Correlated Cluster-Based Randomized Experiments: Robust Variance Minimization

(*Online Appendix*)

Ozan Candogan[1], Chen Chen[2], and Rad Niazadeh[1]

[1]The University of Chicago Booth School of Business
[2]New York University Shanghai
ozan.candogan@chicagobooth.edu, cc8029@nyu.edu, rad.niazadeh@chicagobooth.edu

First version: August 16, 2021
This version: June 20, 2023

EC1 More on Related Literature

Foundations of Experimental Design Experimental design has found far-reaching applications to guide decision making in areas ranging from medical trials to social sciences, and recently in online marketplaces and social networks. It is grounded in causal inference, but instead of inferring causality from purely observational data, a decision maker can choose how to gather data to gain more statistical efficiency and more convincing empirical evidence of causality. The process often involves randomization, and a more careful design of randomization based on optimization helps to maximize the statistical power. Great expositions of related topics include Owen (2020), Kohavi et al. (2020), and Imbens and Rubin (2015).

Regression Adjustment in Experimental Design There is a rich literature on variance reduction via regression adjustment with covariate information in randomized experiments (e.g., Lin 2013, Jin and Ba 2021, and the references therein). These works fix the joint treatment assignment to be completely random and consider the best analysis (i.e., the estimator) to minimize the variance in the asymptotic regime (under certain settings). They show that agnostic regression can reduce variance. Our work takes the opposite direction. Specifically, we fix the estimator (i.e., we deliberately use a common estimator—the simple unadjusted Horvitz-Thompson estimator) and study the best design. We view this as a building block that might eventually enable combining the two steps and optimizing them jointly. Another distinction is that we provide a bound on performance loss for any problem instance, and such a bound implies that our design is asymptotically optimal under the adversarial model. By contrast, the regression adjustment literature has mainly focused on the asymptotic analysis.

Pair-Matching Experiments The pair-matching experiment pairs similar (in terms of size and/or related covariates) clusters and randomly assigns one cluster from each pair to treatment. It is a special case of the IBR experiments where each block contains two clusters and the marginal assignment probability $q = \frac{1}{2}$. Imai et al. (2009) recommend a pair-matching experiment if cluster-based randomization is used. While the authors do not provide a theoretical justification for this recommendation, they offer empirical evidence for the usefulness of this design based on comparison with other heuristic experiments. Our work, by contrast, provides a theoretical foundation for the aforementioned design. Specifically, we show that an optimal IBR experiment can achieve much of the benefit from the optimal (correlated) randomized assignment across clusters, and is asymptotically optimal when the number of clusters grows large under a mild regularity assumption on cluster sizes. On the other hand, although the pair-matching experiment is not asymptotically optimal, its worst-case variance is guaranteed to be within $\frac{5}{2}$ of the variance of the optimal experiment (Remark 4.4), which is slightly larger than that of the optimal IBR experiment. Finally, Bai (2022) shows that certain pair-matching experiments are optimal under stratified randomization (our IBR experiments take the same form; please refer to Remark 3.1) using a different sampling-based model where units' potential outcomes are independently sampled.

Alternative Approaches to Experimental Design Other recent papers have considered models in which a covariate of a unit is correlated with the potential outcome in a certain way. The optimal experiment usually involves covariate balancing, such that the treatment and control groups are similar in terms of the covariates (Bertsimas et al. 2015, Bertsimas et al. 2019, Kallus 2018, Bhat et al. 2020, and Harshaw et al. 2019). We instead work with a model of potential outcomes with minimal assumptions on their ranges (nevertheless, these ranges can still be inferred from the covariate information).

The rich literature on online learning and multi-armed bandits (see, e.g., Lattimore and Szepesvári 2020 and Slivkins 2019 for surveys) can also be viewed as a form of adaptive and sequential experimental design. There, the decision maker is allowed to switch between variants (i.e., arms), the system is assumed to be stationary and have rapid feedback (i.e., no carryover effect), and the objective is to find the best variant with minimum cumulative regret or number of trials; see, e.g., Hadad et al. (2019) and Bibaut et al. (2021). Finally, we point out that the experimental design problem of minimizing the variance of an (unbiased) estimator is also very related to various variance-reduction methods in the simulation literature (e.g., Asmussen and Glynn 2007).

EC2 Proofs

EC2.1 Details of Remark 2.2

Suppose there are n Bernoulli random variables $Z = (Z_i)_{i=1}^n$, each with a marginal probability $\mathbb{P}[Z_i] = \frac{1}{2}$. The joint distribution of Z that minimizes the variance $\mathbb{V}\text{ar}\Big[\sum_{i=1}^n w_i Z_i\Big]$ (with known weights w_i) is as follows: partition the weights $\{w_i\}$ into two groups as even as possible, then randomly select one group with probability $\frac{1}{2}$, let $Z_i = 1$ for all units in that group and $Z_i = 0$ for the rest units.

Proof. For any joint distribution of $Z = (Z_i)_{i\in[n]}$ with $\mathbb{P}[Z_i] = \frac{1}{2}$ for all $i \in [n]$, the mean value is fixed, i.e., $\mathbb{E}[\sum_{i\in[n]} w_i Z_i] = \frac{1}{2}\sum_{i\in[n]} w_i$. For a set $S \subseteq [n]$, let $w(S) \triangleq \sum_{i\in S} w_i$ denote the sum of weights of units in set S and $d(S) \triangleq \Big|w(S) - \frac{1}{2}\sum_{i\in[n]} w_i\Big|$ denote the absolute difference between the weight of set S and the mean value $\frac{1}{2}\sum_{i\in[n]} w_i$. Let $S^* \triangleq \text{argmin}_{S\subseteq[n]} d(S)$ and $d^* = d(S^*)$; clearly, we have $w(S^*) = w\big([n]\backslash S^*\big) \le w(S)$ for any set $S \subseteq [n]$, and $\big(S^*, [n]\backslash S^*\big)$ is a balanced partition of $\{w_i\}$.

On the other hand, the variance satisfies

$$\mathbb{V}\text{ar}\Big[\sum_{i=1}^n w_i Z_i\Big] = \mathbb{E}\left[d\big(\{i : Z_i = 1\}\big)^2\right] \ge d^*,$$

where the inequality holds by the definition of d^*. Furthermore, if all units $i \in S^*$ take $Z_i = 1$ and the rest take $Z_i = 0$ with probability $\frac{1}{2}$, and vice versa with probability $\frac{1}{2}$, the equality is attained; thus, this joint random assignment satisfies $\mathbb{P}[Z_i] = \frac{1}{2}$ for all units $i \in [n]$, and it minimizes the variance. $\square$

EC2.2 Details of Remark 2.3

Suppose that the uncertainty set of each cluster i's potential outcomes is $y_{i1} \in [-w_{i1}, w_{i1}]$ and $y_{i0} \in [-w_{i0}, w_{i0}]$; then the constraint in (1) requires $y_i \in [-w_i, w_i]$ for each cluster i, with $w_i = \sqrt{q(1-q)} \cdot \left(\frac{w_{i1}}{q} + \frac{w_{i0}}{1-q}\right)$. In this case, we claim that the optimal experiment is to simply assign treatment to each cluster independently with probability q. The correlation matrix with such independent assignment is the identity matrix, i.e., $\Sigma^* = I$. The worst-case potential outcome is $y_i = w_i$ for each cluster i, and the worst-case variance is $y^\mathrm{T}\Sigma^* y = \sum_{i\in[n]} w_i^2$.

We now show no experiment can achieve a worst-case variance strictly smaller than $\sum_{i\in[n]} w_i^2$. To see this, consider any feasible experiment and let Σ denote the corresponding correlation matrix. Consider the following randomized potential outcomes with Y_i being either w_i or $-w_i$ with equal probability, and let these Y_i be independent. Then, $\mathbb{E}\big[Y_i^2\big] = w_i^2$ and $\mathbb{E}\big[Y_i Y_k\big] = 0$ for any $i \neq k$. Thus, letting $Y = (Y_i)_{i\in[n]} \in \mathbb{R}^n$ be the concatenation, we have

$$\max_{y \in \times_{i\in[n]}[0,w_i]} y^{\mathrm{T}}\Sigma y \geq \mathbb{E}\Big[Y^{\mathrm{T}}\Sigma Y\Big] = \sum_{i\in[n]} w_i^2,$$

which clearly shows the optimality of independently assigning treatments to each cluster.

EC2.3 Proof of Proposition 3.1

Since the set of joint assignment distributions $\mathcal{P}_q$ is a polyhedron and $\Sigma(P)$ is a linear map of a distribution $P \in \mathcal{P}_q$, we can write (1) as a linear program with exponentially many decision variables and constraints as in (EC-1), i.e.,

$$\begin{aligned} &\underset{P\in\mathcal{P}_q,\, z\in\mathbb{R}}{\text{minimize}} && z \\ &\text{subject to} && z \geq y^{\mathrm{T}}\Sigma(P)y,\ \forall\, y \in \bigtimes_{i\in[n]} \{0, w_i\}, \end{aligned} \tag{EC-1}$$

where we have polyhedral constraints of $\mathcal{P}_q$ and we have one constraint for each extreme point of the potential outcomes' uncertainty set. This formulation also implies that a convex combination of optimal experiments is an optimal experiment as well.

Since clusters have equal sizes, this observation implies that given any optimal experiment, we can always construct another optimal experiment that treats clusters in an identical way. As a result, it is without loss of generality to assume that $\sigma_{ik} = \sigma$ for any two clusters. Thus, the variance of the estimator with a given set of potential outcomes becomes

$$y^{\mathrm{T}}\Sigma y = \sum_{i\in[n]} y_i^2 + \sigma \sum_{i\in[n]} \sum_{[n]\ni k\neq i} y_i y_k.$$

Since all of the potential outcomes are nonnegative, the optimal experiment tries to make σ as small as possible.

Let $S = \sum_{i\in[n]} Z_i$ denote the total number of clusters that receive treatments. We have

$$\mathbb{V}\mathrm{ar}[S] = q(1-q)\Big[n + n(n-1)\sigma\Big],$$

simply because $\mathbb{C}\mathrm{ov}\,[Z_i, Z_j] = q(1-q)\sigma$ for any two clusters $i \neq j$. Thus, to minimize the correlation σ, it is equivalent to minimizing the variance of the summation S. Since the mean value $\mathbb{E}[S] = qn$ is fixed, to minimize the variance $\mathbb{V}\mathrm{ar}[S]$, the (exchangeable) joint assignment should let the summation S concentrate around the mean as much as possible.[EC1]

Case One: Suppose that qn is an integer. Since $\mathbb{V}\mathrm{ar}[S] \geq 0$, we have $\sigma \geq -\frac{1}{n-1}$. The equality is attained when S is constant with probability one, and this can be achieved by assigning qn clusters to treatment uniformly at random. We next study the worst-case potential outcomes and the worst-case variance. Specifically, let $h \in \mathbb{N}$ be the number of clusters that have an outcome

[EC1]Similar problems are studied under more general marginal distributions (i.e., beyond Bernoulli distributions) in Rüschendorf and Uckelmann (2002) and Section 3.6 of Rachev and Rüschendorf (1998).

$y_i = 1$, and suppose that the other clusters have the outcome $y_i = 0$. Then,

$$V^{\text{OPT}} = \max_{h\in[0:n]} h + h(h-1)\sigma. \tag{EC-2}$$

Denote by $h^* \in [0:n]$ the integer that maximizes this quantity; it is easy to see that h^* is the integer closest to $-\frac{1}{2\sigma} + \frac{1}{2} = \frac{n}{2}$. Hence, when n is even, we have $h^* = \frac{n}{2}$ and $V^{\text{OPT}} = \frac{1}{4}\frac{n^2}{n-1}$; and when n is odd, h^* is either $\frac{n+1}{2}$ or $\frac{n-1}{2}$, and $V^{\text{OPT}} = \frac{n+1}{4}$.

Case Two: Suppose that qn is not an integer. Let $p = \mathbb{P}[S < qn]$, $\underline{s} = \mathbb{E}[S|S < qn]$, and $\bar{s} = \mathbb{E}[S|S > qn]$. Since $\mathbb{E}[S] = qn$, by the law of total expectation, we have

$$p \cdot (qn - \underline{s}) = (1-p) \cdot (\bar{s} - qn). \tag{EC-3}$$

Moreover, by the law of total variance,

$$\mathbb{V}\text{ar}[S] \geq p(qn - \underline{s})^2 + (1-p)(\bar{s} - qn)^2 = (qn - \underline{s})(\bar{s} - qn), \tag{EC-4}$$

where the equality follows from (EC-3), and an equality is attained at the inequality if S is a constant conditioning on $S < qn$ and $S > qn$, respectively.

Since the number of clusters in treatment S takes integral values, we have $\underline{s} \leq \lfloor qn \rfloor$ and $\bar{s} \geq \lceil qn \rceil$. As a result, from (EC-4), the optimal experiment picks $\underline{s} = \lfloor qn \rfloor$ and $\bar{s} = \lceil qn \rceil = \underline{s} + 1$, and it uniformly at random treats $\underline{s}$ clusters (and thus let $S = \underline{s}$) with probability p and uniformly at random treats $\bar{s}$ units (and thus let $S = \bar{s}$) with probability $1-p$. Moreover, by (EC-3), $p = \frac{\bar{s} - qn}{\bar{s} - \underline{s}} = \lceil qn \rceil - qn$. The correlation of any two assignments is

$$\sigma = \frac{p \cdot \frac{\underline{s}}{n} \cdot \frac{\underline{s}-1}{n-1} + (1-p) \cdot \frac{\bar{s}}{n} \cdot \frac{\bar{s}-1}{n-1} - q^2}{q(1-q)} = -\frac{nq(1-q) - p(1-p)}{n(n-1)q(1-q)}. \tag{EC-5}$$

To see that $\sigma < 0$, it suffices to show that $nq(1-q) - p(1-p) > 0$. First, if $qn < 1$, then $\underline{s} = 0$, $\bar{s} = 1$, and $p = 1 - qn$; thus, $nq(1-q) - p(1-p) = n(n-1)q^2 > 0$. On the other hand, suppose that $qn > 1$. Then, since $qn > 1 > \max\{p, 1-p\}$ and $1 - q \geq \frac{1}{2} \geq \min\{p, 1-p\}$, we again have $nq(1-q) > p(1-p)$. The fact that $\sigma > -\frac{1}{n-1}$ follows from (EC-5) as $p = \lceil qn \rceil - qn \in (0,1)$. Also note that in (EC-5), $\sigma = -\frac{1}{n-1}$ if $p = \lceil qn \rceil - qn = 0$, which happens when qn is an integer; thus, (EC-5) also covers Case One.

Finally, let h^* denote the integer closest to $\min\left\{-\frac{1}{2\sigma} + \frac{1}{2}, n\right\}$. By (EC-2), in the worst case, h^* clusters have the outcome $y_i = 1$ and the other clusters have the outcome $y_i = 0$. When $n \to \infty$, we have $\sigma \to -\frac{1}{n-1}$ by (EC-5) and therefore following a similar analysis to Case One, $V^{\text{OPT}} \to \frac{n}{4}$.

EC2.4 Proof of Lemma 3.3

Since the objective of $\max_{y_i \in [0, w_i]} y^{\mathrm{T}} \Sigma y$ is jointly convex in y, in the worst case, $y_i \in \{0, w_i\}$ for each cluster $i \in [k]$. We first prove that there exists a worst-case potential outcome y such that $y_i = w_i$ for $i \leq r$ for some integer $r \in [k]$, and $y_i = 0$ for $i > r$. If not, then for any worst-case potential outcome y, there exists indices $i < j$ such that $y_i = 0$ and $y_j = w_j$, whereas $w_i \geq w_j$. Since all the off-diagonals of the correlation matrix Σ have the same value of σ, the objective does not change if we instead swap y_i and y_j, and let $y_i = w_j \in [0, w_i]$ and $y_j = 0$. We can further (weakly) increase the objective by setting y_i to one of its extreme values, i.e., $y_i = w_i$ or $y_i = 0$. By iterating this process, we end up with a worst-case outcome that satisfies our desired property.

It remains to determine the value of r. Let y denote the worst-case potential outcome vector such that $y_i = w_i$ for $i \le r$ and $y_i = 0$ for $i > r$. Note that

$$y^{\mathrm{T}}\Sigma y = \sum_{i\in[k]} y_i^2 + \sigma \cdot \sum_{i\in[k]} \sum_{[k]\ni j\neq i} y_i y_j.$$

Observe that if we update $y_i = 0$ for some cluster $i \le r$, the variance changes by

$$-w_i^2 - 2\sigma \sum_{j\le r \text{ and } j\neq i} w_i w_j \le 0.$$

Similarly, if we set $y_i = w_i$ for some cluster $i > r$, the variance changes by

$$w_i^2 + 2\sigma \sum_{j=1}^{r} w_i w_j \le 0.$$

Together these imply that

$$\forall\, i \le r: \qquad w_i \ge -2\sigma \sum_{j\le r \text{ and } j\neq i} w_j, \tag{EC-6}$$

$$\forall\, i > r: \qquad w_i \le -2\sigma \sum_{j=1}^{r} w_j. \tag{EC-7}$$

Let $r^\star$ be the largest index such that $w_{r^\star} \ge -2\sigma \sum_{i\le r^\star-1} w_i$. Since the cluster sizes are decreasing, $r^\star$ satisfies (EC-6) and (EC-7). Moreover, if $w_{r^\star} > -2\sigma \sum_{i\le r^\star-1} w_i$, $r = r^\star$ is the only integer that satisfies both (EC-6) and (EC-7), and hence corresponds to a worst-case potential outcome. If $w_{r^\star} = -2\sigma \sum_{i\le r^\star-1} w_i$, then both $r = r^\star$ and $r = r^\star - 1$ satisfy (EC-6) and (EC-7), and they both constitute a worst-case potential outcome.

EC2.5 Proof of Lemma 4.2

The lower bound $w_1^2 \le V^{\mathrm{LB}}$ in Lemma 4.2 is trivial because $y^{\mathrm{T}}\Sigma y = w_1^2$ with the outcome vector y such that $y_1 = w_1$ and $y_i = 0$ for all $i \ge 2$, for any correlation matrix Σ. Thus, we only need to prove the inequality $\frac{1}{4}\sum_{i\in[n]} w_i^2 \le V^{\mathrm{LB}}$.

To prove this, consider the following randomized potential outcomes where each Y_i is either 0 or w_i with equal probability, and these Y_i are independent. Then, we have $\mathbb{E}\big[Y_i^2\big] = w_i^2/2$ and $\mathbb{E}\big[Y_i Y_k\big] = w_i w_k/4$ for any $i \neq k$. Let $Y = (Y_i)_{i\in[n]}$ be the concatenation of these randomized potential outcomes and $w = (w_i)_{i\in[n]}$ be the vector of cluster sizes. For any correlation matrix $\Sigma = (\sigma_{ik})_{i,k\in[n]} \in \Re$, we have

$$\max_{y\in \times_{i\in[n]}[0,w_i]} y^{\mathrm{T}}\Sigma y \ge \mathbb{E}\Big[Y^{\mathrm{T}}\Sigma Y\Big] = \frac{1}{2}\sum_{i\in[n]} w_i^2 + \frac{1}{4}\sum_{i\in[n]}\sum_{[n]\ni k\neq i} w_i w_k \sigma_{ik} = \frac{1}{4}\sum_{i\in[n]} w_i^2 + \frac{1}{4} w^{\mathrm{T}}\Sigma w \ge \frac{1}{4}\sum_{i\in[n]} w_i^2.$$

Thus, $V^{\mathrm{LB}} \ge \frac{1}{4}\sum_{i\in[n]} w_i^2$, which completes the proof of the lemma.

EC2.6 Proof of Lemma 4.3

We start by finding an upper bound on the worst-case variance within a group. First, it is clear that the worst-case variance increases with the size w_i of any cluster i in the group because increasing w_i

enlarges the potential outcomes' uncertainty set. Thus, for a group with at most k clusters where the largest cluster size is equal to w, the worst-case variance is largest when there are exactly k clusters and all cluster sizes are equal to w. By Proposition 3.1, the optimal partition is to have all the k clusters in one block, and the corresponding worst-case variance is $f_q(k) \cdot w^2$.

Next, let $N = \lceil \frac{n}{k} \rceil$ be the number of groups. By the above analysis, we have

$$V^k \leq f_q(k) S_k, \tag{EC-8}$$

where $S_k = \sum_{i=1}^N w_{(i-1)k+1}^2$ is the sum of squares of the largest cluster sizes in each group. We claim that

$$4V^{\text{LB}} \geq \sum_{i \in [n]} w_i^2 \geq k \cdot (S_k - w_1^2) + w_1^2. \tag{EC-9}$$

The first inequality in (EC-9) follows from Lemma 4.2. The second inequality in (EC-9) holds, because

$$\begin{aligned} \sum_{i \in [n]} w_i^2 \geq w_1^2 + \sum_{i=2}^{(N-1)k+1} w_i^2 = w_1^2 + \sum_{h=1}^{N-1} \sum_{i=(h-1)k+2}^{hk+1} w_i^2 \\ \overset{(*)}{\geq} w_1^2 + k \cdot \sum_{h=1}^{N-1} w_{hk+1}^2 = w_1^2 + k \cdot (S_k - w_1^2), \end{aligned}$$

where the above inequality $(*)$ follows from the fact that clusters $(h-1)k+2$ to hk have weakly larger sizes than the cluster $hk+1$. Now, by rearranging the terms in (EC-9), we have

$$S_k \leq \frac{4V^{\text{LB}}}{k} + \frac{k-1}{k} w_1^2.$$

Combining this with (EC-8) yields

$$\frac{V^k}{V^{\text{LB}}} \leq f_q(k) \cdot \frac{S_k}{V^{\text{LB}}} \leq f_q(k) \left(\frac{4}{k} + \frac{k-1}{k} \frac{w_1^2}{V^{\text{LB}}} \right) \leq f_q(k) \left(\frac{4}{k} + \frac{k-1}{k} \right),$$

where the last inequality follows from $w_1^2 \leq V^{\text{LB}}$ by Lemma 4.2.

EC2.7 Proof of Lemma 4.7

It suffices to prove Lemma 4.7 only for $k \geq k_0 \triangleq \frac{1}{2q(1-q)} \geq 2$, because when $k \leq k_0$, $f_q(k)$ can be uniformly bounded from above by a constant. By Proposition 3.1, $\sigma = -\frac{nq(1-q)-p(1-p)}{n(n-1)q(1-q)} < 0$ with $p = \lceil qn \rceil - qn$; hence, we have

$$f_q(k) = \max_{h \in [k]} \{h + h(h-1)\sigma\} \leq \max_{h \in \mathbb{R}} \{h + h(h-1)\sigma\} = \frac{k^3 q(1-q)}{4(k-1)\big[kq(1-q) - p(1-p)\big]},$$

where the second equality follows by taking $h = -\frac{1}{2\sigma} + \frac{1}{2}$ that maximizes the quadratic objective. Since $p(1-p) \leq \frac{1}{4} = \frac{k_0}{2} \cdot q(1-q)$, we have

$$4f_q(k) - k \leq \frac{k^3}{(k-1)\big(k - \frac{k_0}{2}\big)} - k = \frac{k^2\big(1 + \frac{k_0}{2}\big) - \frac{k_0}{2}k}{(k-1)\big(k - \frac{k_0}{2}\big)} \leq \left(1 + \frac{k_0}{2}\right) \frac{1}{1 - \frac{1}{k}} \frac{1}{1 - \frac{k_0}{2k}}.$$

By the fact that $\frac{1}{1-x} \leq 1 + 2x$ for $x \in [0, \frac{1}{2}]$ and that $k \geq k_0 \geq 2$, we have

$$4f_q(k) - k \leq \left(1 + \frac{k_0}{2}\right)\left(1 + \frac{2}{k}\right)\left(1 + \frac{k_0}{k}\right) \leq 2\left(1 + \frac{k_0}{2}\right)\left(1 + \frac{2}{k_0}\right) = 4 + 8q(1-q) + \frac{1}{2q(1-q)},$$

which is bounded from above by a constant.

EC2.8 Details and Proof of Correctness of Example 4.1

Throughout this section, we focus on the setting of Example 4.1. For notational convenience, we index the blocks of an IBR experiment in decreasing order of size from the largest cluster in each block. Also, for simplicity, we assume an even number n of clusters. We first show in Lemma EC2.1 that when a block contains four or more clusters, at least $p \geq 2$ clusters take positive values (i.e., non-zero outcomes) in the worst-case potential outcome.

Lemma EC2.1. *Suppose that a block contains an even number $k \geq 4$ clusters, with cluster sizes $w_i = \beta^{k-i} w_k$ for any $i \in [k]$. Then, at least $p \geq 2$ clusters take positive values in the worst-case potential outcome $y = (y_i)_{i\in[k]}$.*

Proof. Since the number of clusters k is even, the correlation between any two clusters is $\sigma = -\frac{1}{k-1}$ by Corollary 3.2. By Lemma 3.3, p is the largest integer that satisfies

$$\beta^{k-p} w_k = w_p \geq \frac{2}{k-1} \cdot \sum_{i=1}^{p-1} w_i = \frac{2}{k-1} \cdot \sum_{i=1}^{p-1} \beta^{k-i} w_k = \frac{2}{k-1} \frac{\beta^p - \beta}{\beta - 1} \cdot \beta^{k-p} w_k.$$

This implies that $p = \left\lfloor \frac{\ln\left(\frac{k-1}{2}(\beta-1)+\beta\right)}{\ln \beta} \right\rfloor \geq 2$ when $k \geq 4$. □

We next show in Lemma EC2.2 that with the optimal partition, all blocks contain either two or four clusters.

Lemma EC2.2. *Suppose that the number of clusters n is even. All blocks in an optimal partition contain either 2 or 4 clusters.*

Proof. By Lemma EC3.3, each block contains an even number of clusters. Suppose that a block instead contains k clusters where k is an even number satisfying $k \geq 6$. Without loss of generality, we assume that $w_i = \beta^{k-i}$ for each cluster $i \in [k]$ (since we can always normalize cluster sizes by the size of the smallest cluster). We claim that we can further partition this block into two smaller blocks to reduce the worst-case variance. Specifically, the first block contains the first $k-2$ clusters of the original block, and the second block contains the other two blocks.

First, consider the worst-case variance associated with the two new blocks, denoted by V_a. Let p be the number of positive values in the worst-case potential outcome of the first block. By Lemma EC2.1, $p \geq 2$ because block one contains $k - 2 \geq 4$ clusters. We have

$$V_a = \left(\sum_{i\in[p]} w_i^2 - \sum_{i\in[p]} \sum_{[p]\ni j\neq i} \frac{1}{k-3} w_i w_j\right) + \beta^2.$$

Here β^2 is simply the worst-case variance of the second block, because in the worst case, cluster $k-1$ has a positive outcome and cluster k has outcome zero (this follows from Lemma EC3.2).

Let V_b denote the worst-case variance of the original block. It satisfies

$$V_b \geq \sum_{i\in[p]} w_i^2 - \sum_{i\in[p]} \sum_{[p]\ni j\neq i} \frac{1}{k-1} w_i w_j,$$

because the correlation between any two clusters in the original block is $-\frac{1}{k-1}$, and p is not necessarily the number of positive values in the worst-case potential outcome of the original block. Thus,

$$\begin{aligned} V_b - V_a &\geq \sum_{i\in[p]} \sum_{[p]\ni j\neq i} \left(\frac{1}{k-3} - \frac{1}{k-1}\right) w_i w_j - \beta^2 \\ &\geq 2\left(\frac{1}{k-3} - \frac{1}{k-1}\right) w_1 w_2 - \beta^2 \\ &= \frac{4\beta^{2k-3}}{(k-1)(k-3)} - \beta^2, \end{aligned}$$

which is nonnegative when $k \geq 6$. Hence, splitting the large block into two smaller blocks reduces the worst-case variance. □

Finally, we show in Lemma EC2.3 that all blocks contain four clusters in the optimal partition.

Lemma EC2.3. *Suppose that the number of clusters n is even. Then the optimal partition of an IBR experiment satisfies the following:*

- *If n is divisible by 4, all blocks contain exactly 4 clusters;*
- *Otherwise, the last block contains 2 clusters, and all the other blocks contain exactly 4 clusters.*

Proof. By Lemma EC2.2, each block contains either two or four clusters. For a two-cluster block, the worst-case outcome is simply the large cluster having a positive outcome and the small cluster having the zero outcome. For a four-cluster block, by the proof of Lemma EC2.1, the worst-case potential outcome is only the two largest clusters taking positive outcomes.

Let K be the number of blocks. It suffices to show that all of the first $K-1$ blocks contain four clusters. Suppose by way of a contradiction that block $h \leq K-1$ contains two clusters, k and $k+1$. If block $h+1$ contains two clusters $(k+2,\, k+3)$ as well, then the above observations imply that merging blocks h and $h+1$ decreases the worst-case variance by

$$w_{k+3}^2 \cdot \left\{ (\beta^6 + \beta^2) - \left(\beta^6 + \beta^4 - \frac{2}{3}\cdot\beta^5\right) \right\} > 0.$$

Now suppose that block $h+1$ contains four clusters. Suppose that we reconstruct blocks h and $h+1$ by assigning clusters k to $k+3$ to block h and clusters $k+4$ and $k+5$ to block $h+1$. Then, the worst-case variance decreases by

$$w_{k+5}^2 \cdot \left\{ \left(\beta^{10} + \beta^6 + \beta^4 - \frac{2}{3}\cdot\beta^5\right) - \left(\beta^{10} + \beta^8 - \frac{2}{3}\cdot\beta^9 + \beta^2\right) \right\} > 0.$$

Thus, it follows that it is not optimal for any block $h \leq K-1$ to contain two clusters. □

We now show that the optimal IBR experiment is asymptotically suboptimal. Consider a block with four clusters with sizes $w_i = \beta^{4-i}$ for $i \in [4]$. The worst-case variance from randomly assigning half of the clusters to treatment is $v^{\text{half}} = \beta^6 + \beta^4 - \frac{2}{3}\cdot\beta^5 = 4.222$. The worst-case variance from the optimal randomized joint assignment (given in Table EC-1) is $v^{\text{OPT}} = 3.815$. Now, let us revisit Example 4.1. For simplicity, we assume that the number of clusters n is divisible by four. Since every block in the optimal partition contains 4 clusters by Lemma EC2.3 and the blocks are identical up to a scaling, the worst-case variance from the optimal IBR experiment is

$$V^{\text{DP}} = \sum_{i=1}^{n/4} w_{4i-3}^2 \cdot v^{\text{half}}.$$

For an experiment that assigns clusters in a block to treatment in an optimal way, and does so independently across blocks, the worst-case variance, denoted by V, is

$$V = \sum_{i=1}^{n/4} w_{4i-3}^2 \cdot v^{\text{OPT}}.$$

Thus,

$$\frac{V^{\text{DP}} - V^{\text{OPT}}}{V^{\text{OPT}}} \geq \frac{V^{\text{DP}} - V}{V} = \frac{v^{\text{half}} - v^{\text{OPT}}}{v^{\text{OPT}}} = 10.7\%,$$

which implies that the optimal IBR experiment is asymptotically strictly suboptimal.

	1	2	3	4	probability
1	$\surd$	$\times$	$\surd$	$\times$	0.1700
2	$\times$	$\surd$	$\times$	$\surd$	0.1700
3	$\surd$	$\surd$	$\times$	$\times$	0.1500
4	$\times$	$\times$	$\surd$	$\surd$	0.1500
5	$\surd$	$\times$	$\times$	$\surd$	0.0954
6	$\times$	$\surd$	$\surd$	$\times$	0.0954
7	$\surd$	$\times$	$\times$	$\times$	0.0846
8	$\times$	$\surd$	$\surd$	$\surd$	0.0846

Table EC-1: The optimal randomized joint assignment of treatment to four clusters, with cluster sizes $w_i = \beta^{4-i}$ for $i \leq 4$ and $\beta = \frac{5}{4}$. Each of the 8 rows corresponds to an assignment, where $\surd$ denotes treatment and $\times$ denotes control.

EC2.9 Proof of Theorem 4.9

Throughout the proof, we refer to a cluster of size w_k as a cluster of type k. First, using a similar argument to the one in the proof of Proposition 3.1, we have that any convex combination of optimal correlation matrices in (5) (which is the optimization problem that defines V^{LB}) is an optimal correlation matrix as well. Thus, there exists an optimal correlation matrix attaining V^{LB} that takes the following form:

$$\Sigma = \begin{pmatrix} \sigma_{11}\mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\sigma_{11})I & \sigma_{12}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \cdots & \sigma_{1K}\mathbf{1}\mathbf{1}^{\mathrm{T}} \\ \sigma_{12}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \sigma_{22}\mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\sigma_{22})I & & \sigma_{2K}\mathbf{1}\mathbf{1}^{\mathrm{T}} \\ \vdots & & \ddots & \vdots \\ \sigma_{1K}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \sigma_{2K}\mathbf{1}\mathbf{1}^{\mathrm{T}} & \cdots & \sigma_{KK}\mathbf{1}\mathbf{1}^{\mathrm{T}} + (1-\sigma_{KK})I \end{pmatrix}, \qquad \text{(EC-10)}$$

where σ_{kk} is the correlation coefficient of the treatment assignments for any two different clusters of the same type k, and $\sigma_{k\ell}$ is the correlation coefficient of the treatment assignments for any two clusters of types k and ℓ with $k \neq \ell$, respectively. Since the correlation matrix Σ needs to be positive semidefinite, for any scalars $x_k \in \mathbb{R}$ for $k \in [K]$, we have

$$
\begin{aligned}
0 &\leq \begin{pmatrix} x_1\mathbf{1}^{\mathrm{T}} & x_2\mathbf{1}^{\mathrm{T}} & \cdots & x_K\mathbf{1}^{\mathrm{T}} \end{pmatrix} \Sigma \begin{pmatrix} x_1\mathbf{1} \\ x_2\mathbf{1} \\ \vdots \\ x_K\mathbf{1} \end{pmatrix} \\
&= \sum_{k=1}^{K} \Big[\sigma_{kk} x_k^2 n_k^2 + (1-\sigma_{kk}) x_k^2 n_k\Big] + \sum_{k=1}^{K} \sum_{\ell \neq k} \sigma_{k\ell} x_k x_\ell n_k n_\ell \\
&= \begin{pmatrix} x_1 n_1 & x_2 n_2 & \cdots & x_K n_K \end{pmatrix} \begin{pmatrix} \frac{1+(n_1-1)\sigma_{11}}{n_1} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{12} & \frac{1+(n_2-1)\sigma_{22}}{n_2} & & \sigma_{2K} \\ \vdots & & \ddots & \vdots \\ \sigma_{1K} & \sigma_{2K} & \cdots & \frac{1+(n_K-1)\sigma_{KK}}{n_K} \end{pmatrix} \begin{pmatrix} x_1 n_1 \\ x_2 n_2 \\ \vdots \\ x_K n_K \end{pmatrix}.
\end{aligned}
$$

Thus, the matrix

$$
\tilde{\Sigma} = \begin{pmatrix} \frac{1+(n_1-1)\sigma_{11}}{n_1} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{12} & \frac{1+(n_2-1)\sigma_{22}}{n_2} & & \sigma_{2K} \\ \vdots & & \ddots & \vdots \\ \sigma_{1K} & \sigma_{2K} & \cdots & \frac{1+(n_K-1)\sigma_{KK}}{n_K} \end{pmatrix} \succeq 0 \tag{EC-11}
$$

needs to be positive semidefinite.

For the IBR experiment with one block for each set of clusters of the same type, the corresponding correlation matrix takes the form of (EC-10) as well, with σ_{kk} as given in Proposition 3.1 and $\sigma_{k\ell} = 0$ for all $k \neq \ell$. For each block k, let $\bar{h}_k$ denote the number of clusters in the block that take values w_k in the worst-case potential outcome of the IBR experiment (the other k-type clusters take value 0). Now, let Σ^* denote an optimal correlation matrix attaining V^{LB} that takes the form of (EC-10) and let $(\sigma^*_{k\ell})_{k,\ell\in[K]}$ denote the corresponding correlation coefficients between cluster types. We have

$$
\begin{aligned}
V^{\mathrm{LB}} &= \max_{h_k \in [0:n_k]} \left\{ \sum_{k\in[K]} w_k^2 \Big[h_k + h_k(h_k-1)\sigma^*_{kk}\Big] + \sum_{k\in[K]} \sum_{\ell\neq k} w_k w_\ell h_k h_\ell \sigma^*_{k\ell} \right\} \\
&\geq \sum_{k\in[K]} w_k^2 \Big[\bar{h}_k + \bar{h}_k(\bar{h}_k-1)\sigma^*_{kk}\Big] + \sum_{k\in[K]} \sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma^*_{k\ell},
\end{aligned}
$$

where we plug in the worst-case potential outcome of the IBR experiment to obtain the inequality.

As a result,

$$
\begin{aligned}
V - V^{\text{LB}} &\le \sum_{k\in[K]} w_k^2 \bar{h}_k \left(\bar{h}_k - 1\right)\left(\sigma_{kk} - \sigma_{kk}^*\right) - \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma_{k\ell}^* \\
&= \sum_{k\in[K]} w_k^2 \left(\bar{h}_k \left(\bar{h}_k - 1\right)\left(\sigma_{kk} - \sigma_{kk}^*\right) + \frac{1+(n_k-1)\sigma_{kk}^*}{n_k}\bar{h}_k^2\right) \\
&\quad - \left\{\sum_{k\in[K]} \frac{1+(n_k-1)\sigma_{kk}^*}{n_k} w_k^2 \bar{h}_k^2 + \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma_{k\ell}^*\right\} \\
&\le \sum_{k\in[K]} w_k^2 \left(\bar{h}_k \left(\bar{h}_k - 1\right)\left(\sigma_{kk} - \sigma_{kk}^*\right) + \frac{1+(n_k-1)\sigma_{kk}^*}{n_k}\bar{h}_k^2\right) \qquad \text{(EC-12)} \\
&= \sum_{k\in[K]} w_k^2 \left[\left(\bar{h}_k - \frac{\bar{h}_k^2}{n_k}\right)\sigma_{kk}^* + \left(\underbrace{\bar{h}_k^2\left(\frac{1}{n_k} + \sigma_{kk}\right) - \bar{h}_k\sigma_{kk}}_{(a)}\right)\right] \\
&\le \sum_{k\in[K]} w_k^2 \left(\frac{1}{4} n_k \sigma_{kk}^* + \frac{1}{2q(1-q)} + 1\right),
\end{aligned}
$$

where the second inequality follows from the fact that with $u = \left(w_k \bar{h}_k\right)_{k\in[K]} \in \mathbb{R}^K$,

$$
\sum_{k\in[K]} \frac{1+(n_k-1)\sigma_{kk}^*}{n_k} w_k^2 \bar{h}_k^2 + \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell \bar{h}_k \bar{h}_\ell \sigma_{k\ell}^* = u^{\mathrm{T}} \tilde{\Sigma} u \ge 0,
$$

because $\tilde{\Sigma}$ is positive semidefinite by (EC-11), and the last inequality follows from the fact that $\bar{h}_k - \frac{\bar{h}_k^2}{n_k} \le \frac{n_k}{4}$ (the equality is attained when $\bar{h}_k = \frac{n_k}{2}$) and that $(a) \le \frac{1}{2q(1-q)} + 1$ by Lemma EC2.4. To bound the value of $n_k \sigma_{kk}^*$ from above, note that by Lemma 4.2,

$$
V^{\text{LB}} \le \sum_{k\in[K]} n_k w_k^2 \le n w_1^2. \qquad \text{(EC-13)}
$$

Also, note that

$$
\begin{aligned}
V^{\text{LB}} &= \max_{h_k\in[0:n_k]} \left\{\sum_{k\in[K]} w_k^2 \Big[h_k + h_k(h_k-1)\sigma_{kk}^*\Big] + \sum_{k\in[K]}\sum_{\ell\neq k} w_k w_\ell h_k h_\ell \sigma_{k\ell}^*\right\} \\
&\ge w_k^2 \Big[n_k + n_k(n_k-1)\sigma_{kk}^*\Big] \qquad \text{(EC-14)} \\
&\ge w_k^2 n_k^2 \sigma_{kk}^*,
\end{aligned}
$$

where we take $h_k = n_k$ and $h_\ell = 0$ for all $\ell \neq k$ to obtain the first inequality. Combining the inequalities (EC-13) and (EC-14) gives

$$
n_k \sigma_{kk}^* \le \min\left\{n_k, \frac{w_1^2 n}{w_k^2 n_k}\right\}.
$$

Plugging this last inequality into the right-hand side of the chain of inequalities in (EC-12) yields the desired bound, and hence completes the proof.

Lemma EC2.4. $(a) \leq \frac{1}{2q(1-q)} + 1$.

Proof. First, if $n_k = 1$, we have $\bar{h}_k = 1$ and hence $(a) = 1 \leq \frac{1}{2q(1-q)} + 1$. In what follows, we assume that $n_k \geq 2$.

Note that $\sigma_{kk} = -\frac{n_k q(1-q) - p(1-p)}{n_k(n_k-1)q(1-q)} \in [-\frac{1}{n_k-1}, 0)$ with $p = \lceil qn_k \rceil - qn_k$ by Proposition 3.1. If $p(1-p) \leq q(1-q)$, then we have $\sigma \leq -\frac{1}{n_k}$ and, as a result, $(a) \leq -\bar{h}_k \sigma_{kk} \leq \frac{n_k}{n_k-1} \leq 2 \leq \frac{1}{2q(1-q)} + 1$.

Otherwise, if $p(1-p) \geq q(1-q)$, then we have $\sigma_{kk} \geq -\frac{1}{n_k}$ and $\sigma_{kk} + \frac{1}{n_k} = \frac{p(1-p) - q(1-q)}{n_k(n_k-1)q(1-q)} \leq \frac{1}{4n_k(n_k-1)q(1-q)}$. As a result,

$$(a) \leq n_k^2 \cdot \frac{1}{4n_k(n_k-1)q(1-q)} + n_k \cdot \frac{1}{n_k} = \frac{n_k}{n_k - 1} \frac{1}{4q(1-q)} + 1 \leq \frac{1}{2q(1-q)} + 1. \qquad \square$$

EC2.10 Proof of Theorem 4.10

First, note that we have

$$K \leq \left\lceil \frac{\ln w_1/\bar{w}}{\ln \alpha} \right\rceil \leq \frac{\ln w_1/\bar{w}}{\ln \alpha} + 1 \leq \frac{(1+\delta_1)\ln n}{2\ln(1+n^{-\delta_2})} + 1 = O\left(n^{\delta_2} \cdot \ln n\right).$$

Let $S_k \subseteq [n]$ denote the set of clusters in block k and let $n_k = |S_k|$ denote this block's cardinality. Consider a new problem instance in which we first drop block zero, and then for each block $k \in [K]$ we decrease all of the cluster sizes to the smallest cluster size in the block. We denote the smallest cluster size in the block k by $\underline{w}_k \triangleq \min_{i \in S_k} w_i$. For this new problem instance, let $\tilde{V}^{\text{LB}}$ denote the lower bound on the worst-case variance of an optimal experiment (i.e., the optimal value of (5)) and let $\tilde{V}$ denote the worst-case variance of the IBR experiment. We proceed by bounding the differences of the original and the new problem instances in terms of the worst-case variances of their IBR experiments and the corresponding lower bounds.

Lemma EC2.5. $\tilde{V}^{\text{LB}} \leq V^{\text{LB}}$ *and* $0 \leq V - \tilde{V} \leq n^{-\delta_1} \sum_{i=1}^n w_i^2 + (\alpha^2 - 1) \sum_{i=1}^n w_i^2$.

Proof. Since cluster sizes are weakly smaller in the new problem instance, the set of the potential outcomes is more restricted, and hence $\tilde{V}^{\text{LB}} \leq V^{\text{LB}}$ and $\tilde{V} \leq V$. For an IBR experiment, the worst-case variance is the sum of the worst-case variances of each block. Let V_k and $\tilde{V}_k$ denote the worst-case variances of block k in the original and the new problem instances, respectively. We have

$$\begin{aligned} V - \tilde{V} &= V_0 + \sum_{k \in [K]} \left(V_k - \tilde{V}_k\right) \\ &\leq \sum_{i \in S_0} w_i^2 + \sum_{k \in [K]} \left(\alpha^2 \tilde{V}_k - \tilde{V}_k\right) \\ &\leq n^{-\delta_1} \sum_{i=1}^n w_i^2 + (\alpha^2 - 1)\tilde{V} \\ &\leq n^{-\delta_1} \sum_{i=1}^n w_i^2 + (\alpha^2 - 1) \sum_{i=1}^n w_i^2. \end{aligned} \tag{EC-15}$$

Here, the first inequality uses the fact that for the specific logarithmic partition, we have $w_i \le \alpha \underline{w}_k$ for any cluster $i \in S_k$. This implies that the worst-case variance V_k is no larger than the worst-case variance when all cluster sizes in block k are equal to $\alpha \underline{w}_k$. The latter quantity is $\alpha^2 \tilde{V}_k$, and hence $V_k \le \alpha^2 \tilde{V}_k$. The first and third inequalities also make use of the inequalities $V_0 \le \sum_{i \in S_0} w_i^2$ and $\tilde{V} \le V \le \sum_{i=1}^n w_i^2$, respectively, by Lemma 4.2. These hold because an IBR experiment always has a smaller worst-case variance than the naive experiment that assigns treatment to each cluster independently. Finally, the second inequality follows from the fact that $w_i \le \bar{w}$ for all clusters $i \in S_0$ and that $n_0 = |S_0| \le n$. □

We next analyze the new problem instance, and bound the gap $\tilde{V} - \tilde{V}^{\mathrm{LB}}$. In the new problem instance, clusters in a block are of equal size. Therefore, we can adopt the analysis for Theorem 4.9. Specifically, by (EC-12) we have

$$\tilde{V} - \tilde{V}^{\mathrm{LB}} \le \frac{1}{4} \sum_{k \in [K]} \underline{w}_k^2 \cdot \left(n_k \sigma_{kk}^* + C(q) \right),$$

where $C(q) \triangleq \frac{2}{q(1-q)} + 4$ is a constant and σ_{kk}^* is the optimal correlation coefficient of any two clusters in block k obtained from the solution of (5) (that takes the form (EC-10)) for the new problem instance. By (EC-14),

$$\underline{w}_k^2 n_k \sigma_{kk}^* \le \min \left\{ \underline{w}_k^2 n_k, \frac{\tilde{V}^{\mathrm{LB}}}{n_k} \right\} \le \underline{w}_k \sqrt{\tilde{V}^{\mathrm{LB}}} \le \underline{w}_k \sqrt{V^{\mathrm{LB}}}.$$

Thus, we have

$$\tilde{V} - \tilde{V}^{\mathrm{LB}} \le \frac{1}{4} \sum_{k \in [K]} \left\{ \underline{w}_k \sqrt{V^{\mathrm{LB}}} + C(q) \cdot \underline{w}_k^2 \right\} \le \frac{K}{4} \cdot \left(w_1 \sqrt{V^{\mathrm{LB}}} + C(q) \cdot w_1^2 \right) \le \frac{C(q)+1}{4} K \cdot w_1 \sqrt{V^{\mathrm{LB}}}, \tag{EC-16}$$

where the last inequality follows from $w_1^2 \le V^{\mathrm{LB}}$ by Lemma 4.2.

Combining (EC-15) and (EC-16), we have

$$\begin{aligned}
\frac{V - V^{\mathrm{LB}}}{V^{\mathrm{LB}}} &\le \frac{V - \tilde{V} + \tilde{V} - \tilde{V}^{\mathrm{LB}}}{V^{\mathrm{LB}}} \\
&\le \frac{4n^{-\delta_1} \sum_{i=1}^n w_i^2 + 4(\alpha^2 - 1) \sum_{i=1}^n w_i^2 + (C(q)+1)\, K \cdot w_1 \sqrt{V^{\mathrm{LB}}}}{4 V^{\mathrm{LB}}} \\
&\le \frac{4n^{-\delta_1} \sum_{i=1}^n w_i^2 + 4(\alpha^2 - 1) \sum_{i=1}^n w_i^2 + (C(q)+1)\, K \cdot w_1 \sqrt{\sum_{i=1}^n w_i^2}}{\sum_{i=1}^n w_i^2} \\
&= 4n^{-\delta_1} + 4(\alpha^2 - 1) + (C(q)+1)\, K \cdot \sqrt{\frac{w_1^2}{\sum_{i=1}^n w_i^2}} \\
&= O\left(n^{-\delta_1}\right) + O\left(n^{-\delta_2}\right) + O\left(n^{-\frac{c}{2}+\delta_2} \cdot \ln n\right),
\end{aligned}$$

where the third inequality follows from the lower and upper bounds of V^{LB} in Lemma 4.2. Taking

$\delta_1 = \delta_2 = \frac{c}{4}$ results in

$$\frac{V - V^{\mathrm{LB}}}{V^{\mathrm{LB}}} = O\left(n^{-\frac{c}{4}} \ln n\right),$$

which completes the proof of the theorem.

EC3 Structural Properties of IBR Experiments

We first show in Lemma EC3.1 that the correlation of any two clusters in a block is nondecreasing as the size of the block increases.

Lemma EC3.1. *Suppose that there are n clusters, and consider the experiment that selects a fraction q of the clusters uniformly at random and assigns them to treatment, as described in Proposition 3.1. The correlation coefficient σ of any two assignments is nondecreasing in the number of clusters n.*

Proof. Let σ_n denote the correlation coefficient of any two clusters when there are n clusters; by Proposition 3.1, we have $\sigma_n = -\frac{nq(1-q)-p_n(1-p_n)}{n(n-1)q(1-q)}$, with $p_n \triangleq \lceil qn \rceil - qn$.

To show that $\sigma_{n+1} \geq \sigma_n$, it is equivalent to show

$$(*) \triangleq (n+1)q(1-q) + (n-1)p_{n+1}(1-p_{n+1}) - (n+1)p_n(1-p_n) \geq 0.$$

In what follows, we check the four possible cases regarding the values of qn and $q(n+1)$ relative to $\lfloor qn \rfloor$ and $\lceil qn \rceil$ (as illustrated in Figure EC-1), and we validate the nonnegativity of $(*)$ for each case.

Case One: $qn \in \mathbb{N}$. In this case, $qn = \lfloor qn \rfloor = \lceil qn \rceil$ and hence, $p_n = 0$. Thus, $(*) > 0$ and hence $\sigma_{n+1} > \sigma_n$.

Case Two: $\lfloor qn \rfloor < qn < q(n+1) < \lceil qn \rceil$. In this case, $(*) = 2\lfloor qn \rfloor\big(\lceil qn \rceil - q(n+1)\big) \geq 0$ and hence $\sigma_{n+1} \geq \sigma_n$. Specifically, if furthermore $qn < 1$, $(*) = 0$ and hence $\sigma_{n+1} = \sigma_n$; otherwise, $\sigma_{n+1} > \sigma_n$.

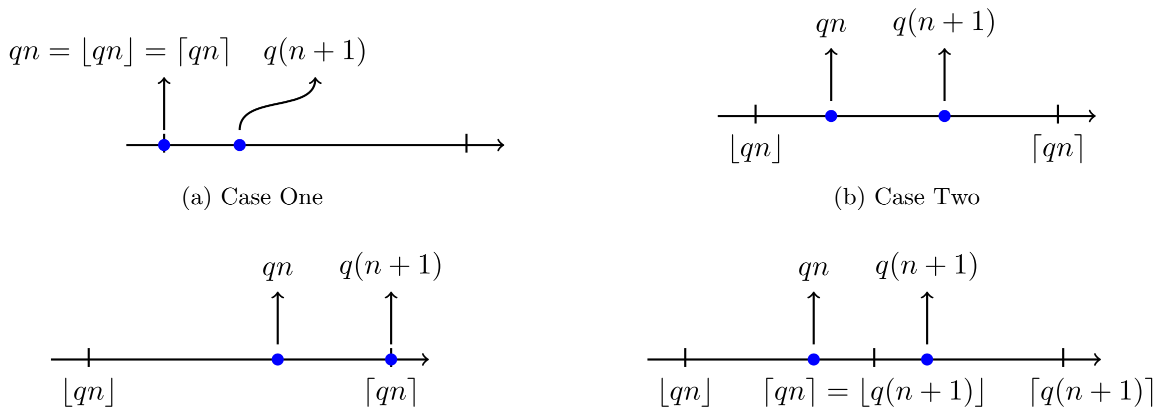
Case Three: $\lfloor qn \rfloor < qn < q(n+1) = \lceil qn \rceil$. In this case, $p_{n+1} = 0$ and $p_n = q$. Hence, $(*) = 0$ and $\sigma_{n+1} = \sigma_n$.

Case Four: $\lfloor qn \rfloor < qn < \lceil qn \rceil < q(n+1)$. In this case, $(*) = 2\big(n - \lceil qn \rceil\big)\big((n+1)q - \lceil qn \rceil\big) > 0$; hence, $\sigma_{n+1} > \sigma_n$.

□

In the rest of the section, we assume that the marginal assignment probability q is equal to $\frac{1}{2}$, and we shed light on two structural properties of the IBR experiments. Our first property establishes that under an IBR experiment, at most half of the clusters in a block can take a positive value in the worst-case potential outcome.

Lemma EC3.2. *Suppose that the marginal assignment probability q is equal to $\frac{1}{2}$. Consider a block with k clusters, and let r be the number of clusters that take a positive value in the worst-case potential outcome. Then, $r \leq \frac{k}{2}$ if k is even, and $r \leq \frac{k+1}{2}$ if k is odd.*



(a) Case One (b) Case Two

(c) Case Three (d) Case Four

Figure EC-1: Illustration of the four cases.

Proof. To see this, note that when the correlation σ is negative, Lemma 3.3 implies that

$$w_r \geq -2\sigma \sum_{i \leq r-1} w_i \geq -2(r-1)\sigma \cdot w_r,$$

which in turn implies that $r \leq -\frac{1}{2\sigma} + 1$. When the block contains an even number k of clusters, $\sigma = -\frac{1}{k-1}$ by Corollary 3.2. Thus, r is an integer no larger than $\frac{k}{2}$. Analogously, when the number of clusters k is odd, $\sigma = -\frac{1}{k}$, and hence, r is an integer no larger than $\frac{k+1}{2}$. □

For the next result, we index the blocks of an IBR experiment in decreasing order of size from the largest cluster in the block (and we break ties using the size of the smallest cluster in the block). We show in Lemma EC3.3 that in the optimal partition obtained from solving the DP, all but the last block contain an even number of clusters.

Lemma EC3.3. *Suppose that the marginal assignment probability q is equal to $\frac{1}{2}$. There exists an optimal cluster partition obtained from solving the DP such that (i) if the number of clusters n is even, then all blocks contain an even number of clusters, and (ii) if n is odd, then all but the last block contain an even number of clusters.*

Proof. We first focus on a single block with k clusters and cluster sizes $w_1 \geq w_2 \geq \cdots \geq w_k$. We highlight three observations. First, clearly, if the vector of cluster sizes entry-wise decreases, the worst-case variance of this block will weakly decrease. Second, if we drop any of the clusters from the block, the worst-case variance of the block will weakly decrease as well because assignments to the remaining clusters become more negatively correlated. Third, if the number of clusters k is odd, adding a cluster of size $w' \leq w_k$ does not change the worst-case variance of this block. To see this, note that after the addition, the correlation between the assignments to any two clusters in the block does not change (and remains $-\frac{1}{k}$ by Corollary 3.2). Moreover, the worst-case potential outcome does not change either, due to Lemma 3.3 and Lemma EC3.2.

We now turn to the optimal partition by solving the DP. Let K denote the number of blocks, and $S_k = \left\{w_1^k \geq w_2^k \geq \cdots \geq w_{n_k}^k\right\}$ denote the set of clusters (sorted in decreasing order of size) in block $k \in [K]$. If a block $h \leq K-1$ has an odd number of clusters, consider a new partition $\{S'_k\}_{k\in[K]}$ as follows: $S'_k = S_k$ for any $k \leq h-1$, $S'_h = S_h \cup \left\{w_1^{h+1}\right\}$, $S'_k = S_k \setminus \left\{w_1^k\right\} \cup \left\{w_1^{k+1}\right\}$ for any $h+1 \le k \le K-1$, and $S'_K = S_K \setminus \{w_1^K\}$. By the former discussion, the worst-case variance of each block weakly decreases. Thus, $\{S'_k\}_{k\in[K]}$ is a weakly better partition than $\{S_k\}_{k\in[K]}$. Iterating this last step completes the proof. $\square$

EC4 More on the Airbnb Example in Section 5.2

In this section, we elaborate more on the Airbnb example in Section 5.2.

EC4.1 Visualization and More Information about the Partition

We visualize the geographic locations of the listings in Figure EC-2. Figure EC-3 exhibits the ten largest clusters from the Louvain algorithm, and Table EC-2 provides more details of the ten clusters. The ten clusters cover 97% of the entire listings.

Seventy-seven percent of the ten clusters' listings connect only to listings in the same cluster. For the remaining listings, Figure EC-4 presents a histogram of the fraction of neighbors that are in a different cluster. From Figure EC-4, it can be seen that most of the listings have the majority of their connections in the same cluster.

cluster	color	description	size
1	orange	entire home/apts in San Francisco (excluding the northwestern area)	2566
2	cyan	entire home/apts to the south of South San Francisco and north of San Jose	2100
3	green	private rooms in San Francisco	2093
4	purple	entire home/apts in San Jose	1908
5	magenta	private rooms in San Jose	1629
6	gold	entire home/apts in the northwestern area of San Francisco	1535
7	pink	entire home/apts in Oakland	1390
8	red	private rooms to the south of San Francisco and north of San Jose	1181
9	brown	private rooms in Oakland	590
10	blue	entire home/apts in Daly City and South San Francisco	518

Table EC-2: Description of the ten clusters visualized in Figure EC-3. The size of a cluster is the number of listings in the cluster, and the color of a cluster corresponds to its color in Figure EC-3.

EC4.2 The Optimal Cluster-Based Experiment

An optimal cluster-based experiment for this example is provided in Table EC-3. The induced correlation structure was reported in Section 5.2. The worst-case value of w_i under the experiment is $y^{\text{OPT}} = [0, 2100, 2093, 0, 1629, 1535, 1390, 0, 0, 0]^{\text{T}}$.

Note that when the marginal assignment probability q equals $\frac{1}{2}$, there always exists an optimal experiment that is symmetric across treatment and control assignments; i.e., letting $P(S)$ denote the probability that clusters in set $S \subseteq [n]$ receive the treatment whereas clusters not in set S receive the control, we obtain that $P(S) = P(S^c)$ for any subset $S \subseteq [n]$ of clusters.[EC2] Such an optimal

[EC2] To see this, note that given an optimal experiment, we can always construct another optimal experiment by flipping the treatment and control assignments by Remark 2.1. Then, since the optimal experimental design problem

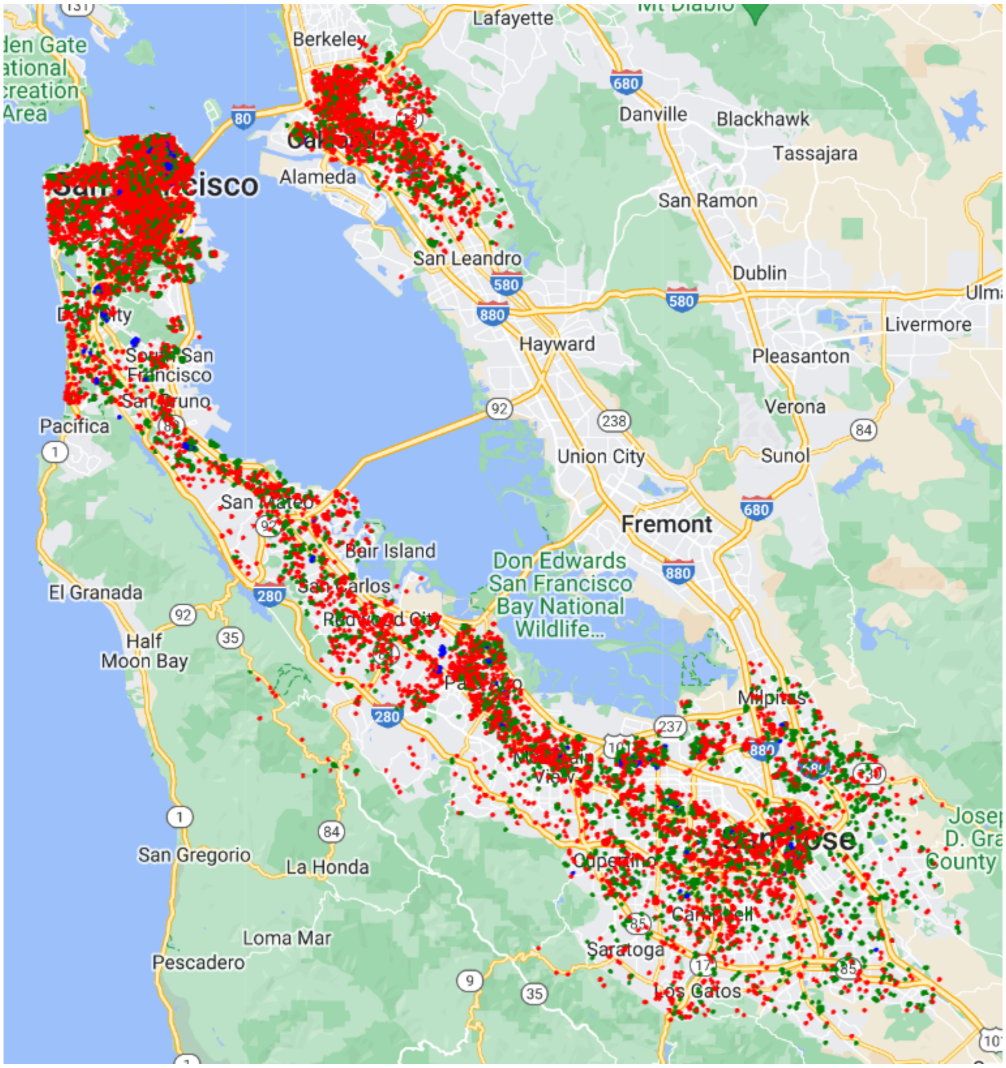


Figure EC-2: The geographical distributions of Airbnb listings in the Bay area. Red nodes represent listings whose room types are "entire home/apartment" (62.7%), green nodes represent listings whose room types are "private room" (34.5%), and blue nodes represent listings whose room types are either "shared room" or "hotel room" (altogether 2.8%).

symmetric experiment still has a complex randomized assignment. Specifically, it randomizes over 88 different possible assignment vectors (where the number of treated clusters varies between 4 and 6), and chooses different probabilities for these vectors without following any clear patterns.

can be formulated as a linear program as in (EC-1), randomizing over the two optimal experiments with equal probability is still optimal, and is symmetric across the treatment and control assignments.

	1	2	3	4	5	6	7	8	9	10	probability
1	√	√	×	×	√	√	×	×	×	×	0.0565
2	√	×	√	√	×	×	√	×	×	×	0.0499
3	√	√	√	×	×	×	×	×	√	√	0.0434
4	×	×	√	√	√	√	×	×	×	√	0.0417
5	×	√	√	×	×	√	√	√	×	×	0.0386
6	√	×	√	×	√	×	×	√	×	√	0.0351
7	×	√	√	×	√	×	×	√	√	×	0.0349
8	√	×	×	×	√	×	√	√	√	×	0.0349
9	√	×	×	√	×	√	×	√	√	×	0.0337
10	×	×	√	√	×	√	×	√	√	√	0.0323
11	×	√	√	√	×	×	×	√	×	×	0.0322
12	×	√	√	×	×	√	√	×	√	×	0.0309
13	√	×	×	×	×	√	√	√	√	√	0.0296
14	×	×	√	×	√	√	√	×	√	√	0.0296
15	×	√	×	√	×	√	√	×	×	√	0.0286
16	√	×	√	×	√	×	√	×	×	√	0.0270
17	×	√	×	√	√	×	√	×	√	√	0.0266
18	√	×	×	×	×	√	√	√	×	√	0.0256
10	√	√	×	√	×	×	×	√	×	√	0.0245
20	×	×	√	√	√	√	×	×	√	×	0.0239
21	√	√	×	√	×	×	×	×	√	√	0.0232
22	×	√	√	×	×	×	√	√	×	√	0.0182
23	√	×	√	√	×	×	×	×	√	×	0.0181
24	√	×	×	√	√	×	√	×	×	×	0.0172
25	×	√	×	√	×	×	√	√	√	√	0.0167
26	×	×	×	√	√	√	×	√	√	√	0.0155
27	√	√	×	×	√	√	×	×	√	×	0.0155
28	×	×	×	√	√	√	√	√	×	√	0.0151
29	×	√	×	√	√	×	√	×	√	×	0.0148
30	×	√	×	√	√	×	√	√	×	×	0.0140
31	√	×	×	√	×	√	√	×	√	×	0.0139
32	×	√	×	×	√	×	√	√	√	√	0.0127
33	√	×	√	×	×	√	×	√	×	×	0.0126
34	×	√	×	×	√	√	×	√	×	√	0.0114
35	√	√	×	×	×	√	×	×	√	√	0.0097
36	×	×	√	×	√	×	√	√	√	×	0.0097
37	√	×	×	√	√	×	×	√	√	×	0.0096
38	×	√	×	√	√	√	×	√	×	×	0.0085
39	×	√	×	√	√	×	×	√	×	√	0.0080
40	×	×	×	√	√	√	√	√	×	×	0.0079
41	×	√	×	√	√	×	√	√	√	×	0.0076
42	√	√	×	×	×	√	√	×	×	√	0.0064
43	×	√	√	×	×	√	√	×	√	√	0.0061
44	×	×	√	√	√	×	√	×	×	×	0.0060
45	×	×	√	√	√	×	√	×	√	√	0.0055
46	√	×	×	×	√	×	√	√	×	√	0.0041
47	√	√	×	√	×	√	×	×	×	×	0.0035
48	√	√	×	×	√	×	×	√	×	√	0.0033
49	×	√	√	×	√	√	×	√	×	×	0.0028
50	√	√	×	√	×	×	√	×	×	×	0.0014
51	√	×	√	×	×	×	√	√	√	×	0.0008
52	√	×	√	×	√	×	√	×	√	×	0.0005
53	√	×	√	×	×	√	×	×	√	×	0.0001

Table EC-3: The randomized joint assignment of the optimal cluster-based experiment. Each of the 53 rows corresponds to a possible assignment, where √ denotes treatment and × denotes control.

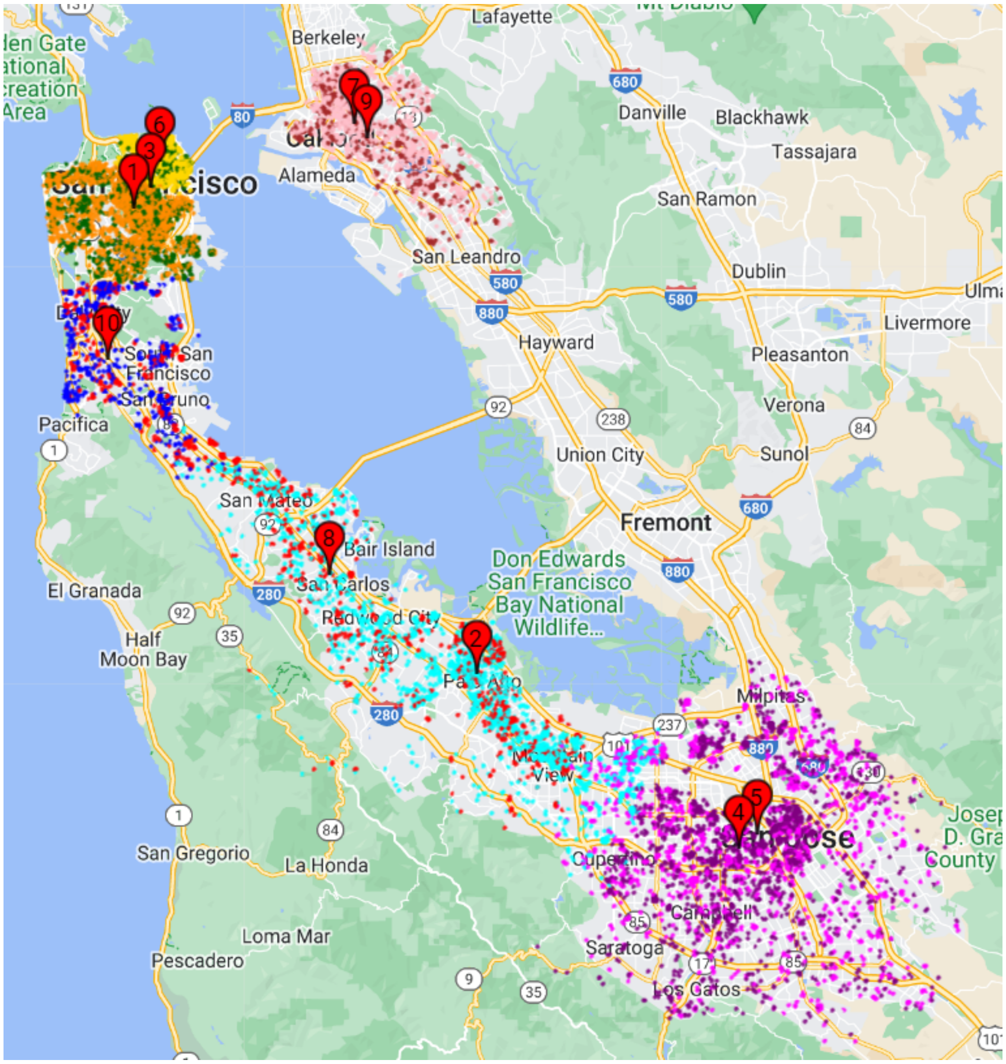


Figure EC-3: The partition of listings into $n = 10$ clusters, with each color representing a different cluster.

EC4.3 The Optimal IBR Experiment and Other Heuristics

The correlation matrix of the optimal IBR experiment (computed by solving a DP) is

$$\Sigma^{\text{DP}} = \begin{pmatrix} 1 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & 1 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{3} & 1 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & -\frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

The worst-case value of w_i for this experiment is $y^{\text{DP}} = [2566, 2100, 0, 0, 1629, 1535, 0, 0, 590, 0]^{\text{T}}$.

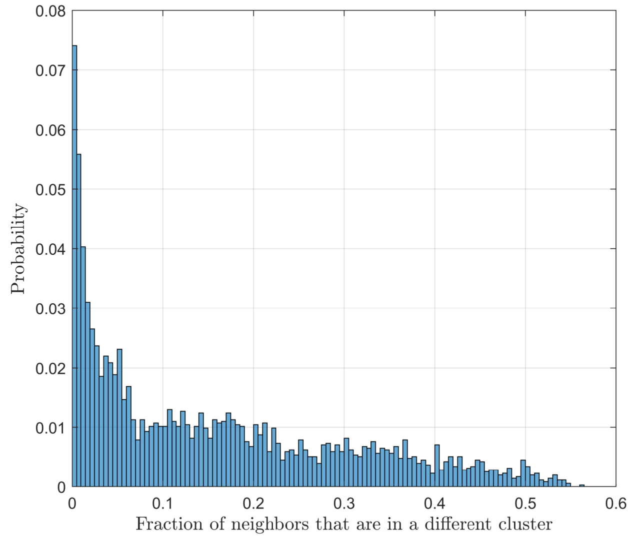


Figure EC-4: Histogram of the fraction of connections that are in a different cluster, among those listings that ever have a connection to a different cluster (which amount to 23% of the entire listings in the ten clusters).

The correlation matrix of the `HALF` experiment is

$$\Sigma^{\mathrm{half}} = \frac{10}{9} I - \frac{1}{9} \mathbf{1}\mathbf{1}^{\mathrm{T}};$$

i.e., all the diagonal entries are one and all the off-diagonal entries are $-\frac{1}{9}$. The worst-case value of w_i for this experiment is $y^{\mathrm{half}} = [2566, 2100, 2093, 1908, 0, 0, 0, 0, 0, 0]^{\mathrm{T}}$.

The correlation matrix of the `PAIR` experiment is

$$\Sigma^{\mathrm{pair}} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

The worst-case value of w_i for this experiment is $y^{\mathrm{pair}} = [2566, 0, 2093, 0, 1629, 0, 1390, 0, 590, 0]^{\mathrm{T}}$.

Finally, the correlation matrix of the `IND` experiment is the identity matrix $\Sigma^{\mathrm{ind}} = I$. The worst-case value of w_i for this experiment is $y^{\mathrm{ind}} = [2566, 2100, 2093, 1908, 1629, 1535, 1390, 1181, 590, 518]^{\mathrm{T}}$.

EC5 Numerical Example: Facebook Subnetworks of US Universities

In this section, we consider a numerical example based on Facebook subnetworks of one hundred US universities. Specifically, we leverage the data described in Section 2 of Traud et al. (2012), which can be accessed from Rossi and Ahmed (2015). We consider cluster-based experiments over these subnetworks, with the users from each university constituting one cluster. We assume that users from different universities are only loosely connected (in contrast with the dense connection structure within each subnetwork), and that the interference among these subnetworks is negligible. Analogous to Section 5.2, we again focus on the case where the marginal assignment probability q is set to be $\frac{1}{2}$, and we assume that the upper bounds w_{i1} and w_{i0} of the cluster-level treatment and control potential outcomes are both proportional to the size (i.e., number of users) of cluster i; please refer to Traud et al. (2012) for the sizes of these one hundred clusters.

With $n = 100$ clusters, it is computationally prohibitive to obtain the optimal cluster-based experiment. The optimal IBR experiment, on the other hand, is fairly easy to compute by solving the DP in Section 3.2. Specifically, the optimal IBR experiment partitions the clusters into 11 blocks, with the number of clusters in each block being[EC3]

$$8, 10, 10, 12, 10, 12, 12, 10, 8, 4, 4.$$

We again consider the three natural experiments in Section 5.2 for comparison. The `HALF` experiment increases the worst-case variance by $\frac{V^{\text{half}}-V^{\text{DP}}}{V^{\text{DP}}} = 78.3\%$ relative to our IBR experiment. The `PAIR` experiment increases the worst-case variance by $\frac{V^{\text{pair}}-V^{\text{DP}}}{V^{\text{DP}}} = 62.9\%$. The `IND` experiment increases the worst-case variance by $\frac{V^{\text{ind}}-V^{\text{DP}}}{V^{\text{DP}}} = 210.0\%$. Thus, the IBR experiment again reduces the worst-case variance considerably when compared to other commonly used heuristic experiments.

EC5.1 Average-Case Analysis of the Facebook Example

In this section, we conduct an average-case analysis of the Facebook subnetwork example by comparing different experiments' variances under the same potential outcomes (in contrast to their respective worst-case outcomes as considered earlier) with the potential outcomes drawn randomly from a given distribution. We focus on the case where the marginal assignment probability q equals $\frac{1}{2}$, and we compare the optimal IBR experiment and the `HALF`, `PAIR`, and `IND` experiments as described in Section 5.2. With some abuse of notation, we let V^{DP}, V^{half}, V^{pair}, and V^{ind} denote the variances of the Horvitz–Thompson estimator for the optimal IBR experiment, the `HALF` experiment, the `PAIR` experiment, and the `IND` experiment, respectively; these are random variables that depend on the value of the potential outcomes.

We consider the following three cases for the underlying distribution of the cluster-level treatment and control potential outcomes y_{i1} and y_{i0}:

$$\text{Case 1: Sample } y_{i1}, y_{i0} \sim \text{Unif}[0, w_i] \text{ for each cluster } i, \text{ all } i.i.d.,$$

$$\text{Case 2: Sample } y_{i0} \sim \text{Unif}\left[0, \tfrac{w_i}{2}\right], \ i.i.d., \text{ and let } y_{i1} = y_{i0} + 0.2w_i \text{ for each cluster } i,$$

$$\text{Case 3: Sample } y_{i0} \sim \text{Unif}\left[0, \tfrac{w_i}{2}\right], \ i.i.d., \text{ and let } y_{i1} = y_{i0} + 0.4w_i \text{ for each cluster } i,$$

where w_i is the number of users in cluster i. In all three cases, the sampling distribution is independent across clusters; we do so to avoid assuming a specific stylized correlation structure across clusters.

Now, let $\tau_a \triangleq \frac{\tau}{m}$ denote the average treatment effect, where τ is the total market effect and $m = \sum_{i\in[n]} w_i$ is the total number of users. Note that τ_a is approximately zero in Case 1, $\tau_a = 0.2$ in Case 2, and $\tau_a = 0.4$ in Case 3. We further let $\sigma_{\text{DP}} \triangleq \frac{\sqrt{V^{\text{DP}}}}{m}$ denote the standard deviation of the Horvitz–Thompson estimator for τ_a under the optimal IBR experiment. Analogously, we let $\sigma_{\text{half}} \triangleq$

[EC3]Blocks are sorted in decreasing order of size from the largest cluster in each block.

$\frac{\sqrt{V^{\text{half}}}}{m}$, $\sigma_{\text{pair}} \triangleq \frac{\sqrt{V^{\text{pair}}}}{m}$, and $\sigma_{\text{ind}} \triangleq \frac{\sqrt{V^{\text{ind}}}}{m}$ denote the standard deviations of the Horvitz–Thompson estimator for τ_a under the `HALF`, `PAIR`, and `IND` experiments, respectively. In the simulation, we randomly draw the potential outcomes 10^4 times for each case and we present the box plots of the values σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind} and the ratios $\sigma_{\text{half}}/\sigma_{\text{DP}}$ and $\sigma_{\text{pair}}/\sigma_{\text{DP}}$ in Figure EC-5.

As can be seen from Figure EC-5, the optimal IBR experiment reduces the variance substantially compared to the `HALF` and `IND` experiments in all of the cases and under all realizations of the potential outcomes. The `PAIR` experiment has a marginally smaller variance on average under the three sampling distributions[EC4] (see the right column of Figure EC-5). On the other hand, the optimal IBR experiment attains a smaller worst-case variance, its variance is more concentrated around the median, and hence it is more robust to the unknown potential outcomes. We also highlight that the observation that the `PAIR` experiment has a smaller variance (on average) than our IBR experiment is substantially an artifact of the choice of the sampling distribution. In particular, the assumption that the potential outcomes are independent across clusters is indeed the source of this observation. When, for example, the potential outcomes are negatively correlated between clusters of similar size, the variance of the `PAIR` experiment is in general larger than the variance of the optimal IBR experiment.[EC5] From all three cases in this example, it can be seen that although the optimal IBR experiment is designed with the goal of minimizing the worst-case variance, it maintains the same performance or even reduces the variance on average in comparison to other heuristic experiments.

Finally, we present the box plot of the ratio $\sigma_{\text{DP}}/\tau_a$, which is the standard deviation of the Horvitz–Thompson estimator of the average treatment effect over the true value, for Cases 2 and 3 in Figure EC-6. In Figure EC-6, the standard deviation is relatively small compared to the true average treatment effect, and this demonstrates the statistical power of the estimator. We elaborate more on this point in Section EC6.

EC6 Statistical Inference from IBR Experiments

In this section, we discuss a way to construct the confidence interval for estimating the total market effect with an IBR experiment. The Horvitz–Thompson estimator $\hat{\tau}$ is unbiased. Now fix the potential outcomes for treatment and control. When the number of blocks is large, by the central limit theorem and the fact that assignments are independent across blocks, the distribution of the estimator $\hat{\tau}$ is approximately normal. When the number of blocks is small, we assume that $\hat{\tau}$ is approximately normal as well. Hence, an α-level confidence interval for the total market effect can be given by $\left[\hat{\tau} - z_{\alpha/2}\sqrt{\mathbb{V}\text{ar}[\hat{\tau}]}, \hat{\tau} + z_{\alpha/2}\sqrt{\mathbb{V}\text{ar}[\hat{\tau}]}\right]$, where $z_{\alpha/2} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$, with $\Phi(\cdot)$ being the CDF of a standard normal distribution.

The problem, however, is that we are not able to compute the variance $\mathbb{V}\text{ar}[\hat{\tau}]$, as it depends on *all* the values of potential outcomes, which cannot be observed simultaneously (recall that by Lemma 2.1, $\mathbb{V}\text{ar}[\hat{\tau}] = q(1-q)y^{\mathrm{T}}\Sigma y$ with $y_i = \frac{y_{i1}}{q} + \frac{y_{i0}}{1-q}$ for each cluster $i \in [n]$). We may use the worst-case variance of the IBR experiment as a proxy for $\mathbb{V}\text{ar}[\hat{\tau}]$, but this can be quite loose especially when some of the potential outcomes are observed after the experiment.

Analogously to Section 4.3 of Imai et al. (2009) and Section 4.2 of Bojinov et al. (2020), we consider a conservative estimator for the variance $\mathbb{V}\text{ar}[\hat{\tau}]$ (which is a variant of a Neymanian conservative variance estimator; see Imbens and Rubin 2015 and Aronow and Middleton 2013). Specifically,

[EC4] Specifically, the median of the ratio $\sigma_{\text{pair}}/\sigma_{\text{DP}}$ is 0.98 in Cases 1 and 2 and 0.97 in Case 3.

[EC5] This can happen when the potential outcomes are correlated through (possibly unobserved) covariates and these covariates are quite different across clusters of similar size.

for each cluster $i \in [n]$, we let $y_i^{\text{obs}} \triangleq \frac{\sqrt{1-q}}{q} y_{i1} \cdot \mathbb{1}\left[Z_i = 1\right] + \frac{\sqrt{q}}{1-q} y_{i0} \cdot \mathbb{1}\left[Z_i = 0\right]$ denote the weighted observed outcome for cluster i. We use the following estimator $\hat{\sigma}^2$ for the variance $\mathbb{V}\text{ar}[\hat{\tau}]$, with

$$\hat{\sigma}^2 \triangleq 2 \sum_{i \in [n]} \left(y_i^{\text{obs}}\right)^2 + \sum_{i \in [n]} \sum_{[n] \ni k \neq i} \sigma_{ik} \left(\left(y_i^{\text{obs}}\right)^2 + \left(y_k^{\text{obs}}\right)^2 \right),$$

where σ_{ik} is the correlation between clusters i and k. The mean of $\hat{\sigma}^2$ provides an upper bound on the variance $\mathbb{V}\text{ar}[\hat{\tau}]$ because

$$\begin{aligned}
\mathbb{E}\left[\hat{\sigma}^2\right] &= q(1-q) \left\{ 2 \sum_{i \in [n]} \left[\left(\frac{y_{i1}}{q}\right)^2 + \left(\frac{y_{i0}}{1-q}\right)^2 \right] + \sum_{i \in [n]} \sum_{k \neq i} \sigma_{ik} \left[\left(\frac{y_{i1}}{q}\right)^2 + \left(\frac{y_{i0}}{1-q}\right)^2 + \left(\frac{y_{k1}}{q}\right)^2 + \left(\frac{y_{k0}}{1-q}\right)^2 \right] \right\} \\
&\geq q(1-q) \left\{ \sum_{i \in [n]} y_i^2 + \sum_{i \in [n]} \sum_{[n] \ni k \neq i} \sigma_{ik} y_i y_k \right\} \\
&= \mathbb{V}\text{ar}[\hat{\tau}],
\end{aligned}$$

where the inequality simply follows from the basic inequality $2xy \leq x^2 + y^2$ and the fact that $y_i = \frac{y_{i1}}{q} + \frac{y_{i0}}{1-q}$.

Following Imai et al. (2009) and Bojinov et al. (2020), we suggest using $\left[\hat{\tau} - z_{\alpha/2}\sqrt{\hat{\sigma}^2}, \hat{\tau} + z_{\alpha/2}\sqrt{\hat{\sigma}^2}\right]$ for an α-level confidence interval of the total market effect τ. This is a common heuristic, and we leave its formal analysis for future work.

References

Aronow, P. M. and Middleton, J. A. (2013), 'A class of unbiased estimators of the average treatment effect in randomized experiments', *Journal of Causal Inference* **1**(1), 135–154.

Asmussen, S. and Glynn, P. W. (2007), *Stochastic simulation: algorithms and analysis*, Vol. 57, Springer Science & Business Media.

Bai, Y. (2022), 'Optimality of matched-pair designs in randomized controlled trials', *Available at SSRN 3483834* .

Bertsimas, D., Johnson, M. and Kallus, N. (2015), 'The power of optimization over randomization in designing experiments involving small samples', *Operations Research* **63**(4), 868–876.

Bertsimas, D., Korolko, N. and Weinstein, A. M. (2019), 'Covariate-adaptive optimization in online clinical trials', *Operations Research* **67**(4), 1150–1161.

Bhat, N., Farias, V. F., Moallemi, C. C. and Sinha, D. (2020), 'Near-optimal ab testing', *Management Science* **66**(10), 4477–4495.

Bibaut, A., Chambaz, A., Dimakopoulou, M., Kallus, N. and van der Laan, M. (2021), 'Post-contextual-bandit inference', *arXiv preprint arXiv:2106.00418* .

Bojinov, I., Simchi-Levi, D. and Zhao, J. (2020), 'Design and analysis of switchback experiments', *Available at SSRN 3684168* .

Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S. and Athey, S. (2019), 'Confidence intervals for policy evaluation in adaptive experiments', *arXiv preprint arXiv:1911.02768* .

Harshaw, C., Sävje, F., Spielman, D. and Zhang, P. (2019), 'Balancing covariates in randomized experiments using the gram-schmidt walk', *arXiv preprint arXiv:1911.03071* .

Imai, K., King, G. and Nall, C. (2009), 'The essential role of pair matching in cluster-randomized experiments, with application to the mexican universal health insurance evaluation', *Statistical Science* **24**(1), 29–53.

Imbens, G. W. and Rubin, D. B. (2015), *Causal inference in statistics, social, and biomedical sciences*, Cambridge University Press.

Jin, Y. and Ba, S. (2021), 'Towards optimal variance reduction in online controlled experiments', *arXiv preprint arXiv:2110.13406* .

Kallus, N. (2018), 'Optimal a priori balance in the design of controlled experiments', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**(1), 85–112.

Kohavi, R., Tang, D. and Xu, Y. (2020), *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, Cambridge University Press.

Lattimore, T. and Szepesvári, C. (2020), *Bandit algorithms*, Cambridge University Press.

Lin, W. (2013), 'Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique', *The Annals of Applied Statistics* **7**(1), 295–318.

Owen, A. B. (2020), 'A first course in experimental design notes from stat 263/363', *Lecture notes* . Available at https://statweb.stanford.edu/~owen/courses/363/doenotes.pdf.

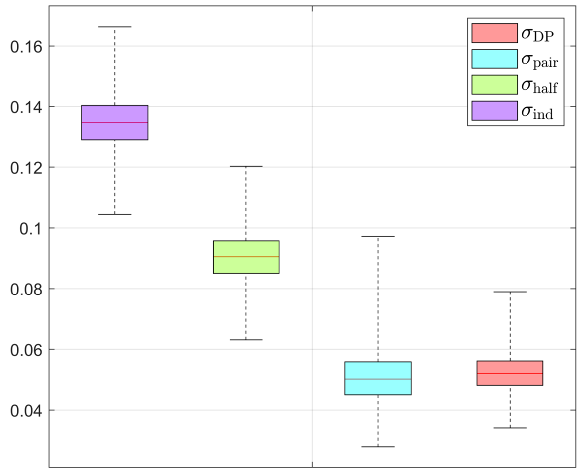
Rachev, S. T. and Rüschendorf, L. (1998), *Mass Transportation Problems: Volume I: Theory*, Vol. 1, Springer Science & Business Media.

Rossi, R. A. and Ahmed, N. K. (2015), 'The network data repository'. Available at http://networkrepository.com/socfb.php.

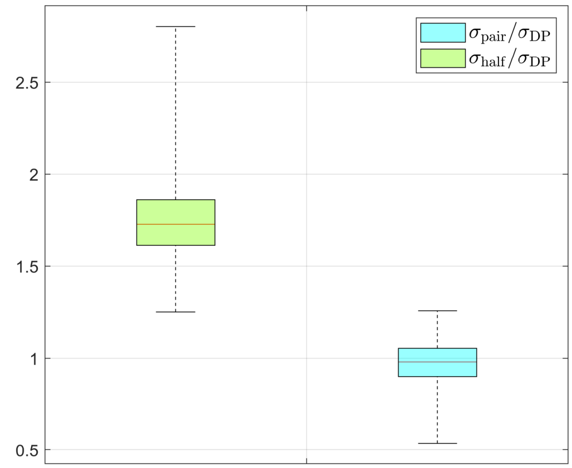
Rüschendorf, L. and Uckelmann, L. (2002), Variance minimization and random variables with constant sum, *in* 'Distributions with given marginals and statistical modelling', Springer, pp. 211–222.

Slivkins, A. (2019), 'Introduction to multi-armed bandits', *Foundations and Trends® in Machine Learning* **12**(1-2), 1–286.

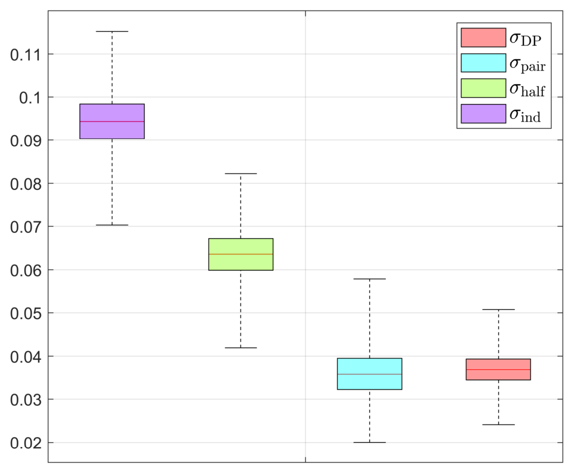
Traud, A. L., Mucha, P. J. and Porter, M. A. (2012), 'Social structure of facebook networks', *Physica A: Statistical Mechanics and its Applications* **391**(16), 4165–4180.



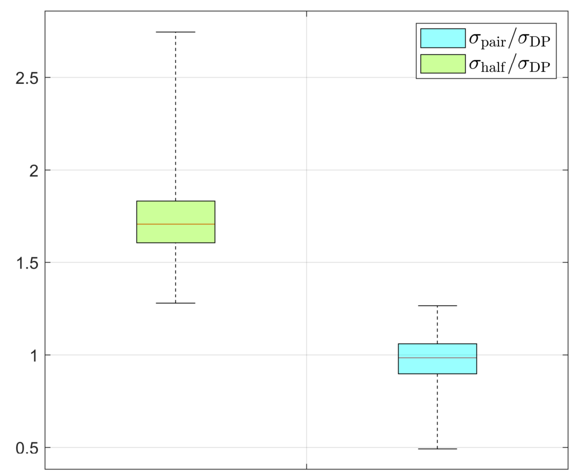
(a) Case 1: σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind}



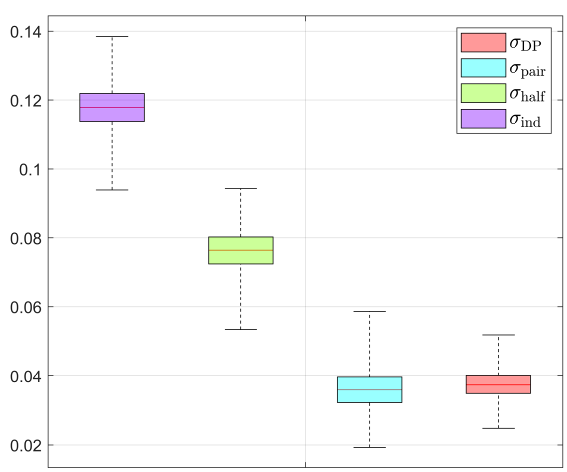
(b) Case 1: $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$



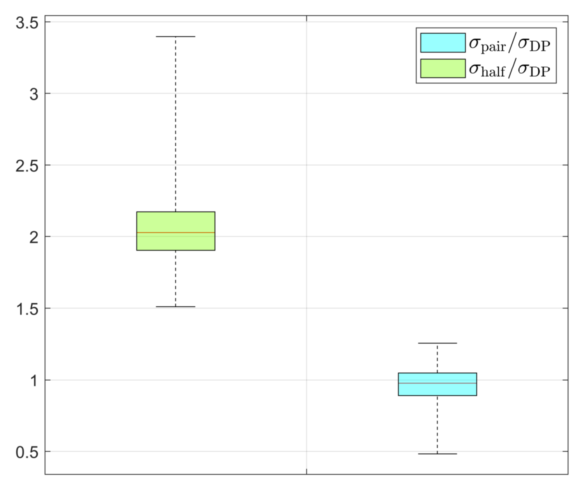
(c) Case 2: σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind}



(d) Case 2: $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$



(e) Case 3: σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind}



(f) Case 3: $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$

Figure EC-5: Box plots of the values σ_{DP}, σ_{half}, σ_{pair}, and σ_{ind} (left column) and the ratios $\sigma_{\mathrm{half}}/\sigma_{\mathrm{DP}}$ and $\sigma_{\mathrm{pair}}/\sigma_{\mathrm{DP}}$ (right column) over 10^4 samples for each case. The interpretation of the box plots is the same as in Figure 3.

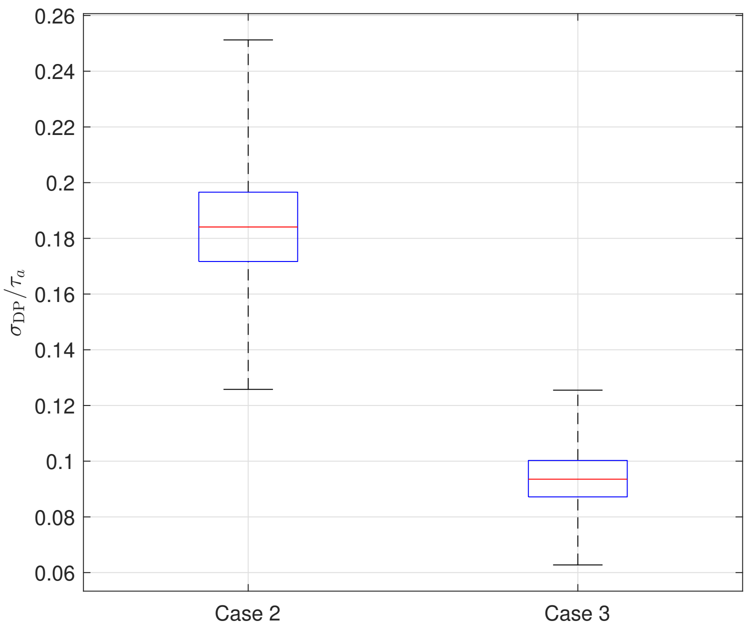


Figure EC-6: Box plot of the ratio $\sigma_{\mathrm{DP}}/\tau_a$ over 10^4 samples for Cases 2 and 3. The interpretation of the box plot is the same as in Figure 3.