Inference for ATE with Small Strata and Cluster-Level Treatment Assignment

Juri Trifonov

Mar 2024

1 Individual-level treatment assignment

1.1 Notation and Setup

Consider n strata, each of fixed size k. In the case of individual-level treatment assignment, strata are formed by matching individuals according to some function. For cluster-level treatment assignment, strata consist of clusters. Within each stratum, $k(d)$ units are assigned to treatment d, and $k(0)$ units to control. The total number of units is $N = n \times k$ in the individual-level assignment case, and the total number of clusters is $G = n \times k$ in the cluster-level case. We define the share of units assigned to treatment d as $\pi_d = \frac{k(d)}{k}$, and the share assigned to control as $\pi_0 = \frac{k(0)}{k}$, where $k(d)$ and $k(0)$ are coprime.

1.2 ATE Estimator

For every treatment $d \in \mathcal{D}$, define the following objects:

$$N_d := \sum_{1 \leqslant i \leqslant N} \mathbb{I}\{D_i = d\}, \text{and } N_0 := \sum_{1 \leqslant i \leqslant N} \mathbb{I}\{D_i = 0\}.$$

Then, we can define the ATE estimator for every treatment $d \in \mathcal{D}$ as follows:

$$\begin{aligned} \hat{\theta}_d^{adj} = & \frac{1}{N_d} \sum_{1 \leqslant i \leqslant N} Y_i \mathbb{I}\{D_i = d\} - \frac{1}{N_0} \sum_{1 \leqslant i \leqslant N} Y_i \mathbb{I}\{D_i = 0\} \\ & - \left[\frac{1}{N_d} \sum_{1 \leqslant i \leqslant N} \left(\psi_i - \bar{\psi}_N\right) \mathbb{I}\{D_i = d\} - \frac{1}{N_0} \sum_{1 \leqslant i \leqslant N} (\psi_i - \bar{\psi}_N) \mathbb{I}\{D_i = 0\} \right]' \hat{\beta}_d, \end{aligned}$$

where $\hat{\beta}_d$ is estimated by running the following regression:

$$\left(\frac{1}{k(d)} \sum_{i \in \lambda_j} Y_i \mathbb{I}\{D_i = d\} - \frac{1}{k(0)} \sum_{i \in \lambda_j} Y_i \mathbb{I}\{D_i = 0\} \right) = \gamma + \left(\frac{1}{k(d)} \sum_{i \in \lambda_j} \psi_i \mathbb{I}\{D_i = d\} - \frac{1}{k(0)} \sum_{i \in \lambda_j} \psi_i \mathbb{I}\{D_i = 0\} \right)' \beta_d + v_i.$$

Remark 1. *Note that this estimator is equivalent to the partialled Lin estimator (PLin) discussed in Cytrynbaum (2024). To see this, observe that we can rewrite the estimator in the general form:*

$$\hat{\theta}_{adj} = \hat{\theta}_{un} - \left(\bar{\psi}_1 - \bar{\psi}_0\right)' \hat{\beta}.$$

Specifically,

$$\frac{1}{N_1} \sum_{i=1}^{N} (\psi_i - \bar{\psi}_N) D_i = \bar{\psi}_1 - \frac{1}{N_1} N_1 \bar{\psi}_N = \bar{\psi}_1 - \bar{\psi}_N,$$

$$\frac{1}{N_0} \sum_{i=1}^{N} (\psi_i - \bar{\psi}_N)(1 - D_i) = \bar{\psi}_0 - \frac{1}{N_0} N_0 \bar{\psi}_N = \bar{\psi}_0 - \bar{\psi}_N.$$

Subtracting the second expression from the first yields exactly the form used in the PLin estimator.

1.3 Variance Estimator

First, define the adjusted outcome Y_i^a:

$$Y_i^a = Y_i - (\psi_i - \bar{\psi}_N)'\hat{\beta}_d.$$

Next, define the following objects to construct the variance estimator:

$$\hat{\Gamma}_n(d) := \frac{1}{n \times k(d)} \sum_{1 \leqslant i \leqslant N} Y_i^a \mathbb{I}\{D_i = d\}; \quad \hat{\Gamma}_n(0) := \frac{1}{n \times k(0)} \sum_{1 \leqslant i \leqslant N} Y_i^a \mathbb{I}\{D_i = 0\},$$

where $k(d) = \sum_{i \in \lambda_j} \mathbb{I}\{D_i = d\}$ and $k(0) = \sum_{i \in \lambda_j} \mathbb{I}\{D_i = 0\}$, i.e., they represent the number of treated (with treatment d) and untreated observations in each stratum (tuple), respectively.

Then,

$$\hat{\sigma}_n^2(d) := \frac{1}{n \times k(d)} \sum_{1 \leqslant i \leqslant N} \left(Y_i^a - \hat{\Gamma}_n(d)\right)^2 \mathbb{I}\{D_i = d\};$$

$$\hat{\sigma}_n^2(0) := \frac{1}{n \times k(0)} \sum_{1 \leqslant i \leqslant N} \left(Y_i^a - \hat{\Gamma}_n(0)\right)^2 \mathbb{I}\{D_i = 0\}.$$

Further, consider the following objects:

$$\hat{\rho}_n(d,0) := \frac{1}{n} \sum_{1 \leqslant j \leqslant n} \frac{1}{k(d) \times k(0)} \left(\sum_{i \in \lambda_j} Y_i^a \mathbb{I}\{D_i = d\}\right) \left(\sum_{i \in \lambda_j} Y_i^a \mathbb{I}\{D_i = 0\}\right);$$

$$\hat{\rho}_n(d,d) := \frac{2}{n} \sum_{1 \leqslant j \leqslant \lfloor n/2 \rfloor} \frac{1}{k(d)^2} \left(\sum_{i \in \lambda_{2j-1}} Y_i^a \mathbb{I}\{D_i = d\}\right) \left(\sum_{i \in \lambda_{2j}} Y_i^a \mathbb{I}\{D_i = d\}\right);$$

$$\hat{\rho}_n(0,0) := \frac{2}{n} \sum_{1 \leqslant j \leqslant \lfloor n/2 \rfloor} \frac{1}{k(0)^2} \left(\sum_{i \in \lambda_{2j-1}} Y_i^a \mathbb{I}\{D_i = 0\}\right) \left(\sum_{i \in \lambda_{2j}} Y_i^a \mathbb{I}\{D_i = 0\}\right);$$

and also the following components of the variance estimator:

$$\hat{\mathbb{V}}_{1,n}(d) := \hat{\sigma}_n^2(d) - \left(\hat{\rho}_n(d,d) - \hat{\Gamma}_n(d)\right)^2, \quad \hat{\mathbb{V}}_{1,n}(0) := \hat{\sigma}_n^2(0) - \left(\hat{\rho}_n(0,0) - \hat{\Gamma}_n(0)\right)^2;$$

$$\hat{\mathbb{V}}_{2,n}(d,d) := \hat{\rho}_n(d,d) - \hat{\Gamma}_n^2(d), \quad \hat{\mathbb{V}}_{2,n}(0,0) := \hat{\rho}_n(0,0) - \hat{\Gamma}_n^2(0);$$

$$\hat{\mathbb{V}}_{2,n}(d,0) := \hat{\rho}_n(d,0) - \hat{\Gamma}_n(d)\hat{\Gamma}_n(0).$$

Finally, we can define the variance estimator as follows:

$$\hat{\mathbb{V}}^{adj} = \frac{1}{\pi_1}\hat{\mathbb{V}}_{1,n}(d) + \frac{1}{\pi_0}\hat{\mathbb{V}}_{1,n}(0) + \hat{\mathbb{V}}_{2,n}(d,d) + \hat{\mathbb{V}}_{2,n}(0,0) - 2\hat{\mathbb{V}}_{2,n}(d,0).$$

Then, the "adjusted" t-test is defined as:

$$\phi_N^{adj}\left(W^{(N)}\right) = \mathbb{I}\left\{\left|\sqrt{N}\left(\hat{\theta}^{adj} - \theta_0\right)/\hat{\mathbb{V}}^{adj}\right| > z_{1-\alpha/2}\right\},$$

and the test is asymptotically exact, i.e., under the true H_0:

$$\lim_{n \to \infty} \mathbb{E}\left[\phi_N^{adj}\left(W^{(N)}\right)\right] = \alpha.$$

■

1.3.1 Mixed Estimator Setup

Consider a mixed estimator that is a weighted average of estimators for observations in "small" and "big" strata, i.e.,

$$\hat{\tau}_{mix} = \frac{N_S}{N} \times \hat{\tau}_S + \frac{N_B}{N} \times \hat{\tau}_B,$$

where N_S and N_B represent the number of observations in "small" and "big" strata respectively.

Then:

$$\begin{aligned}\sqrt{N}(\hat{\tau}_{mix} - \tau) &= \sqrt{N}\left(\left[\frac{N_s}{N} \times \hat{\tau}_S + \frac{N_B}{N} \times \hat{\tau}_B\right] - [p_S \times \tau_S + p_B \times \tau_B]\right) \\ &= \underbrace{\sqrt{N}\left(\frac{N_B}{N}(\hat{\tau}_B - \tau_B) + \frac{N_S}{N}(\hat{\tau}_S - \tau_S)\right)}_{W} \\ &\quad \underbrace{- \sqrt{N}\left(\tau_B\left(\frac{N_B}{N} - p_B\right) + \tau_S\left(\frac{N_S}{N} - p_S\right)\right)}_{Z}\end{aligned}$$

Consider the variance of the first term:

$$\mathbb{V}[W] = N\left(\frac{N_B}{N}\right)^2 \times \mathbb{V}[\hat{\tau}_B] + N\left(\frac{N_S}{N}\right)^2 \times \mathbb{V}[\hat{\tau}_S].$$

Now, consider the variance of Z. Note that $\frac{N_S}{N} = 1 - \frac{N_B}{N}$ and $p_S = 1 - p_B$. Also, denote $X := \frac{N_B}{N} - p_B$. Then clearly:

$$\frac{N_S}{N} - p_S = -\frac{N_B}{N} - p_B = -X.$$

Thus, we can rewrite the second term as:

$$Z := \sqrt{N}\left(X(\tau_B - \tau_S)\right) = \sqrt{N}\left[\tau_B\left(\frac{N_b}{N} - p_B\right) - \tau_S\left(\frac{N_B}{N} - p_B\right)\right].$$

We can now compute the $\mathbb{V}[Z]$:

$$\mathbb{V}[Z] = N(\tau_B - \tau_S)^2 \mathbb{V}\left[\frac{N_B}{N} - p_B\right],$$

where

$$\mathbb{V}\left[\frac{N_B}{N} - p_B\right] = \frac{p_B(1 - p_B)}{N}.$$

Thus, using the plug-in estimator, we get:

$$\hat{\mathbb{V}}[Z] = \frac{N_B \times N_S}{N^2}\left(\hat{\tau}_B - \hat{\tau}_S\right)^2.$$

Finally, combining both terms, the variance estimator of $\hat{\tau}_{mix}$ takes the following form:

$$\begin{aligned}\hat{\mathbb{V}}[\hat{\tau}_{mix}] = &\left(\frac{N_S}{N}\right)^2 \times \hat{\mathbb{V}}_S + \left(\frac{N_B}{N}\right)^2 \times \hat{\mathbb{V}}_B \\ &+ \frac{N_B \times N_S}{N^2} \times (\hat{\tau}_B - \hat{\tau}_S)^2\end{aligned}$$

■

Remark 2. *Note that the estimator for $\mathbb{V}[Z]$ coincides with the estimator $\frac{1}{N}\sum_{i=1}^{N}\hat{\Xi}_2(s)$ in the expression for a "big-strata" variance estimator (or in* **jiang2023***)*

2 Cluster-level treatment assignment

2.1 ATE Estimator

It is natural to generalize the setup for the cluster-level treatment assignment.

First, the one-step procedure assigns treatment on a cluster level. That is, either everyone in a cluster is treated or not. Then, the adjusted ATE estimator can be defined for every treatment $d \in \mathcal{D}$ as follows:

$$\begin{aligned}\hat{\theta}_d^{adj} &= \frac{1}{G_d}\sum_{1\leqslant g\leqslant G} N_g \bar{Y}_g \mathbb{I}\{D_g = d\} - \frac{1}{G_0}\sum_{1\leqslant g\leqslant G} N_g \bar{Y}_g \mathbb{I}\{D_g = 0\} \\ &- \left[\frac{1}{G_d}\sum_{1\leqslant g\leqslant G}\left(\psi_g - \bar{\psi}_G\right)\mathbb{I}\{D_g = d\} - \frac{1}{G_0}\sum_{1\leqslant g\leqslant G}(\psi_g - \bar{\psi}_G)\mathbb{I}\{D_g = 0\}\right]' \hat{\beta}_d,\end{aligned}$$

where $\hat{\beta}_d$ is estimated by running the following regression:

$$\begin{aligned}\left(\frac{1}{k(d)}\sum_{g\in\lambda_j}\bar{Y}_g\bar{N}_G\mathbb{I}\{D_g = d\} - \frac{1}{k(0)}\sum_{g\in\lambda_j}\bar{Y}_g\bar{N}_G\mathbb{I}\{D_g = 0\}\right) &= \gamma \\ &+ \left(\frac{1}{k(d)}\sum_{g\in\lambda_j}\psi_g\mathbb{I}\{D_g = d\} - \frac{1}{k(0)}\sum_{g\in\lambda_j}\psi_g\mathbb{I}\{D_g = 0\}\right)'\beta_d + v_i,\end{aligned}$$

where $G_d = \sum_{1\leqslant g\leqslant G}\mathbb{I}\{D_g = d\}N_g$, $G_0 = \sum_{1\leqslant g\leqslant G}\mathbb{I}\{D_g = 0\}N_g$, and $\bar{N}_G = \sum_{1\leqslant g\leqslant G} N_g/G$.

2.2 Variance Estimator

First, define the adjusted cluster-level outcome $\bar{Y}_g^a$:

$$\bar{Y}_g^a = \frac{N_g}{\frac{1}{G}\sum_{1\leqslant g\leqslant G} N_g}\bar{Y}_g - \frac{(\psi_g - \bar{\psi}_G)'\hat{\beta}_d}{\frac{1}{G}\sum_{1\leqslant g\leqslant G} N_g}.$$

Next, define the following objects to construct the variance estimator:

$$\hat{\Gamma}_n(d) := \frac{1}{n\times k(d)}\sum_{1\leqslant g\leqslant G}\bar{Y}_g^a\mathbb{I}\{D_g = d\};\ \hat{\Gamma}_n(0) := \frac{1}{n\times k(0)}\sum_{1\leqslant g\leqslant G}\bar{Y}_g^a\mathbb{I}\{D_g = 0\},$$

where $k(d) = \sum_{g\in\lambda_j}\mathbb{I}\{D_g = 1\}$ and $k(0) = \sum_{g\in\lambda_j}\mathbb{I}\{D_g = 0\}$, i.e., represent the number of treated with treatment d/untreated clusters in each strata (tuple).

Then,

$$\hat{\sigma}_n^2(d) := \frac{1}{n\times k(d)}\sum_{1\leqslant g\leqslant G}\left(\bar{Y}_g^a - \hat{\Gamma}_n(d)\right)^2\mathbb{I}\{D_g = d\};$$

$$\hat{\sigma}_n^2(0) := \frac{1}{n\times k(0)}\sum_{1\leqslant g\leqslant G}\left(\bar{Y}_g^a - \hat{\Gamma}_n(0)\right)^2\mathbb{I}\{D_g = 0\}.$$

Further, consider the following objects:

$$\hat{\rho}_n(d,0) := \frac{1}{n}\sum_{1\leqslant j\leqslant n}\frac{1}{k(d)\times k(0)}\left(\sum_{g\in\lambda_j}\bar{Y}_g^a\mathbb{I}\{D_g = d\}\right)\left(\sum_{g\in\lambda_j}\bar{Y}_g^a\mathbb{I}\{D_g = 0\}\right);$$

$$\hat{\rho}_n(d,d) := \frac{2}{n}\sum_{1\leqslant j\leqslant\lfloor n/2\rfloor}\frac{1}{k(d)^2}\left(\sum_{g\in\lambda_{2j-1}}\bar{Y}_g^a\mathbb{I}\{D_g = d\}\right)\left(\sum_{g\in\lambda_{2j}}\bar{Y}_g^a\mathbb{I}\{D_g = d\}\right);$$

$$\hat{\rho}_n(0,0) := \frac{2}{n}\sum_{1\leqslant j\leqslant\lfloor n/2\rfloor}\frac{1}{k(0)^2}\left(\sum_{g\in\lambda_{2j-1}}\bar{Y}_g^a\mathbb{I}\{D_g = 0\}\right)\left(\sum_{g\in\lambda_{2j}}\bar{Y}_g^a\mathbb{I}\{D_g = 0\}\right);$$

and also the following components of the variance estimator:

$$\hat{\mathbb{V}}_{1,n}(d) := \hat{\sigma}_n^2(d) - \left(\hat{\rho}_n(d,d) - \hat{\Gamma}_n(d)\right)^2, \ \hat{\mathbb{V}}_{1,n}(0) := \hat{\sigma}_n^2(0) - \left(\hat{\rho}_n(0,0) - \hat{\Gamma}_n(0)\right)^2;$$

$$\hat{\mathbb{V}}_{2,n}(d,d) := \hat{\rho}_n(d,d) - \hat{\Gamma}_n^2(d), \ \hat{\mathbb{V}}_{2,n}(0,0) := \hat{\rho}_n(0,0) - \hat{\Gamma}_n^2(0);$$

$$\hat{\mathbb{V}}_{2,n}(d,0) := \hat{\rho}_n(d,0) - \hat{\Gamma}_n(d)\hat{\Gamma}_n(0).$$

Using these expressions, the variance estimator can be constructed in exactly same way as in the case of individual-level treatment assignment.

$$\hat{\mathbb{V}}^{adj} = \frac{1}{\pi_d}\hat{\mathbb{V}}_{1,n}(d) + \frac{1}{\pi_0}\hat{\mathbb{V}}_{1,n}(0) + \hat{\mathbb{V}}_{2,n}(d,d) + \hat{\mathbb{V}}_{2,n}(0,0) - 2\hat{\mathbb{V}}_{2,n}(d,0).$$

Then, the “adjusted“ t-test is defined as:

$$\phi_G^{adj}\left(W^{(G)}\right) = \mathbb{I}\left\{\left|\sqrt{G}\left(\hat{\theta}^{adj} - \theta_0\right)/\hat{\mathbb{V}}^{adj}\right| > z_{1-\alpha/2}\right\},$$

and the test is asymptotically exact, i.e. under the true H_0:

$$\lim_{n\to\infty} \mathbb{E}\left[\phi_G^{adj}\left(W^{(G)}\right)\right] = \alpha$$

■