Efficient switchback experiments via multiple randomization designs

Lorenzo Masoero   Guido Imbens   Thomas Richardson   James McQueen   Suhas Vijaykumar   Ido Rosen

Amazon.com

November 1, 2023

Introduction Online A/B tests have become an indispensable tool across all the technology industry: if performed correctly, "online" experiments can inform effective decision making and product development. It should therefore not be surprising that Gupta et al. [2019] estimates that online businesses alone collectively run hundreds of thousands of experiments annually.

Modern online experiments are often run in marketplaces where multiple populations of units (e.g., buyers and sellers) with competing interests and strategic responses interact and dynamically adapt their behavior to the treatment over time. Despite the modern setting in which online randomized experiments are performed, the industry still heavily relies on assumptions and corresponding designs that closely resemble those of classical randomized experiments dating back to Neyman [1923/1990] and Fisher [1937]. A natural concern in these settings is that the presence of cross-unit interference (spillovers) might invalidate the analysis. To address these shortcomings, a rapidly growing literature on experimental design in settings with interference or spillovers has been developed over the last few decades [Hudgens and Halloran, 2008, Rosenbaum, 2007, Aronow, 2012, VanderWeele et al., 2014, Ogburn and VanderWeele, 2014, Aronow and Samii, 2017]. In this paper we show that even in settings where interference is absent, multiple randomization designs can lead to greater efficiency to estimate causal effects.

Our contribution Among other proposals, recently Bajari et al. [2021, 2023], Johari et al. [2022] introduced a rich novel class of experimental designs — "multiple randomization designs" (in short, MRDs). MRDs allow estimation of complex spillover effects by suitably randomizing across the collection of populations involved in the experiment. In this work, we show how estimators derived from MRDs can be more efficient than standard ones, even in the absence of any interference. In particular, we consider a dynamic context in which the experimenter has the ability to run a switchback experiment, whereby treatment exposure of units is allowed to change over time. Via the lens of MRDs, we show how a switchback design can be provably more efficient than a standard experiment, even in the absence of cross-unit interference. Through our analysis we provide, to the best of our knowledge, the first finite-sample randomization-based inference results for crossover designs. Our work complements corresponding recent model-based results in Bojinov et al. [2020], Xiong et al. [2019], and randomization-based results for a simple two-period case in Shi and Ye [2023].

Methodology Consider a binary randomized experiment where the experimenter can expose experimental units to different treatments over time, and record the corresponding (potential) outcome at each time stamp. Such designs have been referred to as rotation [Cochran, 1939], crossover [Brown Jr, 1980], or switchback experiments [Bojinov et al., 2020]. We here consider randomization-based inference for such crossover designs in the absence of interference. We consider the case in which we have measurements for N individuals over S periods of time. An example of such an experiment with $N = 6$ units and $S = 8$ time periods is given in assignment matrix (1):

Time Period
Individual Unit ↓ →

$$\mathbf{W} = \begin{array}{c} 1\\2\\3\\4\\5\\6 \end{array} \overset{\begin{array}{cccccccc}1&2&3&4&5&6&7&8\end{array}}{\begin{pmatrix} T & T & C & C & C & T & C & T \\ T & T & T & C & T & C & C & C \\ C & C & C & T & T & T & C & T \\ C & C & T & C & C & T & T & T \\ C & T & C & T & T & C & T & C \\ T & C & T & T & C & C & T & C \end{pmatrix}} . \tag{1}$$

In our development we consider a balanced design, where each unit is in the treatment group for four periods, and in every period exactly three units are in the treatment group (like in assignment matrix (1)). The experimental designs we consider are represented via an assignment mechanism that randomizes uniformly over the space $\mathfrak{W}$ of $N \times S$ binary matrices with constant row and column sum:

$$\mathfrak{W} := \mathfrak{W}(N, S, p) = \left\{ W \in \{0,1\}^{N \times S} : \sum_{s=1}^{S} W_{n,s} = pS,\ \forall n \wedge \sum_{n=1}^{N} W_{n,s} = pN,\ \forall s \right\}. \tag{2}$$

Here $W_{n,s} = 1$ if unit n is exposed to treatment at time step s, and $W_{n,s} = 0$ if unit n is exposed to control at time step s. Under the potential outcome framework without cross-unit interference, for any given unit-timestamp pair (n, s) there exist exactly two potential outcomes, $y_{n,s}(1)$ and $y_{n,s}(0)$. The population average outcome for treatment $w \in \{0, 1\}$ is given by $\overline{\overline{y}}_w := \frac{1}{NS}\sum_{n=1}^{N}\sum_{s=1}^{S} y_{n,s}(w)$. In turn the average treatment effect τ is obtained via $\tau := \frac{1}{NS}\sum_{n=1}^{N}\sum_{s=1}^{S}\{y_{n,s}(1) - y_{n,s}(0)\} = \overline{\overline{y}}_1 - \overline{\overline{y}}_0$. Notice that these population quantities are not directly observable. Given a (random) assignment matrix $W \in \mathfrak{W}$ and $p_w = p1(w = 1) + (1 - p)1(w = 0)$, we define the plug-in estimator for $\overline{\overline{y}}_w$:

$$\widehat{\overline{\overline{Y}}}_w := \frac{\sum_{n,s} 1(W_{n,s} = w)y_{n,s}(w)}{\sum_{n,s} 1(W_{n,s} = w)} = \frac{1}{p_w NS}\sum_{n=1}^{N}\sum_{s=1}^{S} 1(W_{n,s} = w)y_{n,s}(w). \tag{3}$$

In turn the estimator $\hat{\tau} = \widehat{\overline{\overline{Y}}}_1 - \widehat{\overline{\overline{Y}}}_0$ for τ can be derived.

A low rank, parametric assumption on potential outcomes The potential efficiency gains of crossover designs in the absence of interference across units or time can be most easily seen in a parametric framework. Suppose $y_{n,s}(0) = \mu + a_n + b_s + \epsilon_{n,s}$, where $a_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2)$, $b_s \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_b^2)$, $\epsilon_{n,s} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$, and $y_{n,s}(1) = Y_{n,s}(0) + \tau$. It follows from the properties of the Gaussian distribution that, letting $\sigma^2 := \sigma_a^2 + \sigma_b^2 + \sigma_\epsilon^2$:

$$y_{n,s}(w) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu + \tau 1(w = 1), \sigma^2). \tag{4}$$

Balanced crossover designs Under the parametric Gaussian assumption of Equation (31), the estimator $\widehat{\overline{\overline{Y}}}_w$ is unbiased for $\overline{\overline{y}}_w$. The expected value of the variance $\mathrm{Var}_{\mathfrak{W}}\left(\widehat{\overline{\overline{Y}}}_w\right)$ over hypothetical re-draws of the data is given by:

$$\mathbb{E}_{\mathcal{N}}\left\{\mathrm{Var}_{\mathfrak{W}}\left(\widehat{\overline{\overline{Y}}}_w\right)\right\} = \frac{\xi}{p^2 NS}\left\{\sigma_\epsilon^2\left(1 - \frac{1}{N} - \frac{1}{S}\right) - \frac{\sigma_a^2}{N} - \frac{\sigma_b^2}{S}\right\} = O\left(\frac{\sigma_\epsilon^2}{NS}\right). \tag{5}$$

Here $\xi := \mathbb{E}[W_{1,1}^2] - \mathbb{E}[W_{1,1}W_{1,2}] - \mathbb{E}[W_{1,1}W_{2,1}] + \mathbb{E}[W_{1,1}W_{2,2}]$, and the expectation is over draws from $\mathfrak{W}$ in Equation (2).

We next show that the estimator $\widehat{\overline{\overline{Y}}}_w$ that one would obtain by considering natural "standard" designs by either randomizing completely at random over each (unit, time) pair, or just over units, or over timestamps, are less efficient than a balanced switchback design. In turn, this implies that the corresponding plug in estimate for $\hat{\tau}$ is less efficient.

Completely randomized designs In this case, the matrix of assignments is a collection of i.i.d. coin flips, each with probability p. The estimator for the population mean we consider has the same form as in Equation (9); however, the space of assignments $\mathfrak{W}^{\text{CRD}}$ and the weighting distribution is different. We hence use the notation $\widehat{\overline{\overline{Y}}}_{w,\text{CRD}}$ to emphasize that this estimator is obtained from a completely randomized design. $\widehat{\overline{\overline{Y}}}_{w,\text{CRD}}$ is unbiased for $\overline{\overline{Y}}_w$, and has variance:

$$\mathbb{E}_{\mathcal{N}}\left\{\mathrm{Var}_{\mathfrak{W}^{\text{CRD}}}\left(\widehat{\overline{\overline{Y}}}_{w,\text{CRD}}\right)\right\} = \left(\frac{1 - p_w}{p_w}\right)\frac{\sigma^2}{NS} = O\left(\frac{\sigma^2}{NS}\right) \tag{6}$$

Individual-randomized designs In this case, our experiment is equivalent to a simple single randomized design with N units, in which each unit has potential outcome $y_n(w) = S^{-1}\sum_{s=1}^{S} y_{n,s}(w) = \overline{y}_n^{\text{N}}(w)$. While the estimator for the population mean has (again) the same form as in Equation (9), we use $\widehat{\overline{\overline{Y}}}_{w,\rightarrow}$ to denote it, to emphasize the dependence on the space of valid assignments $\mathfrak{W}^{\rightarrow}$ (the space of binary matrices with constant rows, and such that exactly pN rows are equal to one). $\widehat{\overline{\overline{Y}}}_{w,\rightarrow}$ is unbiased for $\overline{\overline{y}}_w$ and has variance:

$$\mathbb{E}_{\mathfrak{W}^{\text{N}}}\left[\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\rightarrow}\right)\right] = \frac{1 - p_w}{p_w}\frac{1}{N - 1}\left(\sigma_a^2 + \frac{\sigma_b^2 + \sigma_\epsilon^2}{S} - 2\left[\frac{\sigma_a^2}{N} + \frac{\sigma_b^2}{S} + \frac{\sigma_\epsilon^2}{NS}\right]\right) = O\left(\frac{\sigma_a^2}{N} + \frac{\sigma_\epsilon^2}{NS}\right).$$

Time-randomized designs Symmetrically to the case above, here our experiment is equivalent to a simple single randomized design with S units, in which each unit has potential outcome $y_s(w) = S^{-1}\sum_{n=1}^{N} y_{n,s}(w) = \overline{y}^{\mathrm{S}}_s(w)$. While the estimator for the population mean has (again) the same form as in Equation (9), we use $\widehat{\overline{\overline{Y}}}_{w,\downarrow}$ to denote it, to emphasize the dependence on the space of valid assignments $\mathfrak{W}^{\downarrow}$ (the space of binary matrices with constant columns, and such that exactly pS columns are equal to one). $\widehat{\overline{\overline{Y}}}_{w,\downarrow}$ is unbiased for $\overline{\overline{y}}_w$ and has variance

$$\mathbb{E}_{\mathfrak{W}^{\mathrm{S}}}\left[\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\downarrow}\right)\right] = \frac{1-p_w}{p_w}\frac{1}{S-1}\left(\sigma_b^2 + \frac{\sigma_a^2+\sigma_\epsilon^2}{N} - 2\left[\frac{\sigma_b^2}{S} + \frac{\sigma_a^2}{N} + \frac{\sigma_\epsilon^2}{NS}\right]\right) = O\left(\frac{\sigma_b^2}{S} + \frac{\sigma_\epsilon^2}{NS}\right).$$

Conclusion From our derivation, we see how the expected variance of the simple estimator $\widehat{\overline{\overline{Y}}}_w$ obtained from a switchback design is smaller than the corresponding variance incurred by adopting other competing designs (completely randomized design, time-randomized, unit-randomized). In particular, this is because — by virtue of the balancing — the effect of the variance components σ_a^2 and σ_b^2 is of second order — whereas these terms appear as leading terms in the other variances. In turn, this directly translates into more precise estimates of the average treatment effect. While we have here focused on the case in which the potential outcomes satisfy the assumption of Equation (31), we expect a similar result to hold beyond this parametric specification.

In this discussion we have focused on the case with no spillovers. In practice many crossover designs are motivated by concerns that treatments have spillovers or carryover effects over time. If the experimenter is confident these carryover effects last only S periods, one may limit changes in treatment status to a set of fixed time points, say every $2S$ periods, and use only outcomes in the periods after carryover effects have been exhausted for estimation. See Bojinov et al. [2020], Shi and Ye [2023].

References

Somit Gupta, Ronny Kohavi, Diane Tang, Ya Xu, Reid Andersen, Eytan Bakshy, Niall Cardin, Sumita Chandran, Nanyu Chen, Dominic Coey, et al. Top challenges from the first practical online controlled experiments summit. *ACM SIGKDD Explorations Newsletter*, 21(1):20–35, 2019.

Jerzey Neyman. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 5(4):465–472, 1923/1990.

Ronald Aylmer Fisher. *The design of experiments*. Oliver And Boyd; Edinburgh; London, 1937.

Michael Hudgens and Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, pages 832–842, 2008.

Paul R. Rosenbaum. Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200, 2007. ISSN 01621459.

Peter M Aronow. A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research*, 41(1):3–16, 2012.

Tyler J VanderWeele, Eric J Tchetgen Tchetgen, and M Elizabeth Halloran. Interference and sensitivity analysis. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 29(4):687, 2014.

Elizabeth L Ogburn and Tyler J VanderWeele. Causal diagrams for interference. *Statistical science*, 29(4):559–578, 2014.

Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.

Patrick Bajari, Brian Burdick, Guido W Imbens, Lorenzo Masoero, James McQueen, Thomas Richardson, and Ido M Rosen. Multiple randomization designs. *arXiv preprint arXiv:2112.13495*, 2021.

Patrick Bajari, Brian Burdick, Guido Imbens, Lorenzo Masoero, James McQueen, Thomas Richardson, and Ido Rosen. Experimental design in marketplaces. *Statistical Science*, 2023.

Ramesh Johari, Hannah Li, Inessa Liskovich, and Gabriel Y Weintraub. Experimental design in two-sided platforms: An analysis of bias. *Management Science*, 2022.

Iavor Bojinov, David Simchi-Levi, and Jinglong Zhao. Design and analysis of switchback experiments. *Available at SSRN 3684168*, 2020.

Ruoxuan Xiong, Susan Athey, Mohsen Bayati, and Guido W Imbens. Optimal experimental design for staggered rollouts. *Available at SSRN*, 2019.

Danni Shi and Ting Ye. Behavioral carry-over effect and power consideration in crossover trials. *arXiv preprint arXiv:2302.01246*, 2023.

WG Cochran. Long-term agricultural experiments. *Supplement to the Journal of the Royal Statistical Society*, 6(2): 104–148, 1939.

Byron Wm Brown Jr. The crossover experiment for clinical trials. *Biometrics*, pages 69–79, 1980.

Ben Johnsen and Eldar Straume. Counting binary matrices with given row and column sums. *Mathematics of computation*, 48(178):737–750, 1987.

Persi Diaconis and Anil Gangolli. *Rectangular arrays with fixed margins*. Springer, 1995.

Jeffrey W. Miller and Matthew T. Harrison. Exact sampling and counting for fixed-margin matrices. *The Annals of Statistics*, 41(3):1569 – 1592, 2013a. doi: 10.1214/13-AOS1131. URL `https://doi.org/10.1214/13-AOS1131`.

Alexander Barvinok. Matrices with prescribed row and column sums. *Linear Algebra and its Applications*, 436(4): 820–844, 2012.

Giovanni Strona, Domenico Nappo, Francesco Boccacci, Simone Fattorini, and Jesus San-Miguel-Ayanz. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature communications*, 5(1):4114, 2014a.

William G Cochran. *Sampling Techniques*. John Wiley & Sons, New York, third edition, September 1977.

Michael M Strand. Estimation of a population total under a "bernoulli sampling" procedure. *The American Statistician*, 33(2):81–84, 1979.

Giovanni Strona, Domenico Nappo, Francesco Boccacci, Simone Fattorini, and Jesus San-Miguel-Ayanz. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature communications*, 5(1):4114, 2014b.

Jeffrey W. Miller and Matthew T. Harrison. Exact sampling and counting for fixed-margin matrices. *The Annals of Statistics*, 41(3):1569 – 1592, 2013b. doi: 10.1214/13-AOS1131. URL `https://doi.org/10.1214/13-AOS1131`.

A Estimation and Inference for Crossover Designs

In this section, we discuss randomized-based inference for crossover designs in the absence of interference. We provide exact finite sample results for balanced crossover designs as the ones in assignment matrix (1). We show that these balanced designs can lead to more efficient estimates than the ones obtained using SRDs, even in the absence of interference. We here consider the case in which we have measurements for N individuals over S periods of time. The assignment mechanism we consider randomizes uniformly over the space $\mathfrak{W}$ of $N \times S$ binary matrices with constant row-wise and column-wise sum:

$$\mathfrak{W} := \mathfrak{W}(N,S,p) = \left\{ W \in \{0,1\}^{N\times S} : \sum_{s=1}^{S} W_{n,s} = pS,\ \forall n \wedge \sum_{n=1}^{N} W_{n,s} = pN,\ \forall s \right\}.$$

Remark 1. *The problem of sampling uniformly from $\mathfrak{W}$ has a long history in the combinatorics and statistics literature. See, e.g., Johnsen and Straume [1987], Diaconis and Gangolli [1995], Miller and Harrison [2013a], Barvinok [2012]. In our simulations in Appendix C, we use the fastball algorithm [Strona et al., 2014a] to accomplish this goal.*

We assume here no interference: for any given unit (n,s) there exist exactly two potential outcomes, $y_{n,s}(1)$ and $y_{n,s}(0)$. The average outcome for treatment $w \in \{0,1\}$, that would be observed if every unit were assigned to treatment w at every time step, is given by:

$$\overline{\overline{y}}_w := \frac{1}{NS} \sum_{n=1}^{N} \sum_{s=1}^{S} y_{n,s}(w). \tag{7}$$

The estimand of interest is the average treatment effect

$$\tau := \frac{1}{NS} \sum_{n=1}^{N} \sum_{s=1}^{S} \{y_{n,s}(1) - y_{n,s}(0)\} = \overline{\overline{y}}_1 - \overline{\overline{y}}_0. \tag{8}$$

Given a (random) assignment matrix $W \in \mathfrak{W}$, we define the plug-in estimator for $\overline{\overline{y}}_w$:

$$\widehat{\overline{\overline{Y}}}_w := \frac{\sum_{n,s} 1(W_{n,s} = w) y_{n,s}(w)}{\sum_{n,s} 1(W_{n,s} = w)} = \frac{1}{p_w NS} \sum_{n=1}^{N} \sum_{s=1}^{S} 1(W_{n,s} = w) y_{n,s}(w), \tag{9}$$

where $p_w = p1(w=1) + (1-p)1(w=0)$. In order to characterize the variance of $\widehat{\overline{\overline{Y}}}_w$ we define the row and column "population" averages of the potential outcomes:

$$\overline{y}_n^{\mathrm{N}}(w) := \frac{1}{S} \sum_{s=1}^{S} y_{n,s}(w) \quad \text{and} \quad \overline{y}_s^{\mathrm{S}}(w) := \frac{1}{N} \sum_{n=1}^{N} y_{n,s}(w).$$

and their sample estimator counterparts:

$$\widehat{\overline{Y}}_n^{\mathrm{N}}(w) := \frac{1}{p_w S} \sum_{s: W_{n,s} = w} y_{n,s}(w) \quad \text{and} \quad \widehat{\overline{Y}}_s^{\mathrm{S}}(w) := \frac{1}{p_w N} \sum_{n: W_{n,s} = w} y_{n,s}(w).$$

Lemma 1. *Without any interference, the estimator $\widehat{\overline{\overline{Y}}}_w$ is unbiased for $\overline{\overline{y}}_w$ and has variance*

$$\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_w\right) = \xi_w \frac{\delta(w)^2}{N^2 S^2}. \tag{10}$$

Here, letting $\tilde{W}_{n,s}^{(w)} = 1(W_{n,s} = w)$ be the matrix identifying units in treatment w,

$$\xi_w := \mathbb{E}[(\tilde{W}_{1,1}^{(w)})^2] - \mathbb{E}[\tilde{W}_{1,1}^{(w)} \tilde{W}_{1,2}^{(w)}] - \mathbb{E}[\tilde{W}_{1,1}^{(w)} \tilde{W}_{2,1}^{(w)}] + \mathbb{E}[\tilde{W}_{1,1}^{(w)} \tilde{W}_{2,2}^{(w)}],$$

and $\delta(w)^2 := \sum_{n=1}^{N} \sum_{s=1}^{S} (\delta_{n,s}^{\mathrm{NS}}(w))^2$ is the sum of squared residuals, where

$$\delta_{n,s}^{\mathrm{NS}}(w) := y_{n,s}(w) - \overline{y}_n^{\mathrm{N}}(w) - \overline{y}_s^{\mathrm{S}}(w) + \overline{\overline{y}}_w.$$

Remark 2. *The potential efficiency gains of balanced crossover designs can be most easily seen in a parametric framework. Suppose $Y_{n,s}(0) = \mu + a_n + b_s + \epsilon_{n,s}$, where $a_n \sim \mathcal{N}(0,\sigma_a^2)$, $b_s \sim \mathcal{N}(0,\sigma_b^2)$, $\epsilon_{n,s} \sim \mathcal{N}(0,\sigma_\epsilon^2)$, and $Y_{n,s}(1) = Y_{n,s}(0) + \tau$. We show in Appendix C.2 that the balanced crossover design is (in expectation) provably more efficient than either a (i) completely randomized design, a (ii) individual-randomized experiment, and a (iii) time-randomized experiment, where at a given time step, all individuals have the same experience, and time steps are randomly assigned to treatment.*

To *estimate* the variance in Equation (10), we introduce unbiased estimators of the variance of the average outcomes along rows and columns of the matrix of outcomes:

$$\widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}^{\mathrm{N}}_n(w)\right) := \frac{1-p_w}{p_w^2}\frac{1}{S(S-1)}\sum_{s:W_{n,s}=w}\left[\left(y_{n,s}(w)-\widehat{\overline{Y}}^{\mathrm{N}}_n(w)\right)^2\right],$$

and

$$\widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}^{\mathrm{S}}_s(w)\right) := \frac{1-p_w}{p_w^2}\frac{1}{N(N-1)}\sum_{n:W_{n,s}=w}\left[\left(y_{n,s}(w)-\widehat{\overline{Y}}^{\mathrm{S}}_s(w)\right)^2\right].$$

Under simple random sampling from $\mathfrak{W}$, as showed in Lemmas 10 and 11,

$$\mathbb{E}\left[\widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}^{\mathrm{N}}_n(w)\right)\right] = \mathrm{Var}\left(\widehat{\overline{Y}}^{\mathrm{N}}_n(w)\right) \quad \text{and} \quad \mathbb{E}\left[\widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}^{\mathrm{S}}_s(w)\right)\right] = \mathrm{Var}\left(\widehat{\overline{Y}}^{\mathrm{S}}_s(w)\right).$$

We can now define an unbiased estimator of the variance term in Equation (10).

Lemma 2. *Define the estimator*

$$\widehat{\mathrm{Var}}\left(\widehat{\overline{\overline{Y}}}_w\right) := \frac{\frac{\xi_w}{p_w^2 NS}}{1+\frac{\xi_w}{p_w^2 NS}}\left\{\frac{\sum_{(n,s):W_{n,s}=w}\left[y_{n,s}(w)^2\right]}{p_w NS} - \frac{1}{N}\sum_{n=1}^{N}\left[\left(\widehat{\overline{Y}}^{\mathrm{N}}_n(w)\right)^2 - \widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}^{\mathrm{N}}_n(w)\right)\right]\right.$$
$$\left. - \frac{1}{S}\sum_{s=1}^{S}\left[\left(\widehat{\overline{Y}}^{\mathrm{S}}_s(w)\right)^2 - \widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}^{\mathrm{S}}_s(w)\right)\right] + \widehat{\overline{\overline{Y}}}^2_w\right\}. \tag{11}$$

It holds $\mathbb{E}\left[\widehat{\mathrm{Var}}\left(\widehat{\overline{\overline{Y}}}_w\right)\right] = \mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_w\right)$.

We then have an unbiased estimator for the average treatment effect τ defined in Equation (8) via the simple plug-in estimator

$$\hat{\tau} := \widehat{\overline{\overline{Y}}}_1 - \widehat{\overline{\overline{Y}}}_0. \tag{12}$$

Its variance can be lower and upper bounded via Cauchy-Schwarz leveraging the result in Lemma 2, so that we obtain a conservative estimator of the variance of $\hat{\tau}$ as follows:

$$\widehat{\mathrm{Var}}^{\mathrm{hi}}(\hat{\tau}) \leq \widehat{\mathrm{Var}}\left(\widehat{\overline{\overline{Y}}}_0\right) + \widehat{\mathrm{Var}}\left(\widehat{\overline{\overline{Y}}}_1\right) + 2\sqrt{\widehat{\mathrm{Var}}\left(\widehat{\overline{\overline{Y}}}_0\right)\widehat{\mathrm{Var}}\left(\widehat{\overline{\overline{Y}}}_1\right)}.$$

Remark 3. *In this discussion we focus on the case with no spillovers. In practice many crossover designs are motivated by concerns that treatments have spillovers or carryover effects over time. If the experimenter is confident these carryover effects last only S periods, one may limit changes in treatment status to a set of fixed time points, say every $2S$ periods, and use only outcomes in the periods after carryover effects have been exhausted for estimation. See Bojinov et al. [2020], Shi and Ye [2023].*

B Proofs for Crossover Designs

B.1 Moment characterization of the simple average

We consider the potential outcomes framework, with no interference. In what follows, we consider outcomes for the treated group only. Analogous derivations apply for the control group and are omitted.

The average outcome (population mean) for the treated is given by:

$$\overline{\overline{y}}_{\mathrm{tr}} := \frac{1}{NS}\sum_{n=1}^{N}\sum_{s=1}^{S} y_{n,s}(\mathrm{tr}). \tag{13}$$

Given a (random) assignment matrix $W \in \mathfrak{W}$, we define the (unadjusted) estimator for the treated as follows:

$$\widehat{\overline{\overline{Y}}}_{\mathrm{tr}} := \frac{\sum_{n,s} 1(W_{n,s}=1)y_{n,s}(\mathrm{tr})}{\sum_{n=1}^{N}\sum_{s=1}^{S} 1(W_{n,s}=1)} = (pNS)^{-1}\sum_{n=1}^{N}\sum_{s=1}^{S} 1(W_{n,s}=1)y_{n,s}(\mathrm{tr}). \tag{14}$$

We now characterize the moments of $\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}$. To make progress, we consider an alternative representation of the estimator in Equation (14). Define the following axis-deviations. For the units

$$\delta_n^{\mathrm{N}}(\mathrm{tr}) := \frac{1}{S}\sum_{s=1}^{S}\left(y_{n,s}(\mathrm{tr}) - \overline{\overline{y}}_{\mathrm{tr}}\right) = \overline{y}_n^{\mathrm{N}}(\mathrm{tr}) - \overline{\overline{y}}_{\mathrm{tr}}.$$

For the time stamps:

$$\delta_s^{\mathrm{S}}(\mathrm{tr}) := \frac{1}{N}\sum_{n=1}^{N}\left(y_{n,s}(\mathrm{tr}) - \overline{\overline{y}}_{\mathrm{tr}}\right) = \overline{y}_s^{\mathrm{S}}(\mathrm{tr}) - \overline{\overline{y}}_{\mathrm{tr}}.$$

Last, for the interactions

$$\delta_{n,s}^{\mathrm{NS}}(\mathrm{tr}) := y_{n,s}(\mathrm{tr}) - \overline{y}_n^{\mathrm{N}}(\mathrm{tr}) - \overline{y}_s^{\mathrm{S}}(\mathrm{tr}) + \overline{\overline{y}}_{\mathrm{tr}}.$$

By construction:

$$\sum_{n=1}^{N}\delta_n^{\mathrm{N}}(\mathrm{tr}) = 0, \qquad \sum_{n=1}^{N}\delta_{n,s}^{\mathrm{NS}}(\mathrm{tr}) = 0.$$

Moreover,

$$\sum_{s=1}^{S}\delta_s^{\mathrm{S}}(\mathrm{tr}) = 0, \qquad \sum_{s=1}^{S}\delta_{n,s}^{\mathrm{NS}}(\mathrm{tr}) = 0.$$

Lemma 3 (Decomposition of potential outcomes)**.** *The potential outcome $y_{n,s}(\mathrm{tr})$ can be decomposed as follows:*

$$y_{n,s}(\mathrm{tr}) = \overline{\overline{y}}_{\mathrm{tr}} + \delta_n^{\mathrm{N}}(\mathrm{tr}) + \delta_s^{\mathrm{S}}(\mathrm{tr}) + \delta_{n,s}^{\mathrm{NS}}(\mathrm{tr}). \tag{15}$$

Proof. It holds

$$\begin{aligned}\overline{\overline{y}}_{\mathrm{tr}} + \delta_n^{\mathrm{N}}(\mathrm{tr}) + \delta_s^{\mathrm{S}}(\mathrm{tr}) + \delta_{n,s}^{\mathrm{NS}} &= \overline{\overline{y}}_{\mathrm{tr}} + \left\{\overline{y}_n^{\mathrm{N}} - \overline{\overline{y}}_{\mathrm{tr}}\right\} + \left\{\overline{y}_s^{\mathrm{S}}(\mathrm{tr}) - \overline{\overline{y}}_{\mathrm{tr}}\right\} \\ &\quad + \left\{y_{n,s}(\mathrm{tr}) - \overline{y}_n^{\mathrm{N}}(\mathrm{tr}) - \overline{y}_s^{\mathrm{S}}(\mathrm{tr}) + \overline{\overline{y}}_{\mathrm{tr}}\right\} \\ &= y_{n,s}(\mathrm{tr}).\end{aligned}$$

□

In light of Equation (15) it is simple to show that $\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}$ has a simple linear representation in terms of the random assignment matrix W and the population mean $\overline{\overline{y}}_{\mathrm{tr}}$.

Lemma 4 (Linear Representation of $\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}$)**.** *It holds*

$$\widehat{\overline{\overline{Y}}}_{\mathrm{tr}} = \overline{\overline{y}}_{\mathrm{tr}} + \frac{1}{pNS}\sum_{n,s} W_{n,s}\delta_{n,s}^{\mathrm{NS}}.$$

Proof of Lemma 4 Leveraging Equation (15), we can write

Proof.

$$\begin{aligned}\widehat{\overline{\overline{Y}}}_{\mathrm{tr}} &= \frac{1}{pNS}\sum_{n,s} W_{n,s} y_{n,s}(\mathrm{tr}) \\ &= \frac{1}{pNS}\sum_{n,s} W_{n,s}\left[\overline{\overline{y}}_{\mathrm{tr}} + \delta_n^{\mathrm{N}}(\mathrm{tr}) + \delta_s^{\mathrm{S}}(\mathrm{tr}) + \delta_{n,s}^{\mathrm{NS}}(\mathrm{tr})\right] \\ &= \overline{\overline{y}}_{\mathrm{tr}}\frac{\sum_{n,s} W_{n,s}}{pNS} + \sum_n \delta_n^{\mathrm{N}}(\mathrm{tr})\sum_s \frac{W_{n,s}}{pNS} \\ &+ \sum_s \delta_s^{\mathrm{S}}(\mathrm{tr})\sum_n \frac{W_{n,s}}{pNS} + \frac{1}{pNS}\sum_n\sum_s W_{n,s}\delta_{n,s}^{\mathrm{NS}}(\mathrm{tr})\end{aligned}$$

and because $\sum_s W_{n,s} = pS$ for all n and $\sum_n W_{n,s} = pN$ for all s,

$$
\begin{aligned}
&= \overline{\overline{y}}_{\text{tr}} + \frac{1}{N}\sum_n \delta_n^{\text{N}}(\text{tr}) + \frac{1}{S}\sum_s \delta_s^{\text{S}}(\text{tr}) + \frac{1}{pNS}\sum_{n,s} W_{n,s}\delta_{n,s}^{\text{NS}}(\text{tr}) \\
&= \overline{\overline{y}}_{\text{tr}} + \frac{1}{pNS}\sum_{n,s} W_{n,s}\delta_{n,s}^{\text{NS}}(\text{tr}),
\end{aligned}
$$

where in the last step we have leveraged the fact that $\sum_n \delta_n^{\text{N}}(\text{tr}) = 0$ and $\sum_t \delta_s^{\text{S}}(\text{tr}) = 0$. □

Lemma 5. $\widehat{\overline{\overline{Y}}}_{\text{tr}}$ *is unbiased for* $\overline{\overline{y}}_{\text{tr}}$.

Proof of Lemma 5

Proof. The proof follows from linearity of the expectation operator and by observing that $\mathbb{E}[W_{n,s}] = p$, and $\sum_{n,s} \delta_{n,s}^{\text{NS}}(\text{tr}) = 0$. □

Lemma 6 (Auxiliary lemma on centered matrices). *Let $A \in \mathbb{R}^{I\times J}$ be a matrix such that $\sum_{j=1}^{J} A_{i,j} = 0$ for all $i \in \{1,\ldots,I\}$ and $\sum_{i=1}^{I} A_{i,j} = 0$ for all $j \in \{1,\ldots,J\}$. It holds:*

$$
\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{i'\neq i} A_{i,j}A_{i',j} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{j'\neq j} A_{i,j}A_{i,j'} = -\sum_{i=1}^{I}\sum_{j=1}^{J} A_{i,j}^2 \tag{16}
$$

and

$$
\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{i'\neq i}\sum_{j'\neq j} A_{i,j}A_{i',j'} = \sum_{i=1}^{I}\sum_{j=1}^{J} A_{i,j}^2. \tag{17}
$$

Proof. For a fixed i, j, it follows that

$$
\sum_{i'\neq i} A_{i',j} = A_{1,j} + A_{2,j} + \ldots + A_{i-1,j} + A_{i+1,j} + \ldots + A_{I,j} = -A_{i,j}
$$

where the second equality is a consequence of the fact that $\sum_j A_{i,j} = 0$ for all i. Hence,

$$
\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{i'\neq i} A_{i,j}A_{i',j} = \sum_{i=1}^{I}\sum_{j=1}^{J} A_{i,j}\sum_{i'\neq i} A_{i',j} = -\sum_{i,j} A_{i,j}^2.
$$

Adopting the same logic, the symmetric results holds when summing over the columns:

$$
\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{j'\neq j} A_{i,j}A_{i,j'} = \sum_{i=1}^{I}\sum_{j=1}^{J} A_{i,j}\sum_{j'\neq j} A_{i,j'} = -\sum_{i,j} A_{i,j}^2.
$$

Now for Equation (17), for fixed i, j:

$$
\begin{aligned}
\sum_{i'\neq i}\sum_{j'\neq j} A_{i',j'} &= A_{1,1} + A_{1,2} + \ldots + A_{1,j-1} + A_{1,j+1} + \ldots + A_{1,J} \\
&+ A_{2,1} + A_{2,2} + \ldots + A_{2,j-1} + A_{2,j+1} + \ldots + A_{2,J} \\
&+ \ldots \\
&+ A_{i-1,1} + A_{i-1,2} + \ldots + A_{i-1,j-1} + A_{i-1,j+1} + \ldots + A_{i-1,J} \\
&+ A_{i+1,1} + A_{i+1,2} + \ldots + A_{i+1,j-1} + A_{i+1,j+1} + \ldots + A_{i+1,J} \\
&+ \cdots + \\
&+ A_{I,1} + A_{I,2} + \ldots + A_{I,j-1} + A_{I,j+1} + \ldots + A_{I,J} \\
&= -A_{1,j} - A_{2,j} - \ldots - A_{i-1,j} - A_{i+1,j} - \ldots - A_{I,j} \\
&= A_{i,j},
\end{aligned}
$$

from which

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{i'\neq i}\sum_{j'\neq j} A_{i,j}A_{i',j'} = \sum_{i,j} A_{i,j}^2.$$

□

Definition 1 (Checkerboard representation of balanced-crossover designs)**.** *For any N, S, p, a balanced-crossover design can be obtained by repeatedly stacking (horizontally and vertically) identity matrices $\boldsymbol{I}$ of dimension $1/p$. Specifically, $W^\star$ can be seen as:*

$$W^\star = \begin{bmatrix} \boldsymbol{I}_{1/p} & \boldsymbol{I}_{1/p} & \dots & \boldsymbol{I}_{1/p} \\ \boldsymbol{I}_{1/p} & \boldsymbol{I}_{1/p} & \dots & \boldsymbol{I}_{1/p} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{I}_{1/p} & \boldsymbol{I}_{1/p} & \dots & \boldsymbol{I}_{1/p} \end{bmatrix},$$

where in this matrix, each "row-block" (aggregating $1/p$ rows) is obtained by horizontally stacking Sp identity matrices $\boldsymbol{I}_{1/p}$ and each "column-block" (aggregating $1/p$ columns) is obtained by vertically stacking Np identity matrices $\boldsymbol{I}_{1/p}$.

Lemma 7 (Cross-moments of balanced crossover designs)**.** *Let $W \in \{0,1\}^{N\times S}$ be a random draw from the space of balanced staggered design (binary) matrices. I.e., for a fixed $p \in [0,1]$, W is a random draw from the space of binary matrices for which $\sum_t W_{n,s} = pS$ for all n, and $\sum_n W_{n,s} = pN$ for all s. Then, for $m \neq n$ and $q \neq s$,*

$$\xi_1 := \mathbb{E}[W_{n,s}^2] = p, \quad \xi_2 := \mathbb{E}[W_{n,s}W_{n,q}] = \frac{p(pS-1)}{S-1}, \quad \xi_3 := \mathbb{E}[W_{n,s}W_{m,t}] = \frac{p(pN-1)}{N-1},$$

and

$$\xi_4 := \mathbb{E}[W_{n,s}W_{m,q}] = p\frac{(pN-1)(pS-1) + pSN(1-p)}{(N-1)(S-1)}.$$

Proof. First, notice that any random matrix $W \in \mathfrak{W}$ satisfies $\sum_{n,s} W_{n,s} = pNS$. That is, $P(W_{n,s} = 1) = \mathbb{E}[W_{n,s}] = p$, hence

$$\xi_1 := \mathbb{E}[W_{n,s}^2] = p.$$

Next, notice that any $W \in \mathfrak{W}$ satisfies for all n that $\sum_s W_{n,s} = pS$. The product $W_{n,s}W_{n,q} = 1$ when $W_{n,s} = 1$ and $W_{n,q} = 1$, and $W_{n,s}W_{n,q} = 0$ otherwise. Because the row of a randomly drawn matrix W from $\mathfrak{W}$ is formed by selecting pS columns without replacement from a set of S candidate columns, the probability that both $W_{n,s} = 1$ and $W_{n,q} = 1$ for $q \neq s$ is simply given by

$$\xi_2 := \mathbb{E}[W_{n,s}W_{n,q}] = p(pS-1)/(S-1).$$

A symmetric argument for the columns yields

$$\xi_3 := \mathbb{E}[W_{n,s}W_{m,s}] = p(pN-1)/(N-1).$$

For the last result, we have to observe the following: in light of Definition 1, any matrix $W \in \mathfrak{W}$ has to be obtained by suitably permuting the rows and the columns of a sequence of stacked identity matrices with dimension $1/p$, such that under the chosen permutation the row-wise and column-wise sums are preserved. This in particular means that for any given entry $W_{n,s}$, given its value, the number of non-zero entries lying on a different row and on a different column is fixed. In particular, to reason about ξ_4, condition on the event that $W_{n,s} = 1$. The number of non-zero entries in the $(N-1)(S-1)$ entries $W_{m,q}$ with $m \neq n$ and $q \neq s$ is given by $(pN-1)(pS-1) + SNp(1-p)$. This can be observed by appealing to the "minimal representor" of Definition 1:

- the identity block to which unit (n,s) belongs to contains $(1/p) - 1$ additional entries (s,m) such that $W_{m,q} = 1$.
- horizontally, there are $Sp - 1$ other matrices, each of which has exactly $(1/p) - 1$ additional entries (s,m) such that $W_{m,q} = 1$.
- vertically, there are $Np - 1$ other matrices, each of which has exactly $(1/p) - 1$ additional entries (s,m) such that $W_{m,q} = 1$.
- in the blocks that do not lie along the same axes as the block of unit (n,s) there are $(Sp - 1 + Np - 1)$ identities, each with $1/p$ ones.

Then, under the condition that $W_{n,s} = 1$, it holds:

$$\sum_{m\neq n}\sum_{q\neq s} W_{m,q} = \left(\frac{1}{p}-1\right)[1+(Sp-1)+(Np-1)] + \left(\frac{1}{p}\right)[(Sp-1)+(Np-1)] \tag{18}$$

$$= SNp - pS - pN + 1, \tag{19}$$

hence

$$\xi_4 := \mathbb{E}[W_{n,s}W_{m,q}] = p\left\{\frac{SNp-pS-pN+1}{(S-1)(N-1)}\right\}.$$

□

Lemma 8. *The variance of* $\hat{\overline{\overline{Y}}}_{\text{tr}}$ *is given by*

$$\text{Var}\left(\hat{\overline{\overline{Y}}}_{\text{tr}}\right) = \frac{\delta(\text{tr})^2}{(pNS)^2}\left(\xi_1 - \xi_2 - \xi_3 + \xi_4\right),$$

where $\delta(\text{tr})^2 := \sum_{n,s}(\delta^{\text{NS}}_{n,s}(\text{tr}))^2$.

Proof. From the result in Lemma 4,

$$\text{Var}\left(\hat{\overline{\overline{Y}}}_{\text{tr}}\right) = \frac{\text{Var}\left(\sum_{n,s} W_{n,s}\delta^{\text{NS}}_{n,s}(\text{tr})\right)}{(pNS)^2} = \frac{\mathbb{E}\left\{\left(\sum_{n,s} W_{n,s}\delta^{\text{NS}}_{n,s}(\text{tr})\right)^2\right\}}{(pNS)^2}. \tag{20}$$

The last equality follows from the fact that for a random variable X with $\mathbb{E}[X]=0$, $\text{Var}(X)=\mathbb{E}[X^2]$ together with the observation that $\mathbb{E}[\sum_{n,s} W_{n,s}\delta^{\text{NS}}_{n,s}(\text{tr})] = p\sum_{n,s}\delta^{\text{NS}}_{n,s}(\text{tr}) = 0$. Hence,

$$\begin{aligned}
\text{Var}\left(\hat{\overline{\overline{Y}}}_{\text{tr}}\right) &= \frac{\sum_{n=1}^N\sum_{m=1}^N\sum_{s=1}^S\sum_{s=1}^S \mathbb{E}[W_{n,s}W_{n,q}]\delta^{\text{NS}}_{n,s}(\text{tr})\delta^{\text{NS}}_{m,q}(\text{tr})}{(pNS)^2} \\
&= \frac{\sum_{n,s}\mathbb{E}[W_{n,s}^2](\delta^{\text{NS}}_{n,s}(\text{tr}))^2}{(pNS)^2} \\
&+ \frac{\sum_{n=1}^N\sum_{s=1}^S\sum_{q\neq s}\mathbb{E}[W_{n,s}W_{n,q}]\delta^{\text{NS}}_{n,s}(\text{tr})\delta^{\text{NS}}_{n,q}(\text{tr})}{(pNS)^2} \\
&+ \frac{\sum_{n=1}^N\sum_{s=1}^S\sum_{m\neq n}\mathbb{E}[W_{n,s}W_{m,t}]\delta^{\text{NS}}_{n,s}(\text{tr})\delta^{\text{NS}}_{m,t}(\text{tr})}{(pNS)^2} \\
&+ \frac{\sum_{n=1}^N\sum_{s=1}^S\sum_{m\neq n}\sum_{q\neq s}\mathbb{E}[W_{n,s}W_{m,q}]\delta^{\text{NS}}_{n,s}\delta^{\text{NS}}_{m,q}(\text{tr})}{(pNS)^2} \\
&= \xi_1\frac{\sum_{n,s}(\delta^{\text{NS}}_{n,s}(\text{tr}))^2}{(pNS)^2} + \xi_2\frac{\sum_{n=1}^N\sum_{s=1}^S\sum_{q\neq s}\delta^{\text{NS}}_{n,s}(\text{tr})\delta^{\text{NS}}_{n,q}(\text{tr})}{(pNS)^2} \\
&+ \xi_3\frac{\sum_{n=1}^N\sum_{s=1}^S\sum_{m\neq n}\delta^{\text{NS}}_{n,s}(\text{tr})\delta^{\text{NS}}_{m,t}(\text{tr})}{(pNS)^2} + \xi_4\frac{\sum_{n=1}^N\sum_{s=1}^S\sum_{m\neq n}\sum_{q\neq s}\delta^{\text{NS}}_{n,s}(\text{tr})\delta^{\text{NS}}_{m,q}(\text{tr})}{(pNS)^2}.
\end{aligned}$$

Now notice that because $\sum_n \delta^{\text{NS}}_{n,s}(\text{tr}) = 0$ and $\sum_s \delta^{\text{NS}}_{n,s}(\text{tr}) = 0$, we can apply the results in Lemma 6, and it holds for all n, s

$$\sum_{q\neq s}\delta^{\text{NS}}_{n,q}(\text{tr}) = -\delta^{\text{NS}}_{n,s}(\text{tr}), \quad \text{and} \quad \sum_{m\neq n}\delta^{\text{NS}}_{m,s}(\text{tr}) = -\delta^{\text{NS}}_{n,s}(\text{tr})$$

Hence,

$$\sum_{n=1}^N\sum_{s=1}^S\sum_{q\neq s}\delta^{\text{NS}}_{n,s}(\text{tr})\delta^{\text{NS}}_{n,q}(\text{tr}) = -\sum_{n=1}^N\sum_{s=1}^S(\delta^{\text{NS}}_{n,s}(\text{tr}))^2,$$

and

$$\sum_{n=1}^N\sum_{s=1}^S\sum_{m\neq n}\delta^{\text{NS}}_{n,s}(\text{tr})\delta^{\text{NS}}_{m,s}(\text{tr}) = -\sum_{n=1}^N\sum_{s=1}^S(\delta^{\text{NS}}_{n,s}(\text{tr}))^2,$$

as well as

$$\sum_{n=1}^{N}\sum_{s=1}^{S}\sum_{m\neq n}\sum_{q\neq s}\delta^{\mathrm{NS}}_{n,s}\delta^{\mathrm{NS}}_{m,q}=\sum_{n=1}^{N}\sum_{s=1}^{S}(\delta^{\mathrm{NS}}_{n,s})^2.$$

Calling $\delta(\mathrm{tr})^2 := \sum_{n,s}\left\{\delta^{\mathrm{NS}}_{n,s}(\mathrm{tr})\right\}^2$ and substituting, we get to the final expression

$$\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}\right)=\frac{\delta(\mathrm{tr})^2}{(pNS)^2}\left(\xi_1-\xi_2-\xi_3+\xi_4\right). \tag{21}$$

□

Proof of Lemma 1. The results presented in Lemma 1 follow from the previous derivations. Specifically, unbiasedness of $\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}$ follows from Lemma 5. The expression for the variance was showed to be true in Lemma 8. □

Lemma 9. *The term* $\delta(\mathrm{tr})^2 := \sum_{n,s}\delta^{\mathrm{NS}}_{n,s}(\mathrm{tr})\delta^{\mathrm{NS}}_{n,s}(\mathrm{tr})$ *can be written as*

$$\delta(\mathrm{tr})^2=\sum_{n,s}y_{n,s}(\mathrm{tr})^2-S\sum_{n=1}^{N}\left\{\overline{y}^{\mathrm{N}}_n(\mathrm{tr})\right\}^2-N\sum_{s=1}^{S}\left\{\overline{y}^{\mathrm{S}}_s(\mathrm{tr})\right\}^2+NS(\overline{\overline{y}}_{\mathrm{tr}})^2$$

Proof. Notice that we can write the individual term $\delta^{\mathrm{NS}}_{n,s}(\mathrm{tr})$ as follows:

$$\begin{aligned}(\delta^{\mathrm{NS}}_{n,s}(\mathrm{tr}))^2 &= \left(y_{n,s}(\mathrm{tr})-\overline{y}^{\mathrm{N}}_n(\mathrm{tr})-\overline{y}^{\mathrm{S}}_s(\mathrm{tr})+\overline{\overline{y}}_{\mathrm{tr}}\right)^2\\
&= (y_{n,s}(\mathrm{tr}))^2+\left\{\overline{y}^{\mathrm{N}}_n(\mathrm{tr})\right\}^2+\left\{\overline{y}^{\mathrm{S}}_s(\mathrm{tr})\right\}^2+(\overline{\overline{y}}_{\mathrm{tr}})^2\\
&- 2y_{n,s}(\mathrm{tr})\overline{y}^{\mathrm{N}}_n(\mathrm{tr})-2y_{n,s}(\mathrm{tr})\overline{y}^{\mathrm{S}}_s(\mathrm{tr})+2y_{n,s}(\mathrm{tr})\overline{\overline{y}}_{\mathrm{tr}}\\
&+ 2\overline{y}^{\mathrm{N}}_n(\mathrm{tr})\overline{y}^{\mathrm{S}}_s(\mathrm{tr})-2\overline{y}^{\mathrm{N}}_n(\mathrm{tr})\overline{\overline{y}}_{\mathrm{tr}}+2\overline{y}^{\mathrm{S}}_s(\mathrm{tr})\overline{\overline{y}}_{\mathrm{tr}}.\end{aligned}$$

Now summing over n, s:

$$\begin{aligned}\delta(\mathrm{tr})^2 &= \sum_{n,s}(\delta^{\mathrm{NS}}_{n,s}(\mathrm{tr}))^2\\
&= \sum_{n,s}(y_{n,s}(\mathrm{tr}))^2+S\sum_{n}\left\{\overline{y}^{\mathrm{N}}_n(\mathrm{tr})\right\}^2+N\sum_{s}\left\{\overline{y}^{\mathrm{S}}_s(\mathrm{tr})\right\}^2+NS(\overline{\overline{y}}_{\mathrm{tr}})^2\\
&- 2S\sum_{n}(\overline{y}^{\mathrm{N}}_n(\mathrm{tr}))^2-2N\sum_{s}(\overline{y}^{\mathrm{S}}_s(\mathrm{tr}))^2+2NS(\overline{\overline{y}}_{\mathrm{tr}})^2\\
&+ 2NS(\overline{\overline{y}}_{\mathrm{tr}})^2-2NS(\overline{\overline{y}}_{\mathrm{tr}})^2+2NS(\overline{\overline{y}}_{\mathrm{tr}})^2\\
&= \sum_{n,s}(y_{n,s}(\mathrm{tr}))^2-S\sum_{n}\left\{\overline{y}^{\mathrm{N}}_n(\mathrm{tr})\right\}^2-N\sum_{s}\left\{\overline{y}^{\mathrm{S}}_s(\mathrm{tr})\right\}^2+NS(\overline{\overline{y}}_{\mathrm{tr}})^2.\end{aligned}$$

□

In light of Lemma 9, letting

$$\xi := \xi_1-\xi_2-\xi_3+\xi_4$$

we can write

$$\begin{aligned}\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}\right) &= \frac{\xi}{(pNS)^2}\delta(\mathrm{tr})^2 \qquad (22)\\
&= \xi\left\{\frac{\sum_{n,s}(y_{n,s}(\mathrm{tr}))^2-S\sum_{n}\left\{\overline{y}^{\mathrm{N}}_n(\mathrm{tr})\right\}^2-N\sum_{s}\left\{\overline{y}^{\mathrm{S}}_s(\mathrm{tr})\right\}^2+NS(\overline{\overline{y}}_{\mathrm{tr}})^2}{(pNS)^2}\right\}\\
&= \frac{\xi}{p^2NS}\left[\underbrace{\frac{\sum_{n,s}(y_{n,s}(\mathrm{tr}))^2}{NS}}_{=:\phi_1}-\underbrace{\frac{\sum_{n}\left\{\overline{y}^{\mathrm{N}}_n(\mathrm{tr})\right\}^2}{N}}_{=:\phi_2}-\underbrace{\frac{\sum_{s}\left\{\overline{y}^{\mathrm{S}}_s(\mathrm{tr})\right\}^2}{S}}_{=:\phi_3}+\underbrace{(\overline{\overline{y}}_{\mathrm{tr}})^2}_{=:\phi_4}\right]\\
&= \frac{\xi}{p^2NS}\left[\phi_1-\phi_2-\phi_3+\phi_4\right].\end{aligned}$$

Proof of Lemma 2 We now turn to the question of how to *estimate* the variance of the sample mean (i.e., estimate Equation (22)). We use the representation in Lemma 9, and estimate each summand separately. In what follows, the expectation operator $\mathbb{E}(\cdot)$ and the variance operator $\mathrm{Var}(\cdot)$ are always defined with respect to uniform random sampling from $\mathfrak{W}$.

Given a random matrix $W \in \mathfrak{W}$, we let $\Xi \subset [N] \times [S]$ be the set of indices (n, s) with $W_{n,s} = 1$.

Estimation of ϕ_1 Notice that the number of matrices in $\mathfrak{W}$ such that $W_{n,s} = 1$ is constant for all n, s. Hence, it follows that under uniform sampling from $\mathfrak{W}$,

$$\frac{1}{p}\mathbb{E}\left[\sum_{(n,s)\in\Xi} y_{n,s}(\mathrm{tr})\right] = \sum_{n=1}^{N}\sum_{s=1}^{S} y_{n,s}(\mathrm{tr}).$$

In turn, this implies that

$$\widehat{\phi}_1 := \frac{1}{pNS}\sum_{(n,s)\in\Xi} y_{n,s}(\mathrm{tr})^2 \quad \text{satisfies} \quad \mathbb{E}\left[\widehat{\phi_1}\right] = \phi_1.$$

Estimation of ϕ_2 Define the sample counterpart of $\overline{y}_n^N$ to be

$$\widehat{\overline{Y}}_n^N = \frac{1}{pS}\sum_{s:W_{n,s}=1} y_{n,s}(\mathrm{tr}).$$

It follows that

$$\mathbb{E}\left[\widehat{\overline{Y}}_n^N\right] = \overline{y}_n^N \quad \text{hence} \quad \mathbb{E}\left[\left\{\widehat{\overline{Y}}_n^N\right\}^2\right] = \left\{\overline{y}_n^N\right\}^2 + \mathrm{Var}\left(\widehat{\overline{Y}}_n^N\right).$$

Lemma 10. *The variance of the mean-along the treated columns of the n-th unit, $\widehat{\overline{Y}}_n^N$, is given by*

$$\mathrm{Var}\left(\widehat{\overline{Y}}_n^N\right) = \frac{1-p}{p}\frac{1}{S}\left\{\frac{1}{S-1}\sum_{s=1}^{S}\left(y_{n,s}(\mathrm{tr}) - \overline{y}_n^N\right)^2\right\}.$$

An unbiased estimator of $\mathrm{Var}\left(\widehat{\overline{Y}}_n^N\right)$ *is given by*

$$\widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}_n^N\right) = \frac{1-p}{p}\frac{1}{S}\left\{\frac{1}{pS-1}\sum_{t:W_{n,s}=1}\left(y_{n,s}(\mathrm{tr}) - \widehat{\overline{Y}}_n^N\right)^2\right\}.$$

Proof of Lemma 10. The proof of Lemma 10 stems from the simple observation that, due to the balancedness constrain, we can view every row in the matrix as a simple "single" randomized experiment with S units, in which exactly pS are treated. In this case, the variance of the sample mean (along the columns of a fixed row) and its estimator are well known. See e.g. Cochran [1977, Theorem 2.4]. □

In light of Lemma 10, it follows that

$$\widehat{\phi}_2 := \frac{1}{N}\sum_{n=1}^{N}\left[\left\{\widehat{\overline{Y}}_n^N\right\}^2 - \widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}_n^N\right)\right] \quad \text{satisfies} \quad \mathbb{E}\left[\widehat{\phi}_2\right] = \phi_2.$$

Estimation of ϕ_3 The estimation of ϕ_3 is symmetric to that of ϕ_2:

$$\mathbb{E}\left[\widehat{\overline{Y}}_s^{\mathrm{S}}\right] = \overline{y}_s^{\mathrm{S}} \quad \text{which implies} \quad \mathbb{E}\left[\left\{\widehat{\overline{Y}}_s^{\mathrm{S}}\right\}^2\right] = \left\{\overline{y}_s^{\mathrm{S}}\right\}^2 + \mathrm{Var}\left(\widehat{\overline{Y}}_s^{\mathrm{S}}\right).$$

Lemma 11. *The variance of $\widehat{\overline{Y}}_s^{\mathrm{S}}$ is given by*

$$\mathrm{Var}\left(\widehat{\overline{Y}}_s^{\mathrm{S}}\right) = \frac{1-p}{p}\frac{1}{N}\left\{\frac{1}{N-1}\sum_{n=1}^{N}\left(y_{n,s}(\mathrm{tr}) - \overline{y}_s^{\mathrm{S}}\right)^2\right\}.$$

An unbiased estimator of $\mathrm{Var}\left(\widehat{\overline{Y}}_s^{\mathrm{S}}\right)$ *is given by*

$$\widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}_s^{\mathrm{S}}\right) = \frac{1-p}{p}\frac{1}{N}\left\{\frac{1}{pN-1}\sum_{n:W_{n,s}=1}\left(y_{n,s}(\mathrm{tr}) - \widehat{\overline{Y}}_s^{\mathrm{S}}\right)^2\right\}.$$

Proof of Lemma 11. The same argument used in Lemma 10 applies in this context, where we are averaging over treated rows for a fixed column. □

Hence, it follows that

$$\widehat{\phi}_3 := \frac{1}{S}\sum_{s=1}^{S}\left[\left\{\widehat{\overline{Y}}_s^{\mathrm{S}}\right\}^2 - \widehat{\mathrm{Var}}\left(\widehat{\overline{Y}}_s^{\mathrm{S}}\right)\right] \quad \text{satisfies} \quad \mathbb{E}\left[\widehat{\phi}_3\right] = \phi_3.$$

Estimation of ϕ_4 For the last term,

$$\mathbb{E}\left[\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}\right] = \overline{\overline{y}}_{\mathrm{tr}} \quad \text{implying} \quad \mathbb{E}\left[\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}^2\right] = \phi_4 + \mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}\right).$$

Notice that $\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}\right)$ is itself the object we are interested in estimating.

Letting $\widehat{\eta}_4 := \widehat{\overline{\overline{Y}}}_{\mathrm{tr}}^2$, it is straightforward to observe that

$$\mathbb{E}\left[\hat{\phi}_1 - \hat{\phi}_2 - \hat{\phi}_3 + \hat{\eta}_4\right] = \phi_1 - \phi_2 - \phi_3 + \phi_4 + \frac{\xi}{p^2 NS}\left(\phi_1 - \phi_2 - \phi_3 + \phi_4\right) \tag{23}$$

$$= \left(1 + \frac{\xi}{p^2 NS}\right)\left(\phi_1 - \phi_2 - \phi_3 + \phi_4\right). \tag{24}$$

Last, defining

$$\left\{\widehat{\phi}_1 - \widehat{\phi}_2 - \widehat{\phi}_3 + \widehat{\eta}_4\right\} = \frac{\xi}{\xi + p^2 NS}\left\{\widehat{\phi}_1 - \widehat{\phi}_2 - \widehat{\phi}_3 + \widehat{\eta}_4\right\}.$$

it holds

$$\mathbb{E}\left\{\widehat{\mathrm{Var}}\left(\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}\right)\right\} = \mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}\right).$$

B.2 Efficiency of crossover designs

Consider the setting discussed in Equation (31). Outcomes are given by real-valued matrix of potential outcomes indexed by a binary treatment, $Y(w) \in \mathbb{R}^{N\times S}$, where

$$y_{n,s}(w) = \mu + \tau 1(w=1) + a_n + b_s + \epsilon_{n,s},$$

where $\epsilon_{n,s} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$, drawn independent of other quantities.

To illustrate how the balanced crossover designs we consider can lead to significant efficiency gains even in the absence of interference, we consider the case in which potential outcomes are fixed, but drawn from the following distribution:

$$a_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2) \text{ for } n = 1,\ldots,N, \quad \text{and} \quad b_s \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_b^2) \text{ for } s = 1,\ldots,S.$$

It follows from the properties of the Gaussian distribution that, letting $\sigma^2 := \sigma_a^2 + \sigma_b^2 + \sigma_\epsilon^2$:

$$y_{n,s}(w) \sim \mathcal{N}(\mu + \tau 1(w=1), \sigma^2). \tag{25}$$

Moreover, for each row, the corresponding mean over the columns follows a Gaussian distribution:

$$\begin{aligned}\overline{y}_n^{\mathrm{N}}(w) &= \frac{1}{S}\sum_s y_{n,s}(w)\\ &= \mu + \tau 1(w=1) + a_n + \frac{1}{S}\sum_{s=1}^{S}(b_s + \epsilon_{n,s})\\ &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu + \tau 1(w=1), \sigma_a^2 + \frac{\sigma_b^2 + \sigma_\epsilon^2}{S}\right).\end{aligned}$$

Similarly, for each column the corresponding mean over the rows follows a Gaussian distribution:

$$\begin{aligned}\overline{y}_s^{\mathrm{S}}(w) &= \frac{1}{N}\sum_n y_{n,s}(w)\\ &= \mu + \tau 1(w=1) + \frac{1}{N}\sum_{n=1}^{N}(a_n + \epsilon_{n,s}) + b_s\\ &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu + \tau 1(w=1), \frac{\sigma_a^2 + \sigma_\epsilon^2}{N} + \sigma_b^2\right).\end{aligned}$$

Balanced crossover designs First, consider the case in which we randomize using the balanced crossover design. The variance of the simple plug-in estimator $\hat{\overline{\overline{Y}}}_w$ of Equation (12) was given in Lemma 8. Under the parametric Gaussian assumption of Equation (31), the expected value of this quantity over hypothetical re-draws of the data is given by:

$$\mathbb{E}_{\mathcal{N}}\left\{\mathrm{Var}_{\mathfrak{W}}\left(\hat{\overline{\overline{Y}}}_w\right)\right\} = \frac{\xi}{(pNS)^2}\mathbb{E}_{\mathcal{N}}\left[\delta(w)^2\right] \tag{26}$$

and since $\delta(w)^2 = \sum_{n,s}(y_{n,s}(\mathrm{tr}))^2 - S\sum_n \left\{\overline{y}_n^{\mathrm{N}}(\mathrm{tr})\right\}^2 - N\sum_s \left\{\overline{y}_s^{\mathrm{S}}(\mathrm{tr})\right\}^2 + NS(\overline{\overline{y}}_{\mathrm{tr}})^2$ from Lemma 9:

$$= \frac{\xi}{p^2NS}\left\{\sigma_\epsilon^2\left(1 - \frac{1}{N} - \frac{1}{S}\right) - \frac{\sigma_a^2}{N} - \frac{\sigma_b^2}{S}\right\} \tag{27}$$

$$= O\left(\frac{\sigma_\epsilon^2}{NS}\right). \tag{28}$$

By contrast, we consider as alternative designs the cases in which (i) we are assignment treatment to the (n, s)-th entry by means of a random Bernoulli coin flip with probability p — a completely randomized design, or (ii) a unit-randomized design, in which pN units are assigned (at random) to either control or treatment, and their assignment is constant over time and (iii) a time-randomized design, in which pS time-stamps are assigned (at random) to treatment, and every unit n gets to see the same experience at a given time stamp s.

Completely randomized designs In this case, where the matrix of assignments is a collection of i.i.d. coin flips, each with probability p, the estimator for the population mean has the same form as in Equation (9); however, the space of assignments $\mathfrak{W}^{\mathrm{CRD}}$ and the weighting distribution is different. We hence use the notation $\hat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}$ to emphasize that this estimator is obtained from a completely randomized design. We now show that, when the potential outcomes satisfy the distributional requirement of Equation (31), the variance resulting from this randomization scheme is provably larger (in expectation). Before doing so, we state a useful technical lemma.

Lemma 12 (Approximating the variance of a ratio of random variables)**.** *Let X, Y be positive real valued random variables with finite first two moments, $\mathbb{E}[X] = \mu_X, \mathrm{Var}(X) = \sigma_X^2$ and $\mathbb{E}[Y] = \mu_Y, \mathrm{Var}(Y) = \sigma_Y^2$. Let $Z = X/Y$. A first order Taylor expansion approximation yields*

$$\mathrm{Var}(Z) \approx \left(\frac{\mu_X}{\mu_Y}\right)^2\left[\frac{\sigma_X^2}{\mu_X^2} - 2\frac{\mathrm{Cov}(X,Y)}{\mu_X\mu_Y} + \frac{\sigma_Y^2}{\mu_Y^2}\right]. \tag{29}$$

Lemma 13 (Moments of $\hat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}$)**.** *$\hat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}$ is unbiased for $\overline{\overline{y}}_w$. Its variance can be approximated by*

$$\mathrm{Var}\left(\hat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}\right) \approx \left(\frac{1-p}{p}\right)\frac{1}{NS}\left(\frac{\sum_{n,s} y_{n,s}(w)^2}{NS} - \overline{\overline{y}}_w^2\right).$$

Then, using Lemma 15 together with the Gaussian parametric assumption of Equation (31) we see that:

$$\mathbb{E}_{\mathcal{N}}\left\{\mathrm{Var}_{\mathfrak{W}^{\mathrm{CRD}}}\left(\widehat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}\right)\right\} \approx \left(\frac{1-p_w}{p_w}\right)\frac{1}{NS}\left(\frac{\sum_{n,s}\mathbb{E}[y_{n,s}(w)^2]}{NS} - \mathbb{E}\left[\overline{\overline{y}}_w^2\right]\right)$$
$$= \left(\frac{1-p_w}{p_w}\right)\frac{\sigma^2}{NS} = O\left(\frac{\sigma^2}{NS}\right) \tag{30}$$

Exact formulae for the variance of the sample mean under the completely randomized design can be found, e.g., using the results in Strand [1979].

Individual-randomized designs In this case, our experiment is equivalent to a simple single randomized design with N units, in which each unit has potential outcome $y_n(w) = S^{-1}\sum_{s=1}^{S} y_{n,s}(w) = \overline{y}_n^{\mathrm{N}}(w)$. While the estimator for the population mean has (again) the same form as in Equation (9), we use $\widehat{\overline{\overline{Y}}}_{w,\rightarrow}$ to denote it, to emphasize the dependence on the space of valid assignments $\mathfrak{W}^{\mathrm{N}}$ (the space of binary matrices with constant columns, and such that exactly pS columns are equal to one). The variance of $\widehat{\overline{\overline{Y}}}_{w,\rightarrow}$ is

$$\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\rightarrow}\right) = \frac{1-p_w}{p_w}\frac{1}{N-1}\frac{1}{N}\sum_{n=1}^{N}\left(\overline{y}_n^{\mathrm{N}}(w) - \overline{\overline{y}}_w\right)^2.$$

In the Gaussian case of Equation (31), we can write

$$\mathbb{E}\left[\left(\overline{y}_n^{\mathrm{N}}(w) - \overline{\overline{y}}_w\right)^2\right] = \mathbb{E}\left[\left(\overline{y}_n^{\mathrm{N}}(w)\right)^2 - 2\overline{y}_n^{\mathrm{N}}(w)\overline{\overline{y}}_w + (\overline{\overline{y}}_w)^2\right]$$
$$= \mathbb{E}\left[\left(\overline{y}_n^{\mathrm{N}}(w)\right)^2\right] + \mathbb{E}\left[(\overline{\overline{y}}_w)^2\right] - 2\mathbb{E}\left\{\frac{1}{S}\sum_{s=1}^{S} y_{n,s}(w)\left(\frac{1}{NS}\sum_{m=1}^{N}\sum_{q=1}^{S} y_{m,q}(w)\right)\right\}$$

The sum of the first two expectations in the equation above is equal to

$$\sigma_a^2 + \frac{\sigma_b^2 + \sigma_\epsilon^2}{S} + 2\left(\mu + \tau 1(w=1)\right)^2.$$

For the last term in the equation above, it holds

$$\frac{1}{NS^2}\mathbb{E}\left\{\sum_{m=1}^{N}\sum_{s=1}^{S}\sum_{q=1}^{S} y_{n,s}(w)y_{m,q}(w)\right\} = \frac{1}{NS^2}\sum_{s=1}^{S}\mathbb{E}\left[y_{n,s}(w)^2\right]$$
$$+ \frac{1}{NS^2}\sum_{s=1}^{S}\sum_{q\neq s}^{S}\mathbb{E}\left[y_{n,s}(w)y_{n,q}(w)\right]$$
$$+ \frac{1}{NS^2}\sum_{m\neq n}^{N}\sum_{s=1}^{S}\mathbb{E}\left[y_{n,s}(w)y_{m,s}(w)\right]$$
$$+ \frac{1}{NS^2}\sum_{m\neq n}^{N}\sum_{s=1}^{S}\sum_{q\neq s}^{S}\mathbb{E}\left[y_{n,s}(w)y_{m,q}(w)\right],$$

which implies

$$\sum_{m,s,q}\mathbb{E}\left\{\frac{y_{n,s}(w)y_{m,q}(w)}{NS^2}\right\} = (\mu + \tau 1(w=1))^2 + \frac{1}{NS^2}\left[S\sigma^2 + S(S-1)\sigma_a^2 + (N-1)S\sigma_b^2\right]$$
$$= (\mu + \tau 1(w=1))^2 + \frac{1}{NS^2}\left[S^2\sigma_a^2 + SN\sigma_b^2 + S\sigma_\epsilon^2\right]$$
$$= (\mu + \tau 1(w=1))^2 + \frac{\sigma_a^2}{N} + \frac{\sigma_b^2}{S} + \frac{\sigma_\epsilon^2}{NS}.$$

Hence,

$$\mathbb{E}_{\mathfrak{W}^{\mathrm{N}}}\left[\operatorname{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\rightarrow}\right)\right] = \frac{1-p_w}{p_w}\frac{1}{N-1}\left(\sigma_a^2 + \frac{\sigma_b^2+\sigma_\epsilon^2}{S} - 2\left[\frac{\sigma_a^2}{N}+\frac{\sigma_b^2}{S}+\frac{\sigma_\epsilon^2}{NS}\right]\right)$$
$$= O\left(\frac{\sigma_a^2}{N}+\frac{\sigma_\epsilon^2}{NS}\right).$$

Time-randomized designs Symmetrically to the case above, here our experiment is equivalent to a simple single randomized design with S units, in which each unit has potential outcome $y_s(w) = S^{-1}\sum_{n=1}^{N} y_{n,s}(w) = \overline{y}_s^{\mathrm{S}}(w)$. While the estimator for the population mean has (again) the same form as in Equation (9), we use $\widehat{\overline{\overline{Y}}}_{w,\downarrow}$ to denote it, to emphasize the dependence on the space of valid assignments $\mathfrak{W}^{\mathrm{S}}$ (the space of binary matrices with constant rows, and such that exactly pN rows are equal to one). The variance of $\widehat{\overline{\overline{Y}}}_{w,\downarrow}$ is

$$\operatorname{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\downarrow}\right) = \frac{1-p_w}{p_w}\frac{1}{S-1}\frac{1}{S}\sum_{s=1}^{S}\left(\overline{y}_s^{\mathrm{S}}(w) - \overline{\overline{y}}_w\right)^2.$$

A symmetric derivation to the one above for the unit-randomized design yields:

$$\mathbb{E}_{\mathfrak{W}^{\mathrm{S}}}\left[\operatorname{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\downarrow}\right)\right] = \frac{1-p_w}{p_w}\frac{1}{S-1}\left(\sigma_b^2 + \frac{\sigma_a^2+\sigma_\epsilon^2}{N} - 2\left[\frac{\sigma_b^2}{S}+\frac{\sigma_a^2}{N}+\frac{\sigma_\epsilon^2}{NS}\right]\right)$$
$$= O\left(\frac{\sigma_b^2}{S}+\frac{\sigma_\epsilon^2}{NS}\right).$$

Then, we see that the expected variance of the simple estimator $\widehat{\overline{\overline{Y}}}_w$ obtained from a staggered design is smaller than the corresponding variance incurred by adopting other copmeting designs (completely randomized design, time-randomized, unit-randomized). In particular, this is because — by virtue of the balancing — the effect of the variance components σ_a^2 and σ_b^2 is of second order — whereas these terms appear as leading terms in the other variances.

C Simulations for the balanced crossover designs

C.1 Correctness of the variance formulae

In this section, we present simulation results to show the correctness of our derivation for the variance formulae in the balanced crossover design of Appendix A.

In particular, we here show by means of simulation that (a) the formula of the variance of Equation (10) is correct, and that (b) the variance estimator of Equation (11) is unbiased.

Towards this goal, we first draw a real-valued matrix $Y_w \in \mathbb{R}^{N\times S}$ of potential outcomes, which we consider to be fixed (here, this matrix represents the potential outcomes under treatment; it could equally represent the outcomes under control). In practice, we let the entries of this matrix be random draws from a Gaussian distribution with mean and variance equal to 1 — although any matrix of potential outcomes could be used to show these results. Then, for a fixed $p_w \in (0,1)$, we repeatedly draw binary random assignment matrices $W \in \mathfrak{W}$. We do so using the fastball algorithm [Strona et al., 2014b]. Given valid assignment matrices, we compare:

- The analytic formula of the variance $\operatorname{Var}\left(\widehat{\overline{\overline{Y}}}_w\right)$ provided in Equation (10). Notice that in our simulation, this analytic value be obtained exactly since the whole matrix of potential outcomes is known.
- The Monte Carlo estimate of the analytic variance (i.e., the *empirical* variance of $\widehat{\overline{\overline{Y}}}_w$ across the draws W obtained using the fastball algorithm).
- The Monte Carlo distribution of the estimated variances, i.e. the distribution of $\widehat{\operatorname{Var}}\left(\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}\right)$ over the sampled assignments W.

Towards this goal, we start with the simplest case of $N = 4$ and $T = 4$, and draw a matrix of potential outcomes $Y_w \in \mathbb{R}^{4\times 4}$. In this case, for $p_w = 0.5$, the size of the space of valid assignments is $|\mathfrak{W}| = 90$. This quantity can be

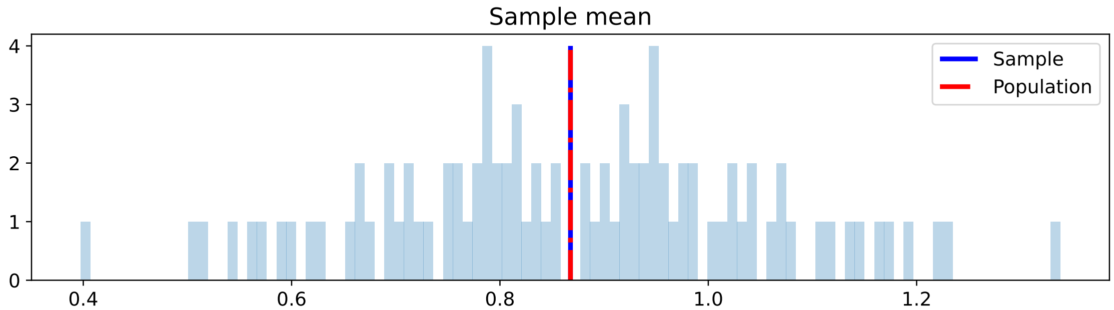

Figure 1: We compare the empirical distribution of $\widehat{\overline{\overline{Y}}}_w$ over valid assignment $W \in \mathfrak{W}$ (blue histogram) and the average over re-randomizations (blue vertical line) with the population mean $\overline{\overline{y}}_w$ (red vertical line).

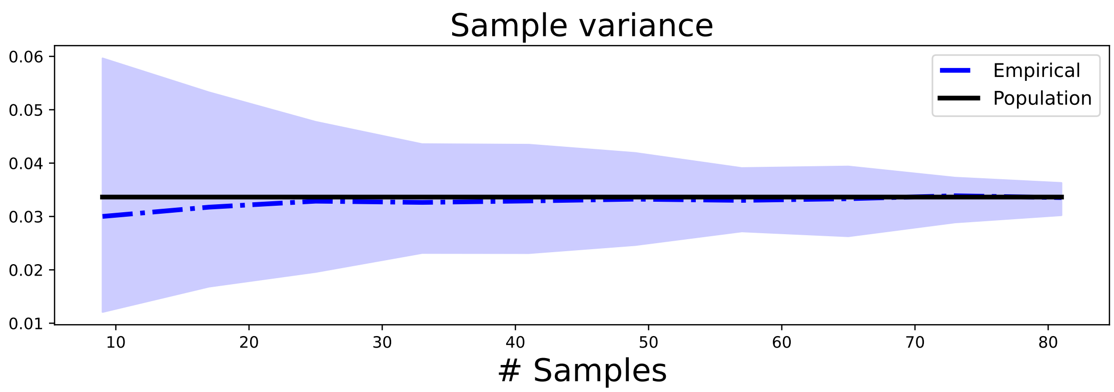

Figure 2: Analytical "population" variance of the sample mean (horizontal solid line) and Monte Carlo estimate of the variance sample mean as a function of the subset size used to obtain such estimate (blue line). For each size $x \leq |\mathfrak{W}|$, 1000 random subsets of size x are used to form an estimate of the variance. We plot the mean across the 1000 random subsets in blue. The shaded region covers the quantiles ranging from 2.5% to 97.5% of the distribution.

obtained explicitly using e.g. the results in [Miller and Harrison, 2013b]. We use the fastball algorithm to enumerate all valid assignment matrices and perform the comparisons discussed above.

We verify empirically the unbiasedness of $\widehat{\overline{\overline{Y}}}_w$ for $\overline{\overline{y}}_w$ as per Lemma 5 in Figure 1.

Next, because sampling from $\mathfrak{W}$ is not trivial, we verify the extent to which we are able to obtain unbiased samples from $\mathfrak{W}$ with the fastball algorithm [Strona et al., 2014b] as follows. For a given size $x \leq |\mathfrak{W}|$ of the sample (horizontal axis), we retain a random subset Ξ of x assignment matrices (without replacement). We compute the sample variance $\widetilde{\mathrm{Var}}_w(\widehat{\overline{\overline{Y}}}_w; \Xi)$ of the sample mean $\widehat{\overline{\overline{Y}}}_w$ on this random subset of x assignment matrices:

$$\widetilde{\mathrm{Var}}_w\left(\widehat{\overline{\overline{Y}}}_w; \Xi\right) = \frac{1}{|\Xi|}\sum_{W\in\Xi}\left(\widehat{\overline{\overline{Y}}}_w(W) - \left[\frac{1}{|\Xi|}\sum_{W'\in\Xi}\widehat{\overline{\overline{Y}}}_w(W')\right]\right)^2.$$

We are here indexing $\widehat{\overline{\overline{Y}}}_w$ additionally by W to emphasize that the estimate depends on the assignment matrix. If we were able to sample uniformly from $\mathfrak{W}$, $\widetilde{\mathrm{Var}}_w(\widehat{\overline{\overline{Y}}}_w; \Xi)$ would be (by the law of large numbers) an unbiased estimate of $\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_w\right)$. We then repeat, for a given value of x, a random selection 1,000 assignment matrices Ξ (each selection comprising x randomly chosen assignment matrices). We plot this comparison in Figure 2 the analytical "population" value of $\mathrm{Var}(\widehat{\overline{\overline{Y}}}_w)$ (horizontal black line) to the Monte Carlo sampling variance of $\widehat{\overline{\overline{Y}}}_{\mathrm{tr}}$ as the number of samples to compute the Monte Carlo estimate increase. Notice that, in the simple case of $N = 4$ and $S = 4$, where the size of $\mathfrak{W} = 90$ and we can enumerate all the sample space, by construction we expect exactly the behavior in Figure 2.

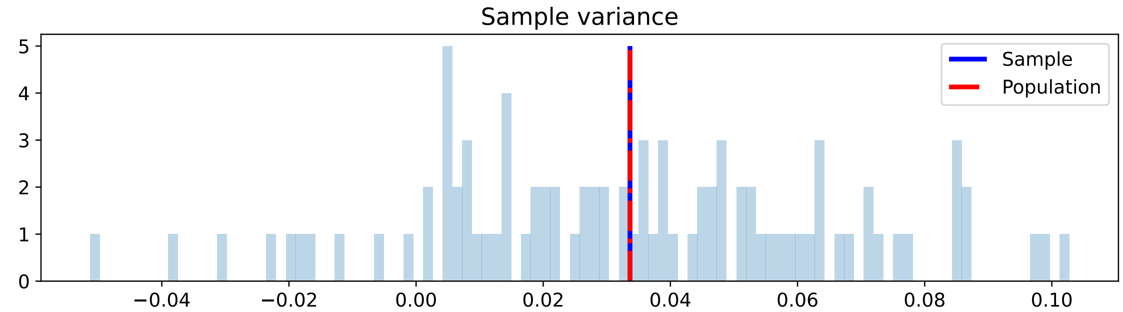

Figure 3: We compare the empirical distribution of the estimated sample variance $\widehat{\mathrm{Var}}(\widehat{\overline{\overline{Y}}}_w)$ over valid assignment $W \in \mathfrak{W}$ (blue histogram) and the average over re-randomizations (blue vertical line) with the variance of the sample mean $\mathrm{Var}(\widehat{\overline{\overline{Y}}}_w)$ (red vertical line).

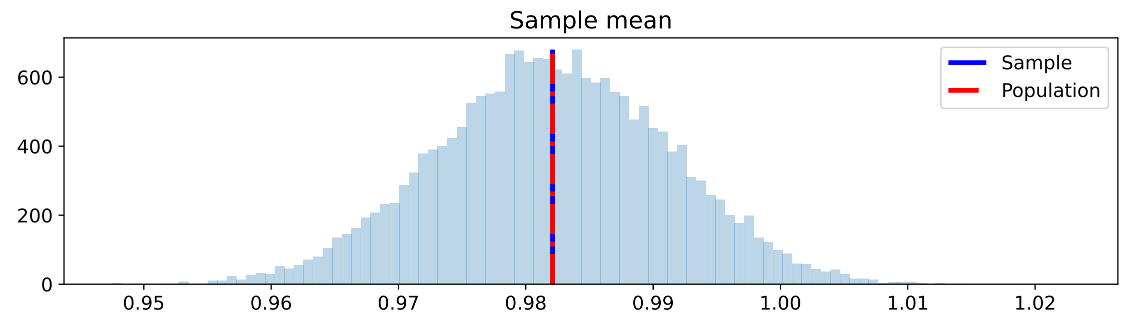

Figure 4: We replicate Figure 1 now for $N = 150$, $S = 80$, and $p = 0.5$.

Last, we compare the empirical distribution of the estimated variance of the sample mean to its analytical value in Figure 3. Again, the fact that the average value of the estimate $\widehat{\mathrm{Var}}(\widehat{\overline{\overline{Y}}}_w)$ coincides with the population value $\mathrm{Var}(\widehat{\overline{\overline{Y}}}_w)$ confirms the calculation in Appendix A.

We replicate the plots above, now for $N = 150$ and $S = 80$. Again we let $p = 0.5$ and the data be drawn from a Gaussian with unit mean and unit variance.

C.2 Efficiency of crossover designs

Consider the setting discussed in Remark 2: we have a real-valued matrix of potential outcomes indexed by a binary treatment, $Y(w) \in \mathbb{R}^{N \times S}$, where

$$y_{n,s}(w) = \mu + \tau 1(w = 1) + a_n + b_s + \epsilon_{n,s},$$

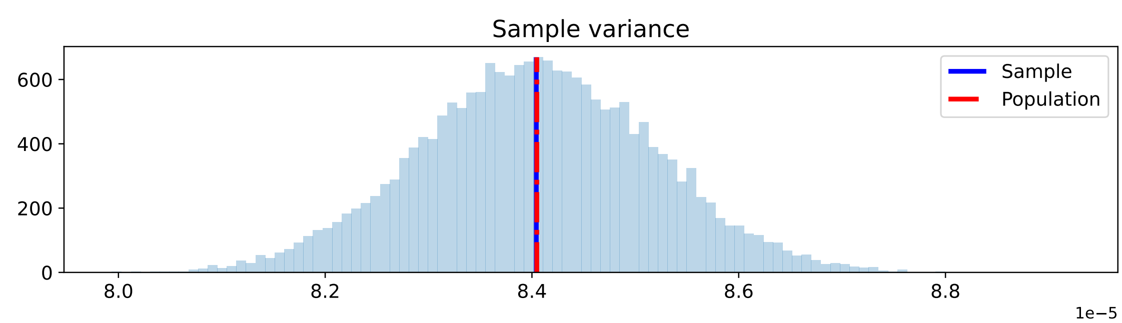

Figure 5: We replicate Figure 3 now for $N = 150$, $S = 80$, and $p = 0.5$.

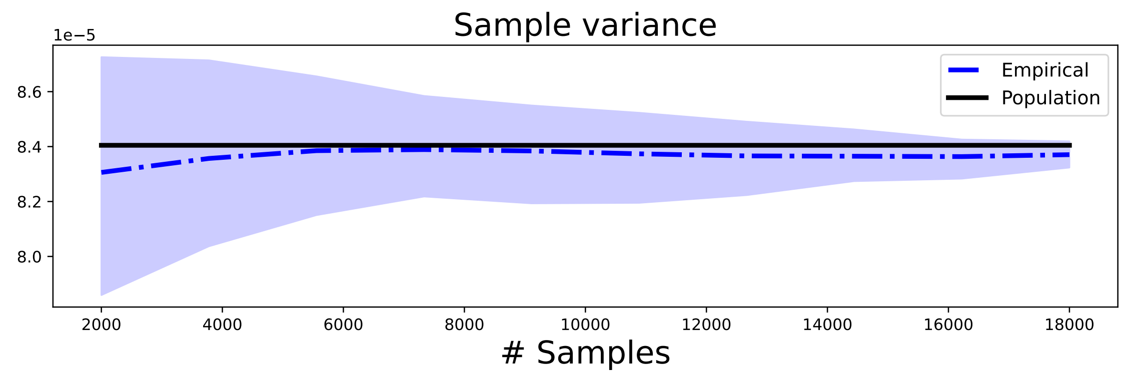

Figure 6: We replicate Figure 2 now for $N = 150$, $S = 80$, and $p = 0.5$.

where $\epsilon_{n,s} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$, drawn independent of other quantities.

To illustrate how the balanced crossover designs we consider in Appendix A can lead to significant efficiency gains even in the absence of interference, we consider the case in which potential outcomes are fixed, but drawn from the following distribution:

$$a_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_a^2) \text{ for } n = 1, \ldots, N, \quad \text{and} \quad b_s \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_b^2) \text{ for } s = 1, \ldots, S.$$

It follows from the properties of the Gaussian distribution that, letting $\sigma^2 := \sigma_a^2 + \sigma_b^2 + \sigma_\epsilon^2$:

$$y_{n,s}(w) \sim \mathcal{N}(\mu + \tau 1(w = 1), \sigma^2). \tag{31}$$

Moreover, for each row, the corresponding mean over the columns follows a Gaussian distribution:

$$\begin{aligned}\overline{y}_n^{\text{N}}(w) &= \frac{1}{S}\sum_s y_{n,s}(w) \\ &= \mu + \tau 1(w = 1) + a_n + \frac{1}{S}\sum_{s=1}^{S}(b_s + \epsilon_{n,s}) \\ &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu + \tau 1(w = 1), \sigma_a^2 + \frac{\sigma_b^2 + \sigma_\epsilon^2}{S}\right).\end{aligned}$$

Similarly, for each column the corresponding mean over the rows follows a Gaussian distribution:

$$\begin{aligned}\overline{y}_s^{\text{S}}(w) &= \frac{1}{N}\sum_n y_{n,s}(w) \\ &= \mu + \tau 1(w = 1) + \frac{1}{N}\sum_{n=1}^{N}(a_n + \epsilon_{n,s}) + b_s \\ &\overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\mu + \tau 1(w = 1), \frac{\sigma_a^2 + \sigma_\epsilon^2}{N} + \sigma_b^2\right).\end{aligned}$$

Balanced crossover designs First, consider the case in which we randomize using the balanced crossover design. The variance of the simple plug-in estimator $\widehat{\overline{\overline{Y}}}_w$ of Equation (12) was given in Lemma 8. Under the parametric Gaussian assumption of Equation (31), the expected value of this quantity over hypothetical re-draws of the data is given by:

$$\mathbb{E}_{\mathcal{N}}\left\{\text{Var}_{\mathfrak{W}}\left(\widehat{\overline{\overline{Y}}}_w\right)\right\} = \frac{\xi}{(pNS)^2}\mathbb{E}_{\mathcal{N}}\left[\delta(w)^2\right] \tag{32}$$

and since $\delta(w)^2 = \sum_{n,s}(y_{n,s}(\text{tr}))^2 - S\sum_n\left\{\overline{y}_n^{\text{N}}(\text{tr})\right\}^2 - N\sum_s\left\{\overline{y}_s^{\text{S}}(\text{tr})\right\}^2 + NS(\overline{\overline{y}}_{\text{tr}})^2$ from Lemma 9:

$$= \frac{\xi}{p^2NS}\left\{\sigma_\epsilon^2\left(1 - \frac{1}{N} - \frac{1}{S}\right) - \frac{\sigma_a^2}{N} - \frac{\sigma_b^2}{S}\right\} \tag{33}$$

$$= O\left(\frac{\sigma_\epsilon^2}{NS}\right). \tag{34}$$

By contrast, we consider as alternative designs the cases in which (i) we are assignment treatment to the (n,s)-th entry by means of a random Bernoulli coin flip with probability p — a completely randomized design, or (ii) a unit-randomized design, in which pN units are assigned (at random) to either control or treatment, and their assignment is constant over time and (iii) a time-randomized design, in which pS time-stamps are assigned (at random) to treatment, and every unit n gets to see the same experience at a given time stamp s.

Completely randomized designs In this case, where the matrix of assignments is a collection of i.i.d. coin flips, each with probability p, the estimator for the population mean has the same form as in Equation (9); however, the space of assignments $\mathfrak{W}^{\mathrm{CRD}}$ and the weighting distribution is different. We hence use the notation $\widehat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}$ to emphasize that this estimator is obtained from a completely randomized design. We now show that, when the potential outcomes satisfy the distributional requirement of Equation (31), the variance resulting from this randomization scheme is provably larger (in expectation). Before doing so, we state a useful technical lemma.

Lemma 14 (Approximating the variance of a ratio of random variables)**.** *Let X, Y be positive real valued random variables with finite first two moments, $\mathbb{E}[X] = \mu_X, \mathrm{Var}(X) = \sigma_X^2$ and $\mathbb{E}[Y] = \mu_Y, \mathrm{Var}(Y) = \sigma_Y^2$. Let $Z = X/Y$. A first order Taylor expansion approximation yields*

$$\mathrm{Var}(Z) \approx \left(\frac{\mu_X}{\mu_Y}\right)^2 \left[\frac{\sigma_X^2}{\mu_X^2} - 2\frac{\mathrm{Cov}(X,Y)}{\mu_X \mu_Y} + \frac{\sigma_Y^2}{\mu_Y^2}\right]. \tag{35}$$

Lemma 15 (Moments of $\widehat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}$)**.** *$\widehat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}$ is unbiased for $\overline{\overline{y}}_w$. Its variance can be approximated by*

$$\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}\right) \approx \left(\frac{1-p}{p}\right)\frac{1}{NS}\left(\frac{\sum_{n,s} y_{n,s}(w)^2}{NS} - \overline{\overline{y}}_w^2\right).$$

Then, using Lemma 15 together with the Gaussian parametric assumption of Equation (31) we see that:

$$\begin{aligned}\mathbb{E}_{\mathcal{N}}\left\{\mathrm{Var}_{\mathfrak{W}^{\mathrm{CRD}}}\left(\widehat{\overline{\overline{Y}}}_{w,\mathrm{CRD}}\right)\right\} &\approx \left(\frac{1-p_w}{p_w}\right)\frac{1}{NS}\left(\frac{\sum_{n,s}\mathbb{E}[y_{n,s}(w)^2]}{NS} - \mathbb{E}\left[\overline{\overline{y}}_w^2\right]\right) \\ &= \left(\frac{1-p_w}{p_w}\right)\frac{\sigma^2}{NS} = O\left(\frac{\sigma^2}{NS}\right) \end{aligned} \tag{36}$$

Exact formulae for the variance of the sample mean under the completely randomized design can be found, e.g., using the results in Strand [1979].

Individual-randomized designs In this case, our experiment is equivalent to a simple single randomized design with N units, in which each unit has potential outcome $y_n(w) = S^{-1}\sum_{s=1}^{S} y_{n,s}(w) = \overline{y}_n^{\mathrm{N}}(w)$. While the estimator for the population mean has (again) the same form as in Equation (9), we use $\widehat{\overline{\overline{Y}}}_{w,\rightarrow}$ to denote it, to emphasize the dependence on the space of valid assignments $\mathfrak{W}^{\mathrm{N}}$ (the space of binary matrices with constant columns, and such that exactly pS columns are equal to one). The variance of $\widehat{\overline{\overline{Y}}}_{w,\rightarrow}$ is

$$\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\rightarrow}\right) = \frac{1-p_w}{p_w}\frac{1}{N-1}\frac{1}{N}\sum_{n=1}^{N}\left(\overline{y}_n^{\mathrm{N}}(w) - \overline{\overline{y}}_w\right)^2.$$

In the Gaussian case of Equation (31), we can write

$$\begin{aligned}\mathbb{E}\left[\left(\overline{y}_n^{\mathrm{N}}(w) - \overline{\overline{y}}_w\right)^2\right] &= \mathbb{E}\left[\left(\overline{y}_n^{\mathrm{N}}(w)\right)^2 - 2\overline{y}_n^{\mathrm{N}}(w)\overline{\overline{y}}_w + (\overline{\overline{y}}_w)^2\right] \\ &= \mathbb{E}\left[\left(\overline{y}_n^{\mathrm{N}}(w)\right)^2\right] + \mathbb{E}\left[(\overline{\overline{y}}_w)^2\right] - 2\mathbb{E}\left\{\frac{1}{S}\sum_{s=1}^{S} y_{n,s}(w)\left(\frac{1}{NS}\sum_{m=1}^{N}\sum_{q=1}^{S} y_{m,q}(w)\right)\right\}\end{aligned}$$

The sum of the first two expectations in the equation above is equal to

$$\sigma_a^2 + \frac{\sigma_b^2 + \sigma_\epsilon^2}{S} + 2\left(\mu + \tau 1(w=1)\right)^2.$$

For the last term in the equation above, it holds

$$\frac{1}{NS^2}\mathbb{E}\left\{\sum_{m=1}^{N}\sum_{s=1}^{S}\sum_{q=1}^{S} y_{n,s}(w)y_{m,q}(w)\right\} = \frac{1}{NS^2}\sum_{s=1}^{S}\mathbb{E}\left[y_{n,s}(w)^2\right] + \frac{1}{NS^2}\sum_{s=1}^{S}\sum_{q\neq s}^{S}\mathbb{E}\left[y_{n,s}(w)y_{n,q}(w)\right] + \frac{1}{NS^2}\sum_{m\neq n}^{N}\sum_{s=1}^{S}\mathbb{E}\left[y_{n,s}(w)y_{m,s}(w)\right] + \frac{1}{NS^2}\sum_{m\neq n}^{N}\sum_{s=1}^{S}\sum_{q\neq s}^{S}\mathbb{E}\left[y_{n,s}(w)y_{m,q}(w)\right],$$

which implies

$$\begin{aligned}\sum_{m,s,q}\mathbb{E}\left\{\frac{y_{n,s}(w)y_{m,q}(w)}{NS^2}\right\} &= (\mu+\tau 1(w=1))^2 + \frac{1}{NS^2}\left[S\sigma^2 + S(S-1)\sigma_a^2 + (N-1)S\sigma_b^2\right] \\ &= (\mu+\tau 1(w=1))^2 + \frac{1}{NS^2}\left[S^2\sigma_a^2 + SN\sigma_b^2 + S\sigma_\epsilon^2\right] \\ &= (\mu+\tau 1(w=1))^2 + \frac{\sigma_a^2}{N} + \frac{\sigma_b^2}{S} + \frac{\sigma_\epsilon^2}{NS}.\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}_{\mathfrak{W}^{\mathrm{N}}}\left[\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\rightarrow}\right)\right] &= \frac{1-p_w}{p_w}\frac{1}{N-1}\left(\sigma_a^2 + \frac{\sigma_b^2+\sigma_\epsilon^2}{S} - 2\left[\frac{\sigma_a^2}{N} + \frac{\sigma_b^2}{S} + \frac{\sigma_\epsilon^2}{NS}\right]\right) \\ &= O\left(\frac{\sigma_a^2}{N} + \frac{\sigma_\epsilon^2}{NS}\right).\end{aligned}$$

Time-randomized designs Symmetrically to the case above, here our experiment is equivalent to a simple single randomized design with S units, in which each unit has potential outcome $y_s(w) = S^{-1}\sum_{n=1}^{N} y_{n,s}(w) = \overline{y}_s^{\mathrm{S}}(w)$. While the estimator for the population mean has (again) the same form as in Equation (9), we use $\widehat{\overline{\overline{Y}}}_{w,\downarrow}$ to denote it, to emphasize the dependence on the space of valid assignments $\mathfrak{W}^{\mathrm{S}}$ (the space of binary matrices with constant rows, and such that exactly pN rows are equal to one). The variance of $\widehat{\overline{\overline{Y}}}_{w,\downarrow}$ is

$$\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\downarrow}\right) = \frac{1-p_w}{p_w}\frac{1}{S-1}\frac{1}{S}\sum_{s=1}^{S}\left(\overline{y}_s^{\mathrm{S}}(w) - \overline{\overline{y}}_w\right)^2.$$

A symmetric derivation to the one above for the unit-randomized design yields:

$$\begin{aligned}\mathbb{E}_{\mathfrak{W}^{\mathrm{S}}}\left[\mathrm{Var}\left(\widehat{\overline{\overline{Y}}}_{w,\downarrow}\right)\right] &= \frac{1-p_w}{p_w}\frac{1}{S-1}\left(\sigma_b^2 + \frac{\sigma_a^2+\sigma_\epsilon^2}{N} - 2\left[\frac{\sigma_b^2}{S} + \frac{\sigma_a^2}{N} + \frac{\sigma_\epsilon^2}{NS}\right]\right) \\ &= O\left(\frac{\sigma_b^2}{S} + \frac{\sigma_\epsilon^2}{NS}\right).\end{aligned}$$

Then, we see that the expected variance of the simple estimator $\widehat{\overline{\overline{Y}}}_w$ obtained from a staggered design is smaller than the corresponding variance incurred by adopting other copmeting designs (completely randomized design, time-randomized, unit-randomized). In particular, this is because — by virtue of the balancing — the effect of the variance components σ_a^2 and σ_b^2 is of second order — whereas these terms appear as leading terms in the other variances.