Staggered Rollout Designs Enable Causal Inference Without Network Knowledge

Mayleen Cortez-Rodrigez, Matthew Eichhorn, Christina Lee Yu

APPENDIX ONLY

note: links may not work in this pdf

A Variance Calculations

In this section, we establish the unbiasedness and variance bounds of the estimators introduced throughout the paper. The following lemma will be useful for some of these calculations.

Lemma 6. *Suppose we have $Y_{i,t}^{\text{obs}} = Y_i(\mathbf{z}^t) + \varepsilon_{i,t}$ for iid noise $\varepsilon_{i,t} \sim N(0, \sigma^2)$ and our estimator has the form*

$$\widehat{\text{TTE}} = \frac{1}{n} \sum_{t=0}^{\beta} \sum_{i=1}^{n} \alpha_{i,t} \cdot Y_{i,t}^{\text{obs}},$$

with each $|\alpha_{i,t}| = O(\alpha)$. Further suppose that for any $t, t' \in 0, \ldots, \beta$ and two subsets $\mathcal{S}, \mathcal{S}'$ of cardinality at most β,

$$\left| \text{Cov}\Big[\prod_{j \in \mathcal{S}} z_j^t, \prod_{j' \in \mathcal{S}'} z_{j'}^{t'} \Big] \right| \le \begin{cases} B_1 & \mathcal{S} \cap \mathcal{S}' \neq \varnothing, \\ B_2 & \mathcal{S} \cap \mathcal{S}' = \varnothing. \end{cases}$$

Then,

$$\text{Var}\Big[\widehat{\text{TTE}}\Big] = O\bigg(\alpha^2 \beta^2 Y_{\max}^2 \Big(\tfrac{d^2}{n} \max\{B_1, B_2\} + B_2 \Big) + \tfrac{\sigma^2 \beta}{n} \alpha^2 \bigg).$$

Proof. By the law of total variance, we have

$$\begin{aligned}
\text{Var}\Big[\widehat{\text{TTE}}\Big] &= \text{Var}\left[\mathbb{E}\left[\widehat{\text{TTE}} \;\middle|\; \mathbf{z}^t \right] \right] + \mathbb{E}\left[\text{Var}\left[\widehat{\text{TTE}} \;\middle|\; \mathbf{z}^t \right] \right] \\
&= \text{Var}\Big[\tfrac{1}{n} \sum_{t=0}^{\beta} \sum_{i=1}^{n} \alpha_{i,t} \cdot Y_i(\mathbf{z}^t) \Big] \\
&\quad + \mathbb{E}\left[\text{Var}\left(\tfrac{1}{n} \sum_{t=0}^{\beta} \sum_{i=1}^{n} \alpha_{i,t} \cdot Y_i(\mathbf{z}^t) + \tfrac{1}{n} \sum_{t=0}^{\beta} \sum_{i=1}^{n} \alpha_{i,t} \cdot \varepsilon_{i,t} \;\middle|\; \mathbf{z}^t \right) \right] \\
&= \text{Var}\Big[\tfrac{1}{n} \sum_{t=0}^{\beta} \sum_{i=1}^{n} \alpha_{i,t} \cdot Y_i(\mathbf{z}^t) \Big] + \mathbb{E}\left[\text{Var}\left(\tfrac{1}{n} \sum_{t=0}^{\beta} \sum_{i=1}^{n} \alpha_{i,t} \cdot \varepsilon_{i,t} \right) \right] \\
&= \text{Var}\Big[\tfrac{1}{n} \sum_{t=0}^{\beta} \sum_{i=1}^{n} \alpha_{i,t} \cdot Y_i(\mathbf{z}^t) \Big] + \mathbb{E}\left[\tfrac{1}{n^2} \sum_{t=0}^{\beta} \sum_{i=1}^{n} \text{Var}(\alpha_{i,t} \cdot \varepsilon_{i,t}) \right] \\
&= \text{Var}\Big[\tfrac{1}{n} \sum_{t=0}^{\beta} \sum_{i=1}^{n} \alpha_{i,t} \cdot Y_i(\mathbf{z}^t) \Big] + O(\tfrac{\sigma^2 \beta}{n} \alpha^2)
\end{aligned}$$

Turning our attention to the first variance term, we introduce the notation $\mathcal{M}_i = \{i' : |\mathcal{N}_i \cap \mathcal{N}_{i'}| \ge 1\}$. Note that $|\mathcal{M}_i| \le d^2$. In addition, for all $i' \notin \mathcal{M}_i$, all $\mathcal{S} \subseteq \mathcal{N}_i$, and all $\mathcal{S}' \subseteq \mathcal{N}_{i'}$, we have $\mathcal{S} \cap \mathcal{S}' = \varnothing$.

We may expand the variance,

$$\begin{aligned}
\mathrm{Var}\Big[\tfrac{1}{n}\sum_{t=0}^{\beta}\sum_{i=1}^{n}\alpha_{i,t}\cdot Y_i(\mathbf{z}^t)\Big] &= \tfrac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{t=0}^{\beta}\sum_{t'=0}^{\beta}\alpha_{i,t}\cdot\alpha_{i',t'}\cdot\mathrm{Cov}\Big[Y_i(\mathbf{z}^t),Y_{i'}(\mathbf{z}^{t'})\Big] \\
&\le \cdot\tfrac{O(\alpha^2)}{n^2}\sum_{i'=1}^{n}\sum_{i=1}^{n}\sum_{t=0}^{\beta}\sum_{t'=0}^{\beta}\Big|\mathrm{Cov}\Big[Y_i(\mathbf{z}^t),Y_{i'}(\mathbf{z}^{t'})\Big]\Big| \\
&\le \tfrac{O(\alpha^2)}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{t=0}^{\beta}\sum_{t'=0}^{\beta}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i\\|\mathcal{S}|\le\beta}}|c_{i,\mathcal{S}}|\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_{i'}\\|\mathcal{S}'|\le\beta}}|c_{i',\mathcal{S}'}|\cdot\Big|\mathrm{Cov}\Big[\prod_{j\in\mathcal{S}}z_j^t,\prod_{j\in\mathcal{S}'}z_{j'}^{t'}\Big]\Big| \\
&\le \tfrac{O(\alpha^2)}{n^2}\Bigg(\sum_{i=1}^{n}\sum_{i'\in\mathcal{M}_i}\sum_{t=0}^{\beta}\sum_{t'=0}^{\beta}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i\\|\mathcal{S}|\le\beta}}|c_{i,\mathcal{S}}|\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_{i'}\\|\mathcal{S}'|\le\beta}}|c_{i',\mathcal{S}'}|\cdot\max\{B_1,B_2\} \\
&\qquad +\sum_{i=1}^{n}\sum_{i'\notin\mathcal{M}_i}\sum_{t=0}^{\beta}\sum_{t'=0}^{\beta}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i\\|\mathcal{S}|\le\beta}}|c_{i,\mathcal{S}}|\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_{i'}\\|\mathcal{S}'|\le\beta}}|c_{i',\mathcal{S}'}|\cdot B_2\Bigg) \\
&\le \tfrac{O(\alpha^2)}{n^2}\Bigg(\sum_{i=1}^{n}\sum_{i'\in\mathcal{M}_i}\beta^2\,Y_{\max}^2\cdot\max\{B_1,B_2\}+\sum_{i=1}^{n}\sum_{i'\notin\mathcal{M}_i}\beta^2\,Y_{\max}^2\cdot B_2\Bigg) \\
&\le \tfrac{O(\alpha^2)}{n^2}\Big(d^2n\cdot\beta^2\,Y_{\max}^2\cdot\max\{B_1,B_2\}+n^2\cdot\beta^2\,Y_{\max}^2\cdot B_2\Big) \\
&= O\Big(\alpha^2\beta^2Y_{\max}^2\Big(\tfrac{d^2}{n}\max\{B_1,B_2\}+B_2\Big)\Big).
\end{aligned}$$

Therefore,

$$\mathrm{Var}\Big[\widehat{\mathrm{TTE}}\Big] = O\Big(\alpha^2\beta^2Y_{\max}^2\Big(\tfrac{d^2}{n}\max\{B_1,B_2\}+B_2\Big)+\tfrac{\sigma^2\beta}{n}\alpha^2\Big).$$

□

A.1 Graph Agnostic with Bernoulli Treatment

By plugging in the Bernoulli treatment probabilities into (3.1), we obtain the estimator:

$$\widehat{\mathrm{TTE}}(\mathbf{p}) := \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{\beta}\Big(\ell_{t,\mathbf{p}}(1)-\ell_{t,\mathbf{p}}(0)\Big)\cdot Y_{i,t}^{\mathrm{obs}}, \qquad \ell_{t,\mathbf{p}}(x)=\prod_{\substack{s=0\\s\ne t}}^{\beta}\frac{x-p_s}{p_t-p_s}.$$

The following lemma will be useful in establishing a bound on the variance of this estimator.

Lemma 7. $\max_{t\in\{0\ldots\beta\}}\big\{|\ell_{t,\mathbf{p}}(1)-\ell_{t,\mathbf{p}}(0)|\big\} = O\Big(\Delta_{\mathbf{p}}^{-\beta}\Big).$

Proof. For each $t\in 0,\ldots,\beta$, we have,

$$\big|\ell_{t,\mathbf{p}}(1)-\ell_{t,\mathbf{p}}(0)\big| \le \Big|\prod_{\substack{s=0\\s\ne t}}^{\beta}\frac{1-p_s}{p_t-p_s}\Big|+\Big|\prod_{\substack{s=0\\s\ne t}}^{\beta}\frac{-p_s}{p_t-p_s}\Big| \le \prod_{\substack{s=0\\s\ne t}}^{\beta}\frac{|1-p_s|}{\Delta_{\mathbf{p}}}+\prod_{\substack{s=0\\s\ne t}}^{\beta}\frac{|p_s|}{\Delta_{\mathbf{p}}} = O\Big(\Delta_{\mathbf{p}}^{-\beta}\Big).$$

Here, the first inequality is an application of the triangle inequality, the second uses the definition of $\Delta_{\mathbf{p}}$, and the third uses the fact that each $p_t\in[0,1]$. □

Proof of Theorem 2. To establish the unbiasedness of the estimator, note that,

$$\begin{aligned}
\mathbb{E}\Big[\widehat{\text{TTE}}(\mathbf{p})\Big] &= \sum_{t=0}^{\beta} \Big(\ell_{t,\mathbf{p}}(1) - \ell_{t,\mathbf{p}}(0)\Big) \cdot \mathbb{E}\Big[\tfrac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}^t)\Big] \\
&= \sum_{t=0}^{\beta} \Big(\ell_{t,\mathbf{p}}(1) - \ell_{t,\mathbf{p}}(0)\Big) \cdot F_B(p_t) \\
&= \Big(\sum_{t=0}^{\beta} \ell_{t,\mathbf{p}}(1) \cdot F_B(p_t)\Big) - \Big(\sum_{t=0}^{\beta} \ell_{t,\mathbf{p}}(0) \cdot F_B(p_t)\Big) \\
&= F_B(1) - F_B(0) \\
&= \text{TTE}.
\end{aligned}$$

Now, we compute a bound on the variance. Since the entries of each $\mathbf{z}^t$ are independent, $\text{Cov}\Big[\prod_{j\in\mathcal{S}} z_j^t, \prod_{j'\in\mathcal{S}'} z_{j'}^{t'}\Big] = 0$ for any disjoint $\mathcal{S}, \mathcal{S}'$. In addition, since both arguments of this covariance are indicator variables, we can upper bound the absolute value of each covariance by 1. We appeal to Lemma 6, with $B_1 = 1$, $B_2 = 0$, and $\alpha = \Delta_{\mathbf{p}}^{-\beta}$ (by Lemma 7), giving,

$$\text{Var}\Big[\widehat{\text{TTE}}(\mathbf{p})\Big] = O\Big(\tfrac{d^2\beta^2}{n}\, Y_{\max}^2\, \Delta_{\mathbf{p}}^{-2\beta} + \tfrac{\sigma^2\beta}{n}\Delta_{\mathbf{p}}^{-2\beta}\Big).$$

□

A.2 Graph Agnostic with Completely Randomized Treatment

We'll make use of the following algebraic lemma to bound the variance; recall the bracket notation introduced in equation (2.3) in Section 2.

Lemma 8. *For any constants $a, b \in \mathbb{N}$ and any $p \in (0, 1]$,*

$$\left| \frac{\left[\frac{pn-a}{n-a}\right]^b}{\left[\frac{pn}{n}\right]^b} - 1 \right| = O\Big(\frac{ab}{pn}\Big).$$

Proof. Expanding the bracket notation, we have,

$$\begin{aligned}
\left| \frac{\left[\frac{pn-a}{n-a}\right]^b}{\left[\frac{pn}{n}\right]^b} - 1 \right| &= \left| \prod_{i=0}^{b-1} \Big(\frac{pn-a-i}{pn-i}\Big)\Big(\frac{n-i}{n-a-i}\Big) - 1 \right| \\
&= \left| \prod_{i=0}^{b-1} \Big(1 - \frac{a}{pn-i}\Big)\Big(1 + \frac{a}{n-a-i}\Big) - 1 \right| \\
&= \left| \prod_{i=0}^{b-1} \Big(1 + O\Big(\tfrac{a(p-1)}{pn}\Big)\Big) - 1 \right| \\
&\le \sum_{j=1}^{b-1} \binom{b}{j} \cdot O\Big(\tfrac{a}{pn}\Big)^j \\
&\le \sum_{j=1}^{b-1} O\Big(\tfrac{ab}{pn}\Big)^j \\
&= O\Big(\tfrac{ab}{pn}\Big).
\end{aligned}$$

□

Proof of Theorem 3. To establish the unbiasedness of the estimator, note that,

$$\begin{aligned}
\mathbb{E}\Big[\widehat{\mathrm{TTE}}(\mathbf{k})\Big] &= \sum_{t=0}^{\beta}\Big(\ell_{t,\mathbf{k}/n}(1) - \ell_{t,\mathbf{k}/n}(0)\Big)\cdot \mathbb{E}\Big[\tfrac{1}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}^t)\Big] \\
&= \sum_{t=0}^{\beta}\Big(\ell_{t,\mathbf{k}/n}(1) - \ell_{t,\mathbf{k}/n}(0)\Big)\cdot F_C(\tfrac{k}{n}) \\
&= \Big(\sum_{t=0}^{\beta}\ell_{t,\mathbf{k}/n}(1)\cdot F_C(\tfrac{k}{n})\Big) - \Big(\sum_{t=0}^{\beta}\ell_{t,\mathbf{k}/n}(0)\cdot F_C(\tfrac{k}{n})\Big) \\
&= F_C(1) - F_C(0) \\
&= \mathrm{TTE}.
\end{aligned}$$

Next, we establish a bound on the variance of this estimator. We consider the covariance term $\Big|\mathrm{Cov}\Big[\prod_{j\in\mathcal{S}} z_j^t, \prod_{j'\in\mathcal{S}'} z_{j'}^{t'}\Big]\Big|$ for various values of $t, t', \mathcal{S}$, and $\mathcal{S}'$. First, note that when t or $t' = 0$, an argument of this covariance is deterministically 0, so the covariance is 0 as well. Otherwise, when $\mathcal{S}\cap\mathcal{S}' \neq \varnothing$, we can bound $\Big|\mathrm{Cov}\Big[\prod_{j\in\mathcal{S}} z_j^t, \prod_{j'\in\mathcal{S}'} z_{j'}^{t'}\Big]\Big| \le 1$ by noting that both arguments are indicator variables. In the case that $\mathcal{S}\cap\mathcal{S}' = \varnothing$, we establish a stronger bound using Lemma 8. We have,

$$\begin{aligned}
\mathrm{Cov}\Big[\prod_{j\in\mathcal{S}} z_j^t, \prod_{j'\in\mathcal{S}'} z_{j'}^{t'}\Big] &= \mathbb{E}\Big[\prod_{j\in\mathcal{S}} z_j^t \prod_{j'\in\mathcal{S}'} z_{j'}^{t'}\Big] - \mathbb{E}\Big[\prod_{j\in\mathcal{S}} z_j^t\Big]\mathbb{E}\Big[\prod_{j'\in\mathcal{S}'} z_{j'}^{t'}\Big] \\
&\le \Big[\tfrac{k_t}{n}\Big]^{|\mathcal{S}|}\Big[\tfrac{k_{t'}}{n}\Big]^{|\mathcal{S}'|}\cdot\left|\frac{\Big[\frac{k_{t'}-|\mathcal{S}|}{n-|\mathcal{S}|}\Big]^{|\mathcal{S}'|}}{\Big[\frac{k_{t'}}{n}\Big]^{|\mathcal{S}'|}} - 1\right| \\
&= O\Big(\tfrac{|\mathcal{S}||\mathcal{S}'|}{k_{t'}}\Big) \\
&= O\Big(\tfrac{\beta^2}{k_1}\Big).
\end{aligned}$$

In the second last line, we bound the first two factors by 1, and use Lemma 8 (with $p = \frac{k_{t'}}{n}$) to bound the third factor. Applying Lemma 6, with $B_1 = 1$, $B_2 = O\Big(\frac{\beta^2}{k_1}\Big)$, and $\alpha = \Big(\frac{n}{\Delta_{\mathbf{k}}}\Big)^{\beta}$ (by Lemma 7 using the substitution $\mathbf{p} = \mathbf{k}/n$), giving,

$$\mathrm{Var}\Big[\widehat{\mathrm{TTE}}(\mathbf{k})\Big] = O\Big(\beta^2\, Y_{\max}^2\Big(\tfrac{d^2}{n} + \tfrac{\beta^2}{k_1}\Big)\cdot\big(\tfrac{n}{\Delta_{\mathbf{k}}}\big)^{2\beta} + \tfrac{\sigma^2\beta}{n}\big(\tfrac{n}{\Delta_{\mathbf{k}}}\big)^{2\beta}\Big).$$

□

A.3 Improved Variance Bounds in the Linear Setting

Proof of Corollary 4. In the linear setting ($\beta = 1$) for $\mathbf{x} = (0, x)$, the Lagrange polynomial coefficients evaluate to $\ell_{0,\mathbf{x}}(1) - \ell_{0,\mathbf{x}}(0) = -\alpha$ and $\ell_{1,\mathbf{x}}(1) - \ell_{1,\mathbf{x}}(0) = \alpha$ for $\alpha = \frac{1}{x}$, so that the estimator $\widehat{\mathrm{TTE}}(\mathbf{x})$ is equal to

$$\begin{aligned}
\widehat{\mathrm{TTE}}(\mathbf{x}) &= \frac{\alpha}{n}\left(\sum_{i=1}^{n} Y_{i,1}^{\mathrm{obs}} - \sum_{i=1}^{n} Y_{i,0}^{\mathrm{obs}}\right) \\
&= \tfrac{\alpha}{n}\sum_{i=1}^{n}\Big(Y_i(\mathbf{z}^1) + \varepsilon_{i,1} - \varepsilon_{i,0} - c_{i,\emptyset}\Big).
\end{aligned}$$

Using the Law of Total Variance, we get

$$\mathrm{Var}\Big[\widehat{\mathrm{TTE}}(\mathbf{x})\Big] = \mathrm{Var}\Big[\mathbb{E}\Big[\widehat{\mathrm{TTE}}\;\Big|\;\mathbf{z}^1\Big]\Big] + \mathbb{E}\Big[\mathrm{Var}\Big[\widehat{\mathrm{TTE}}\;\Big|\;\mathbf{z}^1\Big]\Big] = \mathrm{Var}\Big[\tfrac{\alpha}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}^1)\Big] + \tfrac{2\sigma^2\alpha^2}{n}.$$

Rewriting the first term, we get

$$\begin{aligned}
\mathrm{Var}\Big[\tfrac{\alpha}{n}\sum_{i=1}^{n} Y_i(\mathbf{z}^1)\Big] &= \tfrac{\alpha^2}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\mathrm{Cov}\big[Y_i(\mathbf{z}^1), Y_{i'}(\mathbf{z}^1)\big] \\
&= \tfrac{\alpha^2}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j\in\mathcal{N}_i} c_{ij}\sum_{j'\in\mathcal{N}_{i'}} c_{i'j'}\mathrm{Cov}[z_j, z_{j'}] \qquad \text{(A.1)} \\
&= \tfrac{\alpha^2}{n^2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\left(\sum_{i:j\in\mathcal{N}_i} c_{ij}\right)\left(\sum_{i':j'\in\mathcal{N}_{i'}} c_{i'j'}\right)\mathrm{Cov}[z_j, z_{j'}]. \qquad \text{(A.2)}
\end{aligned}$$

Here, we used the fact that $\mathbf{z}^0 = \mathbf{0}$ deterministically to remove covariance terms, as it has covariance 0 with any other random variable. Under $\mathrm{BRD}(0,p)$ we have $\alpha = \frac{1}{p}$. Additionally, $\mathrm{Var}(z_j^1) = p(1-p)$ for each $j \in [n]$, and $\mathrm{Cov}[z_j, z_{j'}] = 0$ for $j \neq j'$ so we may simplify the variance bound to

$$\begin{aligned}
&= \tfrac{\alpha^2}{n^2}\sum_{j=1}^{n}\left(\sum_{i:j\in\mathcal{N}_i} c_{ij}\right)^2 \cdot \mathrm{Var}(z_j^1) \; + \; \tfrac{2\sigma^2\alpha^2}{n} \\
&\leq \tfrac{\alpha^2}{n^2}\cdot L_{\max}^2 \cdot \sum_{j=1}^{n}\mathrm{Var}(z_j^1) + \tfrac{2\sigma^2\alpha^2}{n} \\
&\leq \tfrac{1-p}{np}\cdot L_{\max}^2 + \tfrac{2\sigma^2}{np^2}.
\end{aligned}$$

The analysis for the completely randomized design setting is presented in pg 32 of [25], and we include it here for convenience. Under $\mathrm{CRD}(0,k)$, we have $\alpha = \frac{n}{k}$. Additionally, $\mathrm{Var}(z_j^1) = \frac{k(n-k)}{n^2}$ for each $j \in [n]$, and

$$\mathrm{Cov}[z_j^1, z_{j'}^1] = \frac{k(k-1)n}{n^2(n-1)} - \frac{k^2(n-1)}{n^2(n-1)} = \frac{-k(n-k)}{n^2(n-1)} \leq 0.$$

Plugging into (A.2), we find that

$$\begin{aligned}
\mathrm{Var}\Big[\widehat{\mathrm{TTE}}(\mathbf{k})\Big] &= \tfrac{1}{k^2}\sum_{j=1}^{n}\left(\sum_{i:j\in\mathcal{N}_i} c_{ij}\right)^2 \cdot \mathrm{Var}(z_j^1) + \tfrac{1}{k^2}\sum_{j\neq j'}\left(\sum_{i:j\in\mathcal{N}_i} c_{ij}\right)\left(\sum_{i:j'\in\mathcal{N}_i} c_{ij'}\right)\cdot \mathrm{Cov}(z_j^1, z_{j'}^1) \; + \; \tfrac{2\sigma^2 n}{k^2} \\
&= \tfrac{1}{k^2}\sum_{j=1}^{n}\left(\sum_{i:j\in\mathcal{N}_i} c_{ij}\right)^2\left(\frac{k(n-k)}{n^2} + \frac{k(n-k)}{n^2(n-1)}\right) + \left(\frac{1}{k}\sum_{j=1}^{n}\sum_{i:j\in\mathcal{N}_i} c_{ij}\right)^2 \frac{-k(n-k)}{n^2(n-1)} \; + \; \tfrac{2\sigma^2 n}{k^2} \\
&\leq \tfrac{nL_{\max}^2}{k^2}\left(\frac{k(n-k)}{n^2} + \frac{k(n-k)}{n^2(n-1)}\right) \; + \; \tfrac{2\sigma^2 n}{k^2} \\
&\leq \tfrac{(n-k)}{(n-1)k} L_{\max}^2 \; + \; \tfrac{2\sigma^2 n}{k^2}.
\end{aligned}$$

□

A.4 Bernoulli Estimator Utilizing Realized Treatment Counts

We will make use of the following lemma to bound the variance of this estimator.

Lemma 9. *Suppose $X \sim Binom(n,p)$, and define*

$$Y = \begin{cases} 0 & X = 0, \\ \frac{1}{X^\beta} & X > 0. \end{cases}$$

Then, $\mathbb{E}\big[Y\big] < (1 + o(1))(np)^{-\beta}$.

Proof. Using the law of total expectation, we can upper bound this expectation,

$$\begin{aligned}\mathbb{E}\big[Y\big] &\leq \Pr\big(X \leq (1-\delta)np\big) + \left(\tfrac{1}{(1-\delta)np}\right)^{\beta} \cdot \Pr\big(X > (1-\delta)np\big) \\ &\leq \Pr\big(X \leq (1-\delta)np\big) + \left(\tfrac{1}{(1-\delta)np}\right)^{\beta}. \end{aligned} \tag{A.3}$$

We apply Bernstein's inequality to compute this probability. Note that we can express

$$X = X_1 + \ldots + X_n,$$

with each $X_i \sim \text{Bernoulli}(p)$. Now, define $Z = Z_1 + \ldots + Z_n$ where each $Z_i = p - X_i$. Note that each $\mathbb{E}\big[Z_i\big] = 0$ and $|Z_i| \leq 1$. Thus,

$$\begin{aligned}\Pr\big(X \leq (1-\delta)np\big) &= \Pr\big(Z \geq \delta np\big) \\ &\leq \exp\left(\frac{-\frac{1}{2}\big(\delta np\big)^2}{\sum_{i=1}^{n}\mathbb{E}\big[Z_i^2\big] + \frac{1}{3}\big(\delta np\big)}\right) \\ &= \exp\left(\frac{-3\delta^2 n^2 p^2}{6np(1-p) + 2\delta np}\right) \\ &\leq \exp\left(\frac{-3\delta^2 np}{6+2\delta}\right).\end{aligned}$$

For $\delta = \log^{-1} n$ and large enough n, $\exp\left(\frac{-3\delta^2 np}{6+2\delta}\right) < (np)^{-2\beta}$, such that plugging into (A.3), we find $\mathbb{E}\big[Y\big] \leq ((1-\delta)np)^{-\beta} + (np)^{-2\beta} = (1+o(1))np^{-\beta}$. □

Proof of Theorem 5. First, we reason about the bias of the estimator. We define the event $\mathcal{E}_1$ be the event $\{k_0 < k_1 < \ldots < k_\beta\}$. By the argument from the proof of Theorem 3, $\widehat{\text{TTE}}(\hat{\mathbf{k}}/n)$ is unbiased on $\mathcal{E}_1$. Thus, we can express the bias as

$$\mathbb{E}\Big[\widehat{\text{TTE}}(\hat{\mathbf{k}}/n) - \text{TTE}\Big] = -\Pr\big(\mathcal{E}_1^c\big) \cdot \text{TTE}.$$

However,

$$\begin{aligned}\Pr\big(\mathcal{E}_1^c\big) &= \Pr\left(\bigcup_{t=1}^{\beta}\big\{\hat{k}_t = \hat{k}_{t-1}\big\}\right) \\ &\leq \sum_{t=1}^{\beta}\Pr\big(\hat{k}_t = \hat{k}_{t-1}\big) && \text{(Union Bound)} \\ &= \sum_{t=1}^{\beta}\Pr\big(\hat{k}_t - \hat{k}_{t-1} \leq 0\big) \\ &\leq \sum_{t=1}^{\beta}\exp\left(\tfrac{-(p_t - p_{t-1})n}{2}\right) && \text{(Chernoff Bound)} \\ &\leq \beta \cdot \exp\left(\tfrac{-\Delta_{\mathbf{p}} n}{2}\right),\end{aligned}$$

so the bias decays exponentially with n.

To bound the variance, we apply the law of total variance:

$$\text{Var}\Big[\widehat{\text{TTE}}\Big] = \text{Var}\Bigg[\mathbb{E}\Big[\widehat{\text{TTE}}\Big|\sum_{j=1}^{n} z_j^t = \hat{k}_t\ \forall t\Big]\Bigg] + \mathbb{E}\Bigg[\text{Var}\Big[\widehat{\text{TTE}}\Big|\sum_{j=1}^{n} z_j^t = \hat{k}_t\ \forall t\Big]\Bigg]. \tag{A.4}$$

We bound these terms individually. For the first term, note that

$$\mathbb{E}\Big[\widehat{\text{TTE}}(\hat{\mathbf{k}}/n)\Big|\sum_{j=1}^{n} z_j^t = \hat{k}_t\ \forall t\Big] = \text{TTE} \cdot \mathbb{I}(\mathcal{E}_1).$$

which implies that,

$$\mathrm{Var}\Big[\mathbb{E}\Big[\widehat{\mathrm{TTE}}(\hat{\mathbf{k}}/n)\Big|\sum_{j=1}^{n} z_j^t = \hat{k}_t\ \forall t\Big]\Big] = \mathrm{TTE}^2 \cdot \mathrm{Var}\Big(\mathbb{I}(\mathcal{E}_1)\Big) = \mathrm{TTE}^2 \cdot \Pr\left(\mathcal{E}_1\right) \cdot \Pr\left(\mathcal{E}_1^c\right).$$

This term decays exponentially as n grows large, so (A.4) will be dominated by the second term.

Next, we define the event

$$\mathcal{E}_2 := \mathcal{E}_1 \cap \bigcap_{t=1}^{\beta} \Big\{|\hat{k}_t - p_t n| \le \delta p_t n\Big\}.$$

Then,

$$\begin{aligned}
\Pr\left(\mathcal{E}_2^c\right) &= \Pr\left(\mathcal{E}_1^c \cup \bigcup_{t=1}^{\beta} \{|\hat{k}_t - p_t n| > \delta p_t n\}\right) \\
&\le \Pr\left(\mathcal{E}_1^c\right) + \sum_{t=1}^{\beta} \Pr\left(|\hat{k}_t - p_t n| \ge \delta p_t n\right) && \text{(Union Bound)} \\
&\le \Pr\left(\mathcal{E}_1^c\right) + \sum_{t=1}^{\beta} \exp\left(\frac{-\delta^2 p_t n}{3}\right) && \text{(Chernoff Bound)} \\
&\le \Pr\left(\mathcal{E}_1^c\right) + \beta \cdot \exp\left(\frac{-\delta^2 p_1 n}{3}\right).
\end{aligned}$$

Notice that by a different application of the law of total variance, we get

$$\begin{aligned}
\mathrm{Var}\Big[\widehat{\mathrm{TTE}}(\hat{\mathbf{k}}/n)\Big] &= \mathrm{Var}\left[\mathbb{E}\left[\widehat{\mathrm{TTE}}(\hat{\mathbf{k}}/n)\ \Big|\ \mathbf{z}^t\right]\right] + \mathbb{E}\left[\mathrm{Var}\left[\widehat{\mathrm{TTE}}(\hat{\mathbf{k}}/n)\ \Big|\ \mathbf{z}^t\right]\right] \\
&= \mathrm{Var}\left[\tfrac{1}{n}\sum_{t=0}^{\beta}\sum_{i=1}^{n}\Big(\ell_{t,\hat{\mathbf{k}}/n}(1) - \ell_{t,\hat{\mathbf{k}}/n}(0)\Big)Y_i(\mathbf{z}^t)\right] + \mathbb{E}\left[\tfrac{\sigma^2}{n}\sum_{t=0}^{\beta}\Big(\ell_{t,\hat{\mathbf{k}}/n}(1) - \ell_{t,\hat{\mathbf{k}}/n}(0)\Big)^2\right].
\end{aligned}$$

We can bound $\ell_{t,\hat{\mathbf{k}}/n}(1) - \ell_{t,\hat{\mathbf{k}}/n}(0) \le n^{\beta}$ independently of the realized treatment counts to get

$$\mathbb{E}\left[\tfrac{\sigma^2}{n}\sum_{t=0}^{\beta}\Big(\ell_{t,\hat{\mathbf{k}}/n}(1) - \ell_{t,\hat{\mathbf{k}}/n}(0)\Big)^2\right] \le \tfrac{\beta\sigma^2}{n} \cdot n^{2\beta}.$$

Let $\widehat{\mathrm{TTE}}_{-\varepsilon} := \frac{1}{n}\sum_{t=0}^{\beta}\sum_{i=1}^{n}\Big(\ell_{t,\hat{\mathbf{k}}/n}(1) - \ell_{t,\hat{\mathbf{k}}/n}(0)\Big)Y_i(\mathbf{z}^t)$.

Using the fact that the variance of Bernoulli random variables is always bounded above by 1, and again using the bound $\ell_{t,\hat{\mathbf{k}}/n}(1) - \ell_{t,\hat{\mathbf{k}}/n}(0) \le n^{\beta}$, we get

$$\begin{aligned}
\mathrm{Var}\Big[\widehat{\mathrm{TTE}}_{-\varepsilon}\Big] &\le \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{t=0}^{\beta}\sum_{t'=0}^{\beta}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i \\ |\mathcal{S}|\le\beta}}\sum_{\substack{\mathcal{S}'\subseteq\mathcal{N}_{i'} \\ |\mathcal{S}'|\le\beta}} \big|c_{i,\mathcal{S}}\big|\cdot\big|c_{i',\mathcal{S}'}\big|\cdot\big|\ell_{t,\frac{\mathbf{k}}{n}}(1) - \ell_{t,\frac{\mathbf{k}}{n}}(0)\big|\cdot\big|\ell_{t',\frac{\mathbf{k}}{n}}(1) - \ell_{t',\frac{\mathbf{k}}{n}}(0)\big| \\
&\le \beta^2 \cdot Y_{\max}^2 \cdot n^{2\beta}.
\end{aligned}$$

Then, to bound the second term of (A.4), we use the unconditional bound

$$\mathrm{Var}\Big[\widehat{\mathrm{TTE}}(\hat{\mathbf{k}}/n)\Big] \le \beta^2 \cdot Y_{\max}^2 \cdot n^{2\beta} \ +\ \tfrac{\beta\sigma^2}{n} \cdot n^{2\beta}. \tag{A.5}$$

Applying the definition of expectation, we have

$$\begin{aligned}
&\mathbb{E}\Big[\mathrm{Var}\Big[\widehat{\mathrm{TTE}}\Big|\sum_{j=1}^{n} z_j^t = \hat{k}_t\Big]\Big] \\
&\le \sum_{\mathbf{k}\in\mathcal{E}_2}\Pr\left(\sum_{j=1}^{n} z_j^t = \hat{k}_t\ \forall t\right)\cdot\mathrm{Var}\Big[\widehat{\mathrm{TTE}}\Big|\sum_{j=1}^{n} z_j^t = \hat{k}_t\Big] + \Pr(\mathcal{E}_2^c)\cdot(\beta^2 Y_{\max}^2 n^{2\beta} + \tfrac{\beta\sigma^2}{n} n^{2\beta}) \\
&\le O\Big(\beta^2\, Y_{\max}^2\Big(\tfrac{d^2}{n} + \tfrac{\beta^2}{(1-\delta)p_1 n}\Big)\cdot\big(\tfrac{n}{(\Delta_{\mathbf{p}}-\delta p)n}\big)^{2\beta} + \tfrac{\sigma^2\beta}{n}\Big(\tfrac{n}{(\Delta_{\mathbf{p}}-\delta p)n}\Big)^{2\beta}\Big) + \Pr(\mathcal{E}_2^c)\cdot(\beta^2 Y_{\max}^2 n^{2\beta} + \tfrac{\beta\sigma^2}{n} n^{2\beta}).
\end{aligned}$$

Here, the first equality makes use of our unconditional bound on the variance, given in inequality A.5. The second inequality plugs the variance bound from Theorem 3 for the most pessimistically perturbed treatment count vector in $\mathcal{E}_2$. The probability $\Pr(\mathcal{E}_2^c)$ decays exponentially in n. Therefore, choosing $\delta = \Theta(\frac{1}{\log(n)})$ and letting n get sufficiently large, the upper bound for this estimator is

$$O\Big(\beta^2\, Y_{\max}^2 \Big(\tfrac{d^2}{n} + \tfrac{\beta^2}{p_1 n}\Big) \cdot \Delta_{\mathbf{p}}^{-2\beta} + \tfrac{\beta\sigma^2}{n} \Delta_{\mathbf{p}}^{-2\beta}\Big).$$

□

B Unbiased Estimation with Additional Observations

A natural question is whether we continue to see improvements in the estimator when we increase the number of estimates beyond $\beta + 1$. Note that we restrict our attention to unbiased estimators, as we desire the asymptotic reduction in mean-squared error as the population grows large. We may thus assess the quality of an estimator by its variance. While in general, with noisy data, more measurements may result in improveed estimates, we show that in the linear setting, under perfect observations (i.e. no observation noise), these extra measurements do not help to reduce variance. In fact, we'll argue that the unbiased estimator with minimum variance is the one that ignores all but its first and last observations and then performs polynomial interpolation on these endpoints. We record this result in Theorem 10.

Theorem 10. *Suppose that the potential outcomes model is linear, and a staggered rollout Bernoulli design is implemented with a set of $T + 1$ distinct treatment probabilities $p_0 < p_1 < \ldots < p_T$. Then, the unbiased estimator for* TTE *of the form*

$$\widehat{\mathrm{TTE}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{T} \alpha_t Y_i(\mathbf{z}^t)$$

that minimizes variance has $\alpha_0 = \frac{-1}{p_T - p_0}$, $\alpha_T = \frac{1}{p_T - p_0}$ and $\alpha_1, \ldots, \alpha_{T-1} = 0$.

On one hand, such a result seems surprising: having more observations seems like it would only lead to a stronger estimator. However, what is overlooked is that there is strong correlation in the different measurements due to the monotonicity of treatments enforced in the staggered rollout design, such that the information in the first and last measurements contain all the useful information one could construct from the intermediate measurements. When random noise is added, the trade-off between the noise-canceling effects of additional measurements and the increased sensitivity of higher-degree interpolating polynomials adds an additional level of complexity.

Proof. To begin, we derive the constraints on $(\alpha_0, \ldots, \alpha_T)$ needed to ensure unbiasedness. We have,

$$\mathbb{E}\Big[\widehat{TTE}\Big] = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=0}^{T}\alpha_t\Big(c_{i,\varnothing} + p_t\sum_{j\in\mathcal{N}_i} c_{ij}\Big) = \frac{1}{n}\sum_{i=1}^{n}\Big[c_{i,\varnothing}\Big(\sum_{t=0}^{T}\alpha_T\Big) + \sum_{j\in\mathcal{N}_i} c_{ij}\Big(\sum_{t=0}^{T}\alpha_T p_t\Big)\Big].$$

Comparing to our expression for TTE in terms of the $c_{i,\mathcal{S}}$ coefficients:

$$\mathrm{TTE} = \tfrac{1}{n}\sum_{i=1}^{n}\sum_{\substack{\mathcal{S}\subseteq\mathcal{N}_i\\ 1\le|\mathcal{S}|\le\beta}} c_{i,\mathcal{S}},$$

we see that we must have,

$$\sum_{t=0}^{T}\alpha_t = 0, \qquad\qquad \sum_{t=0}^{T}\alpha_t p_t = 1. \tag{B.1}$$

Now, we consider the variance of this family of estimators. We have,

$$\begin{aligned}\text{Var}\Big[\widehat{TTE}\Big] &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{t=0}^{T}\sum_{t'=0}^{T}\alpha_t\alpha_{t'}\cdot\text{Cov}\Big[Y_i(\mathbf{z}^t), Y_{i'}(\mathbf{z}^{t'})\Big] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{t=0}^{T}\sum_{t'=0}^{T}\sum_{j\in\mathcal{N}_i\cap\mathcal{N}_{i'}}\alpha_t\alpha_{t'}\cdot c_{ij}c_{i'j}\cdot\big(p_{\min(t,t')}-p_tp_{t'}\big) \\ &= \bigg(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j\in\mathcal{N}_i\cap\mathcal{N}_{i'}}c_{ij}c_{i'j}\bigg)\bigg(\sum_{t=0}^{T}\sum_{t'=0}^{T}\alpha_t\alpha_{t'}\cdot\big(p_{\min(t,t')}-p_tp_{t'}\big)\bigg). \end{aligned} \tag{B.2}$$

Note that the first factor is a constant depending only on the network (i.e. not on the α and p parameters of the estimator). Thus, to minimize the variance, it suffices to locate critical values of this second factor, subject to our unbiasedness constraints. We can rewrite this factor

$$\sum_{t=0}^{T}\alpha_t^2\cdot p_t(1-p_t)+2\sum_{t=0}^{T}\sum_{t'=t+1}^{T}\alpha_t\alpha_{t'}\cdot p_t(1-p_{t'}) \;=\; \sum_{t=0}^{T}\alpha_tp_t\Big(\alpha_t(1-p_t)+2\sum_{t'=t+1}^{T}\alpha_{t'}(1-p_{t'})\Big).$$

Then, we consider the Lagrangian,

$$\mathcal{L} := \sum_{t=0}^{T}\alpha_tp_t\Big(\alpha_t(1-p_t)+2\sum_{t'=t+1}^{T}\alpha_{t'}(1-p_{t'})\Big)+\lambda\sum_{t=0}^{T}\alpha_t+\mu\Big(1-\sum_{t=0}^{T}\alpha_tp_t\Big). \tag{B.3}$$

We compute the partial derivatives of this Lagrangian with respect to each α_t as,

$$\frac{\partial\mathcal{L}}{\partial\alpha_t} = 2(1-p_t)\sum_{t'=0}^{t-1}\alpha_{t'}p_{t'}+2p_t\sum_{t''=t}^{T}\alpha_{t''}(1-p_{t''})+\lambda-p_t\mu.$$

We will set each of these partial derivatives equal to 0 sequentially to fix each of the variables at the critical point. First, we consider the partial derivative with respect to α_0. We have,

$$\frac{\partial\mathcal{L}}{\partial\alpha_0} = 2p_0\sum_{t''=0}^{T}\alpha_{t''}(1-p_{t''})+\lambda-p_0\mu = -p_0(2+\mu)+\lambda.$$

Here, the second inequality uses the unbiasedness constraints. Setting this partial derivative equal to 0, we must have $\lambda = p_0(2+\mu)$. Next, we consider the partial derivative with respect to α_1:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\alpha_1} &= 2\alpha_0p_0(1-p_1)-2p_1\sum_{t''=1}^{T}\alpha_{t''}(1-p_{t''})+\lambda-p_1\mu \\ &= 2\alpha_0p_0(1-p_1)+2p_1\Big(-1-\alpha_0(1-p_0)\Big)+\lambda-p_1\mu && \text{(unbiasedness)} \\ &= 2\alpha_0p_0(1-p_1)-2p_1-2p_1\alpha_0(1-p_0)+p_0(2+\mu)-p_1\mu \\ &= (p_0-p_1)(2\alpha_0+2+\mu).\end{aligned}$$

Note that $p_0-p_1\neq 0$ by our distinct probabilities assumption. Thus, setting this partial derivative equal to 0, we must have $2+\mu = -2\alpha_0$. In addition, combining with the previous constraint, we can re-express $\lambda = -2\alpha_0p_0$. Next, we consider the partial derivative with respect to α_2:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\alpha_2} &= 2\alpha_0p_0(1-p_2)+2\alpha_1p_1(1-p_2)-2p_2-2p_2\alpha_0(1-p_0)-2p_2\alpha_1(1-p_1)+\lambda-p_2\mu \\ &= 2\alpha_0(p_0-p_2)+2\alpha_1(p_1-p_2)-2\alpha_0p_0-p_2(2+\mu) \\ &= 2\alpha_0(p_0-p_2)+2\alpha_1(p_1-p_2)-2\alpha_0(p_0-p_2) \\ &= 2\alpha_1(p_1-p_2).\end{aligned}$$

Setting this partial derivative equal to 0, we must have $\alpha_1 = 0$, since $p_1-p_2\neq 0$. We can iterate this process on the partial derivatives with respect to $\alpha_3,\ldots,\alpha_T$, concluding that $\alpha_2,\ldots,\alpha_{T-1}=0$.

We are left with the system of two linear equations given by the unbiasedness constraints:

$$\alpha_0+\alpha_T = 0, \qquad\qquad \alpha_0p_0+\alpha_Tp_T = 1.$$

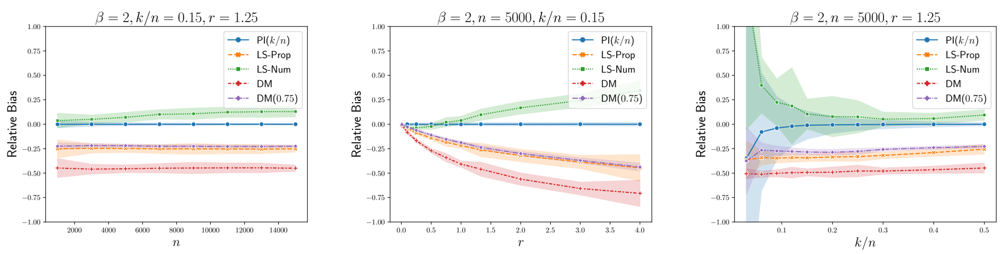
The unique solution to this system is $\alpha_0 = \frac{-1}{p_T-p_0}$, $\alpha_T = \frac{1}{p_T-p_0}$. □

C Experimental Results under a Quadratic Outcomes Model

In this section, we discuss the results of our experiments[1] under a quadratic potential outcomes model ($\beta = 2$). As in the linear setting (see Section 4), for each population size n, we sample G networks from the distribution described in Section 4. For each configuration of parameters in the experiment, we sample N treatment schedules $\{\mathbf{z}^0, \dots, \mathbf{z}^\beta\}$ from our parameterized distribution class (Bernoulli or CRD) and compute the TTE using each estimator. In the experiments for both this setting and the linear setting, we set $G = 30$ and $N = 100$.

For each estimator, we plot the relative bias of the TTE estimates averaged over the results from these GN samples and normalized by the magnitude of the TTE. The width of the shading in the figures depicts the standard deviation across the GN estimates. The experiments in the quadratic setting ran for 29.4 minutes on the same Linux machine.

In Figure 3, we visualize the effect of three network or estimator parameters on the quality of each of the five TTE estimators (the four described in the Other Algorithms paragraph of Section 4, and our CRD estimator with treatment targets $k_t = \frac{tk}{\beta}$). Specifically, we consider the effects of the population size (n), the maximum proportion of treated individuals (k/n) and the degree of the potential outcomes model (β). Each of the plots fixes two of these parameters and varies the third. Specific settings of the parameters are listed on each plot.



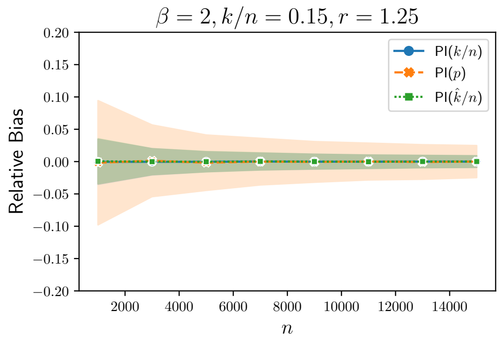
(a) Varying size of the population (b) Varying direct:indirect effects (c) Varying treatment budget

Figure 3: Three graphs visualizing the performance of various TTE estimators as different parameters are adjusted. The height of each graph depicts the experimental relative bias of the estimator and the shaded width depicts the experimental standard deviation.

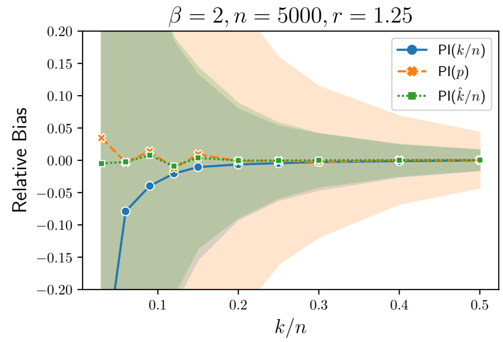
Our estimator is the blue line with blue shading on each of the plots. As expected, the estimator is unbiased and the variance decreases as n or k/n increases. On the other hand, regardless of population size or treatment budget, the rest of the estimators remain biased. In general, the variances of these other estimators remains higher than ours, although it is worth noting that when the treatment budget k/n is lower, the variance of our estimator is higher. As the ratio r increases, the network (aka indirect) effects become greater relative to the direct effect. This is exhibited by the increase in the bias of all the estimators, besides ours, as shown in Figure 3b. As expected, when the ratio is near 0, all estimators are unbiased as this corresponds to the case where there is no network interference.

In Figure 4, we compare the variants of our estimator when $\beta = 2$, evaluating $\widehat{\text{TTE}}_{\text{PI}}(\mathbf{k}/n)$ under CRD and evaluating $\widehat{\text{TTE}}_{\text{PI}}(\mathbf{p})$ and $\widehat{\text{TTE}}_{\text{PI}}(\hat{\mathbf{k}}/n)$ under Bernoulli($\mathbf{p}$) randomized design, where $p_t = tp/\beta$ and $\hat{\mathbf{k}}$ is the vector of realized treatment counts.

The estimators $\widehat{\text{TTE}}_{\text{PI}}(\mathbf{k}/n)$ and $\widehat{\text{TTE}}_{\text{PI}}(\hat{\mathbf{k}}/n)$ perform nearly identically as we vary the size of the population. They differ for lower treatment budgets, with $\widehat{\text{TTE}}_{\text{PI}}(\hat{\mathbf{k}}/n)$ having lower bias than $\widehat{\text{TTE}}_{\text{PI}}(\mathbf{k}/n)$ but about the same variance. As the treatment budget increases, they perform almost identically. $\widehat{\text{TTE}}_{\text{PI}}(\hat{\mathbf{k}}/n)$ has lower variance than $\widehat{\text{TTE}}_{\text{PI}}(\mathbf{p})$, which is intuitive as it performs polynomial interpolation on the realized treatment fraction rather than the expected treatment fraction.



(a) Varying size of the population



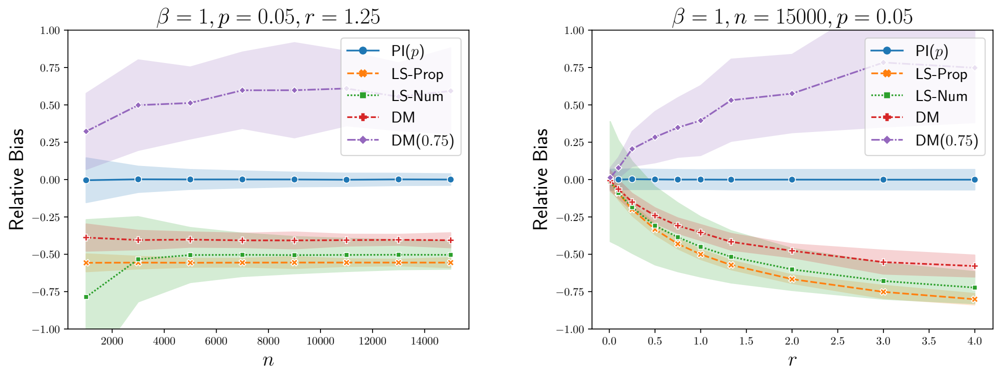
(b) Varying treatment budget

Figure 4: Two graphs visualizing the performance of our proposed TTE estimators as the size of the population (n) or treatment budget (k/n) is varied. The height of each graph depicts the experimental relative bias of the estimator and the shaded width depicts the experimental standard deviation. The blue and the green plots essentially overlap.

D Experimental Results under Bernoulli Design

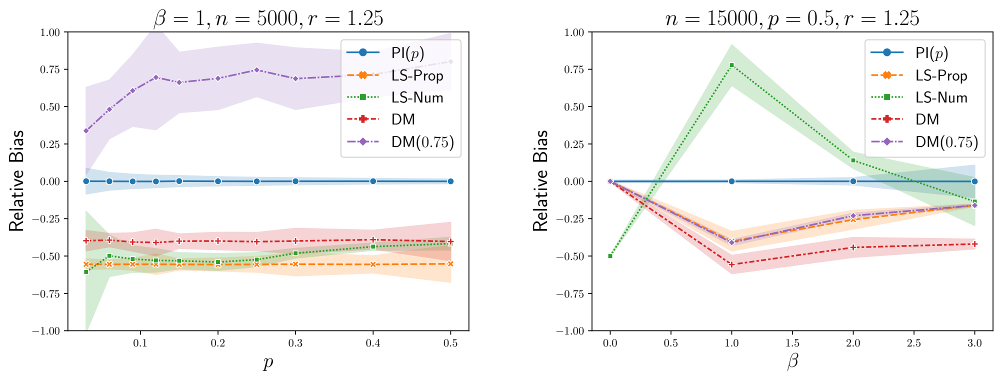
We performed similar experiments to Section 4 and Appendix C for the Bernoulli randomized design setting. The main difference is that our parameterization on the budget in the realized fraction of treated individuals, k/n, has been replaced by an upper threshold on the treatment probability, p. The results we find in this Bernoulli design setting exhibit the same trends as those under completely randomized design. We include these plots for completeness and refer the reader to earlier sections for discussion and analysis.

[1]Code can be found at https://tinyurl.com/kee88h6d



(a) Varying size of the population

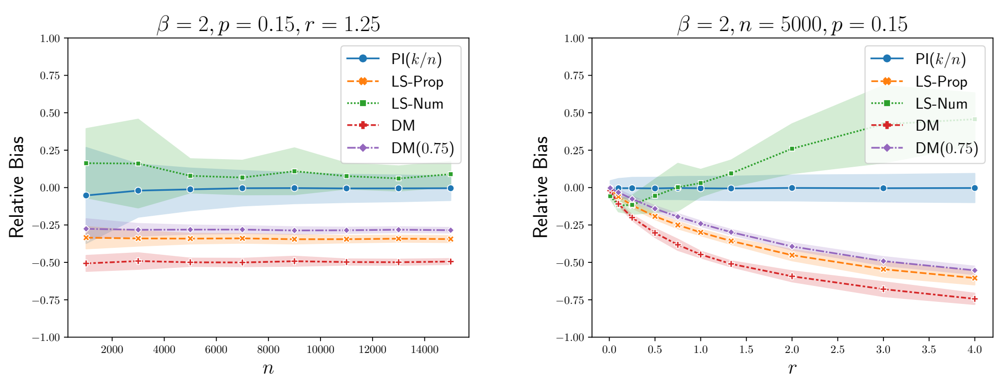
(b) Varying ratio of direct:indirect effects



(c) Varying treatment budget

(d) Varying the model degree

Figure 5: Four graphs visualizing the performance of various TTE estimators, under Bernoulli randomized design, as various parameters are adjusted. The height of each graph depicts the experimental relative bias of the estimator and the shaded width depicts the experimental standard deviation.



(a) Varying size of the population

(b) Varying direct:indirect effects

Figure 6: Two graphs visualizing the performance of our proposed TTE estimators under Bernoulli randomized design as the size of the population (n) or ratio between direct and indirect effects (r) is varied. The height of each graph depicts the experimental relative bias of the estimator and the shaded width depicts the experimental standard deviation.