Statistica Sinica Preprint No: SS-2020-0149	
Title	Causal Inference from Possibly Unbalanced Split-Plot Designs: A Randomization-based Perspective
Manuscript ID	SS-2020-0149
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0149
Complete List of Authors	Rahul Mukerjee and Tirthankar Dasgupta
Corresponding Author	Tirthankar Dasgupta
E-mail	td370@stat.rutgers.edu

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Statistica Sinica

CAUSAL INFERENCE FROM POSSIBLY UNBALANCED SPLIT-PLOT DESIGNS: A RANDOMIZATION-BASED PERSPECTIVE

Rahul Mukerjee and Tirthankar Dasgupta

Indian Institute of Managament Calcutta and Rutgers University

Abstract:

Split-plot designs find wide applicability in multifactor experiments with randomization restrictions. Practical considerations often warrant the use of unbalanced designs. This study investigates randomization-based causal inference in split-plot designs that are possibly unbalanced. An extension of the balanced case yields an expression for the sampling variance of a treatment contrast estimator, as well as a conservative estimator of the sampling variance. However, the bias of this variance estimator does not vanish, even when the treatment effects are strictly additive. A careful and involved matrix analysis is employed to overcome this difficulty, resulting in a new variance estimator that becomes unbiased under milder conditions. We propose a construction procedure that generates such an estimator with a minimax bias. Empirical studies suggest the superiority of the proposed estimator with respect to bias uniformly across different populations. Furthermore, this superiority does not come at the cost of a large inflation of the mean squared error.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Key words and phrases: Bias, Factorial experiment, Finite population, Minimaxity, Treatment-effect additivity.

1. Introduction

Factorial experiments were originally developed in the context of agricultural experiments (Fisher, 1925, 1935; Yates, 1935), and later used extensively in industrial and engineering applications (Wu and Hamada, 2009). Such experiments are currently undergoing a third popularity surge among social, behavioral, and biomedical sciences. However, one of the key challenges of using the standard principles of designing and analyzing factorial experiments in these fields arises from randomization restrictions. Consider a simplified version of the education experiment described in Dasgupta et al. (2015). Suppose the goal is to assess the causal effects on the performances of 40 schools in the state of New York of two interventions (referred to as factors in the experimental design literature) : F_1, a mid-year quality review by a team of experts, and F_2, a bonus scheme for teachers. Each factor has two levels, denoted by 1 (application) and 0 (non-application). A completely randomized assignment of the 40 schools to the four treatment combinations $00, 01, 10, 11$ is likely to disperse the schools assigned to level 1 of factor F_1 (i.e., schools to undergo review) all over the state. Such a

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

design may be prohibitive owing to travel cost and time considerations. A more practical alternative would be to divide these 40 schools by geographic proximity into four groups called whole-plots (WPs). Two of these WPs would then be assigned randomly to level 0, and the other two to level 1 of factor F_1. The teacher bonus scheme can then be applied to half of the schools chosen randomly within each WP. Such a randomization scheme is an example of a classic split-plot design. See Kirk (1982), Cochran and Cox (1957), Box et al. (2005), and Wu and Hamada (2009) for formal definitions.

Randomization-based inference is a useful methodology for drawing an inference on the causal effects of treatments from split-plot experiments in a finite-population setting, as observed by Freedman (2006, 2008). The main advantage of randomization-based inference is the fact that it applies even if the experimental units are not randomly sampled from a larger population, which is the case in most social science experiments and clinical trials (Abadie et al., 2020; Olsen et al., 2013; Rosenberger et al., 2019). Recently, Zhao et al. (2018) developed a framework for a randomization-based estimation procedure of finite-population causal effects for balanced split-plot designs, in which each WP consists of the same number of units or sub-plots (SPs), and any treatment combination of the SP factors occurs equally often in all WPs; see (2.4). However, unbalanced split-plot designs

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

are quite common in the social sciences. Consider the school experiment described earlier. Suppose the 40 schools are spread over four counties with 8, 8, 12, and 12 schools, respectively, in these counties. In this case, each county can be considered as a natural WP. Thus, the design is unbalanced and the estimation methodology proposed by Zhao et al. (2018) is no longer applicable.

In this study, we investigate randomization-based causal inference in split-plot designs that are possibly unbalanced, using the potential outcomes framework originally introduced by Neyman (1923), formalized in Rubin (1974); and extended in the subsequent works of Rubin. We start with a natural unbiased estimator of a typical treatment contrast; and examine how far the approach of Zhao et al. (2018) for the balanced case can be adapted to our more general setup. It is seen that this approach, aided by a variable transformation, yields an expression for the sampling variance of the treatment contrast estimator, but runs into difficulty in variance estimation. Specifically, as in the balanced case and other situations in causal inference, the resulting variance estimator is conservative in the sense of having a nonnegative bias. However, unlike in most standard situations, the bias does not vanish, even under strict additivity or homegeneity of the treatment effects. To overcome this problem, a careful matrix analysis is

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

employed, leading, under wide generality, to a new variance estimator. This estimator is also conservative, but enjoys the nice property of becoming unbiased under between-WP additivity, a condition even milder than strict additivity. We also discuss the issue of minimaxity, with a view to controlling the bias in the variance estimation, and use simulations to explore the bias and mean squared error (MSE) of the two estimators under treatment effect heterogeneity. Proofs of all results appear in the Supplementary Material.

2. Notation and background

Consider a factorial experiment conducted to assess the causal effects of m_1 WP factors $F_{11}, \ldots, F_{1m_1}$ and m_2 SP factors $F_{21}, \ldots, F_{2m_2}$ on a finite population of N units. Each factor has two or more levels. The treatment combinations are denoted by $z = z_1 z_2$, where $z_k \in Z_k$ and Z_k is the set of factor levels of $F_{k1}, \ldots, F_{km_k}$ $(k = 1, 2)$. For $i = 1, \ldots, N$, let $Y_i(z_1 z_2)$ denote the potential outcome of unit i when exposed to treatment combination $z_1 z_2$. This notation assumes the stable unit treatment value assumption (Rubin, 1980), which means that (1) the potential outcome of unit i depends only on the treatment combination it is assigned to, and (2) there are no hidden versions of treatments not represented by all level combinations of

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

the $m_1 + m_2$ factors. A typical treatment contrast for unit i is of the form

$$\tau_i = \sum_{z_1 \in Z_1} \sum_{z_2 \in Z_2} g(z_1 z_2) Y_i(z_1 z_2), \tag{2.1}$$

where $g(z_1 z_2)$, for $z_1 \in Z_1 and z_2 \in Z_2$, are known, not all zeros, and sum to zero. In the school example, unit i has four potential outcomes: $Y_i(00)$, $Y_i(01)$, $Y_i(10)$ and $Y_i(11)$. Following Dasgupta et al. (2015), the unit-level main effect of factor F_1 is defined as $\tau_i^{F_1} = \{-Y_i(00) - Y_i(01) + Y_i(10) + Y_i(11)\}/2$. The contrast coefficients $g(00), g(01), g(10)$ and $g(11)$, are $-1/2, -1/2, 1/2$ and $1/2$, respectively. The contrast coefficients for the main effect of factor F_2 and the interaction F_1F_2 can be similarly defined. Let

$$\overline{Y}(z_1 z_2) = N^{-1} \sum_{i=1}^{N} Y_i(z_1 z_2) \tag{2.2}$$

denote the average potential outcome for treatment combination $z_1 z_2$, and let

$$\overline{\tau} = N^{-1} \sum_{i=1}^{N} \tau_i = \sum_{z_1 \in Z_1} \sum_{z_2 \in Z_2} g(z_1 z_2) \overline{Y}(z_1 z_2) \tag{2.3}$$

denote a treatment contrast for the finite population of N units. We define $\overline{\tau}$ as the finite-population causal estimand of interest, and consider the problem of drawing an inference on $\overline{\tau}$ using the outcomes observed from the experiment.

The observed outcomes are generated through an assignment mecha-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

nism, which is the process of allocating treatment combinations to the N units. Here, we consider the following split-plot assignment mechanism. Suppose there is a partitioning of the N experimental units into $W(\geq 2)$ disjoint sets $\Omega_1, \ldots, \Omega_W$, called WPs, such that Ω_w consists of $M_w(\geq 2)$ units, called SPs, for $w = 1, \ldots, W$, and $M_1 + \ldots M_W = N$. Consider now a two-stage randomization, that assigns $r_1(z_1)$ (≥ 2) WPs to level combination z_1 of $F_{11}, \ldots F_{1m_1}$ and then, for each $w = 1, \ldots, W$, assigns $r_{w2}(z_2)$ SPs within WP Ω_w to level combination z_2 of $F_{21}, \ldots F_{2m_2}$. Here, at each stage, all assignments are equiprobable, $r_1(z_1)$ and $r_{w2}(z_2)$ are fixed positive integers, and $\sum_{z_1 \in Z_1} r_1(z_1) = W$ and $\sum_{z_2 \in Z_2} r_{w2}(z_2) = M_w$, for $w = 1, \ldots, W$.

Note that the above assignment mechanism yields a *balanced* split-plot design if

$$M_1 = \cdots = M_W, \quad r_{12}(z_2) = \cdots = r_{W2}(z_2), \text{ for all } z_2 \in Z_2. \tag{2.4}$$

In the school example described in Section 1, the WPs represent sets of schools within a county, and we have $N = 40$, $W = 4$, $M_1 = M_2 = 8$, $M_3 = M_4 = 12$, and $Z_1 = Z_2 = \{0, 1\}$. Finally, for all $z_2 \in Z_2$, $r_{w2}(z_2) = 4$ for $w = 1, 2$, and $r_{w2}(z_2) = 6$ for $w = 3, 4$. Thus, the design is unbalanced.

To define the observed outcomes of the experiment, we introduce two sets of random treatment assignment indices at the WP and SP levels. Let

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

$T_1(z_1)$ denote the set of indices w such that WP Ω_w is randomly assigned to level combination z_1 of $F_{11}, \ldots, F_{1m_1}$. Similarly, for $z_2 \in Z_2$ and $w = 1, \ldots, W$, let $T_{w2}(z_2)$ be the set of SPs in Ω_w randomly assigned to level combination z_2 of $F_{21}, \ldots, F_{2m_2}$. For any treatment combination z_1z_2, the observed outcomes from the WP Ω_w, for $w \in T_1(z_1)$, are then $Y_i(z_1z_2)$, for $i \in T_{w2}(z_2)$. Let

$$\overline{Y}_w^{\text{obs}}(z_1z_2) = \{r_{w2}(z_2)\}^{-1} \sum_{i \in T_{w2}(z_2)} Y_i(z_1z_2) \tag{2.5}$$

denote the average observed outcome for treatment combination z_1z_2 within WP Ω_w, for $w \in T_1(z_1)$. In the spirit of the usual unbiased estimator of the population mean in two-stage sampling (Cochran, 1977), define

$$\overline{Y}^{\text{obs}}(z_1z_2) = \frac{W}{Nr_1(z_1)} \sum_{w \in T_1(z_1)} M_w \overline{Y}_w^{\text{obs}}(z_1z_2) = \frac{1}{r_1(z_1)} \sum_{w \in T_1(z_1)} \frac{M_w}{\overline{M}} \overline{Y}_w^{\text{obs}}(z_1z_2), \tag{2.6}$$

where $\overline{M} = (M_1 + \ldots + M_W)/W = N/W$ is the average WP size. From (2.5) and (2.6), by conditioning on the randomization at the WP level, it is straightforward to verify that $E\left\{\overline{Y}^{\text{obs}}(z_1z_2)\right\} = \overline{Y}(z_1z_2)$, where $\overline{Y}(z_1z_2)$ is given by (2.2). Using (2.3), an immediate consequence of this fact is Proposition 1.

Proposition 1. *An unbiased estimator of the finite-population treatment*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

contrast $\overline{\tau}$ is given by

$$\widehat{\overline{\tau}} = \sum_{z_1 \in Z_1} \sum_{z_2 \in Z_2} g(z_1 z_2) \overline{Y}^{\text{obs}}(z_1 z_2), \tag{2.7}$$

where $\overline{Y}^{\text{obs}}(z_1 z_2)$ is given by (2.6).

Proposition 1 yields a point estimator of $\overline{\tau}$. However, to quantify the uncertainty associated with the point estimator, one needs to derive and estimate the sampling variance of $\widehat{\overline{\tau}}$ with respect to its distribution induced by the randomization in the split-plot design. In the next two subsections, we briefly describe two areas of research that are related to our problem setting. The first is the case of *unbalanced block designs*, which we discuss in Section 2.1 and explain why the results from this setting do not apply to ours. The second is the case of *balanced split-plot designs*, which motivates our work.

2.1 Why do the results from block designs not work?

The problem of Neymanian variance estimation has been recently investigated for block designs that are unbalanced, that is, they have unequal block sizes, and the results (Pashley and Miratrix, 2019) have been applied to other settings (Schochet et al., 2020). A natural question that arises is whether the ideas used to derive variance estimators for unbalanced block

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

designs that have desirable properties can be applied to the case of split-plot designs. As seen from the subsequent discussion, the fundamentally different setting of the two designs does not permit such adaptation.

Consider a block design in W blocks $\Omega_1, \ldots, \Omega_W$ of sizes $M_1, \ldots, M_W$, such that any treatment combination z_1z_2 is randomly assigned to $r_w(z_1z_2)$ units of Ω_w, for $w = 1, \ldots, W$. The fixed positive integers $r_w(z_1z_2)$ sum to M_w for each w. Subject to this, all treatment assignments within each block are equally likely. In addition, randomization is done independently for different blocks. Similarly to (2.3) (See also (3.2) and (3.3) defined later), let $\overline{\tau}_1, \ldots, \overline{\tau}_W$ be block-level treatment contrasts, leading to the population-level treatment contrast $\overline{\tau} = (1/W)\sum_{w=1}^{W}(M_w/\overline{M})\overline{\tau}_w$, where $\overline{M} = N/W$ and $N = M_1 + \ldots + M_W$. If $\widehat{\overline{\tau}}_w$ is an unbiased estimator of $\overline{\tau}_w$ on the basis of observed responses from Ω_w, then $\overline{\tau}$ is unbiasedly estimated by

$$\widehat{\overline{\tau}} = (1/W)\sum_{w=1}^{W}(M_w/\overline{M})\widehat{\overline{\tau}}_w. \tag{2.8}$$

The sampling variance of this estimator is $(1/W)^2\sum_{w=1}^{W}(M_w/\overline{M})^2\mathrm{var}(\widehat{\overline{\tau}}_w)$, because $\overline{\tau}_1, \ldots, \overline{\tau}_W$ are independent across blocks. In the special case of a single treatment factor with two levels, Pashley and Miratrix (2019) proposed several estimators of this variance, depending on the composition of the blocks. For blocks containing at least two treated and two control units, this estimator is straightforward and is obtained by substituting a conser-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

vative estimator of $\mathrm{var}(\widehat{\overline{\tau}}_w)$ that vanishes under strict additivity. For blocks that contain only one treated or control unit, Pashley and Miratrix (2019) proposed two estimators based on the weighted version of $\sum_{w=1}^{W}\left(\widehat{\overline{\tau}}_w - \widehat{\overline{\tau}}\right)^2$. Both estimators are conservative, in general, but unbiased under the assumption of block-level additivity; that is, the average treatment effect is the same across all blocks. This approach is easily extendable to experiments with multiple treatments and factorial experiments.

To see why *neither of the above estimators can be defined* for split-plot designs, note that in order for them to be defined, the estimator $\widehat{\overline{\tau}}_w$ of the WP-level contrast $\overline{\tau}_w$ needs to be defined for each $w = 1, \ldots, W$. However, in a split-plot design, $\widehat{\overline{\tau}}_w$ cannot be defined for *any* w because of the WP-level randomization. To understand this further, consider a toy example with 12 units and two factors (each at two levels, 0 and 1) in two settings, as shown in Table 2.1: (i) in $W = 2$ blocks of sizes $M_1 = 8$ and $M_2 = 4$ (left half), and (ii) in $W = 2$ WPs of sizes $M_1 = 8$ and $M_2 = 4$ (right half), where the first factor is the WP-factor. Assume that the contrast τ of interest is the main effect of the second factor. The unit-level contrasts $\tau_i = \{-Y_i(00) + Y_i(01) - Y_i(10) + Y_i(11)\}/2$, for $i = 1, \ldots, 12$, and the population-level contrast $\overline{\tau}$ are the same in both situations. To make the two settings comparable, assume that for the block design $r_1(z_1 z_2) = 2$ and

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

$r_2(z_1z_2) = 1$, for $z_1, z_2 \in \{0, 1\}$, and for the split-plot design $r_1(z_1) = 1$, $r_{12}(z_2) = 4$ and $r_{22}(z_2) = 2$. One possible realization of the treatment assignment for each design is shown in Table 2.1. In the split-plot design, the first WP receives level 0 of the first factor, whereas the second WP receives level 1.

For the block design, $\widehat{\overline{\tau}}_w$ is readily defined for each block w. The block-level variance $\text{var}(\widehat{\overline{\tau}}_w)$ can be estimated for $w = 1$, but not for $w = 2$, and hence the estimators based on $\sum_{w=1}^{W} \left(\widehat{\overline{\tau}}_w - \widehat{\overline{\tau}}\right)^2$ can be used. However, for the split-plot design, neither $\widehat{\overline{\tau}}_w$ nor $\text{var}(\widehat{\overline{\tau}}_w)$ can be defined for any of the WPs, because we observe potential outcomes for only two of the four treatment combinations within each WP.

2.2 Sampling variance and its estimation: Prior work on balanced designs

Zhao et al. (2018) derived an expression for the sampling variance of $\widehat{\overline{\tau}}$ for a balanced split-plot design, that is, when conditions (2.4) are satisfied. Under these conditions, we can write $r_{w2}(z_2) = r_2(z_2)$, for $w = 1, \ldots, W$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Table 1: Block design versus split-plot design

Block	Unit	Factor levels 00	01	10	11
1	1	X			
	2	X			
	3		X		
	4		X		
	5			X	
	6			X	
	7				X
	8				X
2	9	X			
	10		X		
	11			X	
	12				X

WP	Unit	Factor levels 00	01	10	11
1	1	X			
	2	X			
	3		X		
	4		X		
	5	X			
	6	X			
	7		X		
	8		X		
2	9			X	
	10			X	
	11				X
	12				X

and $z_2 \in Z_2$. Equations (2.5) and (2.6) respectively reduce to:

$$\begin{aligned}\overline{Y}_w^{\mathrm{obs}}(z_1z_2) &= \{r_2(z_2)\}^{-1} \sum_{i \in T_{w2}(z_2)} Y_i(z_1z_2),\\ \overline{Y}^{\mathrm{obs}}(z_1z_2) &= \{r_1(z_1)\}^{-1} \sum_{w \in T_1(z_1)} \overline{Y}_w^{\mathrm{obs}}(z_1z_2).\end{aligned}$$

Next, Zhao et al. (2018) defined the following quantities for any $z_1, z_1^* \in$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Z_1 and $z_2, z_2^* \in Z_2$:

$$S_{\mathrm{bt}}(z_1 z_2, z_1^* z_2^*) = \frac{M}{W-1} \sum_{w=1}^{W} \left\{ \overline{Y}_w(z_1 z_2) - \overline{Y}(z_1 z_2) \right\} \left\{ \overline{Y}_w(z_1^* z_2^*) - \overline{Y}(z_1^* z_2^*) \right\}, \tag{2.9}$$

$$S_{\mathrm{in}}(z_1 z_2, z_1^* z_2^*) = \frac{\sum_{w=1}^{W} \sum_{i \in \Omega_w} \left\{ Y_i(z_1 z_2) - \overline{Y}_w(z_1 z_2) \right\} \left\{ Y_i(z_1^* z_2^*) - \overline{Y}_w(z_1^* z_2^*) \right\}}{W(M-1)}. \tag{2.10}$$

Here, M is the common size of the WPs in the balanced case and $\overline{Y}_w(z_1 z_2) = M^{-1} \sum_{i \in \Omega_w} Y_i(z_1 z_2)$, for each w and $z_1 z_2$. The quantities $S_{\mathrm{bt}}(z_1 z_2, z_1^* z_2^*)$ and $S_{\mathrm{in}}(z_1 z_2, z_1^* z_2^*)$ represent, respectively, the between- and within-WP mean squares or products in an analysis of variance/covariance decomposition of the potential outcomes, that is, of the quantity

$$\sum_{w=1}^{W} \sum_{i \in \Omega_w} \left\{ Y_i(z_1 z_2) - \overline{Y}(z_1 z_2) \right\} \left\{ Y_i(z_1^* z_2^*) - \overline{Y}(z_1^* z_2^*) \right\}.$$

In addition, they defined the following quantity, which is a function of the observed outcomes and can be computed from experimental data:

$$\widehat{v}(\widehat{\overline{\tau}}) = \sum_{z_1 \in Z_1} \sum_{z_2 \in Z_2} \sum_{z_2^* \in Z_2} \{r_1(z_1)\}^{-1} g(z_1 z_2) g(z_1 z_2^*) s\left(z_1 z_2, z_1 z_2^*\right), \tag{2.11}$$

where

$$s\left(z_1 z_2, z_1 z_2^*\right) = \sum_{w \in T_1(z_1)} \frac{\left\{ \overline{Y}_w^{\mathrm{obs}}(z_1 z_2) - \overline{Y}^{\mathrm{obs}}(z_1 z_2) \right\} \left\{ \overline{Y}_w^{\mathrm{obs}}(z_1 z_2^*) - \overline{Y}^{\mathrm{obs}}(z_1 z_2^*) \right\}}{r_1(z_1) - 1}. \tag{2.12}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

We now summarize the main results of Zhao et al. (2018) in the following theorem on the sampling variance of $\widehat{\overline{\tau}}$ and its estimation.

Theorem 1 (Zhao). (a) *The sampling variance of* $\widehat{\overline{\tau}}$ *is given by*

$$\begin{aligned}\mathrm{var}_{\text{S-P}}(\widehat{\overline{\tau}}) &= \sum_{z_1\in Z_1}\sum_{z_2\in Z_2}\sum_{z_2^*\in Z_2}\frac{g(z_1z_2)g(z_1z_2^*)\left\{S_{\mathrm{bt}}\left(z_1z_2, z_1z_2^*\right)-S_{\mathrm{in}}\left(z_1z_2, z_1z_2^*\right)\right\}}{Mr_1(z_1)}\\ &+ \sum_{z_1\in Z_1}\sum_{z_2\in Z_2}\frac{\left\{g(z_1z_2)\right\}^2 S_{\mathrm{in}}\left(z_1z_2, z_1z_2\right)}{r_1(z_1)r_2(z_2)}-\frac{\sum_{w=1}^{W}\left(\overline{\tau}_w-\overline{\tau}\right)^2}{W(W-1)},\end{aligned}$$

where

$$\overline{\tau}_w = M^{-1}\sum_{i\in\Omega_w}\tau_i = \sum_{z_1\in Z_1}\sum_{z_2\in Z_2} g(z_1z_2)\overline{Y}_w(z_1z_2), \quad (w=1,\ldots,W).$$

(b) $E\left\{\widehat{v}(\widehat{\overline{\tau}})\right\} \geq \mathrm{var}_{\text{S-P}}(\widehat{\overline{\tau}})$, *with equality holding for every treatment contrast* τ *if and only if between-WP additivity holds, which means* $\overline{Y}_w(z_1z_2)-\overline{Y}_w(z_1^*z_2^*)$ *is the same over* $w=1,\ldots,W$, *for every pair of treatment combinations* z_1z_2 *and* $z_1^*z_2^*$.

Thus, to summarize, Zhao et al. (2018) obtained an estimator of the sampling variance that, like most variance estimators in finite-population causal inference (Mukerjee et al., 2018), has a nonnegative bias. Further, they noted that this bias vanishes for every treatment contrast $\overline{\tau}$ if and only if between-WP additivity holds, which means

$$\overline{Y}_1(z_1z_2)-\overline{Y}_1(z_1^*z_2^*) = \cdots = \overline{Y}_W(z_1z_2)-\overline{Y}_W(z_1^*z_2^*), \qquad (2.13)$$

for every pair of treatment combinations z_1z_2 and $z_1^*z_2^*$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

3. Sampling variance and its estimation

In this section, we derive an expression for the sampling variance. Then, we find a variance estimator that generalizes the results in Section 2.2 to the unbalanced case, and examine its properties. Note that these results revolve around the quantities $S_{\mathrm{bt}}(z_1z_2, z_1^*z_2^*)$ and $S_{\mathrm{in}}(z_1z_2, z_1^*z_2^*)$ defined in equations (2.9) and (2.10), respectively, and both depend on the common WP size M. Because M_w varies across whole plots in an unbalanced split-plot design, it is difficult to guess what the counterparts of these two quantities will be in the unbalanced case.

Note that by (2.2), $\overline{Y}(z_1z_2) = N^{-1}\sum_{w=1}^{W} M_w\overline{Y}_w(z_1z_2)$, where $\overline{Y}_w(z_1z_2) = M_w^{-1}\sum_{i\in\Omega_w} Y_i(z_1z_2)$ is the average potential outcome of all units in WP Ω_w for treatment combination z_1z_2. A helpful feature of the balanced case is that $\overline{Y}(z_1z_2)$, is then the simple average of $\overline{Y}_w(z_1z_2)$ for $w = 1,\dots,W$, and similarly, $\overline{Y}^{\mathrm{obs}}(z_1z_2)$ in (2.6) is the simple average of $\overline{Y}_w^{\mathrm{obs}}(z_1z_2)$, for $w \in T_1(z_1)$. We first translate this "simple average" feature to the unbalanced case, with a view to facilitating the derivation there. To that end, we convert the "raw" potential outcomes $Y_i(z_1z_2)$ to "adjusted" potential outcomes

$$U_i(z_1z_2) = (M_w/\overline{M})Y_i(z_1z_2), \tag{3.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

for each $z_1 \in Z_1$, $z_2 \in Z_2$, $i \in \Omega_w$, and $w = 1, \ldots, W$. For each z_1z_2, define

$$\overline{U}_w(z_1z_2) = M_w^{-1} \sum_{i \in \Omega_w} U_i(z_1z_2),\ w = 1, \ldots, W, \quad \text{and} \quad \overline{U}(z_1z_2) = W^{-1} \sum_{w=1}^{W} \overline{U}_w(z_1z_2).$$

By (3.1), $\overline{U}(z_1z_2) = \overline{Y}(z_1z_2)$, that is, $\overline{Y}(z_1z_2)$ is the simple average of $\overline{U}_w(z_1z_2)$, for $w = 1, \ldots, W$, irrespective of whether or not $M_1, \ldots, M_W$ are equal. As seen later in equation (3.5), a similar simple average relationship holds also between their observed counterparts. The points just noted simplify the derivation to some extent, but additional complications remain to be addressed, for example, $r_{w2}(z_2)$ not being constant over w in the unbalanced case.

Next, for $z_1, z_1^* \in Z_1$ and $z_2, z_2^* \in Z_2$, define

$$S_{\text{bt}}(z_1z_2, z_1^*z_2^*) = \frac{\overline{M}}{W-1} \sum_{w=1}^{W} \left\{\overline{U}_w(z_1z_2) - \overline{U}(z_1z_2)\right\} \left\{\overline{U}_w(z_1^*z_2^*) - \overline{U}(z_1^*z_2^*)\right\},$$
$$S_{\text{in},w}(z_1z_2, z_1^*z_2^*) = \frac{1}{M_w - 1} \sum_{i \in \Omega_w} \left\{U_i(z_1z_2) - \overline{U}_w(z_1z_2)\right\} \left\{U_i(z_1^*z_2^*) - \overline{U}_w(z_1^*z_2^*)\right\}.$$

Some additional notation is necessary. First, let

$$\overline{\tau}_w = (1/M_w) \sum_{i \in \Omega_w} \tau_i = \sum_{z_1 \in Z_1} \sum_{z_2 \in Z_2} g(z_1z_2)\overline{Y}_w(z_1z_2),\ w = 1, \ldots, W, \quad (3.2)$$

denote the WP-level treatment contrasts, where $\overline{Y}_w(z_1z_2)$, defined earlier, is the average potential outcome of all units in WP Ω_w for treatment combination z_1z_2. The second equality in (3.2) follows from (2.1). Furthermore,

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

from (2.3) and (3.2), it follows that

$$\overline{\tau} = (1/W) \sum_{w=1}^{W} (M_w/\overline{M}) \overline{\tau}_w. \tag{3.3}$$

Now, define

$$\Delta = \frac{1}{W(W-1)} \sum_{w=1}^{W} \left\{ (M_w/\overline{M}) \overline{\tau}_w - \overline{\tau} \right\}^2, \tag{3.4}$$

where $\overline{\tau}_w$ is given by (3.2). Then, extending the ideas of Zhao et al. (2018), after considerable algebra, we obtain the following result on the sampling variance of $\widehat{\overline{\tau}}$, the unbiased estimator of $\overline{\tau}$.

Theorem 2. *The sampling variance of* $\widehat{\overline{\tau}}$ *is*

$$\begin{aligned} \mathrm{var}(\widehat{\overline{\tau}}) &= \sum_{z_1 \in Z_1} \sum_{z_2 \in Z_2} \sum_{z_2^* \in Z_2} \frac{g(z_1 z_2) g(z_1 z_2^*)}{r_1(z_1)} \left(\frac{S_{\mathrm{bt}}(z_1 z_2, z_1 z_2^*)}{\overline{M}} - \sum_{w=1}^{W} \frac{S_{\mathrm{in},w}(z_1 z_2, z_1 z_2^*)}{W M_w} \right) \\ &+ \sum_{z_1 \in Z_1} \sum_{z_2 \in Z_2} \frac{\{g(z_1 z_2)\}^2}{W r_1(z_1)} \sum_{w=1}^{W} \frac{S_{\mathrm{in},w}(z_1 z_2, z_1 z_2)}{r_{w2}(z_2)} - \Delta. \end{aligned}$$

Next, to obtain an estimator of the sampling variance, we first define the counterparts of $\overline{Y}_w^{\mathrm{obs}}(z_1 z_2)$ and $\overline{Y}^{\mathrm{obs}}(z_1 z_2)$ in (2.5) and (2.6) in terms of the adjusted potential outcomes:

$$\begin{aligned} \overline{U}_w^{\mathrm{obs}}(z_1 z_2) &= \frac{1}{r_{w2}(z_2)} \sum_{i \in T_{w2}(z_2)} U_i(z_1 z_2), \quad w \in T_1(z_1) \quad \text{and} \\ \overline{U}^{\mathrm{obs}}(z_1 z_2) &= \frac{1}{r_1(z_1)} \sum_{w \in T_1(z_1)} \overline{U}_w^{\mathrm{obs}}(z_1 z_2). \end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Then, it is easy to see from (2.5), (2.6), and (3.1) that

$$\overline{Y}^{\text{obs}}(z_1z_2) = \overline{U}^{\text{obs}}(z_1z_2). \tag{3.5}$$

Note that $\overline{U}^{\text{obs}}(z_1z_2)$ is the simple average of $\overline{U}_w^{\text{obs}}(z_1z_2)$, for $w \in T_1(z_1)$, irrespective of whether or not $M_1, \ldots, M_W$ are equal. This is precisely what the relationship between $\overline{Y}^{\text{obs}}(z_1z_2)$ and $\overline{Y}_w^{\text{obs}}(z_1z_2)$ in (2.6) reduces to when $M_1 = \cdots = M_W$, providing us with the intuition to generalize the results of Zhao et al. (2018) by using of the adjusted potential outcomes, in particular, replacing $\overline{Y}^{\text{obs}}(z_1z_2)$ in (2.7) with $\overline{U}^{\text{obs}}(z_1z_2)$, because of (3.5). We now define the following estimator of the sampling variance in Theorem 2:

$$\widehat{V}(\widehat{\overline{\tau}}) = \sum_{z_1 \in Z_1} \sum_{z_2 \in Z_2} \sum_{z_2^* \in Z_2} \frac{g(z_1z_2)g(z_1z_2^*)}{r_1(z_1)} \widehat{S}(z_1z_2, z_1z_2^*), \tag{3.6}$$

where

$$\widehat{S}(z_1z_2, z_1z_2^*) = \frac{1}{r_1(z_1) - 1} \sum_{w \in T_1(z_1)} \left\{ \overline{U}_w^{\text{obs}}(z_1z_2) - \overline{U}^{\text{obs}}(z_1z_2) \right\} \left\{ \overline{U}_w^{\text{obs}}(z_1z_2^*) - \overline{U}^{\text{obs}}(z_1z_2^*) \right\}.$$

Again, considerable algebra yields the following result.

Theorem 3. *The variance estimator $\widehat{V}(\widehat{\overline{\tau}})$ given by (3.6), estimates the sampling variance of $\widehat{\overline{\tau}}$ with a nonnegative bias Δ defined by (3.4), that is,* $E\left\{ \widehat{V}(\widehat{\overline{\tau}}) \right\} = \text{var}(\widehat{\overline{\tau}}) + \Delta.$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Remark 1. Theorem 3 shows that $\widehat{V}(\widehat{\overline{\tau}})$ is a conservative estimator of $\text{var}(\widehat{\overline{\tau}})$ with a nonnegative bias Δ. This property is in line with variance estimators in other situations of randomization-based causal inference. Moreover, in the balanced case, by (3.4), the bias Δ vanishes when $\overline{\tau}_1 = \cdots = \overline{\tau}_W = \overline{\tau}$, which happens for every treatment contrast $\overline{\tau}$ if and only if (2.13), that is, between-WP additivity, holds. This shows how the results of Zhao et al. (2018), presented in Section 2.2, follow from Theorem 3. A disturbing feature of the variance estimator $\widehat{V}(\widehat{\overline{\tau}})$, however, emerges in the unbalanced case, which is the main focus of this study. Then, $\widehat{V}(\widehat{\overline{\tau}})$ remains biased, even if between-WP additivity holds, because by (3.2) and (3.3), condition (2.13) implies $\overline{\tau}_1 = \cdots = \overline{\tau}_W = \overline{\tau}$ and, hence,

$$\Delta = \frac{\overline{\tau}^2}{W(W-1)\overline{M}^2} \sum_{w=1}^{W} (M_w - \overline{M})^2,$$

which is positive when $M_1, \ldots, M_W$ are not all equal, unless $\overline{\tau} = 0$. The situation remains unchanged even under the stronger assumption of strict additivity or homogeneity of treatment effects (Neyman, 1923), which enforces the constancy of $Y_i(z_1z_2) - Y_i(z_1^*z_2^*)$ over $i = 1, \ldots, N$ for every pair of treatment combinations z_1z_2 and $z_1^*z_2^*$.

This property of $\widehat{V}(\widehat{\overline{\tau}})$ described in Remark 1 is a matter of concern because a requirement typically imposed on a variance estimator in causal

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

inference is that it should become unbiased, at least under Neymanian strict additivity, if not under milder versions thereof, such as between-WP additivity in the present context. The estimator $\widehat{V}(\widehat{\overline{\tau}})$, obtained by generalizing the arguments in the balanced case, fails to meet this requirement when $M_1, \ldots, M_W$ are not all equal. In the rest of the paper, we investigate the existence of a variance estimator that overcomes this difficulty and show how, under wide generality, such an estimator can be obtained by appropriately modifying $\widehat{V}(\widehat{\overline{\tau}})$ as given by (3.6).

4. A new variance estimator

We begin our search for an improved variance estimator by expanding the bias term Δ defined in (3.4) as follows:

$$\Delta = (1/N)^2 \left[\sum_{w=1}^{W} M_w^2 \overline{\tau}_w^2 - \sum_{w=1}^{W} \sum_{w^*(\neq w)=1}^{W} \{M_w M_{w^*}/(W-1)\} \overline{\tau}_w \overline{\tau}_{w^*} \right]. \quad (4.1)$$

Note that in (4.1), the term $\overline{\tau}_w^2$ is not unbiasedly estimable, but for $w \neq w^*$, $\overline{\tau}_w \overline{\tau}_{w^*}$ allows unbiased estimation. This is because, by (3.2),

$$\overline{\tau}_w^2 = (1/M_w)^2 \sum_{z_1 \in Z_1} \sum_{z_2 \in Z_2} \sum_{z_1^* \in Z_1} \sum_{z_2^* \in Z_2} \sum_{i \in \Omega_w} \sum_{i^* \in \Omega_w} g(z_1 z_2) g(z_1^* z_2^*) Y_i(z_1 z_2) Y_{i^*}(z_1^* z_2^*). \quad (4.2)$$

The sums over i and i^* in (4.2) include the case $i = i^*$. There is at least one pair of distinct treatment combinations $z_1 z_2$ and $z_1^* z_2^*$ such that

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

$g(z_1z_2)g(z_1^*z_2^*) \neq 0$ and $Y_i(z_1z_2)Y_i(z_1^*z_2^*)$ is never observable, because unit i cannot be assigned simultaneously to both z_1z_2 and $z_1^*z_2^*$. Hence, $\overline{\tau}_w^2$ does not allow unbiased estimation. On the other hand, for $w \neq w^*$, $\overline{\tau}_w\overline{\tau}_{w^*}$ does not involve terms such as $Y_i(z_1z_2)Y_i(z_1^*z_2^*)$, and is unbiasedly estimable. For each w, let z_{1w} denote the level combination of the WP factors assigned to WP Ω_w. Now, define

$$G_w^{\text{obs}} = \sum_{z_2 \in Z_2} g(z_{1w}z_2)\overline{Y}_w^{\text{obs}}(z_{1w}z_2).$$

Proposition 2. *For $w, w^* = 1, \ldots, W$, $w \neq w^*$, an unbiased estimator of $\overline{\tau}_w\overline{\tau}_{w^*}$ is given by*

$$H_{ww^*} = \frac{W(W-1)G_w^{\text{obs}}G_{w^*}^{\text{obs}}}{r_1(z_{1w})\left\{r_1(z_{1w^*}) - \delta(z_{1w}, z_{1w^*})\right\}},$$

where $\delta(z_{1w}, z_{1w^*})$ is an indicator equal to one if $z_{1w} = z_{1w^*}$, and zero otherwise.

We can now use Proposition 2 to construct a new estimator of $\text{var}(\widehat{\overline{\tau}})$. Consider any symmetric matrix $B = ((b_{ww^*}))$ of order W such that $b_{ww} = M_w^2$, for $w = 1, \ldots, W$. Now, define the variance estimator

$$\widetilde{V}(\widehat{\overline{\tau}}) = \widehat{V}(\widehat{\overline{\tau}}) + (1/N^2)\sum_{w=1}^{W}\sum_{w^*(\neq w)=1}^{W}\left[b_{ww^*} + \{M_wM_{w^*}/(W-1)\}\right]H_{ww^*}, \tag{4.3}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

where $\widehat{V}(\widehat{\overline{\tau}})$ is the variance estimator defined in Section 3, and H_{ww^*} is as defined in Proposition 2. Then, from (4.1), (4.3), Theorem 3, and Proposition 2, it is easy to see that

$$\begin{aligned} E\left\{\widetilde{V}(\widehat{\overline{\tau}})\right\} &= \operatorname{var}(\widehat{\overline{\tau}}) + \Delta + (1/N^2)\sum_{w=1}^{W}\sum_{w^*(\neq w)=1}^{W}\left[b_{ww^*} + \{M_w M_{w^*}/(W-1)\}\right]\overline{\tau}_w\overline{\tau}_{w^*} \\ &= \operatorname{var}(\widehat{\overline{\tau}}) + \widetilde{\Delta}, \end{aligned}$$

where

$$\widetilde{\Delta} = (1/N^2)\sum_{w=1}^{W}\sum_{w^*=1}^{W} b_{ww^*}\overline{\tau}_w\overline{\tau}_{w^*}. \tag{4.4}$$

Clearly, if the matrix B is nonnegative definite, then the bias $\widetilde{\Delta}$ is nonnegative, making $\widetilde{V}(\widehat{\overline{\tau}})$ a conservative estimator of $\operatorname{var}(\widehat{\overline{\tau}})$. Furthermore, by (4.4), this bias vanishes if and only if $\overline{\tau}_1 = \cdots = \overline{\tau}_W$, when B has each row sum equal to zero, and is a positive semidefinite (psd) matrix of rank $W-1$. These facts are summarized in Theorem 4, which is the main result of this section.

Theorem 4. *Let there exist a psd matrix $B = ((b_{ww^*}))$ of order W and satisfying the following conditions: (c1) $b_{ww} = M_w^2$, for $w = 1,\ldots,W$, (c2) $\sum_{w^*=1}^{W} b_{ww^*} = 0$, for $w = 1,\ldots,W$, and (c3) rank$(B) = W-1$. Then, the variance estimator $\widetilde{V}(\widehat{\overline{\tau}})$ defined in (4.3) estimates $\operatorname{var}(\widehat{\overline{\tau}})$ with a nonnegative bias $\widetilde{\Delta}$ given by (4.4), which vanishes if and only if $\overline{\tau}_1 = \cdots = \overline{\tau}_W$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Remark 2. Recall that the between-WP additivity condition (2.13) is equivalent to $\overline{\tau}_1 = \cdots = \overline{\tau}_W$ for every treatment contrast. Thus, even when the WP sizes $M_1, \ldots, M_W$ are not all equal, by Theorem 4, the bias $\widetilde{\Delta}$ vanishes for every treatment contrast if and only if between-WP additivity holds. Thus, if a psd matrix B satisfying conditions (c1)–(c3) is available, then Theorem 4 provides us with a variance estimator that possesses properties similar to those derived by Zhao et al. (2018) for the balanced case. However, the issue of the existence of such a matrix turns out to be quite challenging, and is explored in the next section.

Remark 3. In Theorem 4, conditions (c1) and (c2) ensure the "if" part, while (c3) accounts for the "only if" part. To see the role of (c3) in some detail, from (4.4), note that $\widetilde{\Delta}$ vanishes. Hence, $\widetilde{V}(\widehat{\overline{\tau}})$ becomes unbiased for var$(\widehat{\overline{\tau}})$ whenever $(\overline{\tau}_1, \ldots, \overline{\tau}_W) \in \mathcal{R}^{\text{orth}}(B)$, where $\mathcal{R}^{\text{orth}}(B)$ is the orthocomplement of the row space of B. If rank(B) is allowed to be less than W–1 by dropping (c3), then $\mathcal{R}^{\text{orth}}(B)$ is broader than the space of vectors $(\overline{\tau}_1, \ldots, \overline{\tau}_W)$ that have all elements equal. Hence, $\widetilde{V}(\widehat{\overline{\tau}})$ becomes unbiased under a wider variety of situations than $\overline{\tau}_1 = \cdots = \overline{\tau}_W$ alone. However, this gain comes at a cost. Along the lines of Proposition 3 in Section 6, given (c1) and (c2), the largest eigenvalue of B is at least as

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

large as $\sum_{w=1}^{W} M_w^2/\text{rank}(B)$, and hence it can get considerably inflated if $\text{rank}(B) < W{-}1$. As discussed later in Section 6 and illustrated in Example 1, this, in turn, can significantly increase the bias of $\widetilde{V}(\widehat{\overline{\tau}})$ when $(\overline{\tau}_1, \ldots, \overline{\tau}_W) \notin \mathcal{R}^{\text{orth}}(B)$. This is why we retain condition (c3), which not only ensures the "only if" part of Theorem 4, but also helps in controlling the bias when the condition $\overline{\tau}_1 = \cdots = \overline{\tau}_W$ does not hold.

5. Existence and construction

We now study the existence of a psd matrix B satisfying conditions (c1)–(c3) stated in Theorem 4 as a purely mathematical problem. Without loss of generality, we assume hereafter that

$$M_1 \leq M_2 \leq \cdots \leq M_W. \tag{5.1}$$

To motivate the ideas, consider first the case $W = 3$, where conditions (c1) and (c2) determine B uniquely as

$$B = \begin{bmatrix} M_1^2 & (M_3^2 - M_1^2 - M_2^2)/2 & (M_2^2 - M_1^2 - M_3^2)/2 \\ (M_3^2 - M_2^2 - M_1^2)/2 & M_2^2 & (M_1^2 - M_2^2 - M_3^2)/2 \\ (M_2^2 - M_3^2 - M_1^2)/2 & (M_1^2 - M_3^2 - M_2^2)/2 & M_3^2 \end{bmatrix}. \tag{5.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

This matrix is also psd, and satisfies (c3) if and only if its principal minor, given by the first two rows and columns, is positive. A simplification of this condition and an application of (5.1) yields $M_3 < M_1 + M_2$ as the necessary and sufficient condition for B to satisfy (c1)-(c3). This construction of B for $W = 3$ raises the following questions with respect to the general case $W \geq 3$:

(a) Is the condition

$$M_W < M_1 + \cdots + M_{W-1}, \tag{5.3}$$

necessary and sufficient for the existence of a psd matrix B satisfying (c1)–(c3)?

(b) If so, then under (5.3), can one construct such a matrix B using an extension of the form in (5.2) to the general case?

Later in this section, Theorem 5 answers (a) in the affirmative. On the other hand, the question in (b) does not allow a conclusive answer. To see why, observe that the most obvious extension of (5.2) to general $W \geq 3$ is given by $B = ((b_{ww^*}))$, with

$$b_{ww} = M_w^2, w = 1, \ldots, W,$$
$$b_{ww^*} = \frac{M_1^2 + \cdots + M_W^2}{(W-1)(W-2)} - \frac{M_w^2 + M_{w^*}^2}{W-2}, \; w, w^* = 1, \ldots, W, \; w \neq w^*. \tag{5.4}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

The divisors in (5.4) ensure condition (c2) about zero row sums, and make it consistent with (5.2) when $W = 3$. The form (5.4) is also natural because, in keeping with $M_1^2, \ldots, M_W^2$ as the diagonal elements of B, it takes the off-diagonal elements as linear combinations of $M_1^2, \ldots, M_W^2$ in a systematic manner. However, unlike the case of $W = 3$, the matrix B given by (5.4) may not be psd for $W \geq 4$, even when condition (5.3) holds. For instance, if $W = 4$, then this condition holds for both the configurations $(M_1, \ldots, M_4) = (8, 8, 12, 12)$ and $(6, 6, 14, 14)$. The matrix B in (5.4) is psd of rank 3 ($= W - 1$) for the first configuration, but has a negative eigenvalue for the second.

The above discussion makes it clear that, in general, the task of obtaining a psd matrix B satisfying (c1)–(c3) under condition (5.3) can be far more complex than the form (5.2) arising for $W = 3$.Theorem 5 establishes condition (5.3) as a necessary and sufficient condition for the existence of such a matrix.

Theorem 5. *Let $W \geq 3$. Then, condition (5.3), that is, $M_W < M_1 + \ldots + M_{W-1}$, is necessary and sufficient for the existence of a psd matrix $B = ((b_{ww^*}))$ of order W and satisfying the conditions (c1) $b_{ww} = M_w^2, for\ w = 1, \ldots, W$, (c2) $\sum_{w^*=1}^{W} b_{ww^*} = 0, for\ w = 1, \ldots, W$, and (c3) rank$(B) =$*

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

$W-1$.

The sufficiency part of the proof of Theorem 5 leads to a construction procedure of the matrix B satisfying conditions (c1)–(c3). If $M_1 = \ldots = M_W (= M,\ \text{say})$, then one can simply take M^2 at each diagonal position of B, and $-M^2/(W-1)$ at each off-diagonal position. With regard to the case of unequal $M_1 \leq \ldots \leq M_W$, suppose condition (5.3) holds. Let $\mu = (M_1, \ldots, M_{W-1})'$, where the prime denotes transposition, and let e denote the $(W-1) \times 1$ vector of ones. Then, the steps involved in the construction of the matrix B are:

Step 1: Find a vector x with elements ± 1 satisfying the condition

$$|\mu' x| < M_W. \tag{5.5}$$

Step 2: Find nonnegative constants a_1 and a_2, satisfying $a_1 + a_2 < 1$, such that

$$a_1 \left\{ (\mu' x)^2 - \mu'\mu \right\} + a_2 \left\{ (\mu' e)^2 - \mu'\mu \right\} = M_W^2 - \mu'\mu. \tag{5.6}$$

Step 3: Construct the matrix $A = D \left\{ a_1 x x' + a_2 e e' + (1 - a_1 - a_2) I \right\} D$, where x, a_1, and a_2 are obtained from steps 1 and 2 above, I is the identity matrix of order $W-1$, and $D = \text{diag}(M_1, \ldots, M_{W-1})$.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Step 4: Construct matrix B as follows:

$$B = \begin{bmatrix} A & -Ae \\ -e'A & e'Ae \end{bmatrix}.$$

Then, B is psd of order W and satisfies (c1)–(c3) by the proof of the sufficiency part of Theorem 5. A lemma, crucial in this proof, appears in the Supplementary Material and guarantees the existence of the vector x in step 1 and the constants a_1 and a_2 in step 2 under condition (5.3).

Remark 4. It is satisfying that condition (5.3) holds under wide generality. It only requires that the largest WP not be too large compared to the others and holds, in particular, when there is a tie for the largest WP. Moreover, in some situations, it may be possible to adjust the composition of the WPs so as to meet (5.3) without significantly increasing the cost of the experiment. As an illustration, consider a variant of the school example of Section 1, where the 40 schools are spread over four counties with 8, 6, 6, and 20 schools, respectively, so that (5.3) does not hold. In this case, if the third and fourth counties are contagious and two schools of the fourth county are close to their border, then these two may be clubbed with the third county, leading to WPs of sizes 8, 6, 8, and 18 respectively, and ensuring (5.3) is met.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Remark 5. For $W = 3$, one can check that the construction stated above yields the unique B in (5.2). For $W \geq 4$, however, a psd matrix B meeting (c1)–(c3) is non-unique. Indeed, then the above construction can yield a wide class of such matrices B, considering all vectors x that satisfy (5.5), and for each such x, all nonnegative a_1, a_2 satisfying $a_1 + a_2 < 1$ and (5.6). Thus, the issue of discriminating between rival choices of B becomes important. Such a discriminating strategy is discussed in Section 6.

6. Minimax estimators unbiased under between-WP additivity

As seen in Section 5, while condition (5.3) guarantees the existence of the matrix B and, consequently, a variance estimator that is unbiased under between-WP additivity, such a matrix is non-unique. Thus, it is important to define a criterion that can discriminate between possible choices of B. Clearly, a good choice should control the bias $\tilde{\Delta} = (1/N^2) \sum_{w=1}^{W} \sum_{w^*=1}^{W} b_{ww^*} \overline{\tau}_w \overline{\tau}_{w^*}$ given by (4.4) that is associated with the estimation of $\mathrm{var}(\widehat{\overline{\tau}})$. The hurdle here is that $\overline{\tau}_1, \ldots, \overline{\tau}_W$ are unknown. Even the idea of minimaxity does not work without further refinement, because B is psd, and hence $\tilde{\Delta}$ is unbounded with respect to variation of $\overline{\tau}_1, \ldots, \overline{\tau}_W$ in the W-dimensional real space. On the other hand, by (3.3), multiplication of $\overline{\tau}_1, \ldots, \overline{\tau}_W$ by

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

any nonzero constant only rescales the treatment contrast $\overline{\tau}$, without essentially altering it. We, therefore, consider minimization of $\tilde{\Delta}$ subject to $\sum_{w=1}^{W} \overline{\tau}_w^2 = 1$. This is motivated by Mukerjee et al. (2018), who touched upon split-plot designs only in the balanced case. It is easy to see that the above formulation calls for obtaining B, subject to (c1)-(c3), so as to minimize $\lambda_{\max}(B)$, the largest eigenvalue of B. The following proposition provides us with a lower bound for $\lambda_{\max}(B)$.

Proposition 3. *For any psd matrix B satisfying (c1)–(c3), a lower bound for $\lambda_{\max}(B)$ is given by $\lambda_0 = \sum_{w=1}^{W} M_w^2/(W-1)$, but this bound is not sharp whenever $M_1, \ldots, M_W$ are not all equal.*

Given Proposition 3, an analytical solution to the minimaxity problem above seems to be intractable in the unbalanced case. This is anticipated, because a complete characterization of matrices B satisfying (c1)-(c3) is hard, even though in Section 5, we were able to outline a general method for constructing such matrices when condition (5.3) holds. As a practical strategy, therefore, it makes sense to concentrate on matrices B that can be obtained via this method, with a view to minimizing $\lambda_{\max}(B)$ among these matrices. It is reassuring that even then, the class of competing matrices B is quite large, as noted in Remark 5. In practice, these competing matrices

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

B can be generated quite fast by (i) enumerating vectors x with elements ± 1 and satisfying (5.5), and (ii) for each such x, performing a grid search to find nonnegative (a_1, a_2) satisfying $a_1 + a_2 < 1$ and (5.6).To implement (ii), we vary a_1 from 0 to 0.9999 in steps of 0.0001, such that the corresponding a_2, found from (5.6), meets $0 \leq a_2 < 1 - a_1$. Finding the matrix B that has the smallest $\lambda_{\max}$ among all candidate matrices identified through (i) and (ii) is fast and easy. This enumeration becomes even faster noting that, without loss of generality, the first element of x in (i) can be taken as $+1$, because replacing x by $-x$ does not affect Steps 1-4 of Section 5.

Example 1. Returning to the school example in Sections 1 and 2, where we have $N = 40$, $W = 4$, and $(M_1, M_2, M_3, M_4) = (8, 8, 12, 12)$, the smallest $\lambda_{\max}(B)$ obtainable using steps 1 through 4 described in Section 5 is 192, which corresponds to

$$B = \begin{bmatrix} 64 & 32 & -48 & -48 \\ 32 & 64 & -48 & -48 \\ -48 & -48 & 144 & -48 \\ -48 & -48 & -48 & 144 \end{bmatrix},$$

as given by $x = (1, 1, -1)'$, $a_1 = 0.5$, and $a_2 = 0$. Clearly, this B meets (c1)–(c3) of Theorem 4, and quite reassuringly, it has the smallest possible

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

$\lambda_{\max}(B)$ over all psd matrices B satisfying (c1) and (c2), as one can verify numerically. On the other hand, if B satisfies only (c1) and (c2), but not (c3), that is, $\text{rank}(B) < W-1 (= 3)$, then following Remark 3, $\lambda_{\max}(B)$ gets inflated. For example, if $\text{rank}(B) = 2$, then $\lambda_{\max}(B) \geq \sum_{w=1}^{W} M_w^2/2 = 208$. Further, if $\text{rank}(B) = 1$, then by (c1) and (c2), $B = qq'$, where q is equal to $(8, -8, -12, 12)'$ or $(8, -8, 12, -12)'$ and $\lambda_{\max}(B)$ becomes as large as 416.

7. Simulation results and performance comparisons

Whereas Theorem 4 establishes the unbiasedness of $\widetilde{V}(\widehat{\overline{\tau}})$ under (5.3) and between-WP additivity, and minimaxity is expected to provide protection under extreme departures from additivity, it is also important to understand how the bias of $\widetilde{V}(\widehat{\overline{\tau}})$ compares with that of $\widehat{V}(\widehat{\overline{\tau}})$ under different levels of treatment effect heterogeneity. In addition, the theoretical results do not provide any clue about whether the bias adjustment comes at the cost of an undesirable inflation of the MSE. We now conduct some simulations to study these two aspects. We consider the estimation of the main effect of factor F_2 in the setting of Example 1. The unit-level treatment contrast τ_i is equal to $\{-Y_i(00) + Y_i(01) - Y_i(10) + Y_i(11)\}/2$, for $i = 1, \ldots, 40$ (Dasgupta et al., 2015). The finite-population contrast of

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

interest is $\overline{\tau} = \sum_{i=1}^{40} \tau_i/40$. The vector of potential outcomes for unit i, denoted by $Y_i = (Y_i(00), Y_i(01), Y_i(10), Y_i(11))$, is generated using the multivariate normal model $Y_i \sim N_4(\theta_w, \Sigma_w)$, for $i \in \Omega_w$, $w = 1, \ldots, 4$, where $\Sigma_w = \sigma_w^2 \{(1-\rho_w)I_4 + \rho_w J_4\}$ is the covariance matrix for WP Ω_w that depends on two parameters: the variance σ_w^2 and the correlation ρ_w. Matrices I_n and J_n denote the nth-order identity matrix and the matrix of ones, respectively. Seven possible scenarios (listed in Table 2) for generating the potential outcomes are considered.

Table 2: Simulation settings

Population	θ_1	θ_2	θ_3	θ_4	σ_1^2	σ_2^2	σ_3^2	σ_4^2	ρ_1	ρ_2	ρ_3	ρ_4
I	(5,8,7,8)	(7,9,4,6)	(8,11,7,8)	(7,8,6,9)	2.5	2	2	3	.5	.5	.5	.5
II	(10,5,9,8)	(5,9,10,8)	(10,9,8,5)	(10,5,8,9)	2.5	2	2	3	1	1	1	1
III	(10,5,9,8)	(5,9,10,8)	(10,9,8,5)	(10,5,8,9)	2.5	2	2	3	.5	.5	.5	.5
IV	(10,5,9,8)	(5,9,10,8)	(10,9,8,5)	(10,5,8,9)	2.5	2	2	3	.2	.4	.6	.8
V	(10,5,9,8)	(5,9,10,8)	(10,9,8,5)	(10,5,8,9)	2.5	2	2	3	0	0	0	0
VI	(10,5,9,8)	(5,9,10,8)	(10,9,8,5)	(10,5,8,9)	2.5	2	2	3	-.3	-.3	-.3	-.3
VII	(10,5,9,8)	(5,9,10,8)	(10,9,8,5)	(10,5,8,9)	2.5	2	2	3	-.3	.3	-.3	.3

The potential outcomes for population I are forced to ensure, via an appropriate command in R, that the WP means $\overline{\tau}_1, \ldots, \overline{\tau}_4$ are always two. Population II generates different $\overline{\tau}_1, \ldots, \overline{\tau}_4$, but guarantees the same τ_i within each WP. Populations III through VII differ only with respect to the correlation parameters that lead to different types of treatment effect heterogeneity. These include all zero correlations in population V, all neg-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

ative correlations in population VI, and a mix of positive and negative correlations in population VII.

From each population, 25 sets of potential outcomes are generated, and the biases of the variance estimators $\widehat{V}(\widehat{\overline{\tau}})$ and $\widetilde{V}(\widehat{\overline{\tau}})$ are compared. Note that these biases are Δ given by (3.4) and $\tilde{\Delta}$ given by (4.4), respectively. Because we do not have any theoretical expressions for the MSEs of $\widehat{V}(\widehat{\overline{\tau}})$ and $\widetilde{V}(\widehat{\overline{\tau}})$, we estimate them empirically. For each set of potential outcomes, we generate 10000 treatment assignments, and thus 10000 sets of observed outcomes, calculate $\widehat{V}(\widehat{\overline{\tau}})$ and $\widetilde{V}(\widehat{\overline{\tau}})$ for each set, and estimate their MSEs from these 10000 values. The results are summarized in Table 3.

Table 3: Median bias and median MSE of the variance estimators from 25 sets of potential outcomes under seven settings

Population	$\widehat{V}(\widehat{\overline{\tau}})$		$\widetilde{V}(\widehat{\overline{\tau}})$	
	Bias	MSE	Bias	MSE
I	0.0533	0.2317	0.0000	0.2613
II	0.6400	5.9011	0.2800	5.8343
III	0.6403	5.7001	0.2740	5.5989
IV	0.7096	5.5829	0.3306	5.8072
V	0.7284	6.1943	0.3285	5.8122
VI	0.7667	5.2141	0.3719	5.3680
VII	0.5921	5.2011	0.2915	5.5709

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

The results suggest that with respect to bias, the variance estimator $\widetilde{V}(\widehat{\overline{\tau}})$ outperforms the estimator $\widehat{V}(\widehat{\overline{\tau}})$ uniformly across all seven populations. As expected, the bias of $\widetilde{V}(\widehat{\overline{\tau}})$ is zero for population I. However, with respect to the MSE, there appears to be no clear winner. The estimator $\widetilde{V}(\widehat{\overline{\tau}})$ performs slightly better than $\widehat{V}(\widehat{\overline{\tau}})$ in populations II, III, and V, and slightly worse in the other populations with respect to the MSE. However, it is encouraging to note that the bias reduction associated with the estimator $\widetilde{V}(\widehat{\overline{\tau}})$ does not appear to come at the cost of a significant sacrifice of MSE.

8. Discussion

We have seen that the attempt to generalize Neymanian variance estimation from balanced to unbalanced split-plot designs is a highly nontrivial problem. The results from unbalanced block designs cannot be used in such a setting. Following an intricate chain of arguments, it is possible to derive a variance estimator that is unbiased under assumptions of treatment homogeneity, similar to that in the balanced case, and is also robust under departures from such an assumption. Our empirical results suggest that this bias reduction does not come at the cose of an undesirably high in-

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

flation of the MSE. It would be interesting to examine the performance of the two estimators considered in this paper with respect to their coverage of asymptotic confidence intervals for the true contrast $\overline{\tau}$. Such a comparison will, however, entail considerable theoretical work to establish the asymptotic distribution of $\widehat{\overline{\tau}}$, because of the complex setting of the unbalanced split-plot assignment mechanism. We plan to pursue this in future research.

Supplementary Material

The online Supplementary Materials contains the proofs of all results stated in the main manuscript.

Acknowledgements

This work was supported by the J.C. Bose National Fellowship, Government of India, and grants from the Indian Institute of Management Calcutta and National Science Foundation, USA. We are grateful to the two reviewers for their insightful comments.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica 88*, 265–296.

Box, G. E. P., J. S. Hunter, and W. G. Hunter (2005). *Statistics for Experimenters: Design, Innovation, and Discovery* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.

Cochran, W. G. (1977). *Sampling Techniques.* John Wiley & Sons: New York.

Cochran, W. G. and G. M. Cox (1957). *Experimental Designs* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.

Dasgupta, T., N. S. Pillai, and D. B. Rubin (2015). Causal inference for 2^K factorial designs by using potential outcomes. *Journal of the Royal Statistical Society, Series B 77*, 727–753.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh, Scotland: Oliver & Boyd.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Fisher, R. A. (1935). *The Design of Experiments* (1st ed.). Oxford, England: Oliver & Boyd.

Freedman, D. A. (2006). Statistical models for causation: What inferential leverage do they provide? *Evaluation Review 30*, 691–713.

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

Kirk, R. E. (1982). *Experimental Design: Procedures for the Behavioral Sciences*. Brooks/Cole, Monterey, CA.

Mukerjee, R., T. Dasgupta, and D. B. Rubin (2018). Using standard tools from finite population sampling to improve causal inference for complex experiments. *Journal of the American Statistical Association 113*, 868–881.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science 5*, 465–472.

Olsen, R., L. Orr, S. Bell, and E. Stuart (2013). External validity in policy evaluations that choose sites purposively. *Journal of Policy Analysis and Management 32*, 107–121.

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

Pashley, N. and L. Miratrix (2019). Insights on variance estimation for blocked and matched pairs designs. *arXiv:1710.10342v5 [stat.ME]*.

Rosenberger, W. F., D. Uschner, and Y. Wang (2019). Randomization: The forgotten component of the randomized clinical trial. *Statistics in medicine 38*, 1–12.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*, 688–701.

Rubin, D. B. (1980). Comment on "Randomization analysis of experimental data: The Fisher randomization test by D. Basu". *Journal of the American Statistical Association 75*, 591–593.

Schochet, P. S., N. Pashley, L. Miratrix, and T. Kautz (2020). Design-based ratio estimators for clustered, blocked rcts. *arXiv:2002.01146v1 [stat.ME]*.

Wu, C. F. J. and M. S. Hamada (2009). *Experiments: Planning, Analysis, and Optimization* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202020.0149

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

Indian Institute of Management Calcutta, Joka, Diamond Harbour Road, Kolkata 700104, India

E-mail: rmuk0902@gmail.com

Department of Statistics, Rutgers University, 110 Frelinghuysen Road, Piscataway, New Jersey 08901, U.S.A.

E-mail: tirthankar.dasgupta@rutgers.edu

Statistica Sinica