The Annals of Statistics
2022, Vol. 50, No. 2, 1170–1192
https://doi.org/10.1214/21-AOS2144

RECONCILING DESIGN-BASED AND MODEL-BASED CAUSAL INFERENCES FOR SPLIT-PLOT EXPERIMENTS

BY ANQI ZHAO[1,a] AND PENG DING[2,b]

[1]*Department of Statistics and Data Science, National University of Singapore,* [a]*staza@nus.edu.sg*
[2]*Department of Statistics, University of California, Berkeley,* [b]*pengdingpku@berkeley.edu*

The split-plot design arose from agricultural science with experimental units, also known as the subplots, nested within groups known as the whole plots. It assigns different interventions at the whole-plot and subplot levels, respectively, providing a convenient way to accommodate hard-to-change factors. By design, subplots 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 run an ordinary least squares (OLS) regression 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 the parametric assumptions. In contrast, 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, 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 for 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 our theories are either numeric or design-based, and hold regardless of how well the regression equations represent the true data generating process.

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 among the most popular designs in industrial and engineering applications (Jones and Nachtsheim (2009)), and is gaining increasing popularity in social sciences (e.g., Chong et al. (2020), Olken (2007)). It also has deep connections with causal inference with interference (e.g., Basse and Feller (2018), Hudgens and Halloran (2008), Imai, Jiang and Malani (2021)).

Split-plot randomization subjects all units within the same group to the same level of the whole-plot intervention and poses challenges to subsequent inference of the treatment ef-

Received April 2021; revised October 2021.
MSC2020 subject classifications. 62K15, 62J05, 62G05.
Key words and phrases. Cluster randomization, cluster-robust standard error, covariate adjustment, inverse probability weighting, potential outcome, randomization inference.

fects. Model-based analyses often require strong assumptions on the functional forms of the outcome models, for example, Liang and Zeger's (1986) 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., Cox and Reid (2000), Kempthorne (1952), Wu and Hamada (2009)). 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 purely 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 subplot intervention. Mukerjee and Dasgupta (2021) extended the discussion to possibly nonuniform 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 nonuniform 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 nonuniform 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 nonuniform 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 for split-plot experiments.

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 nonuniform 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 regression, as it turns out, ensures the highest asymptotic efficiency when (i) 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 factors of multiple levels. Our paper furthers the growing literature on design-based causal inference with various types of experimental data (e.g., Abadie et al. (2020), Basse and Feller (2018), Box and Andersen (1955), Dasgupta, Pillai and Rubin (2015), Fogarty (2018a, 2018b), Hudgens and Halloran (2008), Imai, King and Nall (2009), Imbens and Rubin (2015), Ji et al. (2017), Kempthorne (1952), Li and Ding (2017), Lin (2013), Liu and Yang (2020), Middleton and Aronow (2015), Miratrix (2013), Mukerjee, Dasgupta and Rubin (2018), Neyman (1923, 1935), Pashley and Miratrix (2021), Rosenbaum (2002), Sabbaghi and Rubin (2014), Schochet (2010), Schochet et al. (2021), Su and Ding (2021), Wu (1981)).

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_∞ and cov_∞ 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, Jiang and Malani (2021), and Su and Ding (2021). The terms "regression" and "cluster-robust covariance" refer to the numeric outputs of the least squares fit without any modeling assumptions; we evaluate their properties under the design-based framework.

2. Setting.

2.1. *Motivating examples.* Consider a study with two interventions or factors of interest and a study population nested in different groups. The split-plot design assigns the two 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. Fricano et al. (2014) conducted a randomized experiment on 14 mice to study the effects of fatty acid delivery and Pten knockdown on soma size of neurons in the brain. The randomization of fatty acid delivery was conducted at the mouse level, and randomly assigned the mice to three levels of exposure. The randomization of Pten knockdown took place at the neuron level, and randomly infected neurons within each mouse with an shRNA against Pten or an mCherry control. The outcome of interest was measured by the soma size of the neurons following the treatments. The number of neurons extracted from each mouse varied depending on the level of infection of each virus. This defines a nonuniform split-plot experiment with fatty acid delivery and Pten knockdown as the whole-plot and subplot factors, respectively.

EXAMPLE 2. Olken (2007) conducted a randomized experiment on 608 villages in Indonesia to study the effects of two interventions on reducing corruption: increasing the probability of external audits ("audits") and increasing participation in accountability meetings ("participation"). 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 participation encouragement measures, on the other hand, was done village by village, and randomly assigned the villages to three levels of encouragement. This defines a nonuniform split-plot experiment with audits and participation constituting the whole-plot and subplot 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, $\mathcal{S} = \{ws : w = 1, \dots, W; s = 1, \dots, M_w\}$, nested in W groups of possibly different sizes, M_w ($w = 1, \dots, W$; $\sum_{w=1}^{W} M_w = N$). A 2^2 split-plot design compounds a cluster randomization with a stratified randomization, and assigns the treatments 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 $w = 1, \dots, W$.

Refer to the first and second steps as the stage (I) and stage (II) randomizations, respectively, with $\{W_a, M_{wb} : a, b = 0, 1;\ w = 1, \dots, W\}$ being some prespecified, fixed 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 $(a, b) \in \mathcal{T} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$; we abbreviate (a, b) as (ab) when no confusion would arise. Refer to each unit as a subplot and each group as a whole plot by convention of the literature on agricultural experiments. Factors A and B become the whole-plot and subplot factors, respectively.

Let $A_{ws} = A_w$ and B_{ws} indicate the levels of factors A and B received by subplot ws, respectively, 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 subplots within the same whole plot. Let $Z_{ws} = (A_w, B_{ws})$ indicate the final treatment for subplot ws with

$$p_{ws}(z) = \mathbb{P}(Z_{ws} = z) = \mathbb{P}(A_w = a) \cdot \mathbb{P}(B_{ws} = b) = p_a q_{wb} \quad \text{for } z = (ab) \in \mathcal{T}.$$

Refer to $p_{ws}(z)$ as the inclusion probability of subplot 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 $\bar{\alpha} = W^{-1} \sum_{w=1}^{W} \alpha_w = 1$. The sample size under treatment $z = (ab)$ equals $N_z = \sum_{w: A_w = a} M_{wb}$ and is in general stochastic unless M_{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 $w = 1, \dots, W$, as specified in the following.

CONDITION 1. *$M_w = M$ and $M_{wb} = M_b$ for all $w = 1, \dots, W$ and $b = 0, 1$.*

Condition 1 ensures that $p_{ws}(z) = p_a q_b = N_z/N$ is identical for all $ws \in \mathcal{S}$, with $z = (ab)$, $q_b = M_b/M$ and $N_z = N p_a q_b$. 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, e.g., Examples 1 and 2).

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

$$\begin{aligned} \tau_{\mathrm{A}} &= 2^{-1}\{\bar{Y}(11) + \bar{Y}(10)\} - 2^{-1}\{\bar{Y}(01) + \bar{Y}(00)\}, \\ \tau_{\mathrm{B}} &= 2^{-1}\{\bar{Y}(11) + \bar{Y}(01)\} - 2^{-1}\{\bar{Y}(10) + \bar{Y}(00)\}, \qquad (1) \\ \tau_{\mathrm{AB}} &= \bar{Y}(11) - \bar{Y}(10) - \bar{Y}(01) + \bar{Y}(00) \end{aligned}$$

define the standard main effects and interaction under 2^2 factorial designs. Sometimes the interaction is also defined as $\tau_{\text{AB}}/2$ (Dasgupta, Pillai and Rubin (2015)); the difference causes no essential change to our discussion. We will discuss inference of the general estimand

$$\tau = G\bar{Y}$$

for arbitrary coefficient matrix G. The standard main effects and interaction in (1) correspond to a special $G = G_0 = (g_{\text{A}}, g_{\text{B}}, g_{\text{AB}})^{\text{T}}$ with $g_{\text{A}} = 2^{-1}(-1, -1, 1, 1)^{\text{T}}$, $g_{\text{B}} = 2^{-1}(-1, 1, -1, 1)^{\text{T}}$ and $g_{\text{AB}} = (1, -1, -1, 1)^{\text{T}}$.

We consider the design-based inference of τ, which conditions on the potential outcomes and views the treatment assignment as the sole source of randomness. The observed outcome equals $Y_{ws} = \sum_{z\in\mathcal{T}} 1(Z_{ws} = z) Y_{ws}(z)$ for subplot ws. We focus on estimators of the form $\hat{\tau} = G\hat{Y}$, where $\hat{Y}$ is some estimator of $\bar{Y}$ based on $(Y_{ws}, Z_{ws})_{ws\in\mathcal{S}}$ and possibly some pretreatment covariates. Assume that $a, b \in \{0, 1\}$ index the levels of factors A and B in treatment combination $z \in \mathcal{T}$ throughout unless specified otherwise.

3. Horvitz–Thompson and Hajek estimators. We review in this section three design-based estimators for estimating $\bar{Y}$. Let $\mathcal{S}(z) = \{ws : Z_{ws} = z,\ ws \in \mathcal{S}\}$ be the set of subplots under treatment z. Let $\mathcal{W}(z)$ be the set of whole plots that contain at least one observation 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. By definition, $\mathcal{W}(z) = \{w : A_w = a\}$ with $|\mathcal{W}(z)| = W_a$ for $z \in \{(a0), (a1)\}$ with level a of factor A.

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

$$\hat{Y}_{\text{sm}}(z) = \left|\mathcal{S}(z)\right|^{-1} \sum_{ws\in\mathcal{S}(z)} Y_{ws} = \left|\mathcal{S}(z)\right|^{-1} \sum_{ws\in\mathcal{S}} 1(Z_{ws} = z) Y_{ws},$$

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

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

$$\hat{Y}_{\text{ht}}(z) = N^{-1} \sum_{ws\in\mathcal{S}(z)} p_{ws}^{-1}(z) Y_{ws} = N^{-1} \sum_{ws\in\mathcal{S}} \frac{1(Z_{ws} = z)}{p_{ws}(z)} Y_{ws}(z). \tag{2}$$

Split-plot randomization ensures $p_{ws}(z) = p_a q_{wb}$ for $z = (ab)$, and simplifies (2) to

$$\hat{Y}_{\text{ht}}(z) = W_a^{-1} \sum_{w\in\mathcal{W}(z)} \alpha_w \hat{Y}_w(z),$$

where $\hat{Y}_w(z) = M_{wb}^{-1} \sum_{s:Z_{ws}=z} Y_{ws}$ is the whole-plot sample mean under treatment $z = (ab)$. 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 $U_w(z) = \bar{M}^{-1} \sum_{s=1}^{M_w} Y_{ws}(z) = \alpha_w \bar{Y}_w(z)$ be the scaled whole-plot total potential outcome with sample analog $\hat{U}_w(z) = \alpha_w \hat{Y}_w(z)$. We have

$$\bar{Y}(z) = W^{-1} \sum_{w=1}^{W} \alpha_w \bar{Y}_w(z) = W^{-1} \sum_{w=1}^{W} U_w(z), \qquad \hat{Y}_{\text{ht}}(z) = W_a^{-1} \sum_{w\in\mathcal{W}(z)} \hat{U}_w(z). \tag{3}$$

This illustrates $\hat{Y}_{\text{ht}}(z)$ as a two-stage sample-mean estimator of $\bar{Y}(z)$ by first using $\hat{U}(z) = W_a^{-1} \sum_{w\in\mathcal{W}(z)} U_w(z)$ to estimate $\bar{Y}(z) = W^{-1} \sum_{w=1}^{W} U_w(z)$ and then using $\hat{U}_w(z)$ to estimate the $U_w(z)$ in $\hat{U}(z)$ for $w \in \mathcal{W}(z)$. Standard results ensure that the two steps are unbiased with regard to the whole-plot and subplot 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)). In contrast, the Hajek estimator

$$\hat{Y}_{\mathrm{haj}}(z) = \frac{\hat{Y}_{\mathrm{ht}}(z)}{\hat{1}_{\mathrm{ht}}(z)}, \quad \text{where } \hat{1}_{\mathrm{ht}}(z) = N^{-1} \sum_{ws \in \mathcal{S}(z)} p_{ws}^{-1}(z),$$

normalizes the Horvitz–Thompson estimator by the sum of the individual weights involved in its definition, and ensures location invariance by construction. We can view $\hat{1}_{\mathrm{ht}}(z)$ as the Horvitz–Thompson estimator of constant 1 when all potential outcomes equal 1. The Hajek estimator is thus a ratio estimator for $\bar{Y}(z) = \bar{Y}(z)/1$ with the numerator and denominator estimated by $\hat{Y}_{\mathrm{ht}}(z)$ and $\hat{1}_{\mathrm{ht}}(z)$, respectively.

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

PROPOSITION 1. *Under Condition 1, we have* $\hat{Y}_{\mathrm{sm}} = \hat{Y}_{\mathrm{ht}} = \hat{Y}_{\mathrm{haj}}$ *with*

$$\hat{Y}_{\mathrm{sm}}(z) = \hat{Y}_{\mathrm{ht}}(z) = \hat{Y}_{\mathrm{haj}}(z) = W_a^{-1} \sum_{w \in \mathcal{W}(z)} \hat{Y}_w(z) \quad \text{for } z = (ab) \in \mathcal{T}.$$

We derive the design-based properties of $\hat{Y}_*$ ($* = \mathrm{sm}, \mathrm{ht}, \mathrm{haj}$) 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 $\{Y_{ws}(z), Y_{ws}(z')\}_{ws \in \mathcal{S}}$ as

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

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

respectively (Mukerjee and Dasgupta (2021)), summarized in $S = (S(z, z'))_{z,z' \in \mathcal{T}}$ and $S_w = (S_w(z, z'))_{z,z' \in \mathcal{T}}$. They measure the between and within whole-plot heterogeneity in potential outcomes after adjusting for the whole-plot sizes. Let

$$H = \mathrm{diag}(p_0^{-1}, p_1^{-1}) \otimes 1_{2\times 2} - 1_{4\times 4}, \qquad H_w = \mathrm{diag}(p_0^{-1}, p_1^{-1}) \otimes \{\mathrm{diag}(q_{w0}^{-1}, q_{w1}^{-1}) - 1_{2\times 2}\}$$

be two symmetric 4×4 matrices defined by the design parameters. Lemma 1 below quantifies the sampling covariance of $\hat{Y}_{\mathrm{ht}}$ in finite samples.

LEMMA 1. *Under the* 2^2 *split-plot randomization, we have*

$$E(\hat{Y}_{\mathrm{ht}}) = \bar{Y}, \qquad \mathrm{cov}(\hat{Y}_{\mathrm{ht}}) = W^{-1}(H \circ S_{\mathrm{ht}} + \Psi)$$

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

Consider Ψ as a summary of $(S_w)_{w=1}^{W}$ after adjusting for the whole-plot sizes. Lemma 1 decomposes the variability in $\hat{Y}_{\mathrm{ht}}$ into that due to the stage (I) randomization, namely $W^{-1}(H \circ S_{\mathrm{ht}})$, and that due to the stage (II) randomization, namely $W^{-1}\Psi$. Lemma S3 in the Supplementary Material (Zhao and Ding (2022)) quantifies this statement rigorously. A direct implication is $\operatorname{var}(g^{\mathrm{T}}\hat{Y}_{\mathrm{ht}}) = W^{-1}g^{\mathrm{T}}(H \circ S_{\mathrm{ht}} + \Psi)g$ for arbitrary $g \in \mathbb{R}^4$. This gives a more compact matrix form of Mukerjee and Dasgupta ((2021), 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 $\hat{Y}_{\mathrm{ht}}$ and $\hat{Y}_{\mathrm{haj}}$ 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

$$\hat{Y}'_{\mathrm{ht}}(z) = N^{-1} \sum_{ws \in \mathcal{S}(z)} p_{ws}^{-1}(z)\, Y'_{ws}(z), \quad \text{where } Y'_{ws}(z) = Y_{ws}(z) - \bar{Y}(z),$$

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

$$\hat{Y}_{\mathrm{haj}}(z) - \bar{Y}(z) = \frac{\hat{Y}_{\mathrm{ht}}(z) - \hat{1}_{\mathrm{ht}}(z)\bar{Y}(z)}{\hat{1}_{\mathrm{ht}}(z)} = \frac{\hat{Y}'_{\mathrm{ht}}(z)}{\hat{1}_{\mathrm{ht}}(z)}. \tag{4}$$

Let $S_{\mathrm{haj}} = (S_{\mathrm{haj}}(z, z'))_{z,z' \in \mathcal{T}}$, where

$$S_{\mathrm{haj}}(z, z') = (W-1)^{-1} \sum_{w=1}^{W} \alpha_w^2 \{\bar{Y}_w(z) - \bar{Y}(z)\}\{\bar{Y}_w(z') - \bar{Y}(z')\}$$

is the scaled between whole-plot covariance of $\{Y'_{ws}(z), Y'_{ws}(z')\}_{ws \in \mathcal{S}}$. Let

$$\overline{\alpha^k} = W^{-1} \sum_{w=1}^{W} \alpha_w^k$$

be the kth moment of $(\alpha_w)_{w=1}^W$ for $k = 1, 2, 4$ with $\bar{\alpha} = W^{-1} \times \sum_{w=1}^W \alpha_w = 1$. Let $\overline{Y_{w\cdot}^4(z)} = M_w^{-1} \sum_{s=1}^{M_w} Y_{ws}^4(z)$ be the uncentered fourth moment of $Y_{ws}(z)$ in whole plot w. We state in Condition 2 below the regularity conditions for asymptotics under split-plot randomization.

CONDITION 2. *As W goes to infinity, for $a, b = 0, 1$ and $z \in \mathcal{T}$:*

(i) $\overline{\alpha^2} = O(1)$; $\overline{\alpha^4} = o(W)$;
(ii) p_a *has a limit in* $(0, 1)$; $\epsilon \le q_{wb} \le 1 - \epsilon$ *for all* $w = 1, \ldots, W$ *for some* $\epsilon \in (0, 1/2]$ *independent of* W;
(iii) $\bar{Y}$, S_{ht}, S_{haj}, *and* Ψ *have finite limits*;
(iv) $\max_{w=1,\ldots,W} |\alpha_w \bar{Y}_w(z) - \bar{Y}(z)|^2 / W = o(1)$;
(v) $W^{-1} \sum_{w=1}^W \alpha_w^2 \overline{Y_{w\cdot}^4(z)} = O(1)$; $W^{-2} \sum_{w=1}^W \alpha_w^4 \overline{Y_{w\cdot}^4(z)} = o(1)$.

For notational simplicity, we will also use p_a, $\bar{Y}$, S_{ht}, S_{haj} and Ψ to denote their respective limiting values when no confusion would arise. The exact meaning should be clear from the context.

With $\bar{\alpha} = 1$, Condition 2(i) requires the finite-population variance of the α_w's 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(ii)–(iii), on the other hand, ensure that $\operatorname{cov}(\hat{Y}_{\mathrm{ht}})$ decays at the rate of W^{-1} and thereby guarantee the consistency of $\hat{Y}_{\mathrm{ht}}$ for estimating $\bar{Y}$. We do not need q_{wb} to converge but only be uniformly bounded as long as Ψ has a finite limit; see Lemma S4 in the Supplementary Material for details. In the neuroscience experiment in Example 1, for example, this imposes bounds on the number of neurons affected by each level of the Pten knockdown intervention. Condition 2(v) stipulates the bounded fourth moment condition peculiar to the split-plot randomization. Provided Condition 2(i), it is satisfied as long as the $\overline{Y^4_{w\cdot}(z)}$'s are uniformly bounded for all w.

Importantly, Condition 2 requires only W goes to infinity and includes both of the following asymptotic regimes as special cases:

(i) M_w goes to infinity for all $w = 1, \dots, W$;
(ii) $\{M_w : w = 1, \dots, W\}$ are uniformly bounded.

Recall from Lemma 1 that $H \circ S_{\text{ht}}$ and $\Psi = W^{-1} \sum_{w=1}^W M_w^{-1}(H_w \circ S_w)$ characterize the variability in $\hat{Y}_{\text{ht}}$ 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 that Ψ has a finite limit. A third asymptotic regime is to have M_w go to infinity for all $w = 1, \dots, W$ 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 for reliable asymptotic approximations under our framework. This requires a large number of mice in Example 1 and a large number of subdistricts in Example 2.

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 $S_{\text{ht}} = O(1)$ and $S_{\text{haj}} = O(1)$ in Condition 2(iii). Consider $N = WM$ units in W equal-sized groups, $\{ws : w = 1, \dots, W;\ s = 1, \dots, M\}$, with $Y_{ws}(z) \sim [\mu_w, \sigma^2]$ and $\mu_w \sim [\mu_0, \sigma_0^2]$. This defines a classical model for characterizing data nested in clusters. Denote by $\mathbb{P}'$ the probability measure induced by the potential outcomes generating process. Standard result shows that $S_{\text{ht}} = O_{\mathbb{P}'}(1)$ as W and M go to infinity, and degenerates to $S_{\text{ht}} = o_{\mathbb{P}'}(1)$ if $\sigma_0 = 0$ and $Y_{ws}(z) \sim [\mu_0, \sigma^2]$.

Theorem 4.1 below states the asymptotic normality of $\hat{Y}_{\text{ht}}$ and $\hat{Y}_{\text{haj}}$.

THEOREM 4.1. *Let* $\Sigma_* = H \circ S_* + \Psi$ *for* $* = \text{ht}, \text{haj}$ *with* $\Sigma_{\text{ht}} = W \text{cov}(\hat{Y}_{\text{ht}})$ *and* $\Sigma_{\text{haj}} = W \text{cov}(\hat{Y}'_{\text{ht}})$ *in finite samples. Under the* 2^2 *split-plot randomization and Condition* 2, *we have*

$$\sqrt{W}(\hat{Y}_* - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_*) \quad \textit{for } * = \text{ht}, \text{haj}.$$

Theorem 4.1 ensures the consistency of $\hat{Y}_{\text{ht}}$ and $\hat{Y}_{\text{haj}}$ for estimating $\bar{Y}$, and establishes Σ_{haj} as the asymptotic sampling covariance of $\sqrt{W}(\hat{Y}_{\text{haj}} - \bar{Y})$. The large-sample relative efficiency between $\hat{Y}_{\text{ht}}$ and $\hat{Y}_{\text{haj}}$ then follows from the comparison of Σ_{ht} and Σ_{haj}.

COROLLARY 1. *Under the* 2^2 *split-plot randomization and Condition* 2, *we have*

$$W\big[\text{var}_\infty\{\hat{Y}_{\text{haj}}(z)\} - \text{var}_\infty\{\hat{Y}_{\text{ht}}(z)\}\big] = \big(p_a^{-1} - 1\big)\big\{S_{\text{haj}}(z, z) - S_{\text{ht}}(z, z)\big\} \quad \textit{for } z = (ab) \in \mathcal{T}$$

with:

(i) $S_{\text{haj}}(z, z) = S_{\text{ht}}(z, z)$ *if* $\bar{Y}(z) = 0$ *or* $\alpha_w = 1$ *for all* w;
(ii) $0 = S_{\text{haj}}(z, z) \le S_{\text{ht}}(z, z)$ *if* $\bar{Y}_w(z)$ *is constant over all* w;
(iii) $0 = S_{\text{ht}}(z, z) \le S_{\text{haj}}(z, z)$ *if* $U_w(z) = \alpha_w \bar{Y}_w(z)$ *is constant over all* w.

Intuitively, $\hat{Y}_{\mathrm{haj}}(z)$ is asymptotically more efficient than $\hat{Y}_{\mathrm{ht}}(z)$ 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 $\hat{Y}_{\mathrm{haj}}(z)$ over $\hat{Y}_{\mathrm{ht}}(z)$.

4.3. *Estimation of the sampling covariances.* The expressions of Σ_* ($* = \mathrm{ht}, \mathrm{haj}$) involve unobserved potential outcomes. We need to estimate them for the Wald-type inference. Let

$$\hat{S}_{\mathrm{ht}}(z, z') = (W_a - 1)^{-1} \sum_{w: A_w = a} \{\alpha_w \hat{Y}_w(z) - \hat{Y}_{\mathrm{ht}}(z)\}\{\alpha_w \hat{Y}_w(z') - \hat{Y}_{\mathrm{ht}}(z')\},$$

$$\hat{S}_{\mathrm{haj}}(z, z') = (W_a - 1)^{-1} \sum_{w: A_w = a} \alpha_w^2 \{\hat{Y}_w(z) - \hat{Y}_{\mathrm{haj}}(z)\}\{\hat{Y}_w(z') - \hat{Y}_{\mathrm{haj}}(z')\}$$

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

$$\hat{V}_* = \begin{pmatrix} W_0^{-1} \begin{pmatrix} \hat{S}_*(00, 00) & \hat{S}_*(00, 01) \\ \hat{S}_*(00, 01) & \hat{S}_*(01, 01) \end{pmatrix} & 0_{2\times 2} \\ 0_{2\times 2} & W_1^{-1} \begin{pmatrix} \hat{S}_*(10, 10) & \hat{S}_*(10, 11) \\ \hat{S}_*(10, 11) & \hat{S}_*(11, 11) \end{pmatrix} \end{pmatrix} \tag{5}$$

to estimate the sampling covariance of $\hat{Y}_*$ for $* = \mathrm{ht}, \mathrm{haj}$, respectively. Mukerjee and Dasgupta (2021) introduced $\hat{V}_{\mathrm{ht}}$, and we introduce $\hat{V}_{\mathrm{haj}}$.

THEOREM 4.2. *Under the 2^2 split-plot randomization and Condition 2, we have*

$$W \hat{V}_* - \Sigma_* = S_* + o_{\mathbb{P}}(1) \quad \textit{for } * = \mathrm{ht}, \mathrm{haj}.$$

Mukerjee and Dasgupta ((2021), Theorem 2) implied $E(\hat{V}_{\mathrm{ht}}) - \mathrm{cov}(\hat{Y}_{\mathrm{ht}}) = W^{-1} S_{\mathrm{ht}} \geq 0$ such that $\hat{V}_{\mathrm{ht}}$ is a conservative estimator of $\mathrm{cov}(\hat{Y}_{\mathrm{ht}})$ in finite samples, extending Zhao et al. (2018) to possibly nonuniform 2^2 split-plot randomization. Theorem 4.2 extends the discussion to finite-population asymptotics, and establishes the asymptotic conservativeness of $\hat{V}_*$ for estimating the true sampling covariance of $\hat{Y}_*$ for $* = \mathrm{ht}, \mathrm{haj}$. This, together with Theorem 4.1, justifies the Wald-type inference of $\tau = G\bar{Y}$ based on $\hat{\tau} = G\hat{Y}_*$ with estimated covariance $G\hat{V}_* G^{\mathrm{T}}$ for $* = \mathrm{ht}, \mathrm{haj}$.

5. Reconciliation with model-based inference.

5.1. *Overview.* Despite the nice theoretical 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

$$Y_{ws} \sim 1(Z_{ws} = 00) + 1(Z_{ws} = 01) + 1(Z_{ws} = 10) + 1(Z_{ws} = 11) \tag{6}$$

TABLE 1

Regression estimators of $\bar{Y}$ 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	(6)	1	$\tilde{\beta}_{\text{ols}}$	$\hat{Y}_{\text{sm}}$	no	no
wls	(6)	$\{p_{ws}(Z_{ws})\}^{-1}$	$\tilde{\beta}_{\text{wls}}$	$\hat{Y}_{\text{haj}}$	no	yes
ag	(7)	1	$\tilde{\beta}_{\text{ag}}$	$\hat{Y}_{\text{ht}}$	yes	yes

as a standard formulation for analyzing 2^2 factorial experiments. We propose two general strategies, namely the *inverse probability weighting* and *aggregate model*, to recover the Hajek and Horvitz–Thompson estimators of $\bar{Y}$ directly as coefficients from (6) and its variant, respectively, and establish the appropriateness of the associated 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 regression equations represent the true outcome generating process.

5.2. *Least squares estimators from unit and aggregate regressions*. We introduce in this subsection three fitting schemes, denoted by "ols," "wls" and "ag," respectively, for estimating $\bar{Y}$ from least squares regressions, 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 $\bar{Y}$. Let $\tilde{\beta}_{\text{ols}}$ denote the resulting estimator.

Next, motivated by the use of inverse probability weighting in constructing $\hat{Y}_{\text{ht}}(z)$ and $\hat{Y}_{\text{haj}}(z)$, 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 $\bar{Y}$ by the resulting WLS coefficients. Let $\tilde{\beta}_{\text{wls}}$ 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 $w \in \mathcal{W}(z)$. 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, resulting in a total of $2W$ whole-plot level observations: $\{\hat{U}_w(A_w b) : w = 1, \dots, W;\ b = 0, 1\}$. We propose to fit

$$\hat{U}_w(A_w b) \sim 1(A_w b = 00) + 1(A_w b = 01) + 1(A_w b = 10) + 1(A_w b = 11) \tag{7}$$

over $\{(w, b) : w = 1, \dots, W;\ b = 0, 1\}$ for these $2W$ observations as an aggregate analog of (6). The "ag" fitting scheme takes the resulting OLS coefficients to estimate $\bar{Y}$. 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, $\{\tilde{\beta}_\dagger : \dagger = \text{ols, wls, ag}\}$, summarized in Table 1. As a convention, we use the tilde symbol to signify outputs from least squares fits. Proposition 2 below states their numeric equivalence with $\hat{Y}_{\text{sm}}$, $\hat{Y}_{\text{haj}}$ 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, {ols, wls, ag}, and the three estimation schemes, {sm, haj, ht}, runs through the following discussion and reconciles the model-based and design-based perspectives.

REMARK 2. There are alternative regression strategies for estimating $\bar{Y}$. First, the least squares fit of $\alpha_w Y_{ws} \sim 1(Z_{ws} = 00) + 1(Z_{ws} = 01) + 1(Z_{ws} = 10) + 1(Z_{ws} = 11)$ with weights $\alpha_w^{-1}\{p_{ws}(Z_{ws})\}^{-1}$ 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. Second, Miratrix, Weiss and Henderson (2021) reviewed an alternative WLS scheme in the context of stratified experiments, which corresponds to the least squares fit of (6) with weights $N_{Z_{ws}}/p_{ws}(Z_{ws})$. Lemma S13 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 $\tilde{V}_\dagger$ the classic cluster-robust covariance for $\tilde{\beta}_\dagger$ ($\dagger = \text{ols, wls, ag}$) from the same least squares fit. Theorem 5.1 below shows the asymptotic equivalence of $\tilde{V}_\dagger$ ($\dagger = \text{wls, ag}$) with $\hat{V}_{\text{haj}}$ and $\hat{V}_{\text{ht}}$, respectively.

THEOREM 5.1. *Define* $\hat{1}_{\text{ht}} = \text{diag}\{\hat{1}_{\text{ht}}(z)\}_{z\in\mathcal{T}}$. *Then*

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

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

Further assume Condition 2. Then $\hat{1}_{\text{ht}} = I_4 + o_{\mathbb{P}}(1)$, *and thus*

$$W(\tilde{V}_{\text{wls}} - \hat{V}_{\text{haj}}) = o_{\mathbb{P}}(1),$$

$$W(\tilde{V}_{\text{ag}} - \hat{V}_{\text{ht}}) = o_{\mathbb{P}}(1).$$

With $\hat{1}_{\text{ht}} = I_4 + o_{\mathbb{P}}(1)$, the asymptotic equivalence between the cluster-robust covariances and their design-based counterparts is a direct consequence of the numeric correspondence, and ensures the asymptotic conservativeness of $\tilde{V}_\dagger$ for estimating the true sampling covariance of $\tilde{\beta}_\dagger$ for $\dagger = \text{wls, ag}$. This, together with Proposition 2, justifies the Wald-type inference of $\tau = G\bar{Y}$ based on point estimator $\tilde{\tau}_\dagger = G\tilde{\beta}_\dagger$ and estimated covariance $G\tilde{V}_\dagger G^{\text{T}}$ for $\dagger = \text{wls, ag}$. Importantly, the cluster-robust covariance is necessary for valid regression-based inferences because the heteroskedasticity-robust covariance can be asymptotically anticonservative.

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 $\tilde{V}_{\text{ag}}$ recovers $\hat{V}_{\text{ht}}$ exactly in finite samples (Basse and Feller (2018), Bell and McCaffrey (2002), Imai, Jiang and Malani (2021)). The difference between the classic and HC2 estimators vanishes as the sample size goes to infinity. We relegate the details to Section S4.2 of the Supplementary Material. With small W, Bell and McCaffrey (2002), Cameron and Miller (2015), Pustejovsky and Tipton (2018) and MacKinnon, Nielsen and Webb (2021) proposed various confidence intervals to achieve better finite-sample coverage properties. They are likely to improve the classic cluster-robust covariance and its HC2 variant under the design-based framework as well. We leave this to future research. Lastly, $\hat{V}_{\text{ht}}$ is unbiased for estimating $\text{cov}(\hat{Y}_{\text{ht}})$ if $\alpha_w \bar{Y}_w(z)$ is identical over $w = 1, \dots, W$ for all $z \in \mathcal{T}$. Modification to $\hat{V}_{\text{ht}}$ is proposed by Mukerjee and Dasgupta (2021) that ensures unbiased estimation of $\text{cov}(\hat{Y}_{\text{ht}})$ 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. *Background*: *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, $\mathcal{T} = \{0, 1\}$, and a study population of N units with potential outcomes $\{Y_i(0), Y_i(1) : i = 1, \dots, N\}$. The finite-population average treatment effect equals $\tau = \bar{Y}(1) - \bar{Y}(0)$, where $\bar{Y}(z) = N^{-1} \sum_{i=1}^{N} Y_i(z)$.

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

We extend below 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 regression 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 normal estimators for τ, 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 $\text{cov}_\infty(\hat{\tau}_1) \le \text{cov}_\infty(\hat{\tau}_2)$ for all possible values of $\{Y_{ws}(z) : z \in \mathcal{T}\}_{ws \in \mathcal{S}}$, and the strict inequality holds for at least one set of $\{Y_{ws}(z) : z \in \mathcal{T}\}_{ws \in \mathcal{S}}$.

6.2. *Additive regressions*. Let $x_{ws} = (x_{ws[1]}, \dots, x_{ws[J]})^{\text{T}}$ be the $J \times 1$ covariate vector for subplot ws. Adding x_{ws} to (6) yields

$$Y_{ws} \sim \sum_{z \in \mathcal{T}} 1(Z_{ws} = z) + x_{ws} \sim d_{ws} + x_{ws} \tag{8}$$

as the *additive unit regression* over $ws \in \mathcal{S}$ with

$$d_{ws} = (1(Z_{ws} = 00), 1(Z_{ws} = 01), 1(Z_{ws} = 10), 1(Z_{ws} = 11))^{\text{T}}.$$

Let $\tilde{\beta}_{\text{ols,F}}$ and $\tilde{\beta}_{\text{wls,F}}$ denote the coefficient vectors of d_{ws} from the OLS and WLS fits of (8), respectively; we use the subscript "F" to signify Fisher (1935).

Let $\hat{v}_w(z) = \alpha_w \hat{x}_w(z)$ be the covariate analog of $\hat{U}_w(z) = \alpha_w \hat{Y}_w(z)$ with $\hat{x}_w(z) = M_{wb}^{-1} \sum_{s:Z_{ws}=z} x_{ws}$. Adding $\hat{v}_w(A_w b)$ to (7) defines

$$\hat{U}_w(A_w b) \sim \sum_{z \in \mathcal{T}} 1(A_w b = z) + \hat{v}_w(A_w b) \sim d_w(A_w b) + \hat{v}_w(A_w b) \tag{9}$$

as the *additive aggregate regression* over $\{(w, b) : w = 1, \ldots, W;\ b = 0, 1\}$ with

$$d_w(z) = (1(z = 00), 1(z = 01), 1(z = 10), 1(z = 11))^{\mathrm{T}} \text{ for } z = (A_w 0), (A_w 1).$$

Let $\tilde{\beta}_{\mathrm{ag,F}}$ denote the coefficient vector of $d_w(A_w b)$ from the OLS fit of (9). This, together with the above $\tilde{\beta}_{\mathrm{ols,F}}$ and $\tilde{\beta}_{\mathrm{wls,F}}$, defines three covariate-adjusted regression estimators of $\bar{Y}$.

REMARK 4. We may sometimes want to include certain whole-plot level attributes to both the unit and aggregate regressions; examples include the weights of the mice in Example 1 and the populations of the subdistricts in Example 2. The definition of x_{ws} 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 jth 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, $\tilde{\beta}_{\dagger,\mathrm{F}}$ ($\dagger$ = ols, wls, ag). To simplify the presentation, we center the covariates to have $\bar{x} = N^{-1} \sum_{ws \in \mathcal{S}} x_{ws} = 0_J$. Let $\hat{x}_*(z)$ ($*$ = sm, ht, haj) be the sample-mean, Horvitz–Thompson and Hajek estimators of $\bar{x}$ based on units under treatment z. Let $\hat{x}_* = (\hat{x}_*(00), \hat{x}_*(01), \hat{x}_*(10), \hat{x}_*(11))^{\mathrm{T}}$ be the $4 \times J$ matrix of $\{\hat{x}_*(z)\}_{z \in \mathcal{T}}$ for $*$ = sm, ht, haj. Let $\tilde{\gamma}_{\mathrm{ols}}$, $\tilde{\gamma}_{\mathrm{wls}}$ and $\tilde{\gamma}_{\mathrm{ag}}$ be the coefficient vectors of x_{ws} or $\hat{v}_w(A_w b)$ from the respective least squares fits of (8) and (9). Proposition 3 below parallels Proposition 2, and states the numeric correspondence between $\tilde{\beta}_{\dagger,\mathrm{F}}$ ($\dagger$ = ols, wls, ag) and $\hat{Y}_*$ ($*$ = sm, ht, haj).

PROPOSITION 3. *$\tilde{\beta}_{\mathrm{ols,F}} = \hat{Y}_{\mathrm{sm}} - \hat{x}_{\mathrm{sm}} \tilde{\gamma}_{\mathrm{ols}}$, $\tilde{\beta}_{\mathrm{wls,F}} = \hat{Y}_{\mathrm{haj}} - \hat{x}_{\mathrm{haj}} \tilde{\gamma}_{\mathrm{wls}}$, and $\tilde{\beta}_{\mathrm{ag,F}} = \hat{Y}_{\mathrm{ht}} - \hat{x}_{\mathrm{ht}} \tilde{\gamma}_{\mathrm{ag}}$.*

Recall that $\hat{Y}_{\mathrm{sm}} = \tilde{\beta}_{\mathrm{ols}}$, $\hat{Y}_{\mathrm{haj}} = \tilde{\beta}_{\mathrm{wls}}$, and $\hat{Y}_{\mathrm{ht}} = \tilde{\beta}_{\mathrm{ag}}$ from Proposition 2. Proposition 3 links the covariate-adjusted $\tilde{\beta}_{\dagger,\mathrm{F}}$'s back to the unadjusted $\tilde{\beta}_{\dagger}$'s, and establishes $\tilde{\beta}_{\mathrm{ols,F}}$, $\tilde{\beta}_{\mathrm{wls,F}}$, and $\tilde{\beta}_{\mathrm{ag,F}}$ as the sample-mean, Hajek, and Horvitz–Thompson estimators based on the covariate-adjusted outcomes $Y_{ws} - x_{ws}^{\mathrm{T}} \tilde{\gamma}_{\dagger}$ for $\dagger$ = ols, wls, ag, respectively. 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 $\hat{x}_{\mathrm{ht}}$ and $\tilde{\gamma}_{\mathrm{ag}}$ in $\tilde{\beta}_{\mathrm{ag,F}} = \hat{Y}_{\mathrm{ht}} - \hat{x}_{\mathrm{ht}} \tilde{\gamma}_{\mathrm{ag}}$. The consistency of $\hat{Y}_*$ and $\hat{x}_*(z)$, where $*$ = ht, haj, for estimating $\bar{Y}$ and $\bar{x}$ under mild conditions, on the other hand, ensures the consistency of $\tilde{\beta}_{\mathrm{wls,F}}$ and $\tilde{\beta}_{\mathrm{ag,F}}$ so long as $\tilde{\gamma}_{\mathrm{wls}}$ and $\tilde{\gamma}_{\mathrm{ag}}$ have finite probability limits. We formalize the intuition in Theorem 6.1 below. The sample-mean analog $\tilde{\beta}_{\mathrm{ols,F}}$, 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 S4.1 of the Supplementary Material.

Let $\bar{x}_w = M_w^{-1} \sum_{s=1}^{M_w} x_{ws}$ and $\overline{\|x_{w\cdot}\|_4^4} = M_w^{-1} \sum_{s=1}^{M_w} \|x_{ws}\|_4^4$ be the whole-plot average and uncentered fourth norm of $(x_{ws})_{s=1}^{M_w}$, 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 $(x_{ws})_{ws \in \mathcal{S}}$ and $\{x_{ws}, Y_{ws}(z)\}_{ws \in \mathcal{S}}$, respectively. Let $S_{xY'(z)}$ be the scaled between whole-plot covariance of $(x_{ws})_{ws \in \mathcal{S}}$ with the centered potential outcomes $\{Y'_{ws}(z)\}_{ws\in\mathcal{S}}$; the corresponding scaled within whole-plot covariance 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 S3.1 of 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 Ψ from Lemma 1. 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 $Q_{xx} = (N-1)^{-1}\sum_{ws\in\mathcal{S}} x_{ws}x_{ws}^{\mathrm{T}}$ and $Q_{xY(z)} = (N-1)^{-1}\sum_{ws\in\mathcal{S}} x_{ws}Y_{ws}(z)$ be the finite-population covariances of $(x_{ws})_{ws\in\mathcal{S}}$ and $\{x_{ws}, Y_{ws}(z)\}_{ws\in\mathcal{S}}$, 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) $\max_{w=1,\dots,W} \|\alpha_w \bar{x}_w\|_2^2 / W = o(1)$;

(iii) $W^{-1}\sum_{w=1}^{W} \alpha_w^2 \overline{\|x_{w\cdot}\|_4^4} = O(1)$; $W^{-2}\sum_{w=1}^{W} \alpha_w^4 \overline{\|x_{w\cdot}\|_4^4} = o(1)$.

Condition 3 is the analog of Condition 2(iii)–(v) for the covariates as potential outcomes unaffected by the treatment. The additional requirement on Q_{xx} and $Q_{xY(z)}$ ensures the convergence of $\tilde{\gamma}_{\mathrm{wls}}$ under split-plot randomization.

Let $\gamma_\dagger$ be the finite probability limit of $\tilde{\gamma}_\dagger$ for $\dagger = \mathrm{wls}, \mathrm{ag}$ under split-plot randomization and Conditions 2–3; we relegate the proof of their existence and explicit forms to Lemma S12 in the Supplementary Material. Let $S_{\mathrm{wls,F}}$ and $\Sigma_{\mathrm{wls,F}}$ be the analogs of S_{haj} and Σ_{haj} defined on the adjusted potential outcomes $Y_{ws}(z;\gamma_{\mathrm{wls}}) = Y_{ws}(z) - x_{ws}^{\mathrm{T}}\gamma_{\mathrm{wls}}$. Let $S_{\mathrm{ag,F}}$ and $\Sigma_{\mathrm{ag,F}}$ be the analogs of S_{ht} and Σ_{ht} defined on the adjusted potential outcomes $Y_{ws}(z;\gamma_{\mathrm{ag}}) = Y_{ws}(z) - x_{ws}^{\mathrm{T}}\gamma_{\mathrm{ag}}$. Let $\tilde{V}_{\dagger,\mathrm{F}}$ be the cluster-robust covariance of $\tilde{\beta}_{\dagger,\mathrm{F}}$ for $\dagger = \mathrm{wls}, \mathrm{ag}$.

THEOREM 6.1. *Under the 2^2 split-plot randomization and Conditions 2–3, we have*

$$\sqrt{W}(\tilde{\beta}_{\dagger,\mathrm{F}} - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_{\dagger,\mathrm{F}}), \qquad W\tilde{V}_{\dagger,\mathrm{F}} - \Sigma_{\dagger,\mathrm{F}} = S_{\dagger,\mathrm{F}} + o_{\mathbb{P}}(1)$$

with $S_{\dagger,\mathrm{F}} \geq 0$ for $\dagger = \mathrm{wls}, \mathrm{ag}$.

Theorem 6.1 states the asymptotic normality of $\tilde{\beta}_{\dagger,\mathrm{F}}$ ($\dagger = \mathrm{wls}, \mathrm{ag}$) under split-plot randomization, and ensures the asymptotic conservativeness of $\tilde{V}_{\dagger,\mathrm{F}}$ for estimating the true sampling covariance. This justifies the regression-based inference of $\tau = G\bar{Y}$ from the additive regressions with point estimator $G\tilde{\beta}_{\dagger,\mathrm{F}}$ and estimated covariance $G\tilde{V}_{\dagger,\mathrm{F}}G^{\mathrm{T}}$ for $\dagger = \mathrm{wls}, \mathrm{ag}$. Neither $\tilde{\beta}_{\mathrm{wls,F}}$ nor $\tilde{\beta}_{\mathrm{ag,F}}$, however, guarantees efficiency gains over their unadjusted counterparts, namely $\tilde{\beta}_\dagger$ for $\dagger = \mathrm{wls}, \mathrm{ag}$, even asymptotically. Similar discussion by Freedman (2008) and Lin (2013) under complete randomization suggests that including the interactions between the treatment indicators and covariates in the regression can be a possible remedy. We quantify its design-based properties in the next two subsections.

6.3. *Fully-interacted regressions*. Modify (8) and (9) with full interactions between the treatment indicators and covariates:

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

$$\hat{U}_w(A_w b) \sim d_w(A_w b) + \sum_{z\in\mathcal{T}} 1(A_w b = z)\, \hat{v}_w(A_w b). \tag{11}$$

TABLE 2

Nine regression estimators from the unadjusted specifications (6)–(7), the additive specifications (8)–(9) and the fully-interacted specifications (10)–(11), respectively, under the "ols," "wls" and "ag" fitting schemes

Fitting scheme	Base model	Weight	Unadjusted	Additive	Fully-interacted
ols	$Y_{ws} \sim d_{ws}$	1	$\tilde{\beta}_{\mathrm{ols}}$	$\tilde{\beta}_{\mathrm{ols,F}}$	$\tilde{\beta}_{\mathrm{ols,L}}$
wls	$Y_{ws} \sim d_{ws}$	$\{p_{ws}(Z_{ws})\}^{-1}$	$\tilde{\beta}_{\mathrm{wls}}$	$\tilde{\beta}_{\mathrm{wls,F}}$	$\tilde{\beta}_{\mathrm{wls,L}}$
ag	$\hat{U}_w(A_w b) \sim d_w(A_w b)$	1	$\tilde{\beta}_{\mathrm{ag}}$	$\tilde{\beta}_{\mathrm{ag,F}}$	$\tilde{\beta}_{\mathrm{ag,L}}$

This defines the *fully-interacted unit* and *aggregate regressions*, respectively. Let $\tilde{\beta}_{\mathrm{ols,L}}$ and $\tilde{\beta}_{\mathrm{wls,L}}$ denote the coefficient vectors of d_{ws} from the OLS and WLS fits of the unit regression (10), respectively. Let $\tilde{\beta}_{\mathrm{ag,L}}$ denote the coefficient of $d_w(A_w b)$ from the OLS fit of the aggregate regression (11). We use the subscript "L" to signify Lin (2013). This gives us three more estimators of $\bar{Y}$, $\{\tilde{\beta}_{\dagger,\mathrm{L}} : \dagger = \mathrm{ols, wls, ag}\}$, summarized in Table 2. We now derive their respective design-based properties.

Let $\tilde{\gamma}_{\dagger,z}$ be the coefficient of $1(Z_{ws} = z)\, x_{ws}$ or $1(A_w b = z)\, \hat{v}_w(A_w b)$ from the corresponding regression for $z \in \mathcal{T}$ and $\dagger \in \{\mathrm{ols, wls, ag}\}$. Let $\tilde{\beta}_{\dagger,\mathrm{L}}(z)$ be the element in $\tilde{\beta}_{\dagger,\mathrm{L}}$ that corresponds to treatment z.

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

Observe that Proposition 3 under the additive regressions implies $\tilde{\beta}_{\mathrm{ols,L}}(z) = \hat{Y}_{\mathrm{sm}}(z) - \hat{x}_{\mathrm{sm}}^{\mathrm{T}}(z)\tilde{\gamma}_{\mathrm{ols}}$, $\tilde{\beta}_{\mathrm{wls,L}}(z) = \hat{Y}_{\mathrm{haj}}(z) - \hat{x}_{\mathrm{haj}}^{\mathrm{T}}(z)\tilde{\gamma}_{\mathrm{wls}}$, and $\tilde{\beta}_{\mathrm{ag,L}}(z) = \hat{Y}_{\mathrm{ht}}(z) - \hat{x}_{\mathrm{ht}}^{\mathrm{T}}(z)\tilde{\gamma}_{\mathrm{ag}}$ for $z \in \mathcal{T}$. Proposition 4 parallels Proposition 3, and establishes $\tilde{\beta}_{\mathrm{ols,L}}(z)$, $\tilde{\beta}_{\mathrm{wls,L}}(z)$ and $\tilde{\beta}_{\mathrm{ag,L}}(z)$ as the sample-mean, Hajek and Horvitz–Thompson estimators based on the covariate-adjusted outcomes $Y_{ws} - x_{ws}^{\mathrm{T}}\tilde{\gamma}_{\dagger,z}$ for $\dagger = \mathrm{ols, wls, ag}$, respectively. A key distinction is that the adjustment is now based on treatment-specific coefficients. We hence state the results in scalar form to avoid additional notation.

The correlation between $\hat{x}_{\mathrm{ht}}(z)$ and $\tilde{\gamma}_{\mathrm{ag},z}$ likewise leaves the covariate-adjusted Horvitz–Thompson estimator, $\tilde{\beta}_{\mathrm{ag,L}}$, biased in finite samples. The fact that $\tilde{\gamma}_{\dagger,z}$ has a finite probability limit, denoted by $\gamma_{\dagger,z}$, under Conditions 2–3, on the other hand, ensures the consistency of $\tilde{\beta}_{\dagger,\mathrm{L}}$ for $\dagger = \mathrm{wls, ag}$; see Lemma S11 in the Supplementary Material. Section S4.1 of the Supplementary Material further gives analogous results about $\tilde{\beta}_{\mathrm{ols,L}}$ under uniform designs.

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

THEOREM 6.2. *Under the* 2^2 *split-plot randomization and Conditions 2–3, we have*

$$\sqrt{W}(\tilde{\beta}_{\dagger,\mathrm{L}} - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_{\dagger,\mathrm{L}}), \qquad W\tilde{V}_{\dagger,\mathrm{L}} - \Sigma_{\dagger,\mathrm{L}} = S_{\dagger,\mathrm{L}} + o_{\mathbb{P}}(1)$$

with $S_{\dagger,\mathrm{L}} \geq 0$ *for* $\dagger = \mathrm{wls, ag}$.

Echoing the comment after Theorem 6.1, Theorem 6.2 justifies the regression-based inference of $\tau = G\bar{Y}$ from the fully-interacted regressions with point estimator $G\tilde{\beta}_{\dagger,\mathrm{L}}$ and estimated covariance $G\tilde{V}_{\dagger,\mathrm{L}}G^{\mathrm{T}}$ for $\dagger = \mathrm{wls, ag}$.

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 specification.

Motivated 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 $\hat{v}_w(A_w b)$; see Remark 4. Intuitively, α_w measures the size of the whole plot and is thus prognostic to $\hat{U}_w(z)$ as the outcome of the aggregate regressions. Let $\tilde{\beta}_{\text{ag,F}}(\alpha, v)$ and $\tilde{\beta}_{\text{ag,L}}(\alpha, v)$ be the resulting OLS coefficient vectors of $d_w(A_w b)$ under the additive and fully-interacted specifications, respectively; we use the suffix "(α, v)" to emphasize the components of the corresponding augmented covariates. Proposition 5 below states the asymptotic efficiency of $\tilde{\beta}_{\text{ag,L}}(\alpha, v)$.

PROPOSITION 5. *Under the 2^2 split-plot randomization and Conditions 2–3, if*

$$\Psi_{xx}(z, z) = o(1) \quad \text{for all } z \in \mathcal{T}, \tag{12}$$

then $\tilde{\beta}_{\text{ag,L}}(\alpha, v)$ has the smallest asymptotic sampling covariance among

$$\mathcal{B} = \{\tilde{\beta}_{\text{wls}}, \tilde{\beta}_{\text{wls},\diamond}; \tilde{\beta}_{\text{ag}}, \tilde{\beta}_{\text{ag},\diamond}, \tilde{\beta}_{\text{ag},\diamond}(\alpha, v) : \diamond = \text{F}, \text{L}\}.$$

Proposition 5 establishes the optimality of $\tilde{\beta}_{\text{ag,L}}(\alpha, v)$ among the eight consistent regression estimators in $\mathcal{B}$, 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 $\hat{Y}_*$ ($* = \text{ht}, \text{haj}$) then follows from the numeric identities in Proposition 2. 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 $\tilde{\beta}_{\text{ag,L}}(\alpha, v)$ 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 regression equations. We leave the more general theory to future work.

Further observe that $x_{ws} = \varnothing$ for all $ws \in \mathcal{S}$ is a special case of homogeneous covariates within whole plots. By Proposition 5, including $\alpha_w - 1$ in the fully-interacted aggregate regression ensures gain in asymptotic efficiency over the unadjusted estimators even without other covariate information. This highlights the benefit of adjusting for whole-plot sizes in analyzing split-plot data.

This guaranteed minimum asymptotic covariance, however, should not be the basis for dismissing the additive regressions completely. In particular, the fully-interacted regression (11) involves $|\mathcal{T}| \times (1 + J)$ parameters compared with the $(|\mathcal{T}| + J)$ parameters in the additive regression (9), subjecting $\tilde{\beta}_{\text{ag,L}}$ 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.* The regressions so far take indicators of the treatments, namely $1(Z_{ws} = ab)$ or $1(A_w b = ab)$, as regressors, and yield coefficients as estimators of $\bar{Y}$. Despite the generality of such 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, regress the outcome on the factors themselves, and estimate the factorial effects directly by the regression coefficients.

With the treatment combinations of interest exhibiting a 2^2 factorial structure, the standard factor-based specification takes the form

$$Y_{ws} \sim 1 + A_w + B_{ws} + A_w B_{ws} \quad (ws \in \mathcal{S}), \tag{13}$$

and interprets the coefficients of the nonintercept terms as the main effects and interaction, respectively. Let $\tilde{\tau}'_{\text{A},0}$, $\tilde{\tau}'_{\text{B},0}$ and $\tilde{\tau}'_{\text{AB},0}$ be the OLS coefficients of A_w, B_{ws} and $A_{ws}B_{ws}$ from (13), respectively. Standard least squares theory ensures

$$\tilde{\tau}'_{\text{A},0} = \hat{Y}_{\text{sm}}(10) - \hat{Y}_{\text{sm}}(00),$$
$$\tilde{\tau}'_{\text{B},0} = \hat{Y}_{\text{sm}}(01) - \hat{Y}_{\text{sm}}(00),$$
$$\tilde{\tau}'_{\text{AB},0} = \hat{Y}_{\text{sm}}(11) - \hat{Y}_{\text{sm}}(10) - \hat{Y}_{\text{sm}}(01) + \hat{Y}_{\text{sm}}(00),$$

equaling the sample-mean estimators of $\tau_{\text{A},0} = \bar{Y}(10) - \bar{Y}(00)$, $\tau_{\text{B},0} = \bar{Y}(01) - \bar{Y}(00)$, and τ_{AB} from (1), respectively. When the goal is to estimate the standard factorial effects $(\tau_{\text{A}}, \tau_{\text{B}}, \tau_{\text{AB}})$ as defined in (1), an algebraic trick is to shift 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 $ws \in \mathcal{S}$. Fitting (14) by OLS recovers the sample-mean estimators of τ_{A}, τ_{B} and τ_{AB}, respectively (Zhao and Ding (2021a)). Denote by $\tilde{\tau}'_{\text{ols}}$ and $\tilde{\Omega}'_{\text{ols}}$ the resulting coefficient vector and cluster-robust covariance of the three nonintercept terms, respectively. As a convention, we use the combination of tilde and prime to signify outputs from factor-based regressions like (13) and (14).

Following the intuition from the OLS fit, let $\tilde{\tau}'_{\text{wls}}$ be the coefficient vector of the nonintercept terms from (14) under the "wls" fitting scheme, with $\tilde{\Omega}'_{\text{wls}}$ denoting the associated cluster-robust covariance. 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 $\{(w, b) : w = 1, \dots, W;\ b = 0, 1\}$, with $\tilde{\tau}'_{\text{ag}}$ and $\tilde{\Omega}'_{\text{ag}}$ denoting the resulting OLS coefficient vector and cluster-robust covariance of the nonintercept terms under fitting scheme "ag."

Recall G_0 as the contrast matrix corresponding to $(\tau_{\text{A}}, \tau_{\text{B}}, \tau_{\text{AB}})^{\text{T}} = G_0 \bar{Y}$. Proposition 6 below follows from the invariance of least squares to nondegenerate linear transformation of the regressors, and ensures the validity of $(\tilde{\tau}'_{\dagger}, \tilde{\Omega}'_{\dagger})$ for the Wald-type inference of the standard factorial effects for $\dagger = \text{wls}, \text{ag}$.

PROPOSITION 6. $\tilde{\tau}'_{\dagger} = G_0 \tilde{\beta}_{\dagger}$ *and* $\tilde{\Omega}'_{\dagger} = G_0 \tilde{V}_{\dagger} G_0^{\text{T}}$ *for* $\dagger = \text{ols}, \text{wls}, \text{ag}$.

Specifications (14)–(15) thus deliver the Hajek and Horvitz–Thompson estimators of the standard factorial effects, namely $\hat{\tau}_{\text{haj}} = G_0 \hat{Y}_{\text{haj}}$ and $\hat{\tau}_{\text{ht}} = G_0 \hat{Y}_{\text{ht}}$, directly as regression coefficients. We thus recommend using (14)–(15) if the goal is the standard factorial effects and switching back to (6)–(7) if otherwise.

The results on covariate adjustment are almost identical to those based on the treatment indicators. 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 the 2^2 split-plot design. We relegate the details to Section 4.3 of the Supplementary Material.

7.2. *$T_{\mathrm{A}} \times T_{\mathrm{B}}$ split-plot design.* All discussion so far concerns the 2^2 split-plot design with the whole-plot factor and the subplot factor both of two levels. We now extend the result to general split-plot designs with factors of multiple levels.

Consider two factors of interest, A and B, of $T_{\mathrm{A}} \geq 2$ and $T_{\mathrm{B}} \geq 2$ levels, respectively, indexed by $a \in \mathcal{T}_{\mathrm{A}} = \{0, \dots, T_{\mathrm{A}} - 1\}$ and $b \in \mathcal{T}_{\mathrm{B}} = \{0, \dots, T_{\mathrm{B}} - 1\}$. A general $T_{\mathrm{A}} \times T_{\mathrm{B}}$ 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 $a \in \mathcal{T}_{\mathrm{A}}$ of factor A with $\sum_{a\in\mathcal{T}_{\mathrm{A}}} W_a = W$. It then conducts an independent randomization within each whole plot and assigns completely at random M_{wb} subplots in whole plot w to receive level $b \in \mathcal{T}_{\mathrm{B}}$ of factor B with $\sum_{b\in\mathcal{T}_{\mathrm{B}}} M_{wb} = M_w$ for $w = 1, \dots, W$. Refer to the cluster and stratified randomizations as stage (I) and stage (II) of the assignment, respectively. The final treatment of a subplot is the combination of its whole-plot factor assignment in stage (I) and its subplot factor assignment in stage (II), taking values from $\mathcal{T} = \{(ab) : a \in \mathcal{T}_{\mathrm{A}},\ b \in \mathcal{T}_{\mathrm{B}}\}$. Example 1 defines a 3×2 split-plot experiment, and Example 2 defines a 2×3 split-plot experiment.

All notation and results from the 2^2 case extend to the current setting with the renewed definition of $\mathcal{T}$. We relegate the details on the design-based inference to Section S4.4 of the Supplementary Material and focus below on the model-based inference from factor-based regressions.

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

$$
\begin{aligned}
\tau_{\mathrm{A}a} &= T_{\mathrm{B}}^{-1} \sum_{b\in\mathcal{T}_{\mathrm{B}}} \{\bar{Y}(ab) - \bar{Y}(0b)\}, \\
\tau_{\mathrm{B}b} &= T_{\mathrm{A}}^{-1} \sum_{a\in\mathcal{T}_{\mathrm{A}}} \{\bar{Y}(ab) - \bar{Y}(a0)\}, \\
\tau_{\mathrm{A}a,\mathrm{B}b} &= \bar{Y}(ab) - \bar{Y}(0b) - \bar{Y}(a0) + \bar{Y}(00)
\end{aligned}
\tag{16}
$$

as the standard main effects and interactions at nonbaseline levels $a = 1, \dots, T_{\mathrm{A}} - 1$ and $b = 1, \dots, T_{\mathrm{B}} - 1$, respectively, vectorized as

$$\tau = \{\tau_{\mathrm{A}a}, \tau_{\mathrm{B}b}, \tau_{\mathrm{A}a,\mathrm{B}b} : a = 1, \dots, T_{\mathrm{A}};\ b = 1, \dots, T_{\mathrm{B}}\} = G_0\bar{Y}.$$

The definitions reduce to τ_{A}, τ_{B}, and τ_{AB} from (1) when $T_{\mathrm{A}} = T_{\mathrm{B}} = 2$.

Motivated by the utility of location-shifted factors in recovering τ_{A} and τ_{B} directly as least squares coefficients from (14) and (15), consider

$$
\begin{aligned}
Y_{ws} \sim 1 &+ \sum_{a=1}^{T_{\mathrm{A}}-1} 1_{\mathrm{c}}(A_w = a) + \sum_{b=1}^{T_{\mathrm{B}}-1} 1_{\mathrm{c}}(B_{ws} = b) \\
&+ \sum_{a=1}^{T_{\mathrm{A}}-1} \sum_{b=1}^{T_{\mathrm{B}}-1} 1_{\mathrm{c}}(A_w = a) 1_{\mathrm{c}}(B_{ws} = b),
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
\hat{U}_w(A_w b) \sim 1 &+ \sum_{a=1}^{T_{\mathrm{A}}-1} 1_{\mathrm{c}}(A_w = a) + \sum_{b'=1}^{T_{\mathrm{B}}-1} 1_{\mathrm{c}}(b = b') \\
&+ \sum_{a=1}^{T_{\mathrm{A}}-1} \sum_{b'=1}^{T_{\mathrm{B}}-1} 1_{\mathrm{c}}(A_w = a) 1_{\mathrm{c}}(b = b')
\end{aligned}
\tag{18}
$$

as two generalizations under the $T_{\mathrm{A}} \times T_{\mathrm{B}}$ design with $1_{\mathrm{c}}(A_w = a) = 1(A_w = a) - T_{\mathrm{A}}^{-1}$, $1_{\mathrm{c}}(B_{ws} = b) = 1(B_{ws} = b) - T_{\mathrm{B}}^{-1}$ and $1_{\mathrm{c}}(b = b') = 1(b = b') - T_{\mathrm{B}}^{-1}$.

Renew $\tilde{\tau}'_{\dagger}$ ($\dagger =$ ols, wls, ag) as the coefficient vectors of the nonintercept terms from (17) and (18) under fitting schemes "ols," "wls" and "ag," respectively. Renew $\hat{Y}_*$ as the sample-mean, Horvitz–Thompson and Hajek estimators of $\bar{Y}$ for $* =$ sm, ht, haj. Proposition 7 below states their numeric correspondence paralleling Proposition 6.

PROPOSITION 7. $\tilde{\tau}'_{\text{ols}} = G_0\hat{Y}_{\text{sm}}$, $\tilde{\tau}'_{\text{wls}} = G_0\hat{Y}_{\text{haj}}$, *and* $\tilde{\tau}'_{\text{ag}} = G_0\hat{Y}_{\text{ht}}$.

The unbiasedness and consistency results then follow from the properties of $\hat{Y}_*$ for $* =$ sm, ht, haj such that $\tilde{\tau}'_{\text{wls}}$ and $\tilde{\tau}'_{\text{ag}}$ are both consistent for estimating τ. The results on the cluster-robust covariances parallel those under the 2^2 design, and ensure asymptotically conservative estimation of the true sampling covariances. This justifies the validity of regression-based inferences from (17)–(18) under fitting schemes $\dagger =$ wls, ag.

The results on covariate adjustment are almost identical to those under the 2^2 case and thus omitted. 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 2–3 and (12). It is thus our recommendation for estimating the standard factorial effects under the general $T_{\text{A}} \times T_{\text{B}}$ 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 (Imbens and Rubin (2015)). The theory under complete randomization further demonstrates that the Fisher randomization test with a robustly-studentized test statistic 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 (2021), Zhao and Ding (2021b)). The theory extends naturally to split-plot randomization.

Renew $\mathcal{B} = \{\tilde{\beta}_{\text{wls}}, \tilde{\beta}_{\text{wls},\diamond}; \tilde{\beta}_{\text{ag}}, \tilde{\beta}_{\text{ag},\diamond}, \tilde{\beta}_{\text{ag},\diamond}(\alpha, \upsilon) : \diamond = \text{F}, \text{L}\}$ as the collection of regression estimators of $\bar{Y}$ that are consistent under the general $T_{\text{A}} \times T_{\text{B}}$ split-plot design. We propose to use the Fisher randomization test with test statistic $t^2(\tilde{\beta}) = (G\tilde{\beta})^{\text{T}}(G\tilde{V}G^{\text{T}})^{-1}G\tilde{\beta}$ for $\tilde{\beta} \in \mathcal{B}$, where $\tilde{V}$ is the associated cluster-robust covariance of $\tilde{\beta}$.

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 $\tilde{\beta} \in \mathcal{B}$. Under (12), the test based on $t^2(\tilde{\beta}_{\text{ag,L}}(\alpha, \upsilon))$ 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 of the Supplementary Material.

8. Numerical example. We now apply the proposed methods to the 3×2 neuroscience experiment from Example 1. Due to the space limit, we relegate the simulation studies to Section S5 of the Supplementary Material.

We use the data set from Moen et al. (2016), which consists of $N = 1{,}143$ neuron level observations nested within $W = 14$ mice. Recall fatty acid delivery ("fa") and Pten knockdown ("pten") as the whole-plot and subplot factors, respectively, with $(T_{\text{A}}, T_{\text{B}}) = (3, 2)$. The treatment sizes at the whole-plot level are $(W_0, W_1, W_2) = (5, 4, 5)$. The whole-plot sizes $(M_w)_{w=1}^{W}$ vary from 13 to 152, with the q_{w1}'s ranging from 0.518 to 0.846.

Denote by "fa1" and "fa2" the standard main effects of fatty acid delivery, by "pten" the standard main effect of Pten knockdown, and by "fa1:pten" and "fa2:pten" their interactions, with definitions given by (16). We apply four factor-based regression schemes as the combinations of two fitting schemes, namely "wls" for the unit specification and "ag" for

TABLE 3
Re-analyzing the data from Moen et al. (2016). "p.normal" and "p.frt" indicate the p-values from large-sample approximations and Fisher randomization tests, respectively

	Unadjusted				Adjusted			
	est	se	p. normal	p. frt	est	se	p. normal	p. frt
(a) regression based on unit data								
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								
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

the aggregate specification, and the presence or absence of covariate adjustment. We use $x_{ws} = \alpha_w = M_w/\bar{M}$ as the covariate for neurons in mouse w, and conduct covariate adjustment using the additive specification due to the small number of whole plots at $W = 14$.

Table 3 shows the point estimators, cluster-robust standard errors, p-values based on large-sample approximations of the t-statistics, and p-values based on Fisher randomization tests, respectively. Covariate adjustment reduces the standard errors of the estimators of "fa1" and "fa2" under both the unit and aggregate specifications, yet has no effect on the estimators of "pten" and the two interactions. The identicalness of the unadjusted and additive regressions 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. The resulting covariate matrix, as the concatenation of $(x_{ws})_{ws\in\mathcal{S}}$, is 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 gives a rigorous statement.

The p-values from large-sample approximations and Fisher randomization tests concur in most cases at significance level 0.05. Two exceptions are the tests for "fa2" under the adjusted unit regression and those for "fa1" under the unadjusted aggregate regression. This is likely due to the 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 is especially important in the current case given its 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 fitting schemes concur in most cases with the two exceptions being the p-values from large-sample approximations for "fa1" under the unadjusted regressions and those for "fa2" under the adjusted regressions. This is again likely due to the 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 regression if otherwise. The point estimator based on the OLS coefficient is consistent for estimating the finite-population average treatment effect, with the associated 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 regression equation represents the true relationship between the outcome, treatments and covariates.

Acknowledgments. We thank the Associate Editor, two referees, Zhichao Jiang and Rahul Mukerjee for most insightful comments.

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

SUPPLEMENTARY MATERIAL

Supplement to "Reconciling design-based and model-based causal inferences for split-plot experiments." (DOI: 10.1214/21-AOS2144SUPP; .pdf). We give the proofs of the results in the main paper, and provide additional results on the special case of uniform designs, extensions and simulation studies.

REFERENCES

ABADIE, A., ATHEY, S., IMBENS, G. W. and WOOLDRIDGE, J. M. (2017). When should you adjust standard errors for clustering? Technical Report, National Bureau of Economic Research.

ABADIE, A., ATHEY, S., IMBENS, G. W. and WOOLDRIDGE, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica* **88** 265–296. MR4084709 https://doi.org/10.3982/ecta12675

BASSE, G. and FELLER, A. (2018). Analyzing two-stage experiments in the presence of interference. *J. Amer. Statist. Assoc.* **113** 41–55. MR3803438 https://doi.org/10.1080/01621459.2017.1323641

BELL, R. M. and MCCAFFREY, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Surv. Methodol.* **28** 169–181.

BOX, G. E. P. and ANDERSEN, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **17** 1–34.

CAMERON, A. and MILLER, D. (2015). A practitioner's guide to cluster-robust inference. *J. Hum. Resour.* **50** 317–372.

CHONG, A., GONZALEZ-NAVARRO, M., KARLAN, D. and VALDIVIA, M. (2020). Do information technologies improve teenagers' sexual education? Evidence from a randomized evaluation in Colombia. *World Bank Econ. Rev.* **34** 371–392.

COX, D. R. and REID, N. (2000). *The Theory of the Design of Experiments*. Chapman and Hall/CRC, London.

DASGUPTA, T., PILLAI, N. S. and RUBIN, D. B. (2015). Causal inference from 2^K factorial designs by using potential outcomes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 727–753. MR3382595 https://doi.org/10.1111/rssb.12085

FISHER, R. A. (1935). *The Design of Experiments*, 1st ed. Oliver and Boyd, Edinburgh, London.

FOGARTY, C. B. (2018a). On mitigating the analytical limitations of finely stratified experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 1035–1056. MR3874309 https://doi.org/10.1111/rssb.12290

FOGARTY, C. B. (2018b). Regression-assisted inference for the average treatment effect in paired experiments. *Biometrika* **105** 994–1000. MR3877880 https://doi.org/10.1093/biomet/asy034

FREEDMAN, D. A. (2008). On regression adjustments to experimental data. *Adv. in Appl. Math.* **40** 180–193. MR2388610 https://doi.org/10.1016/j.aam.2006.12.003

FRICANO, C. J., DESPENZA JR., T., FRAZEL, P. W., LI, M., O'MALLEY, J. A., WESTBROOK, G. L. and LUIKART, B. W. (2014). Fatty acids increase neuronal hypertrophy of Pten knockdown neurons. *Front. Mol. Neuroscience* **7** 30.

FULLER, W. A. (2009). *Sampling Statistics*. Wiley, New York.

HINKELMANN, K. and KEMPTHORNE, O. (2008). *Design and Analysis of Experiments. Vol.* 1. Wiley, New York.

HUDGENS, M. G. and HALLORAN, M. E. (2008). Toward causal inference with interference. *J. Amer. Statist. Assoc.* **103** 832–842. MR2435472 https://doi.org/10.1198/016214508000000292

IMAI, K., JIANG, Z. and MALANI, A. (2021). Causal inference with interference and noncompliance in two-stage randomized experiments. *J. Amer. Statist. Assoc.* **116** 632–644. MR4270009 https://doi.org/10.1080/01621459.2020.1775612

IMAI, K., KING, G. and NALL, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. *Statist. Sci.* **24** 29–53. MR2561126 https://doi.org/10.1214/08-STS274

IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference—for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge Univ. Press, New York. MR3309951 https://doi.org/10.1017/CBO9781139025751

JI, X., FINK, G., ROBYN, P. J. and SMALL, D. S. (2017). Randomization inference for stepped-wedge cluster-randomized trials: An application to community-based health insurance. *Ann. Appl. Stat.* **11** 1–20. MR3634312 https://doi.org/10.1214/16-AOAS969

JONES, B. and NACHTSHEIM, C. J. (2009). Split-plot designs: What, why, and how. *J. Qual. Technol.* **41** 340–361.

KEMPTHORNE, O. (1952). *The Design and Analysis of Experiments*. Wiley, New York; CRC Press, London. MR0045368

LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *J. Amer. Statist. Assoc.* **112** 1759–1769. MR3750897 https://doi.org/10.1080/01621459.2017.1295865

LIANG, K. Y. and ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22. MR0836430 https://doi.org/10.1093/biomet/73.1.13

LIN, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. *Ann. Appl. Stat.* **7** 295–318. MR3086420 https://doi.org/10.1214/12-AOAS583

LIU, L. and HUDGENS, M. G. (2014). Large sample randomization inference of causal effects in the presence of interference. *J. Amer. Statist. Assoc.* **109** 288–301. MR3180564 https://doi.org/10.1080/01621459.2013.844698

LIU, H. and YANG, Y. (2020). Regression-adjusted average treatment effect estimates in stratified randomized experiments. *Biometrika* **107** 935–948. MR4186497 https://doi.org/10.1093/biomet/asaa038

MACKINNON, J. G., NIELSEN, M. and WEBB, M. D. (2021). Cluster-robust inference: A guide to empirical practice. Working Paper No. 1456, Economics Department, Queen's University.

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

MIRATRIX, L. W. (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 369–396. MR3021392 https://doi.org/10.1111/j.1467-9868.2012.01048.x

MIRATRIX, L. W., WEISS, M. J. and HENDERSON, B. (2021). An applied researcher's guide to estimating effects from multisite individually randomized trials: Estimands, estimators, and estimates. *J. Res. Educ. Eff.* **14** 270–308.

MOEN, E. L., FRICANO-KUGLER, C. J., LUIKART, B. W. and O'MALLEY, A. J. (2016). Analyzing clustered data: Why and how to account for multiple observations nested within a study participant? *PLoS ONE* **11** e0146721. https://doi.org/10.1371/journal.pone.0146721

MUKERJEE, R. and DASGUPTA, T. (2021). Causal inference from possibly unbalanced split-plot designs: A randomization-based perspective. *Statist. Sinica*.

MUKERJEE, R., DASGUPTA, T. and RUBIN, D. B. (2018). Using standard tools from finite population sampling to improve causal inference for complex experiments. *J. Amer. Statist. Assoc.* **113** 868–881. MR3832233 https://doi.org/10.1080/01621459.2017.1294076

NEYMAN, J. (1935). Statistical problems in agricultural experimentation (with discussion). *Suppl. J. R. Stat. Soc.* **2** 107–180.

OHLSSON, E. (1989). Asymptotic normality for two-stage sampling from a finite population. *Probab. Theory Related Fields* **81** 341–352. MR0983089 https://doi.org/10.1007/BF00340058

OLKEN, B. A. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. *J. Polit. Econ.* **115** 200–249.

PASHLEY, N. E. and MIRATRIX, L. W. (2021). Insights on variance estimation for blocked and matched pairs designs. *J. Educ. Behav. Stat.* **46** 271–296.

PUSTEJOVSKY, J. E. and TIPTON, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *J. Bus. Econom. Statist.* **36** 672–683. MR3871709 https://doi.org/10.1080/07350015.2016.1247004

ROSENBAUM, P. R. (2002). *Observational Studies*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR1899138 https://doi.org/10.1007/978-1-4757-3692-2

SABBAGHI, A. and RUBIN, D. B. (2014). Comments on the Neyman-Fisher controversy and its consequences. *Statist. Sci.* **29** 267–284. MR3264542 https://doi.org/10.1214/13-STS454

SCHOCHET, P. Z. (2010). Is regression adjustment supported by the Neyman model for causal inference? *J. Statist. Plann. Inference* **140** 246–259. MR2568136 https://doi.org/10.1016/j.jspi.2009.07.008

SCHOCHET, P. Z., PASHLEY, N. E., MIRATRIX, L. W. and KAUTZ, T. (2021). Design-based ratio estimators and central limit theorems for clustered, blocked RCTs. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2021.1906685

NEYMAN, J. (1923). On the application of probability theory to agricultural experiments. *Statist. Sci.* **5** 465–472.

SU, F. and DING, P. (2021). Model-assisted analyses of cluster-randomized experiments. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 994–1015. MR4349125 https://doi.org/10.1111/rssb.12468

WU, C.-F. (1981). On the robustness and efficiency of some randomized designs. *Ann. Statist.* **9** 1168–1177. MR0630100

WU, J. and DING, P. (2021). Randomization Tests for Weak Null Hypotheses in Randomized Experiments. *J. Amer. Statist. Assoc.* **116** 1898–1913. MR4353721 https://doi.org/10.1080/01621459.2020.1750415

WU, C. F. J. and HAMADA, M. S. (2009). *Experiments: Planning, Analysis, and Optimization*, 2nd ed. *Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2583259

YATES, F. (1935). Complex experiments. *Suppl. J. R. Stat. Soc.* **2** 181–247.

YATES, F. (1937). *The Design and Analysis of Factorial Experiments*. Imperial Bureau of Soil Science London, Harpenden.

ZHAO, A. and DING, P. (2021a). Regression-based causal inference with factorial experiments: Estimands, model specifications, and design-based properties. *Biometrika*. arXiv:2101.02400.

ZHAO, A. and DING, P. (2021b). Covariate-adjusted Fisher randomization tests for the average treatment effect. *J. Econometrics* **225** 278–294. MR4328643 https://doi.org/10.1016/j.jeconom.2021.04.007

ZHAO, A., DING, P. (2022). Supplement to "Reconciling design-based and model-based causal inferences for split-plot experiments." https://doi.org/10.1214/21-AOS2144SUPP

ZHAO, A., DING, P., MUKERJEE, R. and DASGUPTA, T. (2018). Randomization-based causal inference from split-plot designs. *Ann. Statist.* **46** 1876–1903. MR3845004 https://doi.org/10.1214/17-AOS1605

Supplementary Material to "Reconciling design-based and model-based causal inferences for split-plot experiments"

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 and proofs for the special case of uniform designs and extensions to the HC2 correction for the cluster-robust covariances, covariate adjustment via factor-based regression, the general $T_{\text{A}} \times T_{\text{B}}$ design, and the Fisher randomization test.

Section S5 gives the results of the simulation studies.

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 useful facts for verifying the results on design-based inference under the 2^2 split-plot design. The results are stated in terms of the general $T_{\text{A}} \times T_{\text{B}}$ design to facilitate generalization.

Let $a \in \mathcal{T}_{\text{A}} = \{0, 1, \dots, T_{\text{A}} - 1\}$ and $b \in \mathcal{T}_{\text{B}} = \{0, 1, \dots, T_{\text{B}} - 1\}$ indicate the levels of factors A and B in treatment combination $z \in \mathcal{T} = \mathcal{T}_{\text{A}} \times \mathcal{T}_{\text{B}}$ throughout unless specified otherwise. The $\mathcal{T}_{\text{A}}$ and $\mathcal{T}_{\text{B}}$ reduce to $\{0, 1\}$ under the 2^2 split-plot design. Assume lexicographical order of z for all vectors and matrices when applicable.

Let $\bar{M} = N/W$ be the average whole-plot size, and let $\alpha_w = M_w/\bar{M}$ be the whole-plot size factor for $w = 1, \dots, W$. Let $Z_{ws} = (A_w, B_{ws}) \in \mathcal{T}$ 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 $\mathcal{W}(z) = \{w : A_w = a\}$ 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 $U_w(z) = \bar{M}^{-1} \sum_{s=1}^{M_w} Y_{ws}(z) = \alpha_w \bar{Y}_w(z)$ be the whole-plot average potential outcome and the scaled whole-plot total potential outcome for whole-plot w under treatment z, respectively. We have

$$\bar{Y}(z) = W^{-1} \sum_{w=1}^{W} \alpha_w \bar{Y}_w(z) = W^{-1} \sum_{w=1}^{W} U_w(z).$$

Let $Y'_{ws}(z) = Y_{ws}(z) - \bar{Y}(z)$ be the centered potential outcome with $\bar{Y}'_w(z) = M_w^{-1} \sum_{s=1}^{M_w} Y'_{ws}(z) = \bar{Y}_w(z) - \bar{Y}(z)$ and $\bar{Y}'(z) = N^{-1} \sum_{ws \in \mathcal{S}} Y'_{ws}(z) = 0$.

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

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

$$S_{\text{haj}}(z, z') = (W-1)^{-1} \sum_{w=1}^{W} \alpha_w^2 \{Y_w(z) - \bar{Y}(z)\}\{Y_w(z') - \bar{Y}(z')\}.$$

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

(S1) $\quad \lambda_W \{S_{\text{haj}}(z, z') - S_{\text{ht}}(z, z')\} = \bar{Y}(z)\bar{Y}(z')(\overline{\alpha^2} + 1) - \bar{Y}(z')\, \overline{\alpha U(z)} - \bar{Y}(z)\, \overline{\alpha U(z')}$

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

Let $\hat{Y}_w(z) = M_{wb}^{-1} \sum_{s:Z_{ws}=z} Y_{ws}$ 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 $\hat{Y}_w(z) = \hat{U}_w(z) = 0$ for $z \notin \{(A_w 0), (A_w 1)\}$. We can decompose $\hat{Y}_{\text{ht}}(z)$ as

$$\begin{aligned}\hat{Y}_{\text{ht}}(z) &= W_a^{-1} \sum_{w \in \mathcal{W}(z)} \hat{U}_w(z) \\ &= W_a^{-1} \sum_{w \in \mathcal{W}(z)} U_w(z) + W_a^{-1} \sum_{w \in \mathcal{W}(z)} \{\hat{U}_w(z) - U_w(z)\} \\ &= \mu(z) + \sum_{w=1}^{W} \delta_w(z) \qquad \text{for } \ z = (ab) \in \mathcal{T},\end{aligned}$$

where

$$\mu(z) = W_a^{-1} \sum_{w \in \mathcal{W}(z)} U_w(z), \qquad \delta_w(z) = 1(A_w = a) \cdot W_a^{-1} \{\hat{U}_w(z) - U_w(z)\}.$$

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 $|\mathcal{T}| \times 1$ vectorizations of $\mu(z)$ and $\delta_w(z)$, respectively. Then

$$\hat{Y}_{\text{ht}} = \mu + \delta, \qquad \text{where} \quad \delta = \sum_{w=1}^{W} \delta_w. \tag{S2}$$

Let $\mathcal{A} = \sigma(A_1, \ldots, A_W)$ be the σ-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 $\mathcal{A}$. As a result, we have $E(\delta_w \mid \mathcal{A}) = E(\delta_w \mid A_w) = 0$, $E(\delta \mid \mathcal{A}) = 0$, $\text{cov}(\delta \mid \mathcal{A}) = \sum_{w=1}^W \text{cov}(\delta_w \mid \mathcal{A})$ with $\text{cov}(\delta_w \mid \mathcal{A}) = \text{cov}(\delta_w \mid A_w)$, and thus

$$\begin{aligned}&E(\delta_w) = 0; \qquad E(\delta) = 0; \\ &\text{cov}(\delta) = \textstyle\sum_{w=1}^W \text{cov}(\delta_w) \qquad \text{with } \ \text{cov}(\delta_w) = E\{\text{cov}(\delta_w \mid A_w)\}; \\ &\text{cov}(\mu, \delta) = E\{\text{cov}(\mu, \delta \mid \mathcal{A})\} + \text{cov}\{E(\mu \mid \mathcal{A}), E(\delta \mid \mathcal{A})\} = 0.\end{aligned} \tag{S3}$$

Further let $U_{ws}(z) = \alpha_w Y_{ws}(z)$ be the scaled potential outcome at the unit level. Then $U_w(z) = M_w^{-1} \sum_{s=1}^{M_w} U_{ws}(z)$ and $S_w(z, z') = (M_w - 1)^{-1} \sum_{s=1}^{M_w} \{U_{ws}(z) - U_w(z)\}\{U_{ws}(z') - U_w(z')\}$ equal the finite-population mean and covariance of $\{U_{ws}(z) : z \in \mathcal{T}\}_{s=1}^{M_w}$ in whole-plot w, respectively, with $\hat{U}_w(z) = M_{wb}^{-1} \sum_{s:Z_{ws}=z} U_{ws}(z)$ as the sample mean under treatment z. Let

$$H_w = \big(H_w(z, z')\big)_{z, z' \in \mathcal{T}} = \text{diag}(p_a^{-1})_{a \in \mathcal{T}_{\text{A}}} \otimes \{\text{diag}(q_{wb}^{-1})_{b \in \mathcal{T}_{\text{B}}} - 1_{T_{\text{B}} \times T_{\text{B}}}\}$$

with $H_w(z, z') = 1(a = a') p_a^{-1} \{q_{wb}^{-1} 1(z = z') - 1\}$ for $z = (ab)$ and $z' = (a'b')$. Then

$$\begin{aligned}\text{cov}\{\hat{U}_w(z), \hat{U}_w(z') \mid A_w = a\} &= M_w^{-1} \{q_{wb}^{-1} 1(z = z') - 1\} S_w(z, z') \\ &= p_a M_w^{-1} H_w(z, z') S_w(z, z')\end{aligned} \tag{S4}$$

for $z = (ab)$ and $z' = (ab')$ that share the same level of factor A. Let $\overline{U_{w\cdot}^4(z)} = M_w^{-1} \sum_{s=1}^{M_w} U_{ws}^4(z) = \alpha_w^4 \overline{Y_{w\cdot}^4(z)}$ be the uncentered fourth moment of $U_{ws}(z)$ in whole-plot w.

Lastly, for two $L \times 1$ vectors $u = (u_1, \ldots, u_L)^{\text{T}}$ and $v = (v_1, \ldots, v_L)^{\text{T}}$, we have

$$(u^{\text{T}} v)^4 \le \|u\|_2^4 \|v\|_2^4; \qquad \bar{u}^4 \le (\overline{u^2})^2 \le \overline{u^4}, \quad \text{where } \ \overline{u^k} = L^{-1} \sum_{l=1}^{L} u_l^k, \tag{S5}$$

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

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 $S_{qq'} = (N-1)^{-1}\sum_{i=1}^N \{Y_i(q) - \bar{Y}(q)\}\{Y_i(q') - \bar{Y}(q')\}^{\mathrm{T}}$ 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 $\hat{\tau} = \sum_{q=1}^Q G_q \hat{Y}(q)$ be the corresponding moment estimator with $\hat{Y}(q) = N_q^{-1}\sum_{i:Z_i=q} Y_i$, where Z_i and Y_i are the treatment indicator and observed outcome for unit i, respectively. Then*

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

where S_τ^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 as $N \to \infty$, for all $1 \leq q, q' \leq Q$, (i) $S_{qq'}$ has a finite limit, (ii) N_q/N has a limit in $(0,1)$, and (iii) $\max_{i=1,\ldots,N} |Y_i(q) - \bar{Y}(q)|^2/N = o(1)$. Then $\sqrt{N}(\hat{\tau} - \tau) \rightsquigarrow \mathcal{N}(0, V)$ with V denoting the limiting value of $N\mathrm{cov}(\hat{\tau})$.

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 $\sup_{W \geq 1} \beta_W < \infty$, we have $E\big[\{\sum_{w=1}^W E(\xi_{W,w}^2 \mid \mathcal{F}_{W,w-1}) - \beta_W^2\}^2\big] = o(1)$.*
(iii) *$\mathcal{L}(X_W) * \mathcal{N}(0, \beta_W^2) \rightsquigarrow \mathcal{L}_0$ for some probability distribution $\mathcal{L}_0$, where $*$ denotes convolution.*

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

S1.3. *New lemmas.* We give in this subsection the key lemmas for verifying the results on design-based inference under the 2^2 split-plot design. The lemmas and their proofs extend to the general $T_{\mathrm{A}} \times T_{\mathrm{B}}$ design with minimal modification.

Decomposition of $\mathbf{cov}(\hat{Y}_{\mathrm{ht}})$ in finite samples. Lemma S3 below separates the parts in $\mathrm{cov}(\hat{Y}_{\mathrm{ht}})$ that are due to μ and δ from (S2), respectively. The decomposition furnishes an alternative proof of Lemma 1 relative to Mukerjee and Dasgupta (2021).

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

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

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

PROOF OF LEMMA S3. The identities $\mathrm{cov}(\hat{Y}_{\mathrm{ht}}) = \mathrm{cov}(\mu) + \mathrm{cov}(\delta)$ and $\mathrm{cov}(\delta) = \sum_{w=1}^W \mathrm{cov}(\delta_w)$ follow from (S3). We verify below the analytic forms of $\mathrm{cov}(\mu)$ and $\mathrm{cov}(\delta_w)$, respectively.

For the analytic form of $\mathrm{cov}(\mu)$, define $U_w(a) = (U_w(a0), U_w(a1))^{\mathrm{T}}$ as the vector potential outcome of whole-plot w under $A_w = a \in \{0,1\}$. The finite-population mean and covariance of $\{U_w(a)\}_{w=1}^W$ equal $\bar{U}(a) = W^{-1}\sum_{w=1}^W U_w(a) = (\bar{Y}(a0), \bar{Y}(a1))^{\mathrm{T}}$ and

$$S_{\mathrm{ht}}(a) = (W-1)^{-1}\sum_{w=1}^W \{U_w(a) - \bar{U}(a)\}^2 = \begin{pmatrix} S_{\mathrm{ht}}(a0,a0) & S_{\mathrm{ht}}(a0,a1) \\ S_{\mathrm{ht}}(a1,a0) & S_{\mathrm{ht}}(a1,a1) \end{pmatrix} \quad (a = 0,1),$$

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

$$\begin{aligned} \mathrm{cov}(\mu) &= W_0^{-1}\begin{pmatrix} I_2 \\ 0_{2\times 2}\end{pmatrix} S_{\mathrm{ht}}(0)\,(I_2, 0_{2\times 2}) + W_1^{-1}\begin{pmatrix} 0_{2\times 2} \\ I_2\end{pmatrix} S_{\mathrm{ht}}(1)\,(0_{2\times 2}, I_2) - W^{-1}S_{\mathrm{ht}} \\ &= W^{-1}(H \circ S_{\mathrm{ht}}). \end{aligned}$$

For the analytic form of $\mathrm{cov}(\delta_w)$, it follows from the definition of δ_w and (S4) that

$$\text{(S6)} \quad \mathrm{cov}(\delta_w \mid A_w = a) = W_a^{-2}p_a M_w^{-1}\{H_w(a) \circ S_w\} = p_a^{-1}W^{-2}M_w^{-1}\{H_w(a)\circ S_w\}$$

with $H_w(0) = \mathrm{diag}(p_0^{-1}, 0) \otimes \{\mathrm{diag}(q_{w0}^{-1}, q_{w1}^{-1}) - 1_{2\times 2}\}$ and $H_w(1) = \mathrm{diag}(0, p_1^{-1}) \otimes \{\mathrm{diag}(q_{w0}^{-1}, q_{w1}^{-1}) - 1_{2\times 2}\}$ corresponding to the upper-left and lower-right 2×2 block matrices of H_w, respectively. It then follows from (S3) and $H_w(0) + H_w(1) = H_w$ that

$$\mathrm{cov}(\delta_w) = E\{\mathrm{cov}(\delta_w \mid A_w)\} = \sum_{a=0,1}\mathbb{P}(A_w = a)\cdot \mathrm{cov}(\delta_w \mid A_w = a) = W^{-2}M_w^{-1}(H_w \circ S_w).$$

□

The weak law of large numbers. We give in this part the weak law of large numbers for quantifying the probability limits of $\hat{Y}_*$ and $\hat{V}_*$ $(* = \mathrm{ht}, \mathrm{haj})$ under some weaker conditions than Condition 2, summarized in Condition S1 below.

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 $(0,1)$; $\epsilon \le \min_{w=1,\ldots,W} q_{wb} \le \max_{w=1,\ldots,W} q_{wb} \le 1-\epsilon$ for some $\epsilon \in (0, 1/2]$ independent of W;
(ii) $\bar{Y}$ has a finite limit; $S = O(1)$ and $\Psi = O(1)$;
(iii) $W^{-2}\sum_{w=1}^W \alpha_w^4 \overline{Y_{w\cdot}^4(z)} = o(1)$.

Condition 2 ensures that $\{Y_{ws}(z) : z \in \mathcal{T}\}_{ws\in\mathcal{S}}$, $\{Y'_{ws}(z) : z \in \mathcal{T}\}_{ws\in\mathcal{S}}$, and the finite population of all ones, namely $\{Y_{ws}(z) = 1 : z \in \mathcal{T}\}_{ws\in\mathcal{S}}$, all satisfy Condition S1(ii)–(iii).

LEMMA S4. *Assume split-plot randomization. Then*

(i) $\hat{Y}_{\mathrm{ht}} - \bar{Y} = o_{\mathbb{P}}(1)$ *provided Condition S1*(i)–(ii);
(ii) $\hat{1}_{\mathrm{ht}} = \mathrm{diag}\{\hat{1}_{\mathrm{ht}}(z)\}_{z\in\mathcal{T}} = I_{|\mathcal{T}|} + o_{\mathbb{P}}(1)$ *provided Condition S1*(i) *and* $\overline{\alpha^2} = O(1)$.

PROOF OF LEMMA S4. Standard result ensures $E(\hat{Y}_{\mathrm{ht}}) = \bar{Y}$. Lemma 1 ensures $\mathrm{cov}(\hat{Y}_{\mathrm{ht}}) = W^{-1}(H \circ S_{\mathrm{ht}} + \Psi) = o(1)$ under Condition S1(i)–(ii). The result for $\hat{Y}_{\mathrm{ht}}$ then follows from Markov's inequality. The result for $\hat{1}_{\mathrm{ht}}$ follows from applying statement (i) to the finite population of all ones. □

LEMMA S5. *Assume split-plot randomization and Condition S1*(i) *and* (iii). *Then*

(i) $W^{-2}\sum_{w=1}^{W} E\{\hat U_w^2(z)\hat U_w^2(z') \mid A_w = a\} = o(1)$ *for* $z = (ab)$ *and* $z' = (ab')$;
(ii) $W^2 \sum_{w=1}^{W} E(\|\delta_w\|_2^4 \mid A_w = a) = o(1)$; $\quad W^2 \sum_{w=1}^{W} E(\|\delta_w\|_2^4) = o(1)$.

PROOF OF LEMMA S5. For statement (i), it follows from $\hat U_w(z) = M_{wb}^{-1}\sum_{s:Z_{ws}=z} U_{ws}(z)$, (S5), and $q_{wb} \geq \epsilon$ by Condition S1(i) that

$$\hat U_w^4(z) \leq M_{wb}^{-1} \sum_{s:Z_{ws}=z} U_{ws}^4(z) \leq M_{wb}^{-1} \sum_{s=1}^{M_w} U_{ws}^4(z) = q_{wb}^{-1}\, \overline{U_{w\cdot}^4(z)} \leq \epsilon^{-1}\, \overline{U_{w\cdot}^4(z)},$$

recalling $\overline{U_{w\cdot}^4(z)} = M_w^{-1}\sum_{s=1}^{M_w} U_{ws}^4(z)$. This ensures

$$W^{-2}\sum_{w=1}^{W} E\{\hat U_w^4(z) \mid A_w = a\} \leq W^{-2}\epsilon^{-1}\sum_{w=1}^{W} \overline{U_{w\cdot}^4(z)} = o(1)$$

by Condition S1(iii). The result then follows from $\hat U_w^2(z)\hat U_w^2(z') \leq 2^{-1}\{\hat U_w^4(z) + \hat U_w^4(z')\}$.

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

$$W^2 \sum_{w=1}^{W} E(\|\delta_w\|_2^4 \mid A_w = a) = o(1). \tag{S7}$$

The second equality then follows from the law of total expectation.

To verify (S7), note that $\|\delta_w\|_2^2 = W_a^{-2}[\{\hat U_w(a1) - U_w(a1)\}^2 + \{\hat U_w(a0) - U_w(a0)\}^2]$ for w with $A_w = a$. This ensures

$$\|\delta_w\|_2^4 = (\|\delta_w\|_2^2)^2 \leq 2W_a^{-4}\{\hat U_w(a1) - U_w(a1)\}^4 + 2W_a^{-4}\{\hat U_w(a0) - U_w(a0)\}^4$$

for w with $A_w = a$ by the Cauchy–Schwarz inequality and hence

$$\begin{aligned} E(\|\delta_w\|_2^4 \mid A_w = a) \leq\ & 2W_a^{-4} E\big[\{\hat U_w(a1) - U_w(a1)\}^4 \mid A_w = a\big] \\ & +2W_a^{-4} E\big[\{\hat U_w(a0) - U_w(a0)\}^4 \mid A_w = a\big]. \end{aligned}$$

A sufficient condition for (S7) is thus

$$W^{-2}\sum_{w=1}^{W} E\big[\{\hat U_w(z) - U_w(z)\}^4 \mid A_w = a\big] = o(1) \qquad \text{for } z = (ab). \tag{S8}$$

With

$$\sum_{w=1}^{W} E\big[\{\hat U_w(z) - U_w(z)\}^4 \mid A_w = a\big] \leq 8\sum_{w=1}^{W} E\{\hat U_w^4(z) \mid A_w = a\} + 8\sum_{w=1}^{W} U_w^4(z)$$

by (S5), (S8) is guaranteed by statement (i) and $W^{-2}\sum_{w=1}^{W} U_w^4(z) \leq W^{-2}\sum_{w=1}^{W} \overline{U_{w\cdot}^4(z)} = o(1)$ by (S5) and Condition S1(iii).

□

Let

$$\hat T_{z,z'} = W_a^{-1}\sum_{w:A_w=a} \hat U_w(z)\hat U_w(z')$$

for $z = (ab)$ and $z' = (ab')$, as the sample analog of

$$\overline{U(z)U(z')} = W^{-1}\sum_{w=1}^{W} U_w(z)U_w(z').$$

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 Ψ. Lemma S6 below states the convergence of $\hat{T}_{z,z'}$ to its expectation, affording the basis for computing the probability limits of $\hat{V}_{\text{ht}}$ and $\hat{V}_{\text{haj}}$.

LEMMA S6. *Assume split-plot randomization and Condition S1. Then*

$$\hat{T}_{z,z'} - E(\hat{T}_{z,z'}) = o_{\mathbb{P}}(1) \qquad \textit{for } \ z = (ab), \ z' = (ab')$$

with $E(\hat{T}_{z,z'}) = \overline{U(z)U(z')} + p_a \Psi(z,z') = (1 - W^{-1}) S_{\text{ht}}(z,z') + \bar{Y}(z)\bar{Y}(z') + p_a \Psi(z,z')$.

PROOF OF LEMMA S6. By Markov's inequality, it suffices to verify the expression of $E(\hat{T}_{z,z'})$ and $\text{cov}(\hat{T}_{z,z'}) = o(1)$. Let $X_w = 1(A_w = a)\,\hat{U}_w(z)\hat{U}_w(z')$ to write $\hat{T}_{z,z'} = W_a^{-1}\sum_{w=1}^{W} X_w$. Let $\mu_w = E(X_w \mid A_w = a) = E\{\hat{U}_w(z)\hat{U}_w(z') \mid A_w = a\}$.

Expression of $E(\hat{T}_{z,z'})$. First, it follows from (S4) that

$$\begin{aligned}
\mu_w &= \text{cov}\{\hat{U}_w(z), \hat{U}_w(z') \mid A_w = a\} + E\{\hat{U}_w(z) \mid A_w = a\}\, E\{\hat{U}_w(z') \mid A_w = a\} \\
&= p_a M_w^{-1} H_w(z,z') S_w(z,z') + U_w(z) U_w(z').
\end{aligned}$$

This, together with

$$E(X_w) = E\{E(X_w \mid A_w)\} = p_a E(X_w \mid A_w = a) = p_a \mu_w, \tag{S9}$$

ensures

$$\begin{aligned}
E(\hat{T}_{z,z'}) = W_a^{-1}\sum_{w=1}^{W} E(X_w) &= W^{-1}\sum_{w=1}^{W}\mu_w = W^{-1}\sum_{w=1}^{W} U_w(z)U_w(z') + p_a\Psi(z,z') \\
&= (1 - W^{-1}) S_{\text{ht}}(z,z') + \bar{Y}(z)\bar{Y}(z') + p_a\Psi(z,z');
\end{aligned}$$

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

Limit of cov$(\hat{T}_{z,z'})$. By (S9),

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

This ensures

$$\begin{aligned}
\text{cov}(X_w) &= E\{\text{cov}(X_w \mid A_w)\} + \text{cov}\{E(X_w \mid A_w)\} \\
&= p_a E(X_w^2 \mid A_w = a) - p_a^2 \mu_w^2, \\
\text{cov}(X_w, X_k) &= \text{cov}\{E(X_w \mid A_w), E(X_k \mid A_k)\} + E\{\text{cov}(X_w, X_k \mid A_w, A_k)\} \\
&= \text{cov}\{E(X_w \mid A_w), E(X_k \mid A_k)\} \\
&= E\{E(X_w \mid A_w)\cdot E(X_k \mid A_k)\} - E(X_w)E(X_k) \\
&= -p_0 p_1 (W-1)^{-1}\mu_w\mu_k \qquad (w \neq k).
\end{aligned}$$

Thus,

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

where $\leq_\infty$ indicates less or equal to as $W \to \infty$. The result then follows from Lemma S5(i). $\square$

S1.4. *Proof of the main results.*

PROOF OF THEOREM 4.1. We verify below the results for $\hat{Y}_{\text{ht}}$ and $\hat{Y}_{\text{haj}}$, respectively.

Asymptotic Normality of $\hat{Y}_{\text{ht}}$. From (S2), $\sqrt{W}(\hat{Y}_{\text{ht}} - \bar{Y}) = \sqrt{W}\left(\mu - \bar{Y} + \delta\right)$. The Cramer–Wold device ensures that $\sqrt{W}(\hat{Y}_{\text{ht}} - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_{\text{ht}})$ as long as

$$\eta^{\mathrm{T}} \sqrt{W}\left(\mu - \bar{Y} + \delta\right) \rightsquigarrow \mathcal{N}(0, \eta^{\mathrm{T}} \Sigma_{\text{ht}} \eta) \tag{S10}$$

for arbitrary non-random 4×1 unit vector η.

Let $X = \eta^{\mathrm{T}} \sqrt{W}(\mu - \bar{Y})$, $\xi_w = \eta^{\mathrm{T}} \sqrt{W} \delta_w$, and $\xi = \sum_{w=1}^{W} \xi_w = \sqrt{W} \eta^{\mathrm{T}} \delta$ to write the left-hand side of (S10) as $X + \xi$. Let $\mathcal{F}_{W,0} = \mathcal{A}$ be the σ-algebra generated by $(A_w)_{w=1}^{W}$, and let $\mathcal{F}_{W,w}$ be the σ-algebra generated by $(A_w)_{w=1}^{W}$ and $\{(B_{vs})_{s=1}^{M_v} : v = 1, \ldots, w\}$ for $w = 1, \ldots, W$. Then $\mathcal{F}_{W,0} \subset \mathcal{F}_{W,1} \subset \cdots \subset \mathcal{F}_{W,W}$ such that $\{\mathcal{F}_{W,w} : w = 0, \ldots, W\}$ is a filtration. Intuitively, $\mathcal{F}_{W,0}$ contains all the information on the stage (I) cluster randomization, whereas $\mathcal{F}_{W,w}$ contains all the information on the stage (I) cluster randomization plus the subset of stage (II) stratified randomization in the first w whole-plots, $v = 1, \ldots, w$. We verify below the sufficient condition (S10) by checking that $(\xi_w)_{w=1}^{W}$ and X satisfy the three conditions of Lemma S2 with $\beta_W^2 = \eta^{\mathrm{T}} \Psi \eta$ with regard to filtration $\{\mathcal{F}_{W,w} : w = 0, \ldots, W\}$. 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.

For Lemma S2 condition (i), (S5) ensures

$$\xi_w^4 = W^2 (\eta^{\mathrm{T}} \delta_w)^4 \leq W^2 \|\eta\|_2^4 \cdot \|\delta_w\|_2^4 = W^2 \|\delta_w\|_2^4. \tag{S11}$$

The result then 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 S5(ii).

For Lemma S2 condition (ii), let $\sigma = \text{var}(\xi \mid \mathcal{F}_{W,0})$ and $\sigma_w = \text{var}(\xi_w \mid \mathcal{F}_{W,0}) = \text{var}(\xi_w \mid A_w)$ with $\sigma = \sum_{w=1}^{W} \sigma_w$. It follows from (S3) and Lemma S3 that $E(\xi \mid \mathcal{F}_{W,0}) = 0$ and $\text{var}(\xi) = \eta^{\mathrm{T}} \Psi \eta = \beta_W^2$. This, together with $E(\xi_w^2 \mid \mathcal{F}_{W,w-1}) = E(\xi_w^2 \mid \mathcal{F}_{W,0}) = \sigma_w$, ensures

$$\sum_{w=1}^{W} E(\xi_w^2 \mid \mathcal{F}_{W,w-1}) = \sigma, \quad \beta_W^2 = \text{var}(\xi) = E\{\text{var}(\xi \mid \mathcal{F}_{W,0})\} + \text{var}\{E(\xi \mid \mathcal{F}_{W,0})\} = E(\sigma)$$

such that $E\left[\{\sum_{w=1}^{W} E(\xi_w^2 \mid \mathcal{F}_{W,w-1}) - \beta_W^2\}^2\right] = E\left[\{\sigma - E(\sigma)\}^2\right] = \text{var}(\sigma)$. Lemma S2 condition (ii) is thus equivalent to

$$\text{var}(\sigma) = o(1). \tag{S12}$$

To verify (S12), view $\sigma_w = \text{var}(\xi_w \mid A_w)$ as the observed value of

$$\sigma_w(a) = \text{var}(\xi_w \mid A_w = a) = W\eta^{\mathrm{T}}\text{cov}\,(\delta_w \mid A_w = a)\,\eta \tag{S13}$$

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

$$\begin{aligned}\text{var}(\sigma) &= \text{var}\{W_0\hat{\sigma}(0) + W_1\hat{\sigma}(1)\}\\ &\le 2\text{var}\{W_0\hat{\sigma}(0)\} + 2\text{var}\{W_1\hat{\sigma}(1)\} = 2Wp_0p_1(S^2_{\sigma(0)} + S^2_{\sigma(1)}).\end{aligned}$$

Sufficient condition (S12) thus holds as long as $WS^2_{\sigma(a)} = o(1)$ for $a = 0, 1$. With $(W-1)S^2_{\sigma(a)} = \sum_{w=1}^{W}\{\sigma_w(a)\}^2 - W\{\bar{\sigma}(a)\}^2$, this is in turn guaranteed by

$$\sum_{w=1}^{W}\{\sigma_w(a)\}^2 = o(1), \qquad W\{\bar{\sigma}(a)\}^2 = o(1) \qquad (a = 0, 1). \tag{S14}$$

We verify below the two sufficient conditions in (S14).

First, (S13) and $E(\xi_w \mid A_w = a) = 0$ together ensure $\sigma_w(a) = E(\xi_w^2 \mid A_w = a)$ and hence

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

by Jensen's inequality and (S11). The first equality in (S14) then follows from

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

by Lemma S5(ii). The second equality in (S14) follows from (S13) and (S6) as

$$\bar{\sigma}(a) = \eta^{\mathrm{T}}\sum_{w=1}^{W}\text{cov}(\delta_w \mid A_w = a)\eta = \eta^{\mathrm{T}}\left[p_a^{-1}W^{-2}\sum_{w=1}^{W}M_w^{-1}\{H_w(a)\circ S_w\}\right]\eta = O(W^{-1})$$

by $\Psi = O(1)$. This verifies (S14) and hence Lemma S2 condition (ii).

Lemma S2 condition (iii) then follows from Lemmas S1 and S3 which ensure $\sqrt{W}(\mu - \bar{Y}) \rightsquigarrow \mathcal{N}(0, H \circ S_{\text{ht}})$ under Condition 2. The convolution of $\mathcal{L}(X)$ with $\mathcal{N}(0, \eta^{\mathrm{T}}\Psi\eta)$ thus converges in distribution to $\mathcal{N}(0, \eta^{\mathrm{T}}\Sigma_{\text{ht}}\eta)$ by the convergence of the characteristic function.

This verifies that $(\xi_w)_{w=1}^{W}$ and X satisfy the three conditions in Lemma S2. The sufficient condition (S10) then follows from Lemma S2 and ensures the result for $\hat{Y}_{\text{ht}}$.

Asymptotic Normality of $\hat{Y}_{\text{haj}}$. Recall from (4) in the main paper that $\hat{Y}_{\text{haj}} - \bar{Y} = \hat{1}_{\text{ht}}^{-1}\hat{Y}'_{\text{ht}}$ with $\hat{1}_{\text{ht}} = \text{diag}\{\hat{1}_{\text{ht}}(z)\}_{z\in\mathcal{T}}$ and $\hat{Y}'_{\text{ht}}$ as the vectorization of $\{\hat{Y}'_{\text{ht}}(z)\}_{z\in\mathcal{T}}$. The asymptotic Normality of $\hat{Y}_{\text{ht}}$ extends to $\hat{Y}'_{\text{ht}}$ as $\sqrt{W}(\hat{Y}'_{\text{ht}} - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_{\text{haj}})$. The result for $\hat{Y}_{\text{haj}}$ then follows from Slutsky's theorem with $\hat{1}_{\text{ht}}^{-1} = I_{|\mathcal{T}|} + o_{\mathbb{P}}(1)$ by Lemma S4. □

PROOF OF COROLLARY 1. By (S1), we have

$$(W-1)\{S_{\text{haj}}(z,z) - S_{\text{ht}}(z,z)\} = \{\bar{Y}(z)\}^2\left(\sum_{w=1}^{W}\alpha_w^2 + W\right) - 2\bar{Y}(z)\sum_{w=1}^{W}\alpha_w^2\bar{Y}_w(z).$$

When $\bar{Y}_w(z) = c$ for all $w = 1, \ldots, W$, we have $\bar{Y}(z) = W^{-1} \sum_{w=1}^{W} \alpha_w \bar{Y}_w(z) = c$ 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 $a_w = 1$ for all w or $c = 0$.

When $U_w(z) = c$ for all $w = 1, \ldots, W$, we have $\bar{Y}(z) = W^{-1} \sum_{w=1}^{W} U_w(z) = c$ 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 $a_w = 1$ for all w or $c = 0$. □

PROOF OF THEOREM 4.2. Let $\hat{V}_*(z,z')$ be the (z,z')th element of $\hat{V}_*$ for $* = \text{ht}, \text{haj}$. Assume $z = (ab)$ and $z' = (ab')$ with the same level of factor A throughout the proof.

Result on $\hat{V}_{\text{ht}}$. Direct 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 $z' = (ab')$. It then follows from Lemmas S4 and S6 that

$$\begin{aligned} W\hat{V}_{\text{ht}}(z,z') &= p_a^{-1}\{E(\hat{T}_{z,z'}) - \bar{Y}(z)\bar{Y}(z')\} + o_{\mathbb{P}}(1) \\ &= p_a^{-1}\{S_{\text{ht}}(z,z') + p_a \Psi(z,z')\} + o_{\mathbb{P}}(1) \\ &= \Sigma_{\text{ht}}(z,z') + S_{\text{ht}}(z,z') + o_{\mathbb{P}}(1), \end{aligned}$$

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 Σ_{ht}. This verifies the probability limit of $W\hat{V}_{\text{ht}}$.

Result on $\hat{V}_{\text{haj}}$. Direct algebra shows that

$$\begin{aligned} \text{(S15)}\quad (W_a - 1)\hat{V}_{\text{haj}}(z,z') = \hat{T}_{z,z'} &+ \hat{Y}_{\text{haj}}(z)\hat{Y}_{\text{haj}}(z')\left(W_a^{-1}\sum_{w:A_w=a}\alpha_w^2\right) \\ &- \hat{Y}_{\text{haj}}(z)\left\{W_a^{-1}\sum_{w:A_w=a}\alpha_w \hat{U}_w(z')\right\} \\ &- \hat{Y}_{\text{haj}}(z')\left\{W_a^{-1}\sum_{w:A_w=a}\alpha_w \hat{U}_w(z)\right\} \quad \text{for } z = (ab),\ z' = (ab'). \end{aligned}$$

We compute below the probability limit of the right-hand side of (S15).

First, apply Lemma S6 to $\{Y_{ws}(z), Y_{ws}(z')\}_{ws \in \mathcal{S}}$ with $Y_{ws}(z') = 1$ for all $ws \in \mathcal{S}$ to see

$$W_a^{-1}\sum_{w:A_w=a}\alpha_w \hat{U}_w(z) - \overline{\alpha U(z)} = o_{\mathbb{P}}(1).$$

This, together with $\overline{\alpha U(z)} = O(1)$ by $|\overline{\alpha U(z)} - \bar{Y}(z)| = W^{-1}|\sum_{w=1}^{W}\alpha_w\{U_w(z) - \bar{Y}(z)\}| \leq \overline{\alpha^2}\, S_{\text{ht}}(z,z)$, ensures

$$\text{(S16)}\qquad \hat{Y}_{\text{haj}}(z')\left\{W_a^{-1}\sum_{w:A_w=a}\alpha_w \hat{U}_w(z)\right\} - \bar{Y}(z')\,\overline{\alpha U(z)} = o_{\mathbb{P}}(1).$$

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

$$\text{(S17)}\qquad \hat{Y}_{\text{haj}}(z)\hat{Y}_{\text{haj}}(z')\left(W_a^{-1}\sum_{w:A_w=a}\alpha_w^2\right) - \bar{Y}(z)\bar{Y}(z')\,\overline{\alpha^2} = o_{\mathbb{P}}(1).$$

Plug (S16)–(S17) and the probability limit of $\hat{T}_{z,z'}$ from Lemma S6 in (S15) to see

$$
\begin{aligned}
(W_a - 1)\hat{V}_{\text{haj}}(z,z') &= \{S_{\text{ht}}(z,z') + \bar{Y}(z)\bar{Y}(z') + p_a\Psi(z,z')\} + \bar{Y}(z)\bar{Y}(z')\,\overline{\alpha^2} \\
&\quad - \bar{Y}(z)\,\overline{\alpha U(z')} - \bar{Y}(z')\,\overline{\alpha U(z)} + o_{\mathbb{P}}(1) \\
&= p_a\Psi(z,z') + (1 - W^{-1})S_{\text{haj}}(z,z') + W^{-1}S_{\text{ht}}(z,z') + o_{\mathbb{P}}(1)
\end{aligned}
$$

by (S1). This ensures $W\hat{V}_{\text{haj}}(z,z') = p_a^{-1}S_{\text{haj}}(z,z') + \Psi(z,z') + o_{\mathbb{P}}(1)$, and the result follows from $p_a^{-1}S_{\text{haj}}(z,z') = H(z,z')S_{\text{haj}}(z,z') + S_{\text{haj}}(z,z')$. □

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 sub-plot 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 WLS fit. We have

$$
\hat{Y}_{\text{ht}}(z) = \sum_{ws \in \mathcal{S}(z)} \pi_{ws}Y_{ws}, \qquad \hat{1}_{\text{ht}}(z) = \sum_{ws \in \mathcal{S}(z)} \pi_{ws} \tag{S18}
$$

with

$$
\pi_{ws} = N^{-1}p_a^{-1}q_{wb}^{-1} = \alpha_w W_a^{-1} M_{wb}^{-1} \tag{S19}
$$

for sub-plots under treatment $z = (ab)$.

Recall $d_{ws} = (1(Z_{ws} = 00), 1(Z_{ws} = 01), 1(Z_{ws} = 10), 1(Z_{ws} = 11))^{\mathrm{T}}$ and $d_w(A_w b) = (1(A_w b = 00), 1(A_w b = 01), 1(A_w b = 10), 1(A_w b = 11))^{\mathrm{T}}$ as the regressor vectors in regressions (6) and (7). Let $\tilde{\beta}_\dagger(z)$ be the zth element in $\tilde{\beta}_\dagger$ corresponding to $1(Z_{ws} = z)$ or $1(A_w b = z)$ for $\dagger = \text{ols}, \text{wls}, \text{ag}$, respectively. Let

$$
Y \sim D, \qquad U \sim D_{\text{ag}}
$$

be the matrix forms of (6) and (7), respectively, where Y and U are the vectorizations of $\{Y_{ws} : ws \in \mathcal{S}\}$ and $\{\hat{U}_w(A_w b) : w = 1,\dots,W;\ b = 0,1\}$, and D and D_{ag} are the matrices with rows $\{d_{ws} : ws \in \mathcal{S}\}$ and $\{d_w(A_w b) : w = 1,\dots,W;\ b = 0,1\}$, respectively. Let $\Pi = \text{diag}(\pi_{ws})_{ws \in \mathcal{S}}$ be the corresponding weighting matrix under fitting scheme "wls". Assume lexicographical orders of ws and (w,b) throughout unless specified otherwise.

Let $Y_w = (Y_{w1},\dots,Y_{wM_w})^{\mathrm{T}}$, $D_w = (d_{w1},\dots,d_{wM_w})^{\mathrm{T}}$, $\Pi_w = \text{diag}(\pi_{ws})_{s=1}^{M_w}$, $U_w = (\hat{U}_w(A_w 0), \hat{U}_w(A_w 1))^{\mathrm{T}}$, and $D_{\text{ag},w} = (d_w(A_w 0), d_w(A_w 1))^{\mathrm{T}}$ be the parts in Y, D, Π, U, and D_{ag} corresponding to whole-plot w, respectively. The cluster-robust covariances equal

$$
\tilde{V}_{\text{wls}} = (D^{\mathrm{T}}\Pi D)^{-1}\left(\sum_{w=1}^{W} D_w^{\mathrm{T}}\Pi_w e_{\text{wls},w}e_{\text{wls},w}^{\mathrm{T}}\Pi_w D_w\right)(D^{\mathrm{T}}\Pi D)^{-1}, \tag{S20}
$$

$$
\tilde{V}_{\text{ag}} = (D_{\text{ag}}^{\mathrm{T}}D_{\text{ag}})^{-1}\left(\sum_{w=1}^{W} D_{\text{ag},w}^{\mathrm{T}}e_{\text{ag},w}e_{\text{ag},w}^{\mathrm{T}}D_{\text{ag},w}\right)(D_{\text{ag}}^{\mathrm{T}}D_{\text{ag}})^{-1}, \tag{S21}
$$

where $e_{\text{wls},w} = (e_{\text{wls},w1},\dots,e_{\text{wls},wM_w})^{\mathrm{T}}$ and $e_{\text{ag},w} = (e_{\text{ag},w}(A_w 0), e_{\text{ag},w}(A_w 1))^{\mathrm{T}}$ are the residuals in whole-plot w from (6) and (7), respectively.

Let $P = \text{diag}(p_0, p_1) \otimes I_2$ and $Q_w = I_2 \otimes \text{diag}(q_{w0}, q_{w1})$ with $\text{diag}\{p_{ws}(z)\}_{z \in \mathcal{T}} = PQ_w$ for all ws. Let $R = \text{diag}(r_z)_{z \in \mathcal{T}}$, where $r_z = N_z/N$. Some useful facts are

$$
\begin{aligned}
&N^{-1}D^{\mathrm{T}}D = R, && D^{\mathrm{T}}\Pi D = \hat{1}_{\text{ht}}, && W^{-1}D_{\text{ag}}^{\mathrm{T}}D_{\text{ag}} = P, \\
&N^{-1}D^{\mathrm{T}}Y = R\hat{Y}_{\text{sm}}, && D^{\mathrm{T}}\Pi Y = \hat{Y}_{\text{ht}}, && W^{-1}D_{\text{ag}}^{\mathrm{T}}U = P\hat{Y}_{\text{ht}}; \\
&D_w^{\mathrm{T}}Y_w = M_w Q_w \hat{Y}_w, && D_w^{\mathrm{T}}\Pi_w Y_w = W^{-1}P^{-1}(\alpha_w \hat{Y}_w), && D_{\text{ag},w}^{\mathrm{T}}U_w = \hat{U}_w,
\end{aligned} \tag{S22}
$$

where $\hat{Y}_w$ and $\hat{U}_w$ are the 4×1 vectors of $\{\hat{Y}_w(z)\}_{z\in\mathcal{T}}$ and $\{\hat{U}_w(z)\}_{z\in\mathcal{T}}$ in lexicographical order of z, respectively, with $\hat{Y}_w(z) = \hat{U}_w(z) = 0$ for $z \notin \{(A_w 0), (A_w 1)\}$ by definition. The proof of (S22) follows from direct algebra and is thus omitted.

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

PROOF OF PROPOSITION 2. The result follows from $\tilde{\beta}_{\text{ols}} = (D^{\mathrm{T}}D)^{-1}D^{\mathrm{T}}Y$, $\tilde{\beta}_{\text{wls}} = (D^{\mathrm{T}}\Pi D)^{-1}D^{\mathrm{T}}\Pi Y$, $\tilde{\beta}_{\text{ag}} = (D_{\text{ag}}^{\mathrm{T}}D_{\text{ag}})^{-1}D_{\text{ag}}^{\mathrm{T}}U$, and (S22). □

PROOF OF THEOREM 5.1. We verify below the numeric expressions of $\tilde{V}_{\text{wls}}$ and $\tilde{V}_{\text{ag}}$ in finite samples. The asymptotic equivalence then follows from Lemma S4.

Numeric expression of $\tilde{V}_{\text{wls}}$. Proposition 2 ensures $e_{\text{wls},ws} = Y_{ws} - \tilde{\beta}_{\text{wls}}(Z_{ws}) = Y_{ws} - \hat{Y}_{\text{haj}}(Z_{ws})$. Let $\hat{e}_{\text{wls},w} = (\hat{e}_{\text{wls},w}(00), \hat{e}_{\text{wls},w}(01), \hat{e}_{\text{wls},w}(10), \hat{e}_{\text{wls},w}(11))^{\mathrm{T}}$ with

$$\hat{e}_{\text{wls},w}(z) = M_{wb}^{-1} \sum_{s:Z_{ws}=z} e_{\text{wls},ws} = \begin{cases} \hat{Y}_w(z) - \hat{Y}_{\text{haj}}(z) & \text{for } z \in \{(A_w 0), (A_w 1)\}, \\ 0 & \text{for } z \notin \{(A_w 0), (A_w 1)\}. \end{cases}$$

Set $Y_w = e_{\text{wls},w}$ in (S22) to see $D_w^{\mathrm{T}}\Pi_w e_{\text{wls},w} = W^{-1}P^{-1}(\alpha_w \hat{e}_{\text{wls},w})$. The "meat" part of (S20) thus equals

$$\sum_{w=1}^{W} D_w^{\mathrm{T}}\Pi_w e_{\text{wls},w} e_{\text{wls},w}^{\mathrm{T}} \Pi_w D_w = W^{-2}P^{-1}\left(\sum_{w=1}^{W} \alpha_w^2 \hat{e}_{\text{wls},w}\hat{e}_{\text{wls},w}^{\mathrm{T}}\right) P^{-1} = \big(\Omega_{\text{wls}}(z,z')\big)_{z,z'\in\mathcal{T}},$$

where

$$\begin{aligned} \Omega_{\text{wls}}(z,z') &= W_a^{-1}W_{a'}^{-1}\sum_{w=1}^{W} \alpha_w^2 \hat{e}_{\text{wls},w}(z)\hat{e}_{\text{wls},w}(z') \\ &= \begin{cases} W_a^{-2}(W_a - 1)\hat{S}_{\text{haj}}(z,z') & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a = a', \\ 0 & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a \neq a'. \end{cases} \end{aligned}$$

The result for $\tilde{V}_{\text{wls}}$ then follows from $D^{\mathrm{T}}\Pi D = \hat{1}_{\text{ht}}$ by (S22).

Numeric expression of $\tilde{V}_{\text{ag}}$. Proposition 2 ensures $e_{\text{ag},w}(z) = \hat{U}_w(z) - \tilde{\beta}_{\text{ag}}(z) = \hat{U}_w(z) - \hat{Y}_{\text{ht}}(z)$ for $z \in \{(A_w 0), (A_w 1)\}$. Set $U_w = e_{\text{ag},w}$ in (S22) to see $D_{\text{ag},w}^{\mathrm{T}} e_{\text{ag},w} = \hat{e}_{\text{ag},w}$ with $\hat{e}_{\text{ag},w} = (e_{\text{ag},w}(00), e_{\text{ag},w}(01), e_{\text{ag},w}(10), e_{\text{ag},w}(11))^{\mathrm{T}}$, where $e_{\text{ag},w}(z) = 0$ for $z \notin \{(A_w 0), (A_w 1)\}$. The "meat" part of (S21) thus equals

$$\sum_{w=1}^{W} D_{\text{ag},w}^{\mathrm{T}} e_{\text{ag},w} e_{\text{ag},w}^{\mathrm{T}} D_{\text{ag},w} = \sum_{w=1}^{W} \hat{e}_{\text{ag},w}\hat{e}_{\text{ag},w}^{\mathrm{T}} = \big(\Omega_{\text{ag}}(z,z')\big)_{z,z'\in\mathcal{T}}, \tag{S23}$$

where

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

The result for $\tilde{V}_{\text{ag}}$ then follows from $(D_{\text{ag}}^{\mathrm{T}}D_{\text{ag}})^{-1} = \text{diag}(W_0^{-1}, W_1^{-1}) \otimes I_2$ by (S22). □

S3. Regression-based covariate adjustment.

S3.1. *Notation and lemmas.* Inherit all notation from Section S2.1. In addition, recall $\bar{x}_w = M_w^{-1}\sum_{s=1}^{M_w} x_{ws}$ as the whole-plot average covariate vector with $W^{-1}\sum_{w=1}^{W}\alpha_w\bar{x}_w = \bar{x} = 0_J$. Let $v_{ws} = \alpha_w x_{ws}$, $v_w = \alpha_w\bar{x}_w$, and $U_w'(z) = \alpha_w\bar{Y}_w'(z) = \alpha_w\bar{Y}_w(z) - \alpha_w\bar{Y}(z)$ be the analogs of $U_{ws}(z) = \alpha_w Y_{ws}(z)$ and $U_w(z) = \alpha_w\bar{Y}_w(z)$ defined on the covariates and centered potential outcomes, respectively. We have

$$S_{xx} = (W-1)^{-1}\sum_{w=1}^{W}\alpha_w^2\bar{x}_w\bar{x}_w^{\mathrm{T}} = (W-1)^{-1}\sum_{w=1}^{W}v_w v_w^{\mathrm{T}},$$

$$S_{xx,w} = (M_w-1)^{-1}\alpha_w^2\sum_{s=1}^{M_w}(x_{ws}-\bar{x}_w)(x_{ws}-\bar{x}_w)^{\mathrm{T}} = (M_w-1)^{-1}\sum_{s=1}^{M_w}(v_{ws}-v_w)(v_{ws}-v_w)^{\mathrm{T}},$$

$$S_{xY(z)} = S_{Y(z)x}^{\mathrm{T}} = (W-1)^{-1}\sum_{w=1}^{W}\alpha_w\bar{x}_w\{\alpha_w\bar{Y}_w(z)-\bar{Y}(z)\} = (W-1)^{-1}\sum_{w=1}^{W}v_w U_w(z),$$

$$S_{xY'(z)} = S_{Y'(z)x}^{\mathrm{T}} = (W-1)^{-1}\sum_{w=1}^{W}\alpha_w\bar{x}_w\{\alpha_w\bar{Y}_w'(z)\} = (W-1)^{-1}\sum_{w=1}^{W}v_w U_w'(z),$$

$$S_{xY(z),w} = S_{Y(z)x,w}^{\mathrm{T}} = S_{xY'(z),w} = S_{Y'(z)x,w}^{\mathrm{T}} = (M_w-1)^{-1}\alpha_w^2\sum_{s=1}^{M_w}(x_{ws}-\bar{x}_w)\{Y_{ws}(z)-\bar{Y}_w(z)\}$$

$$= (M_w-1)^{-1}\sum_{s=1}^{M_w}(v_{ws}-v_w)\{U_{ws}(z)-U_w(z)\}.$$

Let

$$\hat{S}_{xx}(z) = W_a^{-1}\sum_{w\in\mathcal{W}(z)}\hat{v}_w(z)\hat{v}_w^{\mathrm{T}}(z), \qquad \hat{S}_{xY(z)} = W_a^{-1}\sum_{w\in\mathcal{W}(z)}\hat{v}_w(z)\hat{U}_w(z)$$

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

Recall $Q_{xx} = (N-1)^{-1}\sum_{ws\in\mathcal{S}}x_{ws}x_{ws}^{\mathrm{T}}$ and $Q_{xY(z)} = (N-1)^{-1}\sum_{ws\in\mathcal{S}}x_{ws}Y_{ws}(z)$ as the finite-population covariances of $(x_{ws})_{ws\in\mathcal{S}}$ and $\{x_{ws}, Y_{ws}(z)\}_{ws\in\mathcal{S}}$, respectively. Let

$$\hat{Q}_{xx}(z) = \lambda_N^{-1}\sum_{ws\in\mathcal{S}(z)}\pi_{ws}x_{ws}x_{ws}^{\mathrm{T}}, \qquad \hat{Q}_{xY(z)} = \lambda_N^{-1}\sum_{ws\in\mathcal{S}(z)}\pi_{ws}x_{ws}Y_{ws},$$

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

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 Ψ from Lemma 1 with $H_w(z,z) = p_a^{-1}(q_{wb}^{-1}-1)$ for $z = (ab)$. Then

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

by replacing $S_w(z,z)$ with $S_{xx,w}$ and $S_{xY(z),w}$, respectively. Let

$$T_{xx}(z) = S_{xx} + p_a\Psi_{xx}(z,z), \qquad T_{xY}(z) = S_{xY(z)} + p_a\Psi_{xY}(z,z).$$

LEMMA S7. *Under the 2^2 split-plot randomization and Conditions 2–3, we have*

(i) $\hat{x}_{\mathrm{ht}}(z) = \sum_{ws\in\mathcal{S}(z)}\pi_{ws}x_{ws} = o_{\mathbb{P}}(1)$*;*

(ii) $\hat{Q}_{xx}(z) - Q_{xx} = o_{\mathbb{P}}(1)$, $\hat{Q}_{xY(z)} - Q_{xY(z)} = o_{\mathbb{P}}(1)$; $\hat{S}_{xx}(z) - T_{xx}(z) = o_{\mathbb{P}}(1)$, $\hat{S}_{xY(z)} - T_{xY}(z) = o_{\mathbb{P}}(1)$

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

PROOF OF LEMMA S7. The result on $\hat{x}_{\text{ht}}(z)$ follows from applying Lemma S4(i) componentwise. We verify below statement (ii) 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 $Q_{xx} = (N-1)^{-1}\sum_{ws\in\mathcal{S}} \sigma_{ws}$. Lemma S4 ensures that $\hat{Q}_{xx}(z) - Q_{xx} = o_{\mathbb{P}}(1)$ for all $z \in \mathcal{T}$ as long as Condition S1(ii) holds for the finite population of $\{Y_{ws}(z) = \sigma_{ws} : z \in \mathcal{T}\}_{ws\in\mathcal{S}}$. To verify this, let $\bar{\sigma}_w = M_w^{-1}\sum_{s=1}^{M_w} \sigma_{ws}$ with $\bar{\sigma}_w^2 \le M_w^{-1}\sum_{s=1}^{M_w} \sigma_{ws}^2 = M_w^{-1}\sum_{s=1}^{M_w} x_{ws}^4$ by (S5). This ensures

$$W^{-1}\sum_{w=1}^{W} \alpha_w^2 \bar{\sigma}_w^2 \le W^{-1}\sum_{w=1}^{W} \alpha_w^2 \left(M_w^{-1}\sum_{s=1}^{M_w} x_{ws}^4\right) = O(1) \tag{S24}$$

by Condition 3(iii). Condition S1(ii) is thus satisfied with

- $\bar{\sigma} = N^{-1}\sum_{ws\in\mathcal{S}} \sigma_{ws} = \lambda_N Q_{xx}$ having a finite limit by Condition 3(i);
- $S_{\sigma\sigma} = (W-1)^{-1}\sum_{w=1}^{W} (\alpha_w\bar{\sigma}_w - \bar{\sigma})^2 \le (W-1)^{-1}\sum_{w=1}^{W} \alpha_w^2\bar{\sigma}_w^2 = O(1)$ by (S24);
- $\Psi_{\sigma\sigma} = W^{-1}\sum_{w=1}^{W} M_w^{-1}\{H_w \circ (S_{\sigma\sigma,w} 1_{4\times 4})\} = W^{-1}\sum_{w=1}^{W} M_w^{-1} S_{\sigma\sigma,w} H_w = O(1)$ given
$$S_{\sigma\sigma,w} = \frac{1}{M_w-1}\alpha_w^2\left(\sum_{s=1}^{M_w}\sigma_{ws}^2 - M_w\bar{\sigma}_w^2\right) \le \frac{1}{M_w-1}\alpha_w^2\left(\sum_{s=1}^{M_w}\sigma_{ws}\right)^2 = \frac{M_w^2}{M_w-1}\alpha_w^2\bar{\sigma}_w^2$$
and (S24).

This verifies $\hat{Q}_{xx}(z) - Q_{xx} = o_{\mathbb{P}}(1)$. The proof for $\hat{Q}_{xY(z)} - Q_{xY} = o_{\mathbb{P}}(1)$ is almost identical by verifying Condition S1(ii) for $\{x_{ws}Y_{ws}(z) : z \in \mathcal{T}\}_{ws\in\mathcal{S}}$ and thus omitted.

The result on $\hat{S}_{xx}(z)$ follows by applying Lemma S6 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 $\hat{S}_{xY(z)}$ to follow from letting $Y_{ws}(z') = x_{ws}$ in Lemma S6. □

For a set of $J\times 1$ vectors $\gamma = (\gamma_z)_{z\in\mathcal{T}}$, let $Y_{ws}(z;\gamma_z) = Y_{ws}(z) - x_{ws}^{\mathrm{T}}\gamma_z$ be the adjusted potential outcome based on γ_z, and let $S_{*,\gamma}$, $\Sigma_{*,\gamma}$, $S_{w,\gamma}$, Ψ_γ, $\hat{Y}_{*,\gamma} = \{\hat{Y}_{*,\gamma}(z)\}_{z\in\mathcal{T}}$, $\hat{U}_{w,\gamma}(z) = \alpha_w\hat{Y}_{w,\gamma}(z)$, and $\hat{V}_{*,\gamma}$ be the analogs of S_*, Σ_*, S_w, Ψ, $\hat{Y}_* = \{\hat{Y}_*(z)\}_{z\in\mathcal{T}}$, $\hat{U}_w(z) = \alpha_w\hat{Y}_w(z)$, and $\hat{V}_*$ based on $\{Y_{ws}(z;\gamma_z) : z\in\mathcal{T}\}_{ws\in\mathcal{S}}$ for $* = \text{ht}, \text{haj}$, respectively. We have $\Psi_\gamma = W^{-1}\sum_{w=1}^{W} M_w^{-1} H_w \circ S_{w,\gamma}$ and $\Sigma_{*,\gamma} = H\circ S_{*,\gamma} + \Psi_\gamma$. With a slight repetition, let

$$\hat{Y}_*(z;\gamma) = \hat{Y}_{*,\gamma}(z), \qquad \hat{Y}_*(\gamma) = \hat{Y}_{*,\gamma}, \qquad \hat{U}_w(z;\gamma) = \hat{U}_{w,\gamma}(z). \tag{S25}$$

LEMMA S8. *Assume split-plot randomization and Conditions 2–3. For* $\hat{\gamma} = (\hat{\gamma}_z)_{z\in\mathcal{T}}$ *with* $\hat{\gamma}_z = \gamma_z + o_{\mathbb{P}}(1)$, *where* γ_z *is some fixed* $J\times 1$ *vector, we have* $\sqrt{W}(\hat{Y}_{*,\hat{\gamma}} - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_{*,\gamma})$ *and* $W\hat{V}_{*,\hat{\gamma}} - \Sigma_{*,\gamma} = S_{*,\gamma} + o_{\mathbb{P}}(1)$ *for* $* = \text{ht}, \text{haj}$.

PROOF OF LEMMA S8. We verify below the result for $* = \text{ht}$. The proof for $* = \text{haj}$ is almost identical and thus omitted.

First, the (z,z')th elements of $S_{\text{ht},\gamma}$, $S_{w,\gamma}$, and $S_{\text{haj},\gamma}$ equal

$$\begin{aligned} S_{\text{ht}}(z,z';\gamma) &= S_{\text{ht}}(z,z') - \gamma_z^{\mathrm{T}} S_{xY(z')} - S_{Y(z)x}\gamma_{z'} + \gamma_z^{\mathrm{T}} S_{xx}\gamma_{z'}, \\ S_w(z,z';\gamma) &= S_w(z,z') - \gamma_z^{\mathrm{T}} S_{xY(z'),w} - S_{Y(z)x,w}\gamma_{z'} + \gamma_z^{\mathrm{T}} S_{xx,w}\gamma_{z'}, \\ S_{\text{haj}}(z,z';\gamma) &= S_{\text{haj}}(z,z') - \gamma_z^{\mathrm{T}} S_{xY'(z')} - S_{Y'(z)x}\gamma_{z'} + \gamma_z^{\mathrm{T}} S_{xx}\gamma_{z'}, \end{aligned} \tag{S26}$$

respectively, by direct algebra. Conditions 2–3 together imply that Condition 2 holds for the finite population of $\{Y_{ws}(z;\gamma_z) : z \in \mathcal{T}\}_{ws\in\mathcal{S}}$ with fixed $\gamma = (\gamma_z)_{z\in\mathcal{T}}$. This ensures

$$\text{(S27)} \qquad \sqrt{W}(\hat{Y}_{\text{ht},\gamma} - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_{\text{ht},\gamma}), \qquad W\hat{V}_{\text{ht},\gamma} - \Sigma_{\text{ht},\gamma} = S_{\text{ht},\gamma} + o_{\mathbb{P}}(1)$$

by Theorems 4.1–4.2.

Result on $\hat{Y}_{\text{ht},\hat{\gamma}}$. By Slutsky's theorem,

$$\text{(S28)} \qquad \sqrt{W}\{\hat{Y}_{\text{ht}}(z;\hat{\gamma}_z) - \hat{Y}_{\text{ht}}(z;\gamma_z)\} = -(\hat{\gamma}_z - \gamma_z)^{\mathrm{T}}\sqrt{W}\hat{x}_{\text{ht}}(z) = o_{\mathbb{P}}(1)$$

given $\hat{\gamma}_z - \gamma_z = o_{\mathbb{P}}(1)$ and the asymptotic Normality of $\sqrt{W}\hat{x}_{\text{ht}}(z)$ by Theorem 4.1. This ensures $\sqrt{W}(\hat{Y}_{\text{ht},\hat{\gamma}} - \hat{Y}_{\text{ht},\gamma}) = o_{\mathbb{P}}(1)$. The result follows from (S27) and Slutsky's theorem.

Result on $W\hat{V}_{\text{ht},\hat{\gamma}}$. By (S27), it suffices to verify $W(\hat{V}_{\text{ht},\hat{\gamma}} - \hat{V}_{\text{ht},\gamma}) = o_{\mathbb{P}}(1)$. This is in turn guaranteed by

$$\text{(S29)} \qquad \hat{S}_{\text{ht}}(z,z';\hat{\gamma}) - \hat{S}_{\text{ht}}(z,z';\gamma) = o_{\mathbb{P}}(1),$$

where

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

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

for all $z = (ab)$ and $z' = (ab')$ with the same level of factor A. Given $\hat{Y}_{\text{ht}}(z;\hat{\gamma}_z) - \hat{Y}_{\text{ht}}(z;\gamma_z) = o_{\mathbb{P}}(1)$ by (S28), (S29) holds as long as the difference between the first terms satisfies

$$\text{(S30)} \quad \Delta = W_a^{-1}\left\{\sum_{w:A_w=a}\hat{U}_w(z;\hat{\gamma}_z)\hat{U}_w(z';\hat{\gamma}_{z'}) - \sum_{w:A_w=a}\hat{U}_w(z;\gamma_z)\hat{U}_w(z';\gamma_{z'})\right\} = o_{\mathbb{P}}(1).$$

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

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

such that

$$\begin{aligned}
\Delta &= W_a^{-1}\sum_{w:A_w=a}\left\{\Delta_w(z)\Delta_w(z') - \Delta_w(z)\hat{U}_w(z';\gamma_{z'}) - \Delta_w(z')\hat{U}_w(z;\gamma_z)\right\}\\
&= (\hat{\gamma}_z - \gamma_z)^{\mathrm{T}}\left\{W_a^{-1}\sum_{w:A_w=a}\hat{v}_w(z)\hat{v}_w^{\mathrm{T}}(z')\right\}(\hat{\gamma}_{z'} - \gamma_{z'})\\
&\quad -(\hat{\gamma}_z - \gamma_z)^{\mathrm{T}}\left\{W_a^{-1}\sum_{w:A_w=a}\hat{v}_w(z)\hat{U}_w(z';\gamma_{z'})\right\} - (\hat{\gamma}_{z'} - \gamma_{z'})^{\mathrm{T}}\left\{W_a^{-1}\sum_{w:A_w=a}\hat{v}_w(z')\hat{U}_w(z;\gamma_z)\right\}\\
&= o_{\mathbb{P}}(1);
\end{aligned}$$

the last equality follows from $W_a^{-1}\sum_{w:A_w=a}\hat{v}_w(z)\hat{v}_w^{\mathrm{T}}(z') = O_{\mathbb{P}}(1)$ and

$$W_a^{-1}\sum_{w:A_w=a}\hat{v}_w(z)\hat{U}_w(z';\gamma_{z'}) = W_a^{-1}\sum_{w:A_w=a}\hat{v}_w(z)\hat{U}_w(z') - \left\{W_a^{-1}\sum_{w:A_w=a}\hat{v}_w(z)\hat{v}_w^{\mathrm{T}}(z')\right\}\gamma_{z'} = O_{\mathbb{P}}(1)$$

by Lemmas S6–S7. This verifies (S30) and hence the result. □

LEMMA S9. *(Styan, 1973, Theorem 3.1) If G_1 and G_2 are positive semi-definite, then their Hadamard product $G_1 \circ G_2$ is also positive semi-definite.*

Lastly, let

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

be extensions of $\Psi_{xx}(z,z)$ and $\Psi_{xY}(z,z)$ to $z,z' \in \mathcal{T}$.

LEMMA S10. *Under Conditions 2–3 and (12), we have*

$$\Psi_{xx}(z,z') = o(1), \qquad \Psi_{xY}(z,z') = o(1) \qquad (z,z' \in \mathcal{T}).$$

PROOF OF LEMMA S10. We verify below the result for scalar covariate $x_{ws} \in \mathbb{R}$ to simplify the presentation. Let $u_0 > 0$ and $l_0 > 0$ be some uniform upper and lower bounds of $|H_w(z,z')|$ for all $z,z' \in \mathcal{T}$ and $w = 1,\ldots,W$ under Condition 2(ii), both independent of W. That is, $l_0 \le |H_w(z,z')| \le u_0$ for all $z,z' \in \mathcal{T}$ and $w = 1,\ldots,W$.

First, with $H_w(z,z) > 0$ and $l_0^{-1} H_w(z,z) \ge 1$, Condition (12) ensures

$$W^{-1}\sum_{w=1}^{W} M_w^{-1} S_{xx,w} \le W^{-1}\sum_{w=1}^{W} M_w^{-1}\{l_0^{-1} H_w(z,z)\} S_{xx,w} = l_0^{-1}\Psi_{xx}(z,z) = o(1). \tag{S31}$$

The result for $\Psi_{xx}(z,z')$ then follows from $H_w(z,z') = O(1)$.

Second, the Cauchy–Schwarz inequality ensures $|S_{xY(z),w}| \le S_{xx,w}^{1/2}\{S_w(z,z)\}^{1/2}$. This, together with $|H_w(z,z')| \le u_0$, suggests

$$\begin{aligned}
|\Psi_{xY}(z,z')| &\le W^{-1}\sum_{w=1}^{W} M_w^{-1} \left|H_w(z,z')\right| \left|S_{xY(z'),w}\right| \\
&\le u_0 W^{-1}\sum_{w=1}^{W} M_w^{-1} S_{xx,w}^{1/2}\{S_w(z',z')\}^{1/2} \\
&\le u_0 W^{-1}\sum_{w=1}^{W} \left(M_w^{-1} S_{xx,w}\right)^{1/2} \left\{M_w^{-1} S_w(z',z')\right\}^{1/2} \\
&\le u_0 \left(W^{-1}\sum_{w=1}^{W} M_w^{-1} S_{xx,w}\right)^{1/2} \left\{W^{-1}\sum_{w=1}^{W} M_w^{-1} S_w(z',z')\right\}^{1/2} \\
&= o(1);
\end{aligned}$$

the last equality follows from (S31) and the fact that $W^{-1}\sum_{w=1}^{W} M_w^{-1} S_w(z',z') \le l_0^{-1}\Psi(z',z') = O(1)$ by similar reasoning. □

S3.2. *Results under the fully-interacted regressions.* We verify in this part the results under the fully-interacted regressions (10) and (11).

LEMMA S11. *Under the 2^2 split-plot randomization and Conditions 2–3, we have $\gamma_{\text{wls},z} = Q_{xx}^{-1} Q_{xY(z)}$ and $\gamma_{\text{ag},z} = T_{xx}^{-1}(z) T_{xY}(z)$ for $z \in \mathcal{T}$.*

PROOF OF PROPOSITION 4 AND LEMMA S11. We verify below the results for the unit and aggregate regressions, respectively.

Unit regression. For $\dagger = \text{ols}, \text{wls}$, the inclusion of full interactions ensures that $\tilde{\beta}_{\dagger,\text{L}}(z)$ and $\tilde{\gamma}_{\dagger,z}$ from (10) equal the coefficients of 1 and x_{ws} from the treatment-specific regression

$$\text{(S32)} \qquad Y_{ws} \sim 1 + x_{ws} \qquad \text{over } ws \in \mathcal{S}(z),$$

respectively, under their respective fitting schemes. Let Y_z and X_z be the concatenations of Y_{ws} and x_{ws} over $ws \in \mathcal{S}(z)$, respectively. The matrix form of (S32) equals $Y_z \sim 1_{N_z} + X_z$ for $z \in \mathcal{T}$.

That $\tilde{\beta}_{\text{ols,L}}(z) = \hat{Y}_{\text{sm}}(z) - \hat{x}^{\text{T}}_{\text{sm}}(z)\tilde{\gamma}_{\text{ols},z}$ follows from standard results.

Let $\Pi_z = \text{diag}(\pi_{ws})_{ws\in\mathcal{S}(z)}$ be the weighting matrix under fitting scheme "wls". The first-order condition of WLS ensures $G_1(\tilde{\beta}_{\text{wls,L}}(z), \tilde{\gamma}^{\text{T}}_{\text{wls},z})^{\text{T}} = G_2$, where

$$G_1 = (1_{N_z}, X_z)^{\text{T}}\Pi_z(1_{N_z}, X_z) = \begin{pmatrix} 1^{\text{T}}_{N_z}\Pi_z 1_{N_z} & 1^{\text{T}}_{N_z}\Pi_z X_z \\ X^{\text{T}}_z\Pi_z 1_{N_z} & X^{\text{T}}_z\Pi_z X_z \end{pmatrix} = \begin{pmatrix} \hat{1}_{\text{ht}}(z) & \hat{x}^{\text{T}}_{\text{ht}}(z) \\ \hat{x}_{\text{ht}}(z) & \lambda_N \hat{Q}_{xx}(z) \end{pmatrix},$$

$$G_2 = (1_{N_z}, X_z)^{\text{T}}\Pi_z Y_z = \begin{pmatrix} 1^{\text{T}}_{N_z}\Pi_z Y_z \\ X^{\text{T}}_z\Pi_z Y_z \end{pmatrix} = \begin{pmatrix} \hat{Y}_{\text{ht}}(z) \\ \lambda_N \hat{Q}_{xY(z)} \end{pmatrix}$$

by (S18) and $1^{\text{T}}_{N_z}\Pi_z X_z = \hat{x}^{\text{T}}_{\text{ht}}(z)$. Compare the first row to see $\hat{1}_{\text{ht}}(z)\tilde{\beta}_{\text{wls,L}}(z) + \hat{x}^{\text{T}}_{\text{ht}}(z)\tilde{\gamma}_{\text{wls},z} = \hat{Y}_{\text{ht}}(z)$ and hence the numeric result for $\tilde{\beta}_{\text{wls,L}}(z)$. The probability limit then follows from $(\tilde{\beta}_{\text{wls,L}}(z), \tilde{\gamma}^{\text{T}}_{\text{wls},z})^{\text{T}} = G_1^{-1}G_2$ with $G_1 = \text{diag}(1, Q_{xx}) + o_{\mathbb{P}}(1)$ and $G_2 = (\bar{Y}(z), Q^{\text{T}}_{xY(z)})^{\text{T}} + o_{\mathbb{P}}(1)$ by Lemmas S4 and S7.

Aggregate regression. The inclusion of full interactions ensures that $\tilde{\beta}_{\text{ag,L}}(z)$ and $\tilde{\gamma}_{\text{ag},z}$ from (11) equal the OLS coefficients of 1 and $\hat{v}_w(z)$ from the treatment-specific regression

$$\text{(S33)} \qquad \hat{U}_w(z) \sim 1 + \hat{v}_w(z) \qquad \text{over } w \in \mathcal{W}(z),$$

respectively. Let U_z be the vectorization of the W_a observations, namely $\{\hat{U}_w(z) : w \in \mathcal{W}(z)\}$, under treatment $z = (ab)$, and let Λ_z be the concatenation of the corresponding $\hat{v}_w(z)$'s. The matrix form of (S33) equals $U_z \sim 1_{W_a} + \Lambda_z$ for $z \in \mathcal{T}$. The first-order condition of OLS ensures $G_1(\tilde{\beta}_{\text{ag,L}}(z), \tilde{\gamma}^{\text{T}}_{\text{ag},z})^{\text{T}} = G_2$, where

$$G_1 = W_a^{-1}(1_{W_a}, \Lambda_z)^{\text{T}}(1_{W_a}, \Lambda_z) = W_a^{-1}\begin{pmatrix} 1^{\text{T}}_{W_a}1_{W_a} & 1^{\text{T}}_{W_a}\Lambda_z \\ \Lambda^{\text{T}}_z 1_{W_a} & \Lambda^{\text{T}}_z\Lambda_z \end{pmatrix} = \begin{pmatrix} 1 & \hat{x}^{\text{T}}_{\text{ht}}(z) \\ \hat{x}_{\text{ht}}(z) & \hat{S}_{xx}(z) \end{pmatrix},$$

$$G_2 = W_a^{-1}(1_{W_a}, \Lambda_z)^{\text{T}}U_z = W_a^{-1}\begin{pmatrix} 1^{\text{T}}_{W_a}U_z \\ \Lambda^{\text{T}}_z U_z \end{pmatrix} = \begin{pmatrix} \hat{Y}_{\text{ht}}(z) \\ \hat{S}_{xY(z)} \end{pmatrix}.$$

Compare the first row to see $\tilde{\beta}_{\text{ag,L}}(z) + \hat{x}^{\text{T}}_{\text{ht}}(z)\tilde{\gamma}_{\text{ag},z} = \hat{Y}_{\text{ht}}(z)$ and hence the numeric result. The probability limit then follows from $(\tilde{\beta}_{\text{ag,L}}(z), \tilde{\gamma}^{\text{T}}_{\text{ag},z})^{\text{T}} = G_1^{-1}G_2$ with $G_1 = \text{diag}\{1, T_{xx}(z)\} + o_{\mathbb{P}}(1)$ and $G_2 = (\bar{Y}(z), T^{\text{T}}_{xY(z)})^{\text{T}} + o_{\mathbb{P}}(1)$ by Lemmas S4 and S7. □

We next verify the asymptotic Normality of $\tilde{\beta}_{\dagger,\text{L}}$ and the asymptotic conservativeness of $\tilde{V}_{\dagger,\text{L}}$ for $\dagger = \text{wls}, \text{ag}$ in Theorem 6.2, respectively.

PROOF OF THEOREM 6.2, PART I FOR $\tilde{\beta}_{\dagger,\text{L}}$. For $\dagger = \text{wls}, \text{ag}$, let $\tilde{\gamma}_{\dagger,\text{L}} = (\tilde{\gamma}_{\dagger,z})_{z\in\mathcal{T}}$ with $\tilde{\gamma}_{\dagger,z} = \gamma_{\dagger,z} + o_{\mathbb{P}}(1)$ by Lemma S11. Recall the definitions of $\hat{Y}_*(z;\gamma)$ and $\hat{Y}_*(\gamma)$ from (S25). Proposition 4 ensures

$$\text{(S34)} \qquad \tilde{\beta}_{\text{wls,L}}(z) = \hat{Y}_{\text{haj}}(z; \tilde{\gamma}_{\text{wls},z}), \qquad \tilde{\beta}_{\text{wls,L}} = \hat{Y}_{\text{haj}}(\tilde{\gamma}_{\text{wls,L}}), \qquad \tilde{\beta}_{\text{ag,L}} = \hat{Y}_{\text{ht}}(\tilde{\gamma}_{\text{ag,L}}),$$

respectively. The results on $\tilde\beta_{\dagger,\mathrm{L}}$ ($\dagger = \mathrm{wls}, \mathrm{ag}$) then follow from Lemma S8. □

PROOF OF THEOREM 6.2, PART II FOR $\tilde V_{\dagger,\mathrm{L}}$. We verify below the result for $\tilde V_{\mathrm{wls,L}}$ from the unit regression (10). The proof for $\tilde V_{\mathrm{ag,L}}$ from (11) is almost identical and thus omitted.

Let χ be the concatenation of $\chi_{ws} = d_{ws} \otimes x_{ws}$ over $ws \in \mathcal{S}$. The design matrix of (10) equals $C_{\mathrm{L}} = (D, \chi)$. Let $C_{\mathrm{L},w} = (D_w, \chi_w)$ be the sub-matrix of C_{L} corresponding to whole-plot w. Let $\epsilon_w = (\epsilon_{w1}, \dots, \epsilon_{wM_w})^{\mathrm{T}}$, where

$$\epsilon_{ws} = Y_{ws} - \tilde\beta_{\mathrm{wls,L}}(Z_{ws}) - x_{ws}^{\mathrm{T}} \tilde\gamma_{\mathrm{wls},Z_{ws}} \tag{S35}$$

is the residual from the WLS fit of (10). Then $W\tilde V_{\mathrm{wls,L}}$ equals the upper-left 4×4 matrix of

$$(C_{\mathrm{L}}^{\mathrm{T}}\Pi C_{\mathrm{L}})^{-1} \left(W \sum_{w=1}^{W} C_{\mathrm{L},w}^{\mathrm{T}} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w C_{\mathrm{L},w} \right) (C_{\mathrm{L}}^{\mathrm{T}}\Pi C_{\mathrm{L}})^{-1}. \tag{S36}$$

Let

$$\tilde\Omega_{\mathrm{wls,L}} = (D^{\mathrm{T}}\Pi D)^{-1} \left(\sum_{w=1}^{W} D_w^{\mathrm{T}} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w D_w \right) (D^{\mathrm{T}}\Pi D)^{-1}$$

be an intermediate quantity. The result on $\tilde V_{\mathrm{wls,L}}$ holds as long as

$$\text{(i)}\ W\tilde\Omega_{\mathrm{wls,L}} - \Sigma_{\mathrm{wls,L}} = S_{\mathrm{wls,L}} + o_{\mathbb{P}}(1) \quad \text{and} \quad \text{(ii)}\ W(\tilde V_{\mathrm{wls,L}} - \tilde\Omega_{\mathrm{wls,L}}) = o_{\mathbb{P}}(1). \tag{S37}$$

We verify below these two conditions one by one.

Condition (S37)(i). Direct comparison shows that $\tilde\Omega_{\mathrm{wls,L}}$ is an analog of $\tilde V_{\mathrm{wls}}$ from (S20), with the unadjusted $e_{\mathrm{wls},w} = (e_{\mathrm{wls},ws})_{s=1}^{M_w}$ replaced by $\epsilon_w = (\epsilon_{ws})_{s=1}^{M_w}$. By (S34) and (S35),

$$\epsilon_{ws} = (Y_{ws} - x_{ws}^{\mathrm{T}} \tilde\gamma_{\mathrm{wls},Z_{ws}}) - \hat Y_{\mathrm{haj}}(Z_{ws}; \tilde\gamma_{\mathrm{wls},Z_{ws}})$$

is essentially the analog of $e_{\mathrm{wls},ws} = Y_{ws} - \hat Y_{\mathrm{haj}}(Z_{ws})$ defined on the adjusted potential outcomes $Y_{ws}(z; \tilde\gamma_{\mathrm{wls},z}) = Y_{ws}(z) - x_{ws}^{\mathrm{T}} \tilde\gamma_{\mathrm{wls},z}$. This, together with Theorem 5.1, ensures

$$\tilde\Omega_{\mathrm{wls,L}} = \hat 1_{\mathrm{ht}}^{-1} \operatorname{diag}\left(\frac{W_0 - 1}{W_0} I_2, \frac{W_1 - 1}{W_1} I_2 \right) \hat V_{\mathrm{haj}}(\tilde\gamma_{\mathrm{wls,L}})\, \hat 1_{\mathrm{ht}}^{-1}, \tag{S38}$$

where $\hat V_{\mathrm{haj}}(\tilde\gamma_{\mathrm{wls,L}})$ denotes the value of $\hat V_{\mathrm{haj},\gamma}$ at $\gamma = \tilde\gamma_{\mathrm{wls,L}}$. With $\tilde\gamma_{\mathrm{wls},z} - \gamma_{\mathrm{wls},z} = o_{\mathbb{P}}(1)$ by Lemma S11, Lemma S8 ensures $W\hat V_{\mathrm{haj}}(\tilde\gamma_{\mathrm{wls,L}}) - \Sigma_{\mathrm{wls,L}} = S_{\mathrm{wls,L}} + o_{\mathbb{P}}(1)$ by the definitions of $S_{\mathrm{wls,L}}$ and $\Sigma_{\mathrm{wls,L}}$. This, together with (S38), ensures condition (S37)(i).

Condition (S37)(ii). Let $G_1 = W\sum_{w=1}^W D_w^{\mathrm{T}} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w \chi_w$ and $G_2 = W\sum_{w=1}^W \chi_w^{\mathrm{T}} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w \chi_w$. The "meat" and "bread" parts of the sandwich covariance (S36) satisfy

$$\begin{aligned} W \sum_{w=1}^{W} C_{\mathrm{L},w}^{\mathrm{T}} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w C_{\mathrm{L},w} &= W \sum_{w=1}^{W} \begin{pmatrix} D_w^{\mathrm{T}} \\ \chi_w^{\mathrm{T}} \end{pmatrix} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w (D_w, \chi_w) \\ &= W \begin{pmatrix} \sum_{w=1}^W D_w^{\mathrm{T}} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w D_w & \sum_{w=1}^W D_w^{\mathrm{T}} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w \chi_w \\ \sum_{w=1}^W \chi_w^{\mathrm{T}} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w D_w & \sum_{w=1}^W \chi_w^{\mathrm{T}} \Pi_w \epsilon_w \epsilon_w^{\mathrm{T}} \Pi_w \chi_w \end{pmatrix} \\ &= \begin{pmatrix} (D^{\mathrm{T}}\Pi D)(W\tilde\Omega_{\mathrm{wls,L}})(D^{\mathrm{T}}\Pi D) & G_1 \\ G_1^{\mathrm{T}} & G_2 \end{pmatrix}, \end{aligned} \tag{S39}$$

$$C_{\mathrm{L}}^{\mathrm{T}}\Pi C_{\mathrm{L}} = \begin{pmatrix} D^{\mathrm{T}} \\ \chi^{\mathrm{T}} \end{pmatrix} \Pi (D, \chi) = \begin{pmatrix} D^{\mathrm{T}}\Pi D & D^{\mathrm{T}}\Pi\chi \\ \chi^{\mathrm{T}}\Pi D & \chi^{\mathrm{T}}\Pi\chi \end{pmatrix} = \operatorname{diag}(I_{|\mathcal{T}|}, I_{|\mathcal{T}|} \otimes Q_{xx}) + o_{\mathbb{P}}(1),$$

respectively. The last equality follows from $D^{\mathrm{T}}\Pi D = I_{|\mathcal{T}|} + o_{\mathbb{P}}(1)$ by (S22) and Lemma S4, and

$$D^{\mathrm{T}}\Pi\chi = \sum_{ws\in\mathcal{S}} \pi_{ws} d_{ws} \chi_{ws}^{\mathrm{T}} = \mathrm{diag}\left(\sum_{ws\in\mathcal{S}(z)} \pi_{ws} x_{ws}^{\mathrm{T}}\right)_{z\in\mathcal{T}} = \mathrm{diag}\{\hat{x}_{\mathrm{ht}}^{\mathrm{T}}(z)\}_{z\in\mathcal{T}} = o_{\mathbb{P}}(1),$$

$$\chi^{\mathrm{T}}\Pi\chi = \sum_{ws\in\mathcal{S}} \pi_{ws} \chi_{ws} \chi_{ws}^{\mathrm{T}} = \mathrm{diag}\left(\sum_{ws\in\mathcal{S}(z)} \pi_{ws} x_{ws} x_{ws}^{\mathrm{T}}\right)_{z\in\mathcal{T}} = I_{|\mathcal{T}|} \otimes Q_{xx} + o_{\mathbb{P}}(1)$$

by Lemma S7. This, together with (S39), ensures that (S37)(ii) holds as long as

$$G_k = \big(G_k(z,z')\big)_{z,z'\in\mathcal{T}} = O_{\mathbb{P}}(1) \qquad \text{for } k = 1, 2.$$

We verify below $G_2 = O_{\mathbb{P}}(1)$ for scalar covariate $x_{ws} \in \mathbb{R}$ for notational simplicity. The proof for $G_1 = O_{\mathbb{P}}(1)$ is almost identical and thus omitted.

First, recall the expression of π_{ws} from (S19). Direct algebra shows that

$$\begin{aligned}\chi_w^{\mathrm{T}}\Pi_w\epsilon_w &= \sum_{s=1}^{M_w} \pi_{ws}\chi_{ws}\epsilon_{ws} = \sum_{s=1}^{M_w} \pi_{ws}(d_{ws}\otimes x_{ws})(1\otimes\epsilon_{ws}) = \sum_{s=1}^{M_w} \pi_{ws} d_{ws}\otimes(x_{ws}\epsilon_{ws}) \\ &= \alpha_w\left(W_0^{-1}\kappa_w(00),\, W_0^{-1}\kappa_w(01),\, W_1^{-1}\kappa_w(10),\, W_1^{-1}\kappa_w(11)\right)^{\mathrm{T}},\end{aligned}$$

where $\kappa_w(z) = M_{wb}^{-1}\sum_{s:Z_{ws}=z} x_{ws}\epsilon_{ws}$ with $\kappa_w(z) = 0$ if $z \notin \{(A_w0), (A_w1)\}$. This ensures

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

with

$$p_a^2|G_2(z,z')| \le \frac{1}{W}\sum_{w=1}^{W}\alpha_w^2|\kappa_w(z)\kappa_w(z')| \le \frac{1}{2W}\sum_{w=1}^{W}\alpha_w^2\{\kappa_w(z)\}^2 + \frac{1}{2W}\sum_{w=1}^{W}\alpha_w^2\{\kappa_w(z')\}^2$$

for $z=(ab)$ and $z'=(ab')$. It thus suffices to verify $W^{-1}\sum_{w=1}^{W}\alpha_w^2\{\kappa_w(z)\}^2 = O_{\mathbb{P}}(1)$.

To this end, let $\overline{x_{w\cdot}^k} = M_w^{-1}\sum_{s=1}^{M_w} x_{ws}^k$ and $\overline{\epsilon_{w\cdot}^k} = M_w^{-1}\sum_{s=1}^{M_w}\epsilon_{ws}^k$ for $k=2,4$. Then

$$|\kappa_w(z)| \le M_{wb}^{-1}\sum_{s=1}^{M_w}|x_{ws}\epsilon_{ws}| \le 2^{-1}M_{wb}^{-1}\sum_{s=1}^{M_w}(x_{ws}^2+\epsilon_{ws}^2) = 2^{-1}q_{wb}^{-1}\left(\overline{x_{w\cdot}^2}+\overline{\epsilon_{w\cdot}^2}\right)$$

such that, with $q_{wb}\ge\epsilon$ by Condition 2,

$$\{\kappa_w(z)\}^2 \le 4^{-1}\epsilon^{-2}\left(\overline{x_{w\cdot}^2}+\overline{\epsilon_{w\cdot}^2}\right)^2 \le 2^{-1}\epsilon^{-2}\left\{\left(\overline{x_{w\cdot}^2}\right)^2+\left(\overline{\epsilon_{w\cdot}^2}\right)^2\right\} \le 2^{-1}\epsilon^{-2}\left(\overline{x_{w\cdot}^4}+\overline{\epsilon_{w\cdot}^4}\right),$$

where the last inequality follows from $\left(\overline{x_{w\cdot}^2}\right)^2 \le \overline{x_{w\cdot}^4}$ and $\left(\overline{\epsilon_{w\cdot}^2}\right)^2 \le \overline{\epsilon_{w\cdot}^4}$ by (S5). This ensures

$$W^{-1}\sum_{w=1}^{W}\alpha_w^2\{\kappa_w(z)\}^2 \le 2^{-1}\epsilon^{-2}W^{-1}\sum_{w=1}^{W}\alpha_w^2\left(\overline{x_{w\cdot}^4}+\overline{\epsilon_{w\cdot}^4}\right) = O_{\mathbb{P}}(1);$$

the last equality follows from $W^{-1}\sum_{w=1}^{W}\alpha_w^2\overline{x_{w\cdot}^4} = O_{\mathbb{P}}(1)$ by Condition 3(iii) and $W^{-1}\sum_{w=1}^{W}\alpha_w^2\overline{\epsilon_{w\cdot}^4} = O_{\mathbb{P}}(1)$ by $\epsilon_{ws}^4 \le 27Y_{ws}^4 + 27\{\tilde{\beta}_{\mathrm{wls,L}}(Z_{ws})\}^4 + 27(x_{ws}^{\mathrm{T}}\tilde{\gamma}_{\mathrm{wls},Z_{ws}})^4$ from (S35) and (S5).

□

S3.3. *Results under the additive regressions.*

LEMMA S12. *Under the 2^2 split-plot randomization and Conditions 2–3, we have*

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

Further assume Condition 1. Then $\tilde\gamma_{\text{ols}} = \gamma_{\text{ols}} + o_{\mathbb{P}}(1)$ with $\gamma_{\text{ols}} = Q_{xx}^{-1} \sum_{z\in\mathcal{T}} r_z Q_{xY(z)}$, recalling $r_z = N_z/N$.

PROOF OF PROPOSITION 3 AND LEMMA S12. We verify below the numeric results in Proposition 3 and the asymptotic results in Lemma S12 together. Let X and Λ be the concatenations of $\{x_{ws} : ws \in \mathcal{S}\}$ and $\{\hat v_w(A_w b) : w = 1, \dots, W;\ b = 0, 1\}$, respectively. The design matrices of (8) and (9) equal $C_{\text{F}} = (D, X)$ and $C_{\text{ag,F}} = (D_{\text{ag}}, \Lambda)$, respectively. Recall that $R = \text{diag}(r_z)_{z\in\mathcal{T}}$ with $r_z = N_z/N$ and $\hat x_* = (\hat x_*(00), \hat x_*(01), \hat x_*(10), \hat x_*(11))^{\text{T}}$ for $* = \text{sm}, \text{ht}, \text{haj}$. Direct algebra shows that

$$N^{-1} D^{\text{T}} X = R \hat x_{\text{sm}}, \qquad D^{\text{T}} \Pi X = \hat x_{\text{ht}}, \qquad W^{-1} D_{\text{ag}}^{\text{T}} U = P \hat x_{\text{ht}} \tag{S40}$$

analogous to (S22). Lemma S7 further ensures

$$X^{\text{T}} \Pi X = |\mathcal{T}|\, Q_{xx} + o_{\mathbb{P}}(1), \qquad X^{\text{T}} \Pi Y = \sum_{z\in\mathcal{T}} Q_{xY(z)} + o_{\mathbb{P}}(1), \tag{S41}$$

$$W^{-1} \Lambda^{\text{T}} \Lambda = \sum_{z=(ab)\in\mathcal{T}} p_a T_{xx}(z) + o_{\mathbb{P}}(1), \qquad W^{-1} \Lambda^{\text{T}} U = \sum_{z=(ab)\in\mathcal{T}} p_a T_{xY}(z) + o_{\mathbb{P}}(1).$$

Results on $(\tilde\beta_{\text{ols,F}}, \tilde\gamma_{\text{ols}})$. The first-order condition of OLS ensures $G_1(\tilde\beta_{\text{ols,F}}^{\text{T}}, \tilde\gamma_{\text{ols}}^{\text{T}})^{\text{T}} = G_2$, where

$$G_1 = C_{\text{F}}^{\text{T}} C_{\text{F}} = \begin{pmatrix} D^{\text{T}} D & D^{\text{T}} X \\ X^{\text{T}} D & X^{\text{T}} X \end{pmatrix} = \begin{pmatrix} NR & NR\hat x_{\text{sm}} \\ N\hat x_{\text{sm}}^{\text{T}} R & X^{\text{T}} X \end{pmatrix}, \qquad G_2 = C_{\text{F}}^{\text{T}} Y = \begin{pmatrix} D^{\text{T}} Y \\ X^{\text{T}} Y \end{pmatrix} = \begin{pmatrix} NR\hat Y_{\text{sm}} \\ X^{\text{T}} Y \end{pmatrix}.$$

by (S22) and (S40). The numeric result follows by comparing the first row. The probability limit follows from $(\tilde\beta_{\text{ols,F}}^{\text{T}}, \tilde\gamma_{\text{ols}}^{\text{T}})^{\text{T}} = (N^{-1}G_1)^{-1}(N^{-1}G_2)$, where $N^{-1}G_1 = \text{diag}(R, Q_{xx}) + o_{\mathbb{P}}(1)$ and $N^{-1}G_2 = ((R\bar Y)^{\text{T}}, \sum_{z\in\mathcal{T}} r_z Q_{xY(z)}^{\text{T}})^{\text{T}} + o_{\mathbb{P}}(1)$ under Condition 1 by $\hat x_{\text{sm}} = \hat x_{\text{ht}} = o_{\mathbb{P}}(1)$ and

$$N^{-1} X^{\text{T}} X = N^{-1} \sum_{ws\in\mathcal{S}} x_{ws} x_{ws}^{\text{T}} = \sum_{z\in\mathcal{T}} r_z \left(N_z^{-1} \sum_{ws\in\mathcal{S}(z)} x_{ws} x_{ws}^{\text{T}} \right) = Q_{xx} + o_{\mathbb{P}}(1),$$

$$N^{-1} X^{\text{T}} Y = N^{-1} \sum_{ws\in\mathcal{S}} x_{ws} Y_{ws} = \sum_{z\in\mathcal{T}} r_z \left\{ N_z^{-1} \sum_{ws\in\mathcal{S}(z)} x_{ws} Y_{ws}(z) \right\} = \sum_{z\in\mathcal{T}} r_z Q_{xY(z)} + o_{\mathbb{P}}(1)$$

from Lemma S7.

Results on $(\tilde\beta_{\text{wls,F}}, \tilde\gamma_{\text{wls}})$. The first-order condition of WLS ensures $G_1(\tilde\beta_{\text{wls,F}}^{\text{T}}, \tilde\gamma_{\text{wls}}^{\text{T}})^{\text{T}} = G_2$, where

$$G_1 = C_{\text{F}}^{\text{T}} \Pi C_{\text{F}} = \begin{pmatrix} D^{\text{T}} \Pi D & D^{\text{T}} \Pi X \\ X^{\text{T}} \Pi D & X^{\text{T}} \Pi X \end{pmatrix} = \begin{pmatrix} \hat 1_{\text{ht}} & \hat x_{\text{ht}} \\ \hat x_{\text{ht}}^{\text{T}} & X^{\text{T}} \Pi X \end{pmatrix}, \qquad G_2 = C_{\text{F}}^{\text{T}} \Pi Y = \begin{pmatrix} D^{\text{T}} \Pi Y \\ X^{\text{T}} \Pi Y \end{pmatrix} = \begin{pmatrix} \hat Y_{\text{ht}} \\ X^{\text{T}} \Pi Y \end{pmatrix}$$

by (S22) and (S40). The numeric result follows by comparing the first row. The probability limit follows from $(\tilde\beta_{\text{wls,F}}^{\top}, \tilde\gamma_{\text{wls}}^{\top})^{\top} = G_1^{-1}G_2$, where $G_1 = \text{diag}(I_{|\mathcal{T}|}, |\mathcal{T}|\, Q_{xx}) + o_{\mathbb{P}}(1)$ and $G_2 = (\bar Y^{\top}, \sum_{z\in\mathcal{T}} Q_{xY(z)}^{\top})^{\top} + o_{\mathbb{P}}(1)$ by $\hat x_{\text{ht}} = o_{\mathbb{P}}(1)$, $\hat Y_{\text{ht}} = \bar Y + o_{\mathbb{P}}(1)$, and (S41).

Results on $(\tilde\beta_{\text{ag,F}}, \tilde\gamma_{\text{ag}})$. The first-order condition of OLS ensures $G_1(\tilde\beta_{\text{ag,F}}^{\top}, \tilde\gamma_{\text{ag}}^{\top})^{\top} = G_2$, where

$$G_1 = W^{-1}C_{\text{ag,F}}^{\top}C_{\text{ag,F}} = \begin{pmatrix} P & P\hat x_{\text{ht}} \\ (P\hat x_{\text{ht}})^{\top} & \Lambda^{\top}\Lambda \end{pmatrix}, \qquad G_2 = W^{-1}C_{\text{ag,F}}^{\top}U = \begin{pmatrix} P\hat Y_{\text{ht}} \\ \Lambda^{\top}U \end{pmatrix}$$

by (S22) and (S40). The numeric result follows by comparing the first row. The probability limit follows from $(\tilde\beta_{\text{ag,F}}^{\top}, \tilde\gamma_{\text{ag}}^{\top})^{\top} = G_1^{-1}G_2$, where

$$G_1 = \begin{pmatrix} P & \\ & \sum_{z=(ab)\in\mathcal{T}} p_a T_{xx}(z) \end{pmatrix} + o_{\mathbb{P}}(1), \qquad G_2 = \begin{pmatrix} P\bar Y \\ \sum_{z=(ab)\in\mathcal{T}} p_a T_{xY}(z)\} \end{pmatrix} + o_{\mathbb{P}}(1)$$

by $\hat x_{\text{ht}} = o_{\mathbb{P}}(1)$, $\hat Y_{\text{ht}} = \bar Y + o_{\mathbb{P}}(1)$, and (S41). □

PROOF OF THEOREM 6.1. The proof is similar to that of Theorem 6.2 and omitted. □

S3.4. *Guaranteed gains in asymptotic efficiency.* Let $\hat c_w(A_w b) = (\alpha_w - 1, \hat v_w^{\top}(A_w b))^{\top}$ be the augmented whole-plot level covariate vector corresponding to $\tilde\beta_{\text{ag},\diamond}(\alpha, v)$ ($\diamond = \text{F}, \text{L}$). Let $\tilde\gamma_{\text{ag}}(\alpha, v)$ and $\tilde\gamma_{\text{ag},z}(\alpha, v)$ be the analogs of $\tilde\gamma_{\text{ag}}$ and $\tilde\gamma_{\text{ag},z}$, respectively, based on $\hat c_w(A_w b)$. The underlying augmented unit-level covariate vector equals $(1 - \alpha_w^{-1}, x_{ws}^{\top})^{\top}$ for unit ws, and satisfies Condition 3 and (12) as long as $(x_{ws})_{ws\in\mathcal{S}}$ satisfies Condition 3 and (12). All results on $\{\tilde\beta_{\text{ag},\diamond}, \tilde\gamma_{\text{ag}}, \tilde\gamma_{\text{ag},z}\}$ so far thus extend to $\{\tilde\beta_{\text{ag},\diamond}(\alpha, v), \tilde\gamma_{\text{ag}}(\alpha, v), \tilde\gamma_{\text{ag},z}(\alpha, v)\}$ as well.

Recall $\gamma_\dagger$ and $\gamma_{\dagger,z}$ as the probability limits of $\tilde\gamma_\dagger$ and $\tilde\gamma_{\dagger,z}$, respectively, for $\dagger = \text{wls}, \text{ag}$. Let $\Psi_{\dagger,\text{F}}$ and $\Psi_{\dagger,\text{L}}$, be the value of Ψ_γ when $\gamma_z = \gamma_\dagger, \gamma_{\dagger,z}$, respectively, Let $\gamma_{\text{ag}}(\alpha, v)$ and $\gamma_{\text{ag},z}(\alpha, v)$ be the probability limits of $\tilde\gamma_{\text{ag}}(\alpha, v)$ and $\tilde\gamma_{\text{ag},z}(\alpha, v)$, respectively, with $\{S_{\text{ag},\diamond}(\alpha, v), \Psi_{\text{ag},\diamond}(\alpha, v), \Sigma_{\text{ag},\diamond}(\alpha, v)\}$ as the corresponding analogs of $\{S_{\text{ag},\diamond}, \Psi_{\text{ag},\diamond}, \Sigma_{\text{ag},\diamond}\}$ for $\diamond = \text{F}, \text{L}$. They are essentially the analogs of $\{S_{\text{ht}}, \Psi, \Sigma\}$ defined on the adjusted potential outcomes $Y_{ws}(z) - (1 - \alpha_w^{-1}, x_{ws}^{\top})\gamma_{\text{ag}}(\alpha, v)$ and $Y_{ws}(z) - (1 - \alpha_w^{-1}, x_{ws}^{\top})\gamma_{\text{ag},z}(\alpha, v)$, respectively.

PROOF OF PROPOSITION 5. For two sequences of symmetric matrices $(A_N)_{N=1}^{\infty}$ and $(B_N)_{N=1}^{\infty}$, write $A_N \le_\infty B_N$ if the limiting value of $(B_N - A_N)$ is positive semi-definite. The result is equivalent to

$$\Sigma_{\text{ag,L}}(\alpha, v) \le_\infty \Sigma_{\text{ht}}, \Sigma_{\text{haj}}, \Sigma_{\text{wls},\diamond}, \Sigma_{\text{ag},\diamond}, \Sigma_{\text{ag,F}}(\alpha, v) \qquad \text{for } \diamond = \text{F}, \text{L}$$

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

$$\Sigma_{\text{ag},\diamond} = H \circ S_{\text{ag},\diamond} + \Psi_{\text{ag},\diamond}, \qquad \Sigma_{\text{ag},\diamond}(\alpha, v) = H \circ S_{\text{ag},\diamond}(\alpha, v) + \Psi_{\text{ag},\diamond}(\alpha, v).$$

Because $H \ge 0$, Lemma S9 ensures that it suffices to verify

(i) $\Psi_{\text{wls},\diamond} - \Psi = o(1)$, $\Psi_{\text{ag},\diamond} - \Psi = o(1)$, $\Psi_{\text{ag},\diamond}(\alpha, v) - \Psi = o(1)$;
(ii) $S_{\text{ag,L}}(\alpha, v) \le_\infty S_{\text{ht}}, S_{\text{haj}}, S_{\text{wls},\diamond}, S_{\text{ag},\diamond}, S_{\text{ag,F}}(\alpha, v)$

for $\diamond = \text{F}, \text{L}$ under Conditions 2–3 and (12). We verify below (i) and (ii) respectively.

Sufficient condition (i). Let $\Psi(z, z'; \gamma)$ be the (z, z')th element of Ψ_γ. For arbitrary fixed $\gamma = (\gamma_z)_{z\in\mathcal{T}}$, it follows from (S26) and Lemma S10 that

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

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

$$= \Psi(z,z') - \gamma_z^{\mathrm{T}} \Psi_{xY}(z,z') - \gamma_{z'}^{\mathrm{T}} \Psi_{xY}(z',z) + \gamma_z^{\mathrm{T}} \Psi_{xx}(z,z') \gamma_{z'}$$

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

such that $\Psi_\gamma = \Psi + o(1)$. This verifies the result for $\Psi_{\dagger,\diamond}$ ($\dagger = \text{wls}, \text{ag}$; $\diamond = \mathrm{F}, \mathrm{L}$).

The result for $\Psi_{\text{ag},\diamond}(\alpha, v)$ then follows from the same line of reasoning as that for $\Psi_{\text{ag},\diamond}$ given Condition (12) also holds for the augmented unit-level covariates $(1 - \alpha_w^{-1}, x_{ws})$.

Sufficient condition (ii). Let

$$\begin{aligned} e_{1,w}(z) &= U_w(z) - \bar{Y}(z), & e_{2,w}(z) &= U_w(z) - \alpha_w \bar{Y}(z), \\ e_{3,w}(z) &= U_w(z) - \bar{Y}(z) - v_w^{\mathrm{T}} \theta_z, & e_{4,w}(z) &= U_w(z) - \alpha_w \bar{Y}(z) - v_w^{\mathrm{T}} \theta_z, \\ e_{5,w}(z) &= U_w(z) - \bar{Y}(z) - c_w^{\mathrm{T}} \theta_z', \end{aligned}$$

where $v_w = \alpha_w \bar{x}_w$ and $c_w = (\alpha_w - 1, v_w^{\mathrm{T}})^{\mathrm{T}}$ are the population analogs of $\hat{v}_w(A_w b) = \alpha_w \hat{x}_w(A_w b)$ and $\hat{c}_w(A_w b)$, respectively, and $\theta_z \in \mathbb{R}^J$ and $\theta_z' \in \mathbb{R}^{J+1}$ are arbitrary vectors. Let $\gamma_{c,z}$ be the coefficient of c_w from the OLS fit of $U_w(z)$ on $(1, c_w)$ over $w = 1, \ldots, W$ with

$$e_{6,w}(z) = U_w(z) - \bar{Y}(z) - c_w^{\mathrm{T}} \gamma_{c,z}$$

as the corresponding residual. Let $S_k = (S_k(z,z'))_{z,z' \in \mathcal{T}}$ be the finite-population covariance matrix of $\{e_{k,w}(z) : z \in \mathcal{T}\}_{w=1}^{W}$ for $k = 1, \ldots, 6$ with $S_k(z,z') = (W-1)^{-1} \sum_{w=1}^{W} e_{k,w}(z) e_{k,w}(z') = (W-1)^{-1} \{e_k(z)\}^{\mathrm{T}} e_k(z)$, where $e_k(z) = (e_{k,1}(z), \ldots, e_{k,W}(z))^{\mathrm{T}}$.

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 $\{e_k(z) - e_6(z)\}^{\mathrm{T}} e_6(z') = 0$ for all $z, z' \in \mathcal{T}$ and $k = 1, \ldots, 5$. Let $e_{k-6}(z) = e_k(z) - e_6(z)$ be a shorthand for $k = 1, \ldots, 5$. Then $e_k(z) = e_6(z) + e_{k-6}(z)$ with

$$\begin{aligned} S_k(z,z') &= (W-1)^{-1} \{e_6(z) + e_{k-6}(z)\}^{\mathrm{T}} \{e_6(z') + e_{k-6}(z')\} \\ &= S_6(z,z') + (W-1)^{-1} \{e_{k-6}(z)\}^{\mathrm{T}} \{e_{k-6}(z')\}, \end{aligned}$$

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

This, together with

- $S_1 = S_{\text{ht}}$; $S_2 = S_{\text{haj}}$;
- $S_3 = S_{\text{ag},\mathrm{F}}$ and $S_{\text{ag},\mathrm{L}}$ for $\theta_z = \gamma_{\text{ag}}$ and $\gamma_{\text{ag},z}$, respectively;
- $S_4 = S_{\text{wls},\mathrm{F}}$ and $S_{\text{wls},\mathrm{L}}$ for $\theta_z = \gamma_{\text{wls}}$ and $\gamma_{\text{wls},z}$, respectively;
- $S_5 = S_{\text{ag},\mathrm{F}}(\alpha, v)$ and $S_{\text{ag},\mathrm{L}}(\alpha, v)$ for $\theta_z' = \gamma_{\text{ag}}(\alpha, v)$ and $\gamma_{\text{ag},z}(\alpha, v)$, respectively,

ensures

(S42) $$S_6 \leq S_{\text{ht}},\ S_{\text{haj}},\ S_{\text{wls},\diamond},\ S_{\text{ag},\diamond},\ S_{\text{ag},\diamond}(\alpha, v) \qquad \text{for } \diamond = \mathrm{F}, \mathrm{L}.$$

In addition, let $S_{xx}(\alpha, v)$, $S_{xY(z)}(\alpha, v)$, $\Psi_{xx}(z, z; \alpha, v)$, and $\Psi_{xY}(z, z; \alpha, v)$ be the analogs of S_{xx}, $S_{xY(z)}$, $\Psi_{xx}(z,z)$, and $\Psi_{xY}(z,z)$ under the augmented covariates $(1 - \alpha_w^{-1}, x_{ws}^{\mathrm{T}})$, respectively. Then

$$S_{xx}(\alpha, v) = (W-1)^{-1} \sum_{w=1}^{W} c_w c_w^{\mathrm{T}}, \qquad S_{xY(z)}(\alpha, v) = (W-1)^{-1} \sum_{w=1}^{W} c_w U_w(z)$$

with $\Psi_{xx}(z,z;\alpha,v) = o(1)$ and $\Psi_{xY}(z,z;\alpha,v) = o(1)$ by applying Lemma S10 to the augmented covariates. Lemma S11 ensures $\tilde{\gamma}_{\text{ag},z}(\alpha,v) = \{S_{xx}(\alpha,v)\}^{-1}S_{xY(z)}(\alpha,v) + o_{\mathbb{P}}(1)$ and hence

$$\gamma_{\text{ag},z}(\alpha,v) = \{S_{xx}(\alpha,v)\}^{-1}S_{xY(z)}(\alpha,v) + o(1).$$

This, together with $\gamma_{c,z} = (\sum_{w=1}^{W} c_w c_w^{\mathrm{T}})^{-1}\{\sum_{w=1}^{W} c_w U_w(z)\} = \{S_{xx}(\alpha,v)\}^{-1}S_{xY(z)}(\alpha,v)$ by standard OLS results, ensures $\gamma_{\text{ag},z}(\alpha,v) - \gamma_{c,z} = o(1)$ and hence $S_{\text{ag,L}}(\alpha,v) - S_6 = o(1)$. Condition (ii) then follows from (S42).

□

S4. Special case and extensions.

S4.1. *Uniform designs.* We outline in this subsection the unification of the three fitting schemes under Condition 1. The results clarify the theoretical guarantees by the sample-mean estimator and the corresponding "ols" fitting scheme under uniform designs.

Let S, Σ, and $\hat{V}$ be the common values of S_*, Σ_*, and $\hat{V}_*$ for $* = \text{ht}, \text{haj}$ under Condition 1, respectively. Let $\hat{Y}$ be the common value of $\hat{Y}_*$ ($* = \text{sm}, \text{ht}, \text{haj}$) under Condition 1 by Proposition 1. Corollary S1 below justifies the Wald-type inference of τ based on $(\hat{Y}, \hat{V})$.

COROLLARY S1. Assume the 2^2 split-plot randomization and Conditions 1–2. Then $\sqrt{W}(\hat{Y} - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma)$ and $W\hat{V} - \Sigma = S + o_{\mathbb{P}}(1)$ with $S \geq 0$.

Let $\tilde{V}_{\text{ols}}$, $\tilde{V}_{\text{ols,F}}$, and $\tilde{V}_{\text{ols,L}}$ be the cluster-robust covariances of $\tilde{\beta}_{\text{ols}}$, $\tilde{\beta}_{\text{ols,F}}$, and $\tilde{\beta}_{\text{ols,L}}$ from the OLS fits of the unit regressions (6), (8), and (10), respectively. Recall $\tilde{\gamma}_{\text{ols}}$ as the coefficient vector of x_{ws} from the OLS fit of (8). Recall γ_{ols} as the probability limit of $\tilde{\gamma}_{\text{ols}}$ under Conditions 1–3 by Lemma S12. Let $S_{\text{ols,F}}$ and $\Sigma_{\text{ols,F}}$ be the analogs of S and Σ defined on the adjusted potential outcomes $Y_{ws}(z;\gamma_{\text{ols}}) = Y_{ws}(z) - x_{ws}^{\mathrm{T}}\gamma_{\text{ols}}$, respectively.

PROPOSITION S1. Under the 2^2 split-plot experiment and Condition 1, we have

$$\tilde{\beta}_\dagger = \hat{Y}, \qquad \tilde{V}_\dagger = \text{diag}\left(\frac{W_0-1}{W_0}I_2, \frac{W_1-1}{W_1}I_2\right)\hat{V} \qquad (\dagger = \text{ols}, \text{wls}, \text{ag}),$$

and $\{\tilde{\beta}_{\text{ols,L}}, (\tilde{\gamma}_{\text{ols},z})_{z\in\mathcal{T}}\} = \{\tilde{\beta}_{\text{wls,L}}, (\tilde{\gamma}_{\text{wls},z})_{z\in\mathcal{T}}\}$. Further assume Conditions 2–3. Then

$$\sqrt{W}(\tilde{\beta}_{\text{ols},\diamond} - \bar{Y}) \rightsquigarrow \mathcal{N}(0, \Sigma_{\text{ols},\diamond}), \qquad W\tilde{V}_{\text{ols},\diamond} - \Sigma_{\text{ols},\diamond} = S_{\text{ols},\diamond} + o_{\mathbb{P}}(1)$$

for $\diamond = \text{F}, \text{L}$ with $(S_{\text{ols,L}}, \Sigma_{\text{ols,L}}) = (S_{\text{wls,L}}, \Sigma_{\text{wls,L}})$.

PROOF OF PROPOSITION S1. We verify below the numeric and asymptotic results, respectively.

Numeric results. Condition 1 ensures $\alpha_w = 1$ and $q_{wb} = M_b/M = q_b$ for all w such that

$$\text{(S43)} \qquad N = WM, \qquad N_z = W_a M_b, \qquad p_{ws}(z) = p_a q_b = N_z/N, \qquad \pi_{ws} = N_{Z_{ws}}^{-1}$$

for all $ws \in \mathcal{S}$ and $z = (ab) \in \mathcal{T}$. The numeric equivalence between $\tilde{\beta}_\dagger$ follows from Propositions 1 and 2. That between $\{\tilde{\beta}_{\text{ols,L}}, (\tilde{\gamma}_{\text{ols},z})_{z\in\mathcal{T}}\} = \{\tilde{\beta}_{\text{wls,L}}, (\tilde{\gamma}_{\text{wls},z})_{z\in\mathcal{T}}\}$ follows from the equivalence between the "ols" and "wls" fitting schemes under the treatment-specific regression (S32) with $\pi_{ws} = N_{Z_{ws}}^{-1}$ being constant for all units under the same treatment. The results on $\tilde{V}_*$ ($* = \text{wls}, \text{ag}$) follow from Theorem 5.1. We verify below the result on $\tilde{V}_{\text{ols}}$.

First,

$$\tilde{V}_{\text{ols}} = (D^{\mathrm{T}}D)^{-1}\left(\sum_{w=1}^{W} D_w^{\mathrm{T}} e_{\text{ols},w} e_{\text{ols},w}^{\mathrm{T}} D_w\right)(D^{\mathrm{T}}D)^{-1}, \tag{S44}$$

where $e_{\text{ols},w} = (e_{\text{ols},w1}, \dots, e_{\text{ols},wM_w})^{\mathrm{T}}$ with $e_{\text{ols},ws} = Y_{ws} - \tilde{\beta}_{\text{ols}}(Z_{ws}) = Y_{ws} - \hat{Y}(Z_{ws})$. Let $\hat{e}_{\text{ols},w}(z) = M_b^{-1}\sum_{s:Z_{ws}=z} e_{\text{ols},ws}$, with $\hat{e}_{\text{ols},w}(z) = \hat{Y}_w(z) - \hat{Y}(z)$ for $z \in \{(A_w 0), (A_w 1)\}$ and $\hat{e}_{\text{ols},w}(z) = 0$ for $z \notin \{(A_w 0), (A_w 1)\}$. Set $Y_w = e_{\text{ols},w}$ in (S22) to see $D_w^{\mathrm{T}} e_{\text{ols},w} = MQ\hat{e}_{\text{ols},w}$, where $\hat{e}_{\text{ols},w} = (\hat{e}_{\text{ols},w}(00), \hat{e}_{\text{ols},w}(01), \hat{e}_{\text{ols},w}(10), \hat{e}_{\text{ols},w}(11))^{\mathrm{T}}$ and $Q = I_2 \otimes \text{diag}(q_0, q_1)$ is the common value of Q_w over all w under Condition 1. This ensures

$$\sum_{w=1}^{W} D_w^{\mathrm{T}} e_{\text{ols},w} e_{\text{ols},w}^{\mathrm{T}} D_w = M^2 Q\left(\sum_{w=1}^{W} \hat{e}_{\text{ols},w}\hat{e}_{\text{ols},w}^{\mathrm{T}}\right) Q = \big(\Omega_{\text{ols}}(z, z')\big)_{z,z' \in \mathcal{T}} \tag{S45}$$

with

$$\begin{aligned}\Omega_{\text{ols}}(z, z') &= M_b M_{b'} \sum_{w=1}^{W} \hat{e}_{\text{ols},w}(z)\hat{e}_{\text{ols},w}(z') \\ &= \begin{cases} M_b M_{b'}(W_a - 1)\hat{S}(z, z') & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a = a', \\ 0 & \text{for } z = (ab) \text{ and } z' = (a'b') \text{ with } a \neq a'. \end{cases}\end{aligned}$$

The numeric result on $\tilde{V}_{\text{ols}}$ then follows from (S44), (S45), and $D^{\mathrm{T}}D = \text{diag}(W_0M_0, W_0M_1, W_1M_0, W_1M_1)$ by (S22) and (S43).

Asymptotic results. Let $\hat{x}$ and $\hat{Y}_\gamma$ be the common values of $\hat{x}_*$ ($* = \text{sm}, \text{ht}, \text{haj}$) and $\hat{Y}_{*,\gamma}$ ($* = \text{ht}, \text{haj}$) under Condition 1, respectively. Proposition 3 ensures $\tilde{\beta}_{\text{ols,F}} = \hat{Y} - \hat{x}\tilde{\gamma}_{\text{ols}} = \hat{Y}_\gamma$ for $\gamma = (\gamma_z)_{z\in\mathcal{T}}$ where $\gamma_z = \tilde{\gamma}_{\text{ols}}$ for all $z \in \mathcal{T}$. The asymptotic Normality of $\tilde{\beta}_{\text{ols,F}}$ then follows from Lemma S8 with $\tilde{\gamma}_{\text{ols}} = \gamma_{\text{ols}} + o_{\mathbb{P}}(1)$ by Lemma S12. The asymptotic Normality of $\tilde{\beta}_{\text{ols,L}}$ follows from $\tilde{\beta}_{\text{ols,L}} = \tilde{\beta}_{\text{wls,L}}$ as we just showed. The asymptotic conservativeness of $\tilde{V}_{\text{ols},\diamond}$ ($\diamond = \text{F}, \text{L}$) follows from the same reasoning as that for $\tilde{V}_{\text{wls,L}}$ in the proof of Theorem 6.2. □

S4.2. *HC2 correction for the cluster-robust covariance estimators.* The classic cluster-robust covariances recover $\hat{V}_{\text{ht}}$ and $\hat{V}_{\text{haj}}$ only asymptotically by Theorem 5.1. We verify in this subsection the exact recovery of $\hat{V}_{\text{ht}}$ by the HC2 correction in finite samples (Cameron and Miller, 2015). Let

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

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

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

THEOREM S1. $\tilde{V}_{\text{ag2}} = \hat{V}_{\text{ht}}$. *Under Condition 1,* $\tilde{V}_{\text{ols2}} = \tilde{V}_{\text{ag2}} = \hat{V}$.

PROOF OF THEOREM S1. We verify below the results on $\tilde{V}_{\text{ag2}}$ and $\tilde{V}_{\text{ols2}}$, respectively.

Proof of $\tilde{V}_{\text{ag2}} = \hat{V}_{\text{ht}}$. It follows from $D_{\text{ag}}^{\text{T}} D_{\text{ag}} = \text{diag}(W_0, W_1) \otimes I_2$ by (S22) 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 $P_{\text{ag},w} = D_{\text{ag},w}(D_{\text{ag}}^{\text{T}} D_{\text{ag}})^{-1} D_{\text{ag},w}^{\text{T}} = W_{A_w}^{-1} I_2$. This ensures

$$\sum_{w=1}^{W} D_{\text{ag},w}^{\text{T}} (I - P_{\text{ag},w})^{-1/2} e_{\text{ag},w} e_{\text{ag},w}^{\text{T}} (I - P_{\text{ag},w})^{-1/2} D_{\text{ag},w}$$
$$= \sum_{w=1}^{W} (1 - W_{A_w}^{-1})^{-1} D_{\text{ag},w}^{\text{T}} e_{\text{ag},w} e_{\text{ag},w}^{\text{T}} D_{\text{ag},w} = \sum_{a=0,1} \frac{W_a}{W_a - 1} \left(\sum_{w:A_w=a} D_{\text{ag},w}^{\text{T}} e_{\text{ag},w} e_{\text{ag},w}^{\text{T}} D_{\text{ag},w} \right)$$

such that

$$\tilde{V}_{\text{ag2}} = \left\{ \text{diag}\left(\frac{W_0}{W_0 - 1}, \frac{W_1}{W_1 - 1} \right) \otimes I_2 \right\} \tilde{V}_{\text{ag}}$$

by the form of $\sum_{w:A_w=a} D_{\text{ag},w}^{\text{T}} e_{\text{ag},w} e_{\text{ag},w}^{\text{T}} D_{\text{ag},w}$ from (S23). The result then follows from Theorem 5.1.

Proof of $\tilde{V}_{\text{ols2}} = \hat{V}$ under Condition 1. Recall (M, M_0, M_1) as the common value of (M_w, M_{w0}, M_{w1}) for all w under Condition 1. Assume without loss of generality that the first M_0 sub-plots in whole-plot w receive level 0 of factor B. Then $D_w = (D_{w,00}, D_{w,01}, D_{w,10}, D_{w,11})$, where $D_{w,z} = (1(Z_{w1} = z), \ldots, 1(Z_{wM_w} = z))^{\text{T}}$ with

$$D_{w,00} = (1_{M_0}^{\text{T}}, 0_{M_1}^{\text{T}})^{\text{T}}, \quad D_{w,01} = (0_{M_0}^{\text{T}}, 1_{M_1}^{\text{T}})^{\text{T}}, \quad D_{w,10} = D_{w,11} = 0_M \quad \text{if } A_w = 0;$$
$$D_{w,00} = D_{w,01} = 0_M, \quad D_{w,10} = (1_{M_0}^{\text{T}}, 0_{M_1}^{\text{T}})^{\text{T}}, \quad D_{w,11} = (0_{M_0}^{\text{T}}, 1_{M_1}^{\text{T}})^{\text{T}} \quad \text{if } A_w = 1.$$

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

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

Note that the two non-zero columns of D_w, namely $(1_{M_0}^{\text{T}}, 0_{M_1}^{\text{T}})^{\text{T}}$ and $(0_{M_0}^{\text{T}}, 1_{M_1}^{\text{T}})^{\text{T}}$, 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 $(1 - W_{A_w}^{-1})^{-1/2}$, such that $(I - P_{\text{ols},w})^{-1/2} D_w = (1 - W_{A_w}^{-1})^{-1/2} D_w$ (Imai, Jiang and Malani, 2021, Section C.3). This ensures

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

and hence

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

by the form of $\sum_{w:A_w=a} D_w^{\text{T}} e_{\text{ols},w} e_{\text{ols},w}^{\text{T}} D_w$ from (S45). The result for $\tilde{V}_{\text{ols2}}$ then follows from Proposition S1.

□

S4.3. *Covariate adjustment via factor-based regressions.* We give in this subsection the details on covariate adjustment under factor-based regressions.

Let $f_{ws} = (A_w - 1/2, B_{ws} - 1/2, (A_w - 1/2)(B_{ws} - 1/2))^{\mathrm{T}}$ and $f_w(A_w b) = (A_w - 1/2, b - 1/2, (A_w - 1/2)(b - 1/2))^{\mathrm{T}}$ be the vectors of the non-intercept regressors of (14) and (15), respectively. The additive and fully-interacted extensions of (14) and (15) equal

$$Y_{ws} \sim 1 + f_{ws} + x_{ws}, \tag{S46}$$

$$\hat{U}_w(A_w b) \sim 1 + f_w(A_w b) + \hat{v}_w(A_w b), \tag{S47}$$

$$Y_{ws} \sim 1 + f_{ws} + x_{ws} + f_{ws} \otimes x_{ws}, \tag{S48}$$

$$\hat{U}_w(A_w b) \sim 1 + f_w(A_w b) + \hat{v}_w(A_w b) + f_w(A_w b) \otimes \hat{v}_w(A_w b), \tag{S49}$$

respectively. Let $\tilde{\tau}'_{\dagger,\mathrm{F}}$ and $\tilde{\tau}'_{\dagger,\mathrm{L}}$ be the coefficient vectors of f_{ws} and $f_w(A_w b)$ from (S46)–(S49) under fitting schemes $\dagger = \mathrm{ols}, \mathrm{wls}, \mathrm{ag}$, respectively, with cluster-robust covariances $\tilde{\Omega}'_{\dagger,\diamond}$ for $(\dagger, \diamond) \in \{\mathrm{ols}, \mathrm{wls}, \mathrm{ag}\} \times \{\mathrm{F}, \mathrm{L}\}$. Let $\tilde{\tau}'_{\mathrm{ag},\diamond}(\alpha, v)$ and $\tilde{\Omega}'_{\mathrm{ag},\diamond}(\alpha, v)$ be the variants of $\tilde{\tau}'_{\mathrm{ag},\diamond}$ and $\tilde{\Omega}'_{\mathrm{ag},\diamond}$ 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 below follows from the invariance of least squares to non-degenerate transformation of the regressors and, together with Theorems 6.1–6.2 and Proposition 5, ensures the optimality of $\tilde{\tau}'_{\mathrm{ag,L}}(\alpha, v)$ for estimating the standard factorial effects $(\tau_{\mathrm{A}}, \tau_{\mathrm{B}}, \tau_{\mathrm{AB}})^{\mathrm{T}} = G_0 \bar{Y}$ among $\{\tilde{\tau}'_{\mathrm{wls}}, \tilde{\tau}'_{\mathrm{wls},\diamond}; \tilde{\tau}'_{\mathrm{ag}}, \tilde{\tau}'_{\mathrm{ag},\diamond}, \tilde{\tau}'_{\mathrm{ag},\diamond}(\alpha, v) : \diamond = \mathrm{F}, \mathrm{L}\}$.

PROPOSITION S2. For $\dagger = \mathrm{ols}, \mathrm{wls}, \mathrm{ag}$ and $\diamond = \mathrm{F}, \mathrm{L}$, we have

$$\tilde{\tau}'_{\dagger,\diamond} = G_0 \tilde{\beta}_{\dagger,\diamond}, \qquad \tilde{\Omega}'_{\dagger,\diamond} = G_0 \tilde{V}_{\dagger,\diamond} G_0^{\mathrm{T}};$$
$$\tilde{\tau}'_{\mathrm{ag},\diamond}(\alpha, v) = G_0 \tilde{\beta}_{\mathrm{ag},\diamond}(\alpha, v), \qquad \tilde{\Omega}'_{\mathrm{ag},\diamond}(\alpha, v) = G_0 \tilde{V}_{\mathrm{ag},\diamond}(\alpha, v) G_0^{\mathrm{T}}.$$

Recall $\tilde{\tau}'_{\dagger}$ as the coefficient vectors of the non-intercept terms from the unadjusted factor-based regressions (14) and (15), respectively, for estimating $(\tau_{\mathrm{A}}, \tau_{\mathrm{B}}, \tau_{\mathrm{AB}})$. Let $(\tilde{\tau}'_{\dagger,\mathrm{B}}, \tilde{\tau}'_{\dagger,\mathrm{AB}})$ and $(\tilde{\tau}'_{\dagger,\mathrm{F,B}}, \tilde{\tau}'_{\dagger,\mathrm{F,AB}})$ be the elements of $\tilde{\tau}'_{\dagger}$ and $\tilde{\tau}'_{\dagger,\mathrm{F}}$ corresponding to $(\tau_{\mathrm{B}}, \tau_{\mathrm{AB}})$, respectively. Proposition S3 below states the invariance of $(\tilde{\tau}'_{\dagger,\mathrm{B}}, \tilde{\tau}'_{\dagger,\mathrm{AB}})$ to additive covariate adjustment when the x_{ws}'s are identical within each whole-plot for $\dagger = \mathrm{wls}, \mathrm{ag}$. The result extends to the sub-plot effects and interactions under the general $T_{\mathrm{A}} \times T_{\mathrm{B}}$ design with minimal modification of the notation. This illustrates a key difference between the additive and fully-interacted adjustments under factor-based specifications; see Section S5 for examples from simulation studies.

PROPOSITION S3. For $\dagger = \mathrm{wls}, \mathrm{ag}$, $(\tilde{\tau}'_{\dagger,\mathrm{B}}, \tilde{\tau}'_{\dagger,\mathrm{AB}}) = (\tilde{\tau}'_{\dagger,\mathrm{F,B}}, \tilde{\tau}'_{\dagger,\mathrm{F,AB}})$ if $x_{ws} = x_w$ for all $ws \in \mathcal{S}$.

PROOF OF PROPOSITION S3. Let A_{c} and B_{c} be the $N \times 1$ vectors of the centered factor indicators $(A_{ws} - 1/2)_{ws \in \mathcal{S}}$ and $(B_{ws} - 1/2)_{ws \in \mathcal{S}}$, respectively, where $A_{ws} = A_w$. Let $X = (x_{11}, \dots, x_{W,M_W})^{\mathrm{T}}$ with $x_{ws} = x_w$. The design matrices of the unadjusted and additive factor-based unit regressions (14) and (S46) equal

$$C = (1_N, A_{\mathrm{c}}, B_{\mathrm{c}}, A_{\mathrm{c}} \circ B_{\mathrm{c}}), \qquad C_{\mathrm{F}} = (1_N, A_{\mathrm{c}}, X, B_{\mathrm{c}}, A_{\mathrm{c}} \circ B_{\mathrm{c}}),$$

respectively, after a reshuffle of the column order in C_{F}. Recall $\Pi = \mathrm{diag}(\pi_{ws})_{ws \in \mathcal{S}}$ as the weighting matrix under the "wls" fitting scheme. Direct algebra shows that $C_1 = (1_N, A_{\mathrm{c}})$, $C_{\mathrm{F},1} = (1_N, A_{\mathrm{c}}, X)$, and $C_2 = (B_{\mathrm{c}}, A_{\mathrm{c}} \circ B_{\mathrm{c}})$ satisfy

$$C_1^{\mathrm{T}} \Pi C_2 = 0, \qquad C_{\mathrm{F},1}^{\mathrm{T}} \Pi C_2 = 0$$

due to $\sum_{s=1}^{M_w} \pi_{ws}(B_{ws} - 1/2) = 0$ for all w following from (S19). This ensures

$$C^{\mathrm{T}}\Pi C = \begin{pmatrix} C_1^{\mathrm{T}}\Pi C_1 & 0 \\ 0 & C_2^{\mathrm{T}}\Pi C_2 \end{pmatrix}, \qquad C_{\mathrm{F}}^{\mathrm{T}}\Pi C_{\mathrm{F}} = \begin{pmatrix} C_{\mathrm{F},1}^{\mathrm{T}}\Pi C_{\mathrm{F},1} & 0 \\ 0 & C_2^{\mathrm{T}}\Pi C_2 \end{pmatrix},$$

implying $(\tilde{\tau}'_{\text{wls,B}}, \tilde{\tau}'_{\text{wls,AB}})^{\mathrm{T}} = (C_2^{\mathrm{T}}\Pi C_2)^{-1} C_2^{\mathrm{T}}\Pi Y = (\tilde{\tau}'_{\text{wls,F,B}}, \tilde{\tau}'_{\text{wls,F,AB}})^{\mathrm{T}}$.

The proof for $\dagger = \text{ag}$ follows from 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 $T_{\mathrm{A}} \times T_{\mathrm{B}}$ split-plot design. Renew $H = \operatorname{diag}(p_a^{-1})_{a\in\mathcal{T}_{\mathrm{A}}} \otimes 1_{T_{\mathrm{B}}\times T_{\mathrm{B}}} - 1_{|\mathcal{T}|\times|\mathcal{T}|}$ and $H_w = \operatorname{diag}(p_a^{-1})_{a\in\mathcal{T}_{\mathrm{A}}} \otimes \{\operatorname{diag}(q_{wb}^{-1})_{b\in\mathcal{T}_{\mathrm{B}}} - 1_{T_{\mathrm{B}}\times T_{\mathrm{B}}}\}$ with $p_a = W_a/W$ and $q_{wb} = M_{wb}/M_w$. Corollary S2 below extends Lemma 1 and Theorem 4.1 to the general $T_{\mathrm{A}} \times T_{\mathrm{B}}$ split-plot design. The proof is identical to that of the 2^2 case and thus omitted.

COROLLARY S2. Lemma 1 and Theorem 4.1 hold under the $T_{\mathrm{A}} \times T_{\mathrm{B}}$ split-plot randomization and a generalized version of Condition 2 for $a \in \mathcal{T}_{\mathrm{A}}$, $b \in \mathcal{T}_{\mathrm{B}}$, and $z = (ab) \in \mathcal{T}$.

The unit regression (6) extends to the $T_{\mathrm{A}} \times T_{\mathrm{B}}$ split-plot design as

$$Y_{ws} \sim \sum_{z\in\mathcal{T}} 1(Z_{ws} = z). \tag{S50}$$

Lemma S13 below states a numeric result on the invariance of least-squares fits of (S50) 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 S13. *For a general experiment with treatment levels $\mathcal{T} = \{1, \ldots, Q\}$ and units $i = 1, \ldots, N$, let Y_i denote the observed 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 $(w_i)_{i=1}^N$ and $(w_i\rho_i)_{i=1}^N$ yield identical coefficients, heteroskedasticity-robust covariance, and cluster-robust covariance by arbitrary clustering rule.*

S4.5. *Fisher randomization test.* We give in this subsection the details on the Fisher randomization test under split-plot randomization. Assume a general $T_{\mathrm{A}} \times T_{\mathrm{B}}$ split-plot experiment with potential outcomes $\{Y_{ws}(z) : z \in \mathcal{T}\}_{ws\in\mathcal{S}}$. The weak null hypothesis concerns

$$H_{0\mathrm{N}} : G\bar{Y} = 0$$

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

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

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

$$p_{\mathrm{FRT}} = |\mathcal{Z}|^{-1} \sum_{Z'\in\mathcal{Z}} 1\{t(Z', Y, X) \ge t(Z, Y, X)\}$$

TABLE S1

Thirteen regression schemes based on the unadjusted, additive ("F"), and fully-interacted ("L") factor-based specifications. Recall $f_{ws} = (A_w - 1/2, B_{ws} - 1/2, (A_w - 1/2)(B_{ws} - 1/2))^{\mathrm{T}}$ and $f_w(A_w b) = (A_w - 1/2, b - 1/2, (A_w - 1/2)(b - 1/2))^{\mathrm{T}}$ as the vectors of the non-intercept regressors from (14) and (15), 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 regressions, we use "x", "m", and "xm" to indicate three choices of covariate combinations: (i) uses the scaled whole-plot total covariates $\hat{v}_w(A_w b) = \alpha_w \hat{x}_w(A_w b)$ ("x"), (ii) uses the whole-plot size factor α_w ("m"), and (iii) uses both ("xm").

regression scheme	fitting scheme	model specification
ols		$1 + f_{ws}$
ols.x.F	ols	$1 + f_{ws} + x_{ws}$
ols.x.L		$1 + f_{ws} + x_{ws} + f_{ws} \otimes x_{ws}$
wls		$1 + f_{ws}$
wls.x.F	wls	$1 + f_{ws} + x_{ws}$
wls.x.L		$1 + f_{ws} + x_{ws} + f_{ws} \otimes x_{ws}$
ag		$1 + f_w(A_w b)$
ag.x.F		$1 + f_w(A_w b) + \hat{v}_w(A_w b)$
ag.x.L		$1 + f_w(A_w b) + \hat{v}_w(A_w b) + f_w(A_w b) \otimes \hat{v}_w(A_w b)$
ag.m.F	ag	$1 + f_w(A_w b) + (\alpha_w - 1)$
ag.m.L		$1 + f_w(A_w b) + (\alpha_w - 1) + f_w(A_w b)(\alpha_w - 1)$
ag.xm.F		$1 + f_w(A_w b) + (a_w - 1) + \hat{v}_w(A_w b)$
ag.xm.L		$1 + f_w(A_w b) + (a_w - 1) + \hat{v}_w(A_w b) + f_w(A_w b)(\alpha_w - 1) + f_w(A_w b) \otimes \hat{v}_w(A_w b)$

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

Renew $\mathcal{B} = \{\tilde{\beta}_{\text{wls}}, \tilde{\beta}_{\text{wls},\diamond}; \tilde{\beta}_{\text{ag}}, \tilde{\beta}_{\text{ag},\diamond}, \tilde{\beta}_{\text{ag},\diamond}(\alpha, v) : \diamond = \text{F}, \text{L}\}$ as the collection of regression estimators of $\bar{Y}$ that are consistent under the general $T_{\text{A}} \times T_{\text{B}}$ split-plot design. Let $t^2(\tilde{\beta}) = (G\tilde{\beta})^{\text{T}}(G\tilde{V}G^{\text{T}})^{-1}G\tilde{\beta}$ be the robustly studentized test statistic based on $\tilde{\beta} \in \mathcal{B}$, with $\tilde{V}$ as the corresponding cluster-robust covariance. The robust studentization ensures that the resulting p_{FRT} controls the type one error rates asymptotically in the sense of

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

for all $\tilde{\beta} \in \mathcal{B}$ under Conditions 2–3. 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 $\tilde{\beta} \in \mathcal{B}$. The duality between confidence interval and hypothesis testing further ensures that the test based on $t^2(\tilde{\beta}_{\text{ag,L}}(\alpha, v))$ has the highest power asymptotically when the generalized version of (12) holds. The same guarantees also extend to $\tilde{\beta} \in \{\tilde{\beta}_{\text{ols}},\ \tilde{\beta}_{\text{ols,F}},\ \tilde{\beta}_{\text{ols,L}}\}$ under Condition 1.

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 τ_{A}, τ_{B}, and τ_{AB} 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 $(M_{w0}, M_{w1}, M_w)_{w=1}^W$ as $M_{w0} = \max(2, \zeta_{w0})$, $M_{w1} = \max(2, \zeta_{w1})$, and $M_w = M_{w0} + M_{w1}$, respectively, with the ζ_{w0}'s being i.i.d. Poisson(5) and the ζ_{w1}'s being i.i.d. Poisson(3). For each $w = 1, \ldots, W$, we draw a scalar group-level covariate x_w from $\mathcal{N}(0.2, 0.5)$, and set $x_{ws} = x_w$ for $s = 1, \ldots, M_w$. The potential outcomes are then generated as

$$Y_{ws}(00) = \theta_w + 0.5 + 2x_{ws}^2 + \epsilon_{ws}, \qquad Y_{ws}(01) = -0.5\theta_w + 1 + x_{ws}^2 + \epsilon_{ws},$$
$$Y_{ws}(10) = 0.5\theta_w + 1 - x_{ws}^2 + \epsilon_{ws}, \qquad Y_{ws}(11) = \theta_w + 2 + 2x_{ws}^2 + \epsilon_{ws}$$

for $ws \in \mathcal{S}$, where the θ_w's are independent $\mathcal{N}(2M_w/M_{\max}, 0.2)$ with $M_{\max} = \max_{w=1,\ldots,W} M_w$ and the ϵ_{ws}'s are i.i.d. Uniform$(-1, 1)$. Fix $\{Y_{ws}(z), x_{ws} : z \in \mathcal{T}\}_{ws\in\mathcal{S}}$ in simulation. We draw a random permutation of W_1 1's and W_0 0's to assign 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 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 Figure S1. We separate the results into "unadjusted vs. additive regressions" and "unadjusted vs. fully-interacted regressions" for ease of display.

Figure S1(a) shows the comparison between the unadjusted and additive regressions. 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 regressions "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 further 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 4.2–6.1. The last row illustrates the validity of the regression-based Wald-type inference. The overall conservativeness is, again, coherent with Theorems 4.2–6.1.

Figure S1(b) shows the comparison between the unadjusted and fully-interacted regressions. In addition to the same observations as in Figure S1(a), it also illustrates the efficiency gain by covariate adjustment for the sub-plot factor effect and interaction as well. The "a.mx.L" regression scheme, as the theory suggests, secures the highest efficiency overall.

Lastly, Figure S2 gives the results when the covariates are not constant within each whole-plot. Inherit all settings from above except that we now generate x_{ws} as $x_{ws} = x_w + \epsilon_{ws}$, where the ϵ_{ws}'s are i.i.d. $\mathcal{N}(0, 0.5)$. The gain in efficiency by covariate adjustment for estimating the sub-plot factor effect and interaction now surfaces under the additive regressions as well.

REFERENCES

CAMERON, A. and MILLER, D. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources* **50** 317–372.

IMAI, K., JIANG, Z. and MALANI, A. (2021). Causal inference with interference and noncompliance in two-stage randomized experiments. *Journal of the American Statistical Association* **116** 632–644.

LI, X. and DING, P. (2017). General forms of finite population central limit theorems with applications to causal inference. *Journal of the American Statistical Association* **112** 1759–1169.

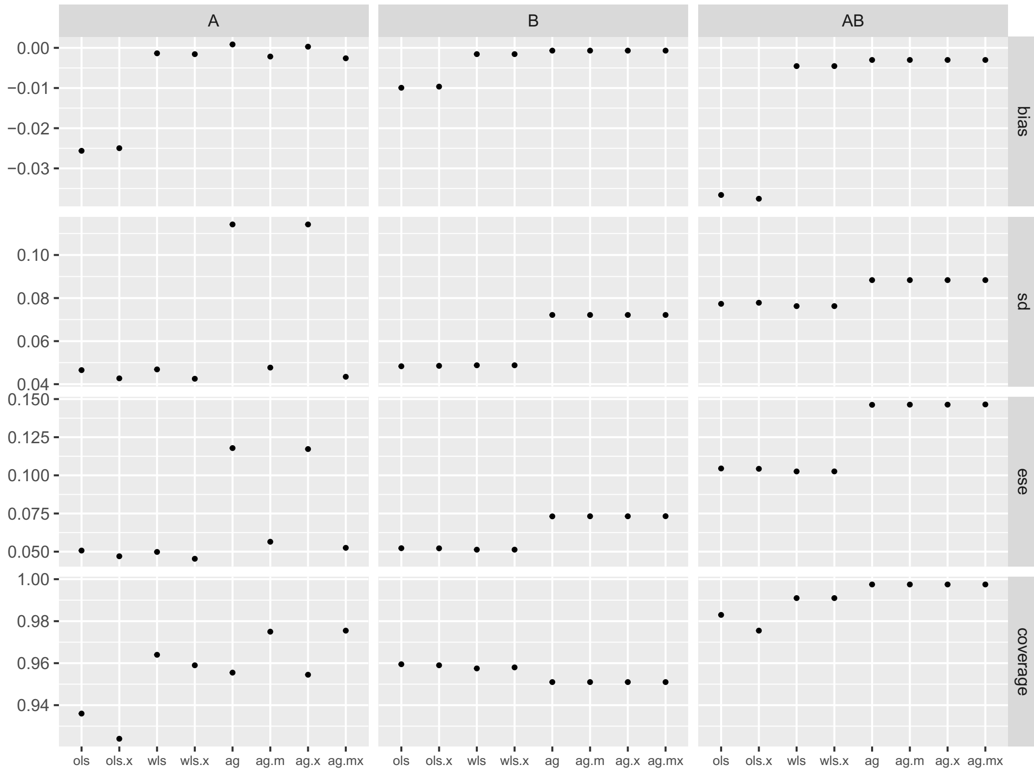
MUKERJEE, R. and DASGUPTA, T. (2021). Causal inference from possibly unbalanced split-plot designs: A randomization-based perspective. *Statistica Sinica*. http://www3.stat.sinica.edu.tw/preprint/SS-2020-0149_Preprint.pdf.

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

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

FIGURE S1. *Comparison of the least-squares estimators with $x_{ws} = x_w$.*

(a) Comparison between the unadjusted and additive regressions. We suppress the suffix ".F" in the names of the additive regressions to save some space.



(b) Comparison between the unadjusted and fully-interacted regressions. We suppress the suffix ".L" in the names of the fully-interacted regressions to save some space.

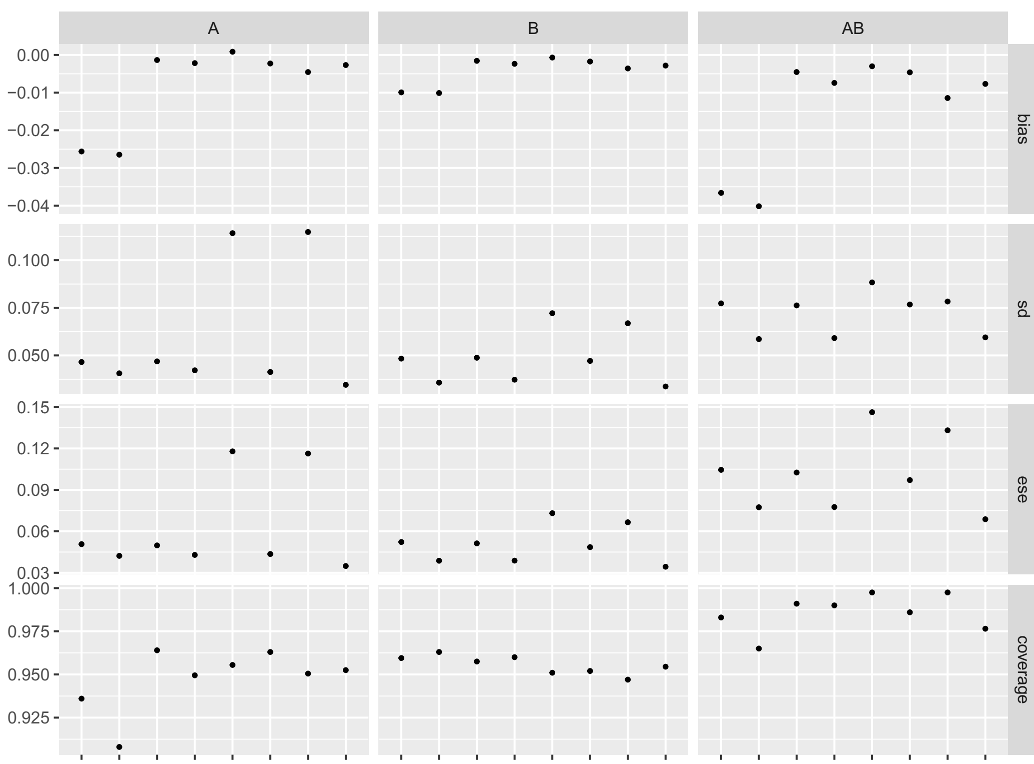
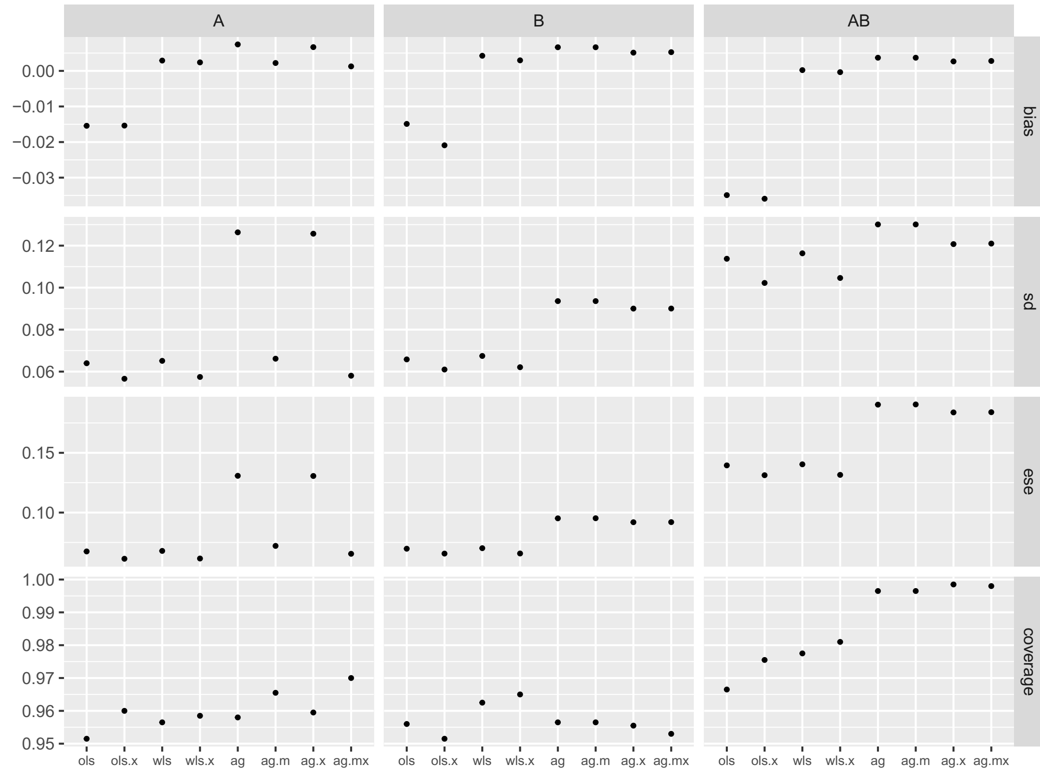


FIGURE S2. *Comparison of the least-squares estimators with varying x_{ws} within each whole-plot.*

(a) Comparison between the unadjusted and additive regressions. We suppress the suffix ".F" in the names of the additive regressions to save some space.



(b) Comparison between the unadjusted and fully-interacted regressions. We suppress the suffix ".L" in the names of the fully-interacted regressions to save some space.