For condition (S31)(i), direct comparison suggests that \( \hat{\Omega}_{wls,l} \) is an analog of \( \check{\nu}_{wls} \) from (S19), with the original unadjusted residuals \( e_{wls,w} \) replaced by the covariate-adjusted counterparts \( e_{w} \).

Proposition \( 4 \) ensures that \( \hat{\beta}_{wls,l}(z) = \check{\nu}_{haj}(z; \check{\gamma}_{wls,z}) \) equals the Hajek estimator defined on the adjusted potential outcomes \( Y_{w,s}(z; \check{\gamma}_{wls,z}) = Y_{w,s}(z) - x_{w,s}^{T} \check{\gamma}_{wls,z} \). This, together with Theorem \( 4 \), ensures

\[
\hat{\Omega}_{wls,l} = \hat{\nu}_{haj}(z; \check{\gamma}_{wls,z}) \cdot \hat{\nu}_{haj}(z; \check{\gamma}_{wls,z})
\]

with \( \hat{\nu}_{haj}(z; \check{\gamma}_{wls,z}) \) being the value of \( \hat{\nu}_{haj}(z) \) at \( \gamma = \check{\gamma}_{wls,l} \). With \( \check{\gamma}_{wls,z} - \gamma_{wls,l} = \omega_{l}(1) \) by Lemma \( 10 \), it follows from Lemma \( 8 \) that \( \hat{\nu}_{haj}(z; \check{\gamma}_{wls,l}) - \Sigma_{wls,l} = \omega_{l}(1) \) by the definitions of \( S_{wls,l} \) and \( \Sigma_{wls,l} \). This verifies condition (S31)(i).

For condition (S31)(ii), a key observation is that

\[
C_{L}^{T} \Pi C_{L} = \left( \frac{D^{T}}{\chi^{T}} \right) \Pi (D, \chi) = \left( \frac{D^{T} \Pi D}{\chi^{T} \Pi \chi} \right) = \text{diag}(I_{4}, I_{4} \otimes Q_{xx}) + \omega_{l}(1); \]

the last equality follows from \( D^{T} \Pi D = I_{4} + \omega_{l}(1) \) by (S21), \( D^{T} \Pi \chi = \sum_{w,s \in S} \pi_{w,s} d_{w,s} x_{w,s}^{T} = \text{diag} \left( \sum_{w,s \in S(z)} \pi_{w,s} x_{w,s}^{T} \right) = \omega_{l}(1), \) and \( \chi^{T} \Pi \chi = \sum_{w,s \in S} \pi_{w,s} x_{w,s} x_{w,s}^{T} = \text{diag} \left( \sum_{w,s \in S(z)} \pi_{w,s} x_{w,s} x_{w,s}^{T} \right) = I_{4} \otimes Q_{xx} + \omega_{l}(1) \) by Lemma \( 7 \). This, together with (S21), ensures that (S31)(ii) holds as long as

\[
G_{1} = W \sum_{w=1}^{W} \frac{D_{w}^{T}}{\chi_{w}^{T}} \Pi_{w} e_{w} e_{w}^{T} \Pi_{w} \chi_{w} = \omega_{l}(1), \quad G_{2} = W \sum_{w=1}^{W} \frac{\chi_{w}^{T}}{w_{w},w} e_{w} e_{w}^{T} \Pi_{w} \chi_{w} = \omega_{l}(1). \quad (S32)
\]
We verify below $G_2 = O_p(1)$ for scalar covariate $x_{ws} \in \mathbb{R}$ for notational simplicity. The proof for $G_1 = O_p(1)$ is almost identical and thus omitted.

First, recall the expression of $\pi_{ws}$ from (S18). Direct algebra shows that

$$\chi^2_w \pi_w e_w = \sum_{s=1}^{M_w} \pi_{ws} \chi_{ws} e_{ws} = \sum_{s=1}^{M_w} \pi_{ws} (d_{ws} \otimes x_{ws}) (1 \otimes \epsilon_{ws}) = \sum_{s=1}^{M_w} \pi_{ws} d_{ws} \otimes (x_{ws} \epsilon_{ws})$$

$$= \alpha_w \left( W_0^{-1} \kappa_w(00), W_0^{-1} \kappa_w(01), W_1^{-1} \kappa_w(10), W_1^{-1} \kappa_w(11) \right)^T,$$

where $\kappa_w(z) = M_{ws}^{-1} \sum_{s:Z_{ws} = z} x_{ws} \epsilon_{ws}$ with $\kappa_w(z) = 0$ if $z \notin \{(A_w0), (A_w1)\}$. This ensures $G_2 = (G_2(z, z'))_{z, z' \in \mathcal{T}}$ with

$$G_2(z, z') = \begin{cases} 
    p^2 W^{-1} \sum_{w: A_w = a} \alpha_w^2 \kappa_w(z) \kappa_w(z') & \text{if } z = (ab) \text{ and } z' = (a'b') \text{ with } a = a', \\
    0 & \text{if } z = (ab) \text{ and } z' = (a'b') \text{ with } a \neq a'.
\end{cases}$$

Second, recall that $q_{wb} \geq \epsilon$ by Condition D. Let $x_{w} = M_{ws}^{-1} \sum_{s=1}^{M_w} x_{ws}^k$ and $\epsilon_{w} = M_{ws}^{-1} \sum_{s=1}^{M_w} \epsilon_{ws}^k$ for $k = 2, 4$. It follows from $|\kappa_w(z)| = M_{ws}^{-1} \sum_{s:Z_{ws} = z} x_{ws} \epsilon_{ws} \leq 2^{-1} q_{wb}^{-1} (x_{w}^2 + \epsilon_{w}^2)$ and $(x_{w}^2 + \epsilon_{w}^2) \leq x_{w}^4$ and $(x_{w}^2 + \epsilon_{w}^2) \leq x_{w}^4$ by (S20) that

$$\{\kappa_w(z)\}^2 \leq 4^{-1} \epsilon^{-2} \left( \frac{x_{w}^4}{x_{w}^2 + \epsilon_{w}^2} \right) \leq 2^{-1} \epsilon^{-2} \left\{ \left( \frac{x_{w}^2}{\epsilon_{w}} \right)^2 + \left( \frac{\epsilon_{w}}{x_{w}} \right)^2 \right\} \leq 2^{-1} \epsilon^{-2} \left( \frac{x_{w}^4}{x_{w}^2 + \epsilon_{w}^2} \right).$$

This ensures

$$W^{-1} \sum_{w: A_w = a} \alpha_w^2 \{\kappa_w(z)\}^2 \leq W^{-1} \sum_{w=1}^{W} \alpha_w^2 \{\kappa_w(z)\}^2 \leq 2^{-1} \epsilon^{-2} W^{-1} \sum_{w=1}^{W} \alpha_w^2 \left( \frac{x_{w}^4}{x_{w}^2 + \epsilon_{w}^2} \right) = O_p(1);$$

the last equality follows from $W^{-1} \sum_{w=1}^{W} \alpha_w^2 x_{w}^4 = O_p(1)$ by Condition D and $W^{-1} \sum_{w=1}^{W} \alpha_w^2 \epsilon_{w}^4 = O_p(1)$ by $\epsilon_{w} = Y_{ws} - \hat{\beta}_{wls}(Z_{ws}) - x_{ws}^T \hat{\gamma}_{wls} Z_{ws}$ with $\epsilon^4_{w} \leq 27 Y^4_{ws} + 27 \{(\hat{\beta}_{wls}(Z_{ws}))^4 + 27 (x_{ws}^t \hat{\gamma}_{wls} Z_{ws})^4.$

As a result,

$$G_2(z, z') = p_a^2 W^{-1} \sum_{w: A_w = a} \alpha_w^2 \kappa_w(z) \kappa_w(z') \leq 2^{-1} p_a^2 W^{-1} \sum_{w: A_w = a} \alpha_w^2 \left[ \{\kappa_w(z)\}^2 + \{\kappa_w(z')\}^2 \right] = O_p(1)$$

for $z = (ab)$ and $z' = (a'b')$ that share the same level of factor A.

\[\square\]

**S3.3. Regression estimators from the additive models**

Recall that $R = \text{diag}(r_z)_{z \in \mathcal{T}}$ with $r_z = N_z/N$ and $\hat{x}_w = (\hat{x}_w(00), \hat{x}_w(01), \hat{x}_w(10), \hat{x}_w(11))^T$ for $* = \text{sm}, \text{ht}, \text{haj}.$

S20
Lemma S11. Under the $2^2$ split-plot randomization and Conditions 2-3, we have

$$
\gamma_{\text{wls}} = |T|^{-1}Q_{xx}^{-1} \sum_{z \in T} Q_{xy}(z), \quad \gamma_{\text{ag}} = \left\{ \sum_{z \in T} p_a T_{xx}(z) \right\}^{-1} \left\{ \sum_{z \in T} p_a T_{xy}(z) \right\}.
$$

Further assume Condition 4. We have $\tilde{\gamma}_{\text{ols}} = Q_{xx}^{-1} \sum_{z \in T} r_z Q_{xy}(z) + o_P(1)$.

Proof of Proposition 4 and Lemma S11. Let $X$ and $\Lambda$ be the concatenations of $x_{ws}$ over $ws \in S$ and $\hat{v}_w(A_w b)$ over $\{(w, b) : w = 1, \ldots, W; b = 0, 1\}$, respectively. The design matrices of (S) and (Q) equal $C_f = (D, X)$ and $C_{ag,f} = (D_{ag}, \Lambda)$. Standard algebra shows that

$$
N^{-1} D^T X = R_{\hat{x}_{\text{sm}}}, \quad D^T \Pi X = \hat{x}_{ht}, \quad W^{-1} D_{ag}^T U = P_{\hat{x}_{ht}} \tag{S33}
$$

analogous to (S21). Lemma S7 further ensures

$$
X^T \Pi X = |T| \cdot Q_{xx} + o_P(1), \quad X^T \Pi Y = \sum_{z \in T} Q_{xy}(z) + o_P(1), \quad W^{-1} \Lambda^T \Lambda = \sum_{z \in T} p_a T_{xx}(z) + o_P(1), \quad W^{-1} \Lambda^T U = \sum_{z \in T} p_a T_{xy}(z) + o_P(1). \tag{S34}
$$

For $\tilde{\beta}_{\text{ols},f}$ and $\tilde{\gamma}_{\text{ols}}$ from the OLS fit of (S), the first-order condition ensures $C_f^T C_f (\tilde{\beta}_{\text{ols},f}, \tilde{\gamma}_{\text{ols}})^T = C_f^T Y$, where

$$
C_f^T C_f = \begin{pmatrix} D^T D & D^T X \\ X^T D & X^T X \end{pmatrix} = \begin{pmatrix} NR & NR_{\hat{x}_{\text{sm}}} \\ N_{\hat{x}_{\text{sm}}}^T R & X^T X \end{pmatrix}, \quad C_f^T Y = \begin{pmatrix} D^T Y \\ X^T Y \end{pmatrix} = \begin{pmatrix} N R Y_{\text{sm}} \\ X^T Y \end{pmatrix}.
$$

Compare the first row to see $R_{\hat{x}_{\text{sm}}}
\tilde{\beta}_{\text{ols},f} + R_{\hat{x}_{\text{sm}}}^T \tilde{\gamma}_{\text{ols}} = R_{\hat{x}_{\text{sm}}} Y_{\text{sm}}$. This verifies the numerical result for $\tilde{\beta}_{\text{ols},f}$. The probability limit of $\tilde{\gamma}_{\text{ols}}$ under uniform designs then follows from $\hat{x}_{\text{sm}} = o_P(1)$, and

$$
\cdot \cdot N^{-1} X^T X = N^{-1} \sum_{ws \in S} x_{ws}^T x_{ws} = \sum_{z \in T} r_z (N_x^{-1} \sum_{ws \in S(z)} x_{ws} x_{ws}^T) = Q_{xx} + o_P(1),
$$

$$
\cdot \cdot N^{-1} X^T Y = N^{-1} \sum_{ws \in S} x_{ws} Y_{ws} = \sum_{z \in T} r_z (N_x^{-1} \sum_{ws \in S(z)} x_{ws} Y_{ws}(z)) = \sum_{z \in T} r_z Q_{xy}(z) + o_P(1)
$$

under Condition 4.

For $\tilde{\beta}_{\text{wls},f}$ and $\tilde{\gamma}_{\text{wls}}$ from the WLS fit of (S), the first-order condition ensures $C_f^T \Pi C_f (\tilde{\beta}_{\text{wls},f}, \tilde{\gamma}_{\text{wls}})^T = C_f^T \Pi Y$, where

$$
C_f^T \Pi C_f = \begin{pmatrix} D^T \Pi D & D^T \Pi X \\ X^T \Pi D & X^T \Pi X \end{pmatrix} = \begin{pmatrix} \hat{1}_{ht} & \hat{x}_{ht} \\ \hat{x}_{ht}^T & X^T \Pi X \end{pmatrix}, \quad C_f^T \Pi Y = \begin{pmatrix} D^T \Pi Y \\ X^T \Pi Y \end{pmatrix} = \begin{pmatrix} \hat{Y}_{ht} \\ X^T \Pi Y \end{pmatrix} \tag{S33}
$$

by (S21) and (S33). The numerical result follows by comparing the first row. The probability limits follow from $(\tilde{\beta}_{\text{wls},f}, \tilde{\gamma}_{\text{wls}})^T = (C_f^T \Pi C_f)^{-1} C_f^T \Pi Y$, where $C_f^T \Pi C_f = \text{diag}(I_{|T|}, |T| \cdot Q_{xx}) + o_P(1)$ and $C_f^T \Pi Y = (\hat{Y}_T, \sum_{z \in T} Q_{xy}(z))^T + o_P(1)$ as a result of $\hat{x}_{ht} = o_P(1)$, $\hat{Y}_T = \hat{Y} + o_P(1)$, and (S33).

For $\tilde{\beta}_{ag,f}$ and $\tilde{\gamma}_{ag}$ from the OLS fit of (Q), the first-order condition ensures $C_{ag,f}^T C_{ag,f} (\tilde{\beta}_{ag,f}, \tilde{\gamma}_{ag})^T = S21$.
\[ C_{ag,f}^T U, \]

where

\[
W^{-1}C_{ag,f}^T C_{ag,f} = \begin{pmatrix}
P & P\hat{\gamma}_{ht} \\
(P\hat{\gamma}_{ht})^T & \Lambda^T \Lambda
\end{pmatrix}, \quad W^{-1}C_{ag,f}^T U = \begin{pmatrix}
P\hat{\gamma}_{ht} \\
\Lambda^T U
\end{pmatrix}
\]

by (S21) and (S33). The numerical result follows by comparing the first row. The probability limits follow from (\( \hat{\beta}_{ag,f}^T, \hat{\gamma}_{ag}^T \)) = (W^{-1}C_{ag,f}^T C_{ag,f})^{-1}W^{-1}C_{ag,f}^T U, where \( W^{-1}C_{ag,f}^T C_{ag,f} = \text{diag}\{P, \sum_{z \in T} p_x T_{x \rho}(z)\} + \alpha_p(1) \) and \( W^{-1}C_{ag,f}^T U = ((P\hat{\gamma})^T, (\sum_{z \in T} p_x T_{x \rho}(z))^T) + \alpha_p(1) \) by (S33).

\[ \square \]

**Proof of Theorem 3.4.** The proof is almost identical to that of Theorem 3 and thus omitted. \( \square \)

### S3.4. Guaranteed gain in asymptotic efficiency

Let \( \hat{c}_w(A_w b) = (\alpha_w - 1, \hat{\nu}_w^T(A_w b))^T \) be the augmented whole-plot level covariate vector corresponding to \( \hat{\beta}_{ag,f}(\alpha, v) \) and \( \hat{\beta}_{ag,L}(\alpha, v) \). Let \( v_\rho = \alpha_w \hat{\nu}_w \) and \( c_w = (\alpha_w - 1, v_\rho^T) \) be the population analogs of \( \hat{v}_w(A_w b) = \alpha_w \hat{\nu}_w(A_w b) \) and \( \hat{c}_w(A_w b) \), respectively.

Recall \( \gamma_{ag} \) and \( \gamma_{ag,z} \) as the probability limits of the OLS coefficients of \( \hat{v}_w(A_w b) \) under the additive and fully-interacted aggregate regressions, respectively, if we use only \( \hat{v}_w(A_w b) \) for covariate adjustment. Let \( \gamma_{ag}(\alpha, v) \) and \( \gamma_{ag,z}(\alpha, v) \) be the probability limits of the OLS coefficients of \( \hat{c}_w(A_w b) \) when we also include \( (\alpha_w - 1) \) as an additional whole-plot level regressor. Let \( S_{ag,f}(\alpha, v) \) and \( S_{ag,L}(\alpha, v) \) be the corresponding scaled between-whole-plot covariances for the adjusted potential outcomes \( Y_{w,\rho}(z; \gamma_{ag}(\alpha, v)) \) and \( Y_{w,\rho}(z; \gamma_{ag,z}(\alpha, v)) \), respectively, with the \((z, z')\)th elements equaling

\[
S_{ag,f}(z, z'; \alpha, v) = (W - 1) \sum_{w=1}^{W} \{ U_{w}(z) - \bar{Y}(z) - c_w^T \gamma_{ag}(\alpha, v) \} \{ U_{w}(z') - \bar{Y}(z') - c_w^T \gamma_{ag}(\alpha, v) \},
\]

\[
S_{ag,L}(z, z'; \alpha, v) = (W - 1) \sum_{w=1}^{W} \{ U_{w}(z) - \bar{Y}(z) - c_w^T \gamma_{ag,z}(\alpha, v) \} \{ U_{w}(z') - \bar{Y}(z') - c_w^T \gamma_{ag,z}(\alpha, v) \}.
\]

Let

\[
\Psi_{xx}(z, z') = W^{-1} \sum_{w=1}^{W} M_w^{-1} H_{w}(z, z') S_{x,x,w}, \quad \Psi_{xY}(z, z') = W^{-1} \sum_{w=1}^{W} M_w^{-1} H_{w}(z, z') S_{x,Y(z'),w}
\]

be generalizations of \( \Psi_{xx}(z, z) \) and \( \Psi_{xY}(z, z) \) for \( z, z' \in T \). A useful fact is that

\[
\Psi_{xx}(z, z') = o(1), \quad \Psi_{xY}(z, z') = o(1)
\]

for \( z = (ab) \) and \( z' = (ab') \) with the same level of factor \( A \) under Conditions 2-4 and (12).

\[ \square \]
be the uniform lower bound of $H_w(z, z) = p_a^{-1}(q_{wb}^{-1} - 1)$ for all $z \in \mathcal{T}$ and $w = 1, \ldots, W$ under Condition (i) independent of $W$.

First, Condition (i) ensures

$$W^{-1} \sum_{w=1}^{W} M_w^{-1} S_{xx,w} \leq \epsilon_0^{-1} \Psi_{xx}(z, z) = o(1). \quad (S36)$$

The result for $\Psi_{xx}(z, z')$ with $z \neq z'$ follows from $H_w(z, z') = -p_a^{-1}$ are uniformly bounded.

Second, the Cauchy–Schwarz inequality ensures $|S_{XY(z),w}| \leq \{(S_{XX,w})^{1/2} \{S_w(z, z)\}^{1/2}$. This, together with $|H_w(z, z')| \leq h_0$, suggests

$$\Psi_{XY}(z, z') = \left| W^{-1} \sum_{w=1}^{W} M_w^{-1} H_w(z, z') S_{XY(z'),w} \right|$$

$$\leq W^{-1} \sum_{w=1}^{W} M_w^{-1} |H_w(z, z')| \{(S_{XX,w})\}^{1/2} \{S_w(z', z')\}^{1/2}$$

$$\leq h_0 \cdot W^{-1} \sum_{w=1}^{W} (M_w^{-1} S_{XX,w})^{1/2} \{M_w^{-1} S_w(z', z')\}^{1/2}$$

$$\leq h_0 \cdot \left( W^{-1} \sum_{w=1}^{W} M_w^{-1} S_{XX,w} \right)^{1/2} \left( W^{-1} \sum_{w=1}^{W} M_w^{-1} S_w(z', z') \right)^{1/2}$$

$$= o(1);$$

the last inequality follows from (S36) and the fact that $W^{-1} \sum_{w=1}^{W} M_w^{-1} S_w(z', z') \leq \epsilon_0^{-1} \Psi(z', z') = O(1)$ given $H_w(z', z') \geq \epsilon_0$ for all $w$.

Proof of Proposition 4. For two sequences of $4 \times 4$ matrices $(A_N)_{N=1}^{\infty}$ and $(B_N)_{N=1}^{\infty}$, write $A_N \leq \infty B_N$ if the limiting value of $(B_N - A_N)$ is positive semi-definite. The result is equivalent to

$$\Sigma_{ag,l}(\alpha, v) \leq \infty \Sigma_{ht}, \Sigma_{haj}, \Sigma_{wls, o}, \Sigma_{ag, o}, \Sigma_{ag,p}(\alpha, v) \quad \text{for } \diamond = F, L$$

with $\Sigma_{ht} = H \circ S_{ht} + \Psi$, $\Sigma_{haj} = H \circ S_{haj} + \Psi$, $\Sigma_{wls,o} = H \circ S_{wls,o} + \Psi_{wls,o}$, and

$$\Sigma_{ag,o} = H \circ S_{ag,o} + \Psi_{ag,o}, \quad \Sigma_{ag,o}(\alpha, v) = H \circ S_{ag,o}(\alpha, v) + \Psi_{ag,o}(\alpha, v).$$

Note that $H = \{\operatorname{diag}(p_0^{-1}, p_1^{-1}) - 1_{2 \times 2}\} \otimes 1_{2 \times 2} \geq 0$ has four positive diagonal elements. Lemma 5 ensures that it suffices to verify

(i) $S_{ag,l}(\alpha, v) \leq \infty S_{ht}, S_{haj}, S_{wls, o}, S_{ag, o}, S_{ag, p}(\alpha, v)$;

(ii) $\Psi_{wls,o} - \Psi = o(1), \Psi_{ag,o} - \Psi = o(1), \Psi_{ag,o}(\alpha, v) - \Psi = o(1)$

for $\diamond = F, L$ under Conditions (i)-(iii) and (iv).
For statement (ii), recall that $S_w(z, z') = S_w(z, z') - \gamma_z^T S_{xY}(z'), w - \gamma_z^T S_{xY}(z), w + \gamma_z^T S_{xx, w} \gamma_z'$ from the proof of Lemma 1. The result follows from

$$
\Psi(z, z'; \gamma) = W^{-1} \sum_{w=1}^{W} M_w^{-1} H_w(z, z') S_w(z, z'; \gamma)
$$

$$
= W^{-1} \sum_{w=1}^{W} M_w^{-1} H_w(z, z') \{S_w(z, z') - \gamma_z^T S_{xY}(z'), w - \gamma_z^T S_{xY}(z), w + \gamma_z^T S_{xx, w} \gamma_z'\}
$$

$$
= \Psi(z, z') - \gamma_z^T \Psi_{xY}(z, z') - \gamma_z^T \Psi_{xY}(z', z) + \gamma_z^T \psi_{x} (z, z') \gamma_z'
$$

$$
= \Psi(z, z') + o(1)
$$

for arbitrary $\gamma = (\gamma_z)_{z \in \mathcal{T}}$ by (S33).

For statement (iii), let

$$
e_{1,w}(z) = U_w(z) - \bar{Y}(z), \quad e_{2,w}(z) = U_w(z) - \alpha_w \bar{Y}(z),
$$

$$
e_{3,w}(z) = U_w(z) - \bar{Y}(z) - v_w^T \theta_z, \quad e_{4,w}(z) = U_w(z) - \alpha_w \bar{Y}(z) - v_w^T \theta_z,
$$

$$
e_{5,w}(z) = U_w(z) - \bar{Y}(z) - c_w^T \theta'_z, \quad e_{6,w}(z) = U_w(z) - \bar{Y}(z) - c_w^T \gamma_{c,z},
$$

where $\theta_z \in \mathbb{R}^J$ and $\theta'_z \in \mathbb{R}^{J+1}$ are arbitrary vectors, and $\gamma_{c,z}$ is the coefficient of $e_w$ from the OLS fit of $U_w(z)$ on $(1, e_w)$ over $w = 1, \ldots, W$ with $e_{6,w}(z)$ as the corresponding residual.

Let $S_k(z, z') = (W - 1)^{-1} \sum_{w=1}^{W} e_{k,w}(z) e_{k,w}(z') = (W - 1)^{-1} \{e_k(z)\}^T e_k(z)$, where $e_k(z) = (e_{k,1}(z), \ldots, e_{k,W}(z))^T$, be the covariance of $\{e_{k,w}(z), e_{k,w}(z')\}_{w=1}^{W}$ for $k = 1, \ldots, 6$, summarized in lexicographical order as $S_k = (S_k(z, z'))_{z, z' \in \mathcal{T}}$.

Note that $e_{k,w}(z) - e_{6,w}(z)$ is a linear combination of $c_w$ for all $k = 1, \ldots, 5$. The theory of least squares ensures that $\{e_k(z) - e_6(z)\}^T e_6(z') = 0$ for $k = 1, \ldots, 5$ with arbitrary $\theta_z$ and $\theta'_z$ such that

$$
S_k(z, z') = (W - 1)^{-1} \{e_k(z) - e_6(z) + e_6(z)\}^T \{e_k(z') - e_6(z') + e_6(z')\}
$$

$$
= S_6(z, z') + (W - 1)^{-1} \{e_k(z) - e_6(z)\}^T \{e_k(z') - e_6(z')\},
$$

$$
S_k - S_6 = (W - 1)^{-1} \begin{pmatrix}
e_{k-6}(00) \\
e_{k-6}(01) \\
e_{k-6}(10) \\
e_{k-6}(11)
\end{pmatrix} \begin{pmatrix}
e_{k-6}(00), e_{k-6}(01), e_{k-6}(10), e_{k-6}(11)\end{pmatrix} \geq 0,
$$

where $e_{k-6}(z) = e_k(z) - e_6(z)$, for $k = 1, \ldots, 5$. This, together with $S_1 = S_{ht}$; $S_2 = S_{haj}$; $S_3 = S_{ag, c}$ for $\phi = F, L$ when $\theta_z = \gamma_{ag}$ and $\gamma_{ag, z}$, respectively; $S_4 = S_{wls, c}$ for $\phi = F, L$ when $\theta_z = \gamma_{wls}$ and $\gamma_{wls, z}$, respectively; and $S_5 = S_{ag, c, \phi}(\alpha, v)$ for $\phi = F, L$ when $\theta'_z = \gamma_{ag}(\alpha, v)$ and $\gamma_{ag, c}(\alpha, v)$, respectively, ensures

$$
S_6 \leq S_{ht}, S_{haj}, S_{wls, c}, S_{ag, c}, S_{ag, c, \phi}(\alpha, v) \quad \text{for} \quad \phi = F, L.
$$

(S37)
In addition, the Frisch–Waugh–Lovell theorem ensures that
\[
\gamma_{c,z} = \left( \frac{1}{W-1} \sum_{w=1}^{W} c_w c_w^T \right)^{-1} \left[ \frac{1}{W-1} \sum_{w=1}^{W} c_w \{ U_w(z) - \bar{Y}(z) \} \right], \quad \tilde{\gamma}_{ag,z}(\alpha, v) = \tilde{T}^{-1}_{cc,z} \tilde{T} U_{c,z}
\]
with
\[
\tilde{T}_{cc,z} = W_a^{-1} \sum_{w \in W(z)} \{ \hat{c}_w(z) - \hat{c}_{ht}(z) \} \{ \hat{c}_w(z) - \hat{c}_{ht}(z) \}^T
\]
\[
= W_a^{-1} \sum_{w \in W(z)} \hat{c}_w(z) \{ \hat{c}_w(z) \}^T - \hat{c}_{ht}(z) \{ \hat{c}_{ht}(z) \}^T = \frac{1}{W-1} \sum_{w=1}^{W} c_w c_w^T + o_P(1),
\]
\[
\tilde{T}_{cU,z} = W_a^{-1} \sum_{w \in W(z)} \{ \hat{c}_w(z) - \hat{c}_{ht}(z) \} \{ \bar{U}_w(z) - \bar{Y}_{ht}(z) \}
\]
\[
= W_a^{-1} \sum_{w \in W(z)} \hat{c}_w(z) \bar{U}_w(z) - \hat{c}_{ht}(z) \bar{Y}_{ht}(z) = \frac{1}{W-1} \sum_{w=1}^{W} c_w \{ U_w(z) - \bar{Y}(z) \} + o_P(1)
\]
by Lemma S2. This ensures \( \tilde{\gamma}_{ag,z}(\alpha, v) - \gamma_{c,z} = o_P(1) \) such that \( \gamma_{ag,z}(\alpha, v) - \gamma_{c,z} = o(1) \) and \( S_{ag,z}(\alpha, v) - S_6 = o(1) \). Statement (ii) then follows from (S27).

\[\square\]

S4. Special case and extensions

S4.1. Uniform designs

We outline in this subsection the unification of the three fitting schemes under Condition 0. The results furnish the theoretical guarantees by the sample-mean estimator and the corresponding "ols" fitting scheme under uniform designs.

Let \( S, \Sigma, \text{ and } \hat{V} \) be the common values of \( S_* \), \( \Sigma_* \), and \( \hat{V}_* \) for * = ht, haj under Condition 0, and let \( \bar{Y} \) be the common value of \( \bar{Y}_{sm} = \bar{Y}_{ht} = \bar{Y}_{haj} \) under Condition 0 by Proposition 0. Corollary S1 states the validity of \( (\hat{Y}, \hat{V}) \) for the large-sample Wald-type inference under Condition 0.

**Corollary S1.** Under the \( 2^2 \) split-plot randomization and Condition 0, we have \( E(\bar{Y}_{sm}) \neq \bar{Y} \) and \( \bar{Y}_{sm} \neq \bar{Y} + o_P(1) \) in general. Further assume Condition 0. We have
\[
\sqrt{W}(\hat{Y} - \bar{Y}) \sim \mathcal{N}(0, \Sigma), \quad W\hat{V} - \Sigma = S + o_P(1)
\]
with \( S \geq 0 \).

Let \( \tilde{V}_{ols}, \tilde{V}_{ols,\ell}, \text{ and } \tilde{V}_{ols,\ell} \) be the cluster-robust covariances of \( \tilde{\beta}_{ols}, \tilde{\beta}_{ols,\ell}, \text{ and } \tilde{\beta}_{ols,\ell} \) from the ols fits of the unit models (0), (S), and (III), respectively. Let \( \tilde{\tau}'_{ols} \) be the coefficients of the non-intercept terms from the ols fit of the factor-based specification (III), and let \( \tilde{\Omega}'_{ols} \) be the corresponding cluster-robust covariance.
Recall \( \tilde{\gamma}_{\text{ols}} \) as the coefficient of \( x_{ws} \) from the \( \text{ols} \) fit of (S). Denote by \( \gamma_{\text{ols}} \) the probability limit of \( \tilde{\gamma}_{\text{ols}} \) under Conditions [1] from Lemma S1. Let \( S_{\text{ols,F}} \) and \( \Sigma_{\text{ols,F}} \) be the analogs of \( S \) and \( \Sigma \) defined on the adjusted potential outcomes \( Y_{ws}(z; \gamma_{\text{ols}}) = Y_{ws}(z) - x_{ws}^T \gamma_{\text{ols}} \), respectively.

**Proposition S1.** For split-plot type data that satisfy Condition [1] we have

\[
\begin{align*}
\tilde{\beta}_{\text{ols}} &= \tilde{\beta}_{\text{wls}} = \hat{Y}, \\
\tilde{V}_{\text{ols}} &= \tilde{V}_{\text{wls}} = \hat{V}_a = \text{diag} \left( \frac{W_0 - 1}{W_0} I_2, \frac{W_1 - 1}{W_1} I_2 \right) \cdot \hat{V}, \\
(\tilde{\beta}_{\text{ols,L}}, \tilde{e}_{\text{ols,L}}) &= (\tilde{\beta}_{\text{wls,L}}, \tilde{e}_{\text{wls,L}}), \\
\hat{e}_{\text{ols}} &= G_0 \tilde{\beta}_{\text{ols}}, \\
\hat{Q}'_{\text{ols}} &= G_0 \tilde{V}_{\text{ols}} G_0^T 
\end{align*}
\]

irrespective of the true data generating process. Further assume split-plot randomization and Conditions [2][3]. We have

\begin{enumerate}[(i)]
\item \( \tilde{\beta}_{\text{ols,F}} = \hat{Y} + o_p(1), \quad \sqrt{W} (\tilde{\beta}_{\text{ols,F}} - \hat{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_{\text{ols,F}}), \quad W \hat{V}_{\text{ols,F}} - \Sigma_{\text{ols,F}} = S_{\text{ols,F}} + o_p(1); \\
\item \( \tilde{\beta}_{\text{ols,L}} = \hat{Y} + o_p(1), \quad \sqrt{W} (\tilde{\beta}_{\text{ols,L}} - \hat{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_{\text{wls,L}}), \quad W \hat{V}_{\text{ols,L}} - \Sigma_{\text{wls,L}} = S_{\text{wls,L}} + o_p(1).
\end{enumerate}

**Proof of Proposition S1.** Condition [1] ensures \( a_w = 1 \) and \( q_{wb} = M_b/M = q_b \) for all \( w = 1, \ldots, W \) such that

\[
N = WM, \quad N_z = WaM_b, \quad p_{ws}(z) = p_aq_b = N_z/N, \quad \pi_{ws} = N_{ws}^{-1}
\]

for all \( ws \in S \) and \( z = (ab) \in T \). The numerical equivalence between \( \tilde{\beta}_{\text{ols}} \) and \( \tilde{\beta}_{\text{wls}} \) and that between \( (\tilde{\beta}_{\text{ols,L}}, \tilde{e}_{\text{ols,L}}) \) and \( (\tilde{\beta}_{\text{wls,L}}, \tilde{e}_{\text{wls,L}}) \) in (S26) follow from the equivalence between the \( \text{ols} \) and \( \text{wls} \) fitting schemes under treatment-specific regressions like (S25) with \( \pi_{ws} = N_{ws}^{-1} \) being constant for all units under the same treatment. The numerical result on the robust covariances in (S28) follows from the definition of \( \hat{V}_{\text{ols}} \) and Theorem 3.

In particular,

\[
\hat{V}_{\text{ols}} = (D^TD)^{-1} \left( \sum_{w=1}^W D_w^T e_{\text{ols,w}} e_{\text{ols,w}}^T D_w \right) (D^TD)^{-1}
\]

by definition, where \( e_{\text{ols,w}} = (e_{\text{ols,w,1}}, \ldots, e_{\text{ols,w,M_w}})^T \) are the \( \text{ols} \) residuals in whole-plot \( w \). Proposition 2 ensures \( e_{\text{ols,ws}} = Y_{ws} - \tilde{\beta}_{\text{ols}}(Z_{ws}) = Y_{ws} - \hat{Y}(Z_{ws}) \). Let \( \hat{e}_{\text{ols,w}}(z) = M_b^{-1} \sum s:Z_{ws}=z e_{\text{ols,ws}} \) be the sample mean residual in whole-plot \( w \), with \( \hat{e}_{\text{ols,w}}(z) = \hat{Y}_w(z) - \hat{Y}(z) \) for \( z \in \{(A_w0), (A_w1)\} \) and \( \hat{e}_{\text{ols,w}}(z) = 0 \) for \( z \not\in \{(A_w0), (A_w1)\} \), vectorized as \( \hat{e}_{\text{ols,w}} = (\hat{e}_{\text{ols,w}(00)}, \hat{e}_{\text{ols,w}(01)}, \hat{e}_{\text{ols,w}(10)}, \hat{e}_{\text{ols,w}(11)})^T \).

Set \( Y_w = e_{\text{ols,w}} \) in (S21) to see \( D_w^T e_{\text{ols,w}} = MQ \hat{e}_{\text{ols,w}} \), where \( Q = I_2 \otimes \text{diag}(q_0, q_1) \) is the common value of \( Q_w \) over all \( w \) under Condition [3]. This ensures

\[
\sum_{w=1}^W D_w^T e_{\text{ols,w}} e_{\text{ols,w}}^T D_w = M^2 Q \left( \sum_{w=1}^W \hat{e}_{\text{ols,w}} \hat{e}_{\text{ols,w}}^T \right) Q = Q_{\text{ols}}(z, z')_{z, z' \in T}
\]
with

\[ \Omega_{\text{ols}}(z, z') = M_{g}M_{y} \sum_{w=1}^{W} \tilde{e}_{\text{ols},w}(z)\tilde{e}_{\text{ols},w}(z') = \begin{cases} M_{q}M_{y}(W_{a} - 1)\tilde{S}(z, z') & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a = a', \\ 0 & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a \neq a'. \end{cases} \]

The result on \( \tilde{V}_{\text{ols}} \) follows from (S30), (S31), and \( D^{T}D = \text{diag}(W_{0}M_{0}, W_{0}M_{1}, W_{1}M_{0}, W_{1}M_{1}) \) by (S21) and (S33).

Recall the equivalence of the sample-mean estimator with the Horvitz–Thompson and Hajek estimators under Condition ~. Statement ~ follows from Proposition ~ and Lemma ~ with \( \tilde{\gamma}_{\text{ols}} = \gamma_{\text{ols}} + o_{p}(1) \) by Lemma ~.

For statement ~, the asymptotic Normality of \( \tilde{\beta}_{\text{ols,l}} \) follows from \( \tilde{\beta}_{\text{ols,l}}, \tilde{\gamma}_{\text{ols,z}} = (\tilde{\beta}_{\text{wls,l}}, \tilde{\gamma}_{\text{wls,z}}) \) under Condition ~. The asymptotic conservativeness of \( \tilde{V}_{\text{ols,l}} \) follows from the same reasoning as that for \( \tilde{V}_{\text{wls,l}} \) in the proof of Theorem ~.

**S4.2. HC2 correction for the cluster-robust covariance estimators**

The classic cluster-robust covariances so far recover \( \tilde{V}_{\text{ht}} \) and \( \tilde{V}_{\text{haj}} \) only asymptotically. We verify in this subsection the exact recovery of \( \tilde{V}_{\text{ht}} \) by the HC2 correction in finite samples. Let

\[
\tilde{V}_{\text{ols2}} = (D^{T}D)^{-1}\left\{ \sum_{w=1}^{W} D_{w}^{T}(I - P_{\text{ols,w}})^{-1/2}e_{\text{ols},w}e_{\text{ols},w}^{T}(I - P_{\text{ols,w}})^{-1/2}D_{w} \right\} (D^{T}D)^{-1},
\]

\[
\tilde{V}_{\text{ag2}} = (D_{\text{ag}}^{T}D_{\text{ag}})^{-1}\left\{ \sum_{w=1}^{W} D_{\text{ag},w}^{T}(I - P_{\text{ag},w})^{-1/2}e_{\text{ag},w}e_{\text{ag},w}^{T}(I - P_{\text{ag},w})^{-1/2}D_{\text{ag},w} \right\} (D_{\text{ag}}^{T}D_{\text{ag}})^{-1}
\]

be the HC2 variants of \( \tilde{V}_{\text{ols}} \) and \( \tilde{V}_{\text{ag}} \), respectively, with \( P_{\text{ols},w} = D_{w}(D^{T}D)^{-1}D_{w}^{T} \) and \( P_{\text{ag},w} = D_{\text{ag},w}(D_{\text{ag}}^{T}D_{\text{ag}})^{-1}D_{\text{ag},w}^{T} \) for \( w = 1, \ldots, W \).

**Theorem S1.** \( \tilde{V}_{\text{ag2}} = \tilde{V}_{\text{ht}} \). Under Condition ~, \( \tilde{V}_{\text{ols2}} = \tilde{V} \).

**Proof of Theorem S1.** For the result on \( \tilde{V}_{\text{ag2}} \), it follows from \( D_{\text{ag}}^{T}D_{\text{ag}} = \text{diag}(W_{0}, W_{1}) \otimes I_{2} \) by (S21) and \( D_{\text{ag},w} = 1(A_{w} = 0) \cdot (I_{2}, 0_{2 \times 2}) + 1(A_{w} = 1) \cdot (0_{2 \times 2}, I_{2}) \) by definition that \( P_{\text{ag},w} = D_{\text{ag},w}(D_{\text{ag}}^{T}D_{\text{ag}})^{-1}D_{\text{ag},w}^{T} \) for \( w = 1, \ldots, W \).

\[
\sum_{w=1}^{W} D_{\text{ag},w}^{T}(I - P_{\text{ag},w})^{-1/2}e_{\text{ag},w}e_{\text{ag},w}^{T}(I - P_{\text{ag},w})^{-1/2}D_{\text{ag},w} = \sum_{w=1}^{W} (1 - W_{A_{w}}^{-1})^{-1} \cdot D_{\text{ag},w}^{T}e_{\text{ag},w}e_{\text{ag},w}^{T}D_{\text{ag},w}
\]

\[
= \sum_{a=0,1}^{W_{a} - 1} \sum_{w:A_{w}=a} D_{\text{ag},w}^{T}e_{\text{ag},w}e_{\text{ag},w}^{T}D_{\text{ag},w}
\]

such that

\[
\tilde{V}_{\text{ag2}} = \left\{ \text{diag} \left( \frac{W_{0}}{W_{0} - 1}, \frac{W_{1}}{W_{1} - 1} \right) \otimes I_{2} \right\} \cdot \tilde{V}_{\text{ag}}
\]

by the form of \( \sum_{w:A_{w}=a} D_{\text{ag},w}^{T}e_{\text{ag},w}e_{\text{ag},w}^{T}D_{\text{ag},w} \) from (S22). The result follows from Theorem ~.

S27
For the result on $\tilde{V}_{\text{ols2}}$ under Condition $\mathbb{II}$, let $D_{w,z} = \{1(Z_{w1} = z), \ldots, 1(Z_{wM_w} = z)\}^T$ be the column in $D_w$ that corresponds to treatment $z$. Condition $\mathbb{II}$ ensures $M_w = M$ and $M_{wb} = M_b$ for all $w = 1, \ldots, M$ and $b = 0, 1$. Without loss of generality, reorganize the units such that the first $M_0$ sub-plots in whole-plot receive level 0 of factor B. We have $D_w = (D_{w,00}, D_{w,01}, D_{w,10}, D_{w,11})$, where

$$D_{w,00} = (1^T_{M_0}, 0^T_{M_1}), \quad D_{w,01} = (0^T_{M_0}, 1^T_{M_1}), \quad D_{w,10} = D_{w,11} = 0_M \quad \text{for } w \in \{w : A_w = 0\};$$

$$D_{w,00} = D_{w,01} = 0_M, \quad D_{w,10} = (1^T_{M_0}, 0^T_{M_1}), \quad D_{w,11} = (0^T_{M_0}, 1^T_{M_1}) \quad \text{for } w \in \{w : A_w = 1\}.$$ 

This, together with $D^T D = \text{diag}(N_z)_{z \in \mathbb{T}}$ where $N_z = W_a M_b$, ensures

$$P_{\text{ols},w} = D_w(D^T D)^{-1} D^T_w = \sum_{z \in \mathbb{T}} N_z^{-1} D_{w,z} D^T_{w,z} = W_{A_w}^{-1} \begin{pmatrix} M_0^{-11} & 0 \\ 0 & M_1^{-11} \end{pmatrix}.$$

Note that $(1^T_{M_0}, 0^T_{M_1})^T$ and $(0^T_{M_0}, 1^T_{M_1})^T$ are both eigen-vectors of $P_{\text{ols},w}$, corresponding to the same eigen-value $W_{A_w}^{-1}$. They are thus also the eigen-vectors of $(I - P_{\text{ols},w})^{-1/2}$, corresponding to the same eigen-value $(1 - W_{A_w}^{-1})^{-1/2}$. With the two non-zero columns of $D_w$ equaling $(1^T_{M_0}, 0^T_{M_1})^T$ and $(0^T_{M_0}, 1^T_{M_1})^T$, respectively, we have $(I - P_{\text{ols},w})^{-1/2} D_w = (1 - W_{A_w}^{-1})^{-1/2} D_w$ such that

$$\sum_{w=1}^{W} D^T_w (I - P_{\text{ols},w})^{-1/2} e_{\text{ols},w} e^T_{\text{ols},w} (I - P_{\text{ols},w})^{-1/2} D_w = \sum_{a=0,1} W_a A_w = \sum_{a=0,1} W_a \left( \sum_{w : A_w = a} D^T_w e_{\text{ols},w} e^T_{\text{ols},w} D_w \right).$$

This ensures

$$\tilde{V}_{\text{ols2}} = \left\{ \text{diag} \left( \frac{W_0}{W_0 - 1}, \frac{W_1}{W_1 - 1} \right) \otimes I_2 \right\} \tilde{V}_{\text{ols}}$$

by the form of $\sum_{w : A_w = a} D^T_w e_{\text{ols},w} e^T_{\text{ols},w} D_w$ from (S44). The result for $\tilde{V}_{\text{ols2}}$ follows from Proposition S11.

\[ \square \]

**S4.3. Covariate adjustment via factor-based linear models**

We furnish in this subsection the details on covariate adjustment under factor-based regressions.

Let $f_{ws} = ((A_w - 1/2), (B_{ws} - 1/2), (A_w - 1/2)(B_{ws} - 1/2))^T$ and $f_w(A_w b) = ((A_w - 1/2), (b - 1/2), (A_w - 1/2)(b - 1/2))^T$ be the vectors of the non-intercept regressors of (II) and (III), respectively. Extend (II) and (III) to the additive specifications

$$Y_{ws} \sim 1 + f_{ws} + x_{ws}, \quad (S42)$$

$$\hat{U}_w(A_w b) \sim 1 + f_w(A_w b) + \hat{v}_w(A_w b), \quad (S43)$$

and the fully-interacted specifications

$$Y_{ws} \sim 1 + f_{ws} + x_{ws} + f_{ws} \otimes x_{ws} \quad (S44)$$

S28
Let $\hat{r}_{\text{wls},\ell}, \hat{r}_{\text{ag},\ell}, \hat{r}_{\text{wls},l}$, and $\hat{r}_{\text{ag},l}$ be the coefficients of $f_{\text{us}}$ and $f_\ell(z)$ from (S12)–(S45) under fitting schemes “wls” and “ag”, respectively, with cluster-robust covariances $\hat{\Omega}_{\ell,\ell}$ for $(\ell, \ell) \in \{\text{wls}, \text{ag}\} \times \{\ell, \ell\}$. Let $\hat{r}'_{\text{ag},\ell}(\alpha, \nu)$ and $\hat{\Omega}'_{\ell,\ell}(\alpha, \nu)$ be the variants of $\hat{r}'_{\text{ag},\ell}$ and $\hat{\Omega}'_{\ell,\ell}$ after including the centered whole-plot size factor, $(a_w - 1)$, as an additional whole-plot level covariate in addition to $\hat{v}_w(A_w b)$. Proposition S2 follows from the invariance of least squares and, together with Theorems S2 and Proposition S3, ensures the optimality of $\hat{r}'_{\text{ag},\ell}(\alpha, \nu)$ for estimating the standard factorial effects $(\tau_\text{A}, \tau_\text{B}, \tau_\text{AB})^T = G_0 \hat{y}$ among $\{\hat{r}'_{\text{wls}}, \hat{r}'_{\text{wls},\ell}; \hat{r}'_{\text{ag},\ell}; \hat{r}'_{\text{ag},\ell}(\alpha, \nu) : \ell = \text{wls, ag}\}$. Proposition S3 follows from almost identical reasoning and is thus omitted.
S4.4. General $T_A \times T_B$ split-plot design

Renew $S_{ht}$, $S_{haj}$, and $S_w$ as the scaled between and within whole-plot covariance matrices for $Y_{ws}(z)$ and $Y_{ws}'(z)$ under the general $T_A \times T_B$ split-plot design, and let

$$\Sigma_* = H \circ S_* + W^{-1} \sum_{w=1}^{W} M_w^{-1}(H_w \circ S_w) \quad \text{for } * = ht, haj$$

with $H = \text{diag}(p_a^{-1})_{a \in T_a} \otimes 1_{T_b \times T_a} - 1_{|T_b| \times |T_a|}$, $H_w = \text{diag}(p_a^{-1})_{a \in T_a} \otimes \{\text{diag}(q_{wb}^{-1})_{b \in T_b} - 1_{T_b \times T_a}\}$, $p_a = W_a/W$, and $q_{wb} = M_{wb}/M_w$. We state in Corollary S2 the asymptotic Normality of $\hat{Y}_{ht}$ and $\hat{Y}_{haj}$ under the general $T_A \times T_B$ split-plot experiment. The proof is identical to that of the 2$^2$ case and thus omitted.

**Corollary S2.** Under the $T_A \times T_B$ split-plot randomization, we have $E(\hat{Y}_{ht}) = \bar{Y}$ and $\text{cov}(\hat{Y}_{ht}) = W^{-1}\Sigma_{ht}$. Further assume a generalized version of Condition 2 for $a \in T_A$, $b \in T_B$, and $z = (ab) \in \mathcal{T}$. We have $\sqrt{W}(\hat{Y}_* - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_*)$ for $* = ht, haj$.

The treatment-based unit model equals

$$Y_{ws} \sim \sum_{z \in \mathcal{T}} 1(Z_{ws} = z) \quad \text{(S46)}$$

under the $T_A \times T_B$ split-plot design. Lemma S12 states a numerical result on the invariance of least-squares fits of (S46) to treatment-specific scaling of the fitting weights. The result ensures that the theory we derived under the “wls” fitting scheme with inverse probability weighting extends to a much larger class of fitting weights with no need of modification. See Remark 2 in the main text for an example. We state the lemma in terms of general multi-armed experiment to highlight its generality. The proof follows from direct algebra and is thus omitted.

**Lemma S12.** For a general experiment with treatment levels $\mathcal{T} = \{1, \ldots, Q\}$ and units $i = 1, \ldots, N$, let $Y_i$ denote the outcome, $Z_i \in \mathcal{T}$ denote the treatment assignment, $w_i > 0$ denote an arbitrary weight, and $\rho_i > 0$ denote a scaling factor that is a function of $Z_i$ only. The least-squares fits of $Y_i \sim \sum_{z \in \mathcal{T}} 1(Z_i = z)$ with weights $(w_i)^N_{i=1}$ and $(w_i\rho_i)^N_{i=1}$ yield identical coefficients, heteroskedasticity-robust covariance, and cluster-robust covariance by arbitrary clustering rule.

S4.5. Fisher randomization test

We furnish in this subsection the details on the Fisher randomization test under split-plot randomization. Assume a general $T_A \times T_B$ split-plot experiment with potential outcomes $\{Y_{ws}(z) : ws \in S, z \in \mathcal{T}\}$. The weak null hypothesis concerns

$$H_{0N} : G\bar{Y} = 0_F$$
for some $F \times |T|$ full row rank contrast matrix $G$ with rows orthogonal to $1_{|T|}$. Given observed data $Z = (Z_{ws})_{ws \in S}$, $Y = (Y_{ws})_{ws \in S}$, $X = (x_{ws})_{ws \in S}$, and $Z$ as the set of all possible values $Z$ can take under the split-plot randomization restriction, we can pretend to be testing

$$H_{0F} : Y_{ws}(z) = Y_{ws} \text{ for all } ws \in S \text{ and } z \in T$$

as a strong null hypothesis that is compatible with $H_{0N}$, and compute the p-value as

$$p_{FRT} = |Z|^{-1} \sum_{z \in Z} 1\{ t(z, Y, X) \geq t(Z, Y, X) \}$$

for some arbitrary test statistic $t(Z, Y, X)$. Of interest is the operating characteristics of $p_{FRT}$ when only $H_{0N}$ holds.

Renew $B = \{ \tilde{\beta}_{wls}, \tilde{\beta}_{wls,0}, \tilde{\beta}_{ag}, \tilde{\beta}_{ag,0}, \tilde{\beta}_{ag,0}(\alpha, v) : \diamond = f, l \}$ as the collection of regression estimators of $\tilde{Y}$ that are consistent under general $T_\Lambda \times T_B$ split-plot design. Let $t^2(\tilde{\beta}) = (G\tilde{\beta})'(G\tilde{V}G')^{-1}G\tilde{\beta}$ be the studentized test statistic based on $\tilde{\beta} \in B$, with $\tilde{V}$ as the corresponding cluster-robust covariance. The studentization ensures that the resulting $p_{FRT}$ controls the type one error rates asymptotically in the sense of

$$\lim_{N \to \infty} \mathbb{P}(p_{FRT} \leq \alpha) \leq \alpha \quad \text{for all } \alpha \in (0, 1)$$

for all $\tilde{\beta} \in B$. The Fisher randomization test with $t^2(\tilde{\beta})$ is therefore finite-sample exact for testing the strong null hypothesis and asymptotically valid for testing the weak null hypothesis under split-plot randomization for all $\tilde{\beta} \in B$. The duality between confidence interval and hypothetical testing further ensures that the test based on $t^2(\tilde{\beta}_{ag,l}(\alpha, v))$ has the highest power asymptotically when the generalized version of $(\text{II2})$ holds. The same theoretical guarantees also hold for $\tilde{\beta} \in \{ \tilde{\beta}_{ols}, \tilde{\beta}_{ols,f}, \tilde{\beta}_{ols,l} \}$ under Condition $\text{II}$.

### S5. Simulation

Define a regression scheme as the combination of model specification and fitting scheme. We illustrate in this section the validity and efficiency of thirteen regression schemes for estimating the standard factorial effects under the $2^2$ split-plot design, summarized in Table S1.

Consider a $2^2$ split-plot experiment with a study population nested in $W = 300$ whole-plots. We set $(W_0, W_1) = (0.7W, 0.3W)$ and generate $(M_w, M_{w1}, M_w)_w = (M_w0, \max(2, \zeta_w0), M_{w1} = \max(2, \zeta_w1)$, and $M_w = M_w0 + M_{w1}$, respectively, with the $\zeta_w0$’s being i.i.d. Pois(5) and the $\zeta_w1$’s being i.i.d. Pois(3). For each $w = 1, \ldots, W$, we draw a scalar group-level covariate $x_w$ from $N(0.2, 0.5)$, and set $x_{ws} = x_w$ for $s = 1, \ldots, M_w$. The potential outcomes are then generated as

$$Y_{ws}(00) = \theta_w + 0.5 + 2x_{ws}^2 + \epsilon_{ws}, \quad Y_{ws}(01) = -0.5\theta_w + 1 + x_{ws}^2 + \epsilon_{ws},$$

$$Y_{ws}(10) = 0.5\theta_w + 1 - x_{ws}^2 + \epsilon_{ws}, \quad Y_{ws}(11) = \theta_w + 2 + 2x_{ws}^2 + \epsilon_{ws}$$
Table S1: Thirteen regression schemes based on the unadjusted, additive (“o”), and fully-interacted (“l”) factor-based models. Recall \(f_{ws} = ((A_w - 1/2), (B_{ws} - 1/2), (A_w - 1/2)(B_{ws} - 1/2))^T\) and \(f_{w}(A_w b) = ((A_w - 1/2), (b - 1/2), (A_w - 1/2)(b - 1/2))^T\) as the vectors of the non-intercept regressors from (1) and (2), respectively. Prefixes “ols”, “wls”, and “ag” indicate the fitting schemes. Suffixes “x.F” and “x.L” denote the additive and fully-interacted specifications for covariate adjustment, respectively. For the aggregate models, we use “x”, “m”, and “xm” to indicate three choices of covariate combinations: (i) uses the scaled whole-plot total covariates \(^T\) as the vectors of the non-intercept regressors from \(\) and \(\) f, respectively. Prefixes “ols”, “wls”, and “ag” indicate the fitting schemes. Suffixes “x.F” and “x.L” denote the additive and fully-interacted specifications for covariate adjustment, respectively. For the aggregate models, we use “x”, “m”, and “xm” to indicate three choices of covariate combinations: (i) uses the scaled whole-plot total covariates \(\) (ii) uses the whole-plot size factor \(\) (iii) uses both (“xm”).

| regression scheme | fitting scheme | model specification with \(z = (A_w b)\) for “ag” |
|-------------------|---------------|--------------------------------------------------|
| ols               | ols           | \(1 + f_{ws}\) |
| ols.x.F           |               | \(1 + f_{ws} + x_{ws}\) |
| ols.x.L           |               | \(1 + f_{ws} + x_{ws} + f_{ws}x_{ws}\) |
| wls               | wls           | \(1 + f_{ws}\) |
| wls.x.F           |               | \(1 + f_{ws} + x_{ws}\) |
| wls.x.L           |               | \(1 + f_{ws} + x_{ws} + f_{ws}x_{ws}\) |
| ag                | ag            | \(1 + f_{w}(z)\) |
| ag.x.F            |               | \(1 + f_{w}(z) + \hat{v}_{w}(z)\) |
| ag.x.L            |               | \(1 + f_{w}(z) + \hat{v}_{w}(z) + f_{w}(z)\hat{v}_{w}(z)\) |
| ag.m.F            |               | \(1 + f_{w}(z) + (\alpha_w - 1)\) |
| ag.m.L            |               | \(1 + f_{w}(z) + (\alpha_w - 1) + f_{w}(z)(\alpha_w - 1)\) |
| ag.xm.F           |               | \(1 + f_{w}(z) + (\alpha_w - 1) + \hat{v}_{w}(z)\) |
| ag.xm.L           |               | \(1 + f_{w}(z) + (\alpha_w - 1) + \hat{v}_{w}(z) + f_{w}(z)(\alpha_w - 1) + f_{w}(z)\hat{v}_{w}(z)\) |

for \(ws \in S\), where the \(\theta_w\)’s are i.i.d. \(\mathcal{N}\{2M_w/ \max(M_w), 0.2\}\) and the \(\epsilon_{ws}\)’s are i.i.d. \(\text{Unif}(-1,1)\). Fix \(Y_{ws}(z), x_{ws} : z \in T, ws \in S\) in simulation. We draw a random permutation of \(W_1\) 1’s and \(W_0\) 0’s to assign levels of factor A at the whole-plot level, and then, for each \(w = 1, \ldots, W\), draw a random permutation of \(M_{w1}\) 1’s and \(M_{w0}\) 0’s to assign levels of factor B in whole plot \(w\).

The procedure is repeated 2,000 times, with the biases (“bias”), true standard deviations (“sd”), average cluster-robust standard errors (“ese”), and coverage rates of the 95% confidence intervals based on the cluster-robust standard errors (“coverage”) for all three standard effects summarized in Figures S31 and S32. We separate the results into “unadjusted vs. additive regressions” and “unadjusted vs. fully-interacted regressions” for ease of display.

In particular, Figure S31 shows the comparison between the unadjusted models and their additive counterparts. The first row illustrates the biases in the ols estimators under non-uniform split-plot designs. The second row illustrates the efficiency gain by covariate adjustment for estimating the whole-plot factor effect. The covariate-adjusted models “ols.x”, “wls.x”, “ag.m”, and “ag.mx” yield less variable estimators than their unadjusted counterparts under all three fitting schemes. The comparison between “ag”, “ag.m”, “ag.x”, and “ag.mx” under the “ag” fitting scheme highlights the importance of whole-plot size adjustment for improving efficiency. The results for the sub-plot factor effect and interaction, on the other hand, remain unchanged under the “wls” and “ag” fitting.
schemes as Proposition $S3$ suggests.

The third row shows the average cluster-robust standard errors for estimating the true standard deviations. Compare it with the second row to see the conservativeness that is coherent with Theorems $2-4$. The last row illustrates the validity of the regression estimators for conducting the Wald-type inference. The overall conservativeness is, again, coherent with Theorems $2-4$.

Figure $S2$ shows the comparison between the unadjusted models with their fully-interacted counterparts. In addition to the same observations as in the additive case, it also illustrates the efficiency gain by covariate adjustment for estimating the sub-plot factor effect and interaction as well. The “a.mx.L” regression scheme, as the theory suggests, secures the highest efficiency overall.

Last but not least, we furnish the results when the covariates are not constant within each whole-plots. Inherit all settings from above except that we now generate $x_{ws}$ as $x_{ws} = x_{w} + \epsilon_{ws}$, where $\epsilon_{ws}$ are i.i.d. $N(0,0.5)$. We summarize the results in Figures $S3-S4$. The gain in efficiency by covariate adjustment for estimating the sub-plot factor effect and interaction surfaces under additive models as well.
Figure S1: Comparison of the least-squares estimators from the unadjusted and additive regressions with $x_{ws} = x_w$. We suppress the suffix “f” in the names of the covariate-adjusted regressions to save some space.

Figure S2: Comparison of the least-squares estimators from the unadjusted and fully-interacted regressions with $x_{ws} = x_w$. We suppress the suffix “L” in the names of the covariate-adjusted regressions to save some space.
Figure S3: Comparison of the least-squares estimators from the unadjusted and additive regressions with varying $x_{wS}$ within each whole-plot. We suppress the suffix “f” in the names of the covariate-adjusted regressions to save some space.

Figure S4: Comparison of the least-squares estimators from the unadjusted and fully-interacted regressions with varying $x_{wS}$ within each whole-plot. We suppress the suffix “l” in the names of the covariate-adjusted regressions to save some space.