At What Level Should One Cluster Standard Errors in Paired and Small-Strata Experiments?

By Clément de Chaisemartin and Jaime Ramirez-Cuellar

Online Appendix

A. Proof of Lemma I

We first introduce some notation. Let $T_p = n_{1p}W_{1p} + n_{2p}W_{2p}$ and $C_p = n_{1p}(1-W_{1p})+n_{2p}(1-W_{2p})$ be the number of treated and untreated observations in pair p. Let $T = \sum_{p=1}^{P} T_p$ and $C = \sum_{p=1}^{P} C_p$ be the total number of treated and untreated observations. Let $SET_p = \sum_{g=1}^{2}\sum_{i=1}^{n_{gp}} W_{gp}\epsilon_{igp}$ and $SEU_p = \sum_{g=1}^{2}\sum_{i=1}^{n_{gp}}(1-W_{gp})\epsilon_{igp}$ respectively be the sum of the residuals ϵ_{igp} for the treated and untreated observations in pair p.

$\widehat{\tau}$ is the well-known difference-in-means estimator:

$$\widehat{\tau} = \sum_{p=1}^{P}\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}} \frac{Y_{igp}W_{gp}}{T} - \sum_{p=1}^{P}\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}} \frac{Y_{igp}(1-W_{gp})}{C}.$$

Remember that $\widehat{\tau}_p = \sum_{g=1}^{2}\left[W_{gp}\sum_{i=1}^{n_{gp}}\frac{Y_{igp}}{n_{gp}} - (1-W_{gp})\sum_{i=1}^{n_{gp}}\frac{Y_{igp}}{n_{gp}}\right]$ is the difference between the average outcome of treated and untreated observations in pair p. It follows from, e.g., Equation (3.3.7) in Angrist and Pischke (2008) and a few lines of algebra that

$$\widehat{\tau}_{fe} = \sum_{p=1}^{P}\omega_p\widehat{\tau}_p, \qquad \text{where} \quad \omega_p = \frac{\left(n_{1p}^{-1}+n_{2p}^{-1}\right)^{-1}}{\sum_{p'=1}^{P}\left(n_{1p'}^{-1}+n_{2p'}^{-1}\right)^{-1}}.$$

Point 1

Proof of $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) = \widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$

It follows from Equations (1) and (2) that

$$\widehat{\alpha} + \widehat{\tau}W_{gp} + \epsilon_{igp} = \widehat{\tau}_{fe}W_{gp} + \sum_{p=1}^{P}\widehat{\gamma}_p\delta_{igp} + u_{igp}.$$

Rearranging and using the fact that under Assumption 2 $\widehat{\tau} = \widehat{\tau}_{fe}$, one obtains that for every p:

$$\epsilon_{igp} = \widehat{\gamma}_p - \widehat{\alpha} + u_{igp}. \tag{A1}$$

Then,

$$\begin{aligned}
\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) &= \frac{1}{T^2}\sum_{p=1}^{P}(SET_p - SEU_p)^2 \\
&= \frac{1}{T^2}\sum_{p}\left[\sum_{g}\sum_{i}(2W_{gp}-1)\epsilon_{igp}\right]^2 \\
&= \frac{1}{T^2}\sum_{p}\left[\sum_{g}\sum_{i}(2W_{gp}-1)(\widehat{\gamma}_p - \widehat{\alpha} + u_{igp})\right]^2 \\
&= \frac{1}{T^2}\sum_{p}\left[\sum_{g}\sum_{i}(2W_{gp}-1)u_{igp} + (\widehat{\gamma}_p - \widehat{\alpha})\sum_{g}\sum_{i}(2W_{gp}-1)\right]^2 \\
&= \frac{4}{T^2}\sum_{p}\left(\sum_{g}\sum_{i}W_{gp}u_{igp}\right)^2. \qquad \text{(A2)}
\end{aligned}$$

The first equality follows from Point 1 of Lemma C.1 and Assumption 2. The third equality follows from Equation (A1). The fifth follows from the following two facts. First, $\sum_g\sum_i(2W_{gp}-1)u_{igp} = 2\sum_g\sum_i W_{gp}u_{igp} - \sum_g\sum_i u_{igp} = 2\sum_g\sum_i W_{gp}u_{igp}$, since $\sum_g\sum_i u_{igp} = 0$ by definition of u_{igp}. Second, $(\widehat{\gamma}_p - \widehat{\alpha})\sum_g\sum_i(2W_{gp}-1) = (\widehat{\gamma}_p - \widehat{\alpha})\left[\sum_g\sum_i W_{gp} - \sum_g\sum_i(1-W_{gp})\right] = (\widehat{\gamma}_p - \widehat{\alpha})[T_p - C_p] = 0$, where the last equality comes from the fact that $n_{1p} = n_{2p}$ by Assumption 2.

Similarly,

$$\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe}) = \frac{4}{T^2}\sum_{p=1}^{P}SET_{p,fe}^2 = \frac{4}{T^2}\sum_{p=1}^{P}\left(\sum_{g}\sum_{i}W_{gp}u_{igp}\right)^2, \tag{A3}$$

where the first equality follows from Equation (H22) in the proof of Lemma C.1 and Assumption 2. Combining Equations (A2) and (A3) yields $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) = \widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$.

Proof of $\mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right] = \mathbb{V}(\widehat{\tau}) + \frac{1}{P(P-1)}\sum_{p=1}^{P}(\tau_p - \tau)^2$

Under Assumption 2, $T = C = n/2$, so

$$
\begin{aligned}
\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) &= \sum_{p=1}^{P}\left(\frac{SET_p}{T} - \frac{SEU_p}{C}\right)^2 \\
&= \frac{4}{n^2}\sum_{p=1}^{P}\left(SET_p - SEU_p\right)^2 \\
&= \frac{4}{n^2}\sum_{p=1}^{P}\left(\sum_g\sum_i(W_{gp}\epsilon_{igp} - (1-W_{gp})\epsilon_{igp})\right)^2 \\
&= \frac{4}{n^2}\sum_{p=1}^{P}\left(\sum_g\sum_i(2W_{gp}-1)\epsilon_{igp}\right)^2 \\
&= \frac{4}{n^2}\sum_{p=1}^{P}\left(\sum_g(2W_{gp}-1)\sum_i(Y_{igp} - \widehat{\tau}W_{gp} - \widehat{\alpha})\right)^2 \\
&= \frac{4}{n^2}\sum_{p=1}^{P}\left(\sum_g(2W_{gp}-1)\left(\sum_i Y_{igp} - \widehat{\tau}W_{gp}\frac{n_p}{2} - \widehat{\alpha}\frac{n_p}{2}\right)\right)^2 \\
&= \frac{4}{n^2}\sum_{p=1}^{P}\left(\sum_g(2W_{gp}-1)\sum_i Y_{igp} - \widehat{\tau}\frac{n_p}{2}\sum_g(2W_{gp}-W_{gp}) - \widehat{\alpha}\frac{n_p}{2}\sum_g(2W_{gp}-1)\right)^2 \\
&= \frac{4}{n^2}\sum_{p=1}^{P}\left(\sum_g(2W_{gp}-1)\sum_i Y_{igp} - \widehat{\tau}\frac{n_p}{2}\sum_g W_{gp}\right)^2 \\
&= \frac{4}{n^2}\sum_{p=1}^{P}\left(\sum_g(2W_{gp}-1)\sum_i Y_{igp} - \widehat{\tau}\frac{n_p}{2}\right)^2 \\
&= \frac{4}{n^2}\sum_{p=1}^{P}\left(\widehat{\tau}_p\frac{n_p}{2} - \widehat{\tau}\frac{n_p}{2}\right)^2
\end{aligned}
$$

(A4)

$$
= \frac{1}{P^2}\sum_{p=1}^{P}\left(\widehat{\tau}_p - \widehat{\tau}\right)^2.
$$

The third equality comes from the definition of SET_p and SEU_p. The fifth equality follows from the Equation (1). The sixth equality follows from $n_{1p} = n_{2p} = n_p/2$, which is a consequence of Assumption 2. The eighth equality comes from the fact that $\sum_g(2W_{gp} - 1) = 0$, which follows from Point 1 of Assumption 1. The ninth equality follows from Point 1 of Assumption 1. The tenth equality follows from $\sum_g (2W_{gp} - 1) \sum_i Y_{igp} = \sum_g W_{gp} \sum_i Y_{igp} - \sum_g (1 - W_{gp}) \sum_i Y_{igp} = n_p \widehat{\tau}_p / 2$. The eleventh equality follows from Assumption 2.

Now, consider Equation (A4). Adding and subtracting τ and $\tau_p = \mathbb{E}[\widehat{\tau}_p]$,

$$\begin{aligned}\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) &= \frac{1}{P^2} \sum_{p=1}^{P} \left((\widehat{\tau}_p - \tau_p) - (\widehat{\tau} - \tau) + (\tau_p - \tau)\right)^2 \\ &= \frac{1}{P^2} \sum_{p=1}^{P} \left[(\widehat{\tau}_p - \tau_p)^2 + (\widehat{\tau} - \tau)^2 + (\tau_p - \tau)^2 - 2(\widehat{\tau}_p - \tau_p)(\widehat{\tau} - \tau) \right. \\ &\qquad \left. + 2(\widehat{\tau}_p - \tau_p)(\tau_p - \tau) - 2(\widehat{\tau} - \tau)(\tau_p - \tau)\right].\end{aligned}$$

Taking the expected value, and given that $\mathbb{E}[\widehat{\tau}] = \tau$ and $\mathbb{E}[\widehat{\tau}_p] = \tau_p$,

$$\begin{aligned}\mathbb{E}[\widehat{\mathbb{V}}_{pair}(\widehat{\tau})] &= \frac{1}{P^2} \sum_{p=1}^{P} \left[\mathbb{V}(\widehat{\tau}_p) + \mathbb{V}(\widehat{\tau}) + (\tau_p - \tau)^2 - 2\text{Cov}(\widehat{\tau}, \widehat{\tau}_p)\right] \\ &= \frac{1}{P^2} \sum_{p=1}^{P} \left[\left(1 - \frac{2}{P}\right) \mathbb{V}(\widehat{\tau}_p) + \mathbb{V}(\widehat{\tau}) + (\tau_p - \tau)^2\right] \\ &= \left(1 - \frac{2}{P}\right) \mathbb{V}(\widehat{\tau}) + \frac{1}{P^2} \sum_{p=1}^{P} \mathbb{V}(\widehat{\tau}) + \frac{1}{P^2} \sum_{p=1}^{P} (\tau_p - \tau)^2 \\ &= \left(1 - \frac{1}{P}\right) \mathbb{V}(\widehat{\tau}) + \frac{1}{P^2} \sum_{p=1}^{P} (\tau_p - \tau)^2.\end{aligned}$$

The second equality follows from the fact that by Point 3 of Assumption 1 and Assumption 2, $\text{Cov}(\widehat{\tau}_p, \widehat{\tau}) = \text{Cov}\left(\widehat{\tau}_p, \sum_{p'} \frac{1}{P} \widehat{\tau}_{p'}\right) = \frac{1}{P} \mathbb{V}(\widehat{\tau}_p)$. The third equality comes from Equation (3). This proves the result.
QED.

Point 2

The result directly follows from Points 3 and 4 of Lemma C.1 and the fact that $n_{1p} = n_{2p} = n_p/2$ under Assumption 2.
QED.

Point 3

Let $\overline{Y}_{gp} \equiv \sum_i Y_{igp}/n_{gp}$, $\widehat{Y}_p(1) \equiv \sum_g W_{gp}\overline{Y}_{gp}$, $\widehat{Y}_p(0) \equiv \sum_g (1-W_{gp})\overline{Y}_{gp}$, and $\widehat{Y}(d) \equiv \sum_p \widehat{Y}_p(d)/P$, for $d \in \{0,1\}$.

$$\mathbb{E}[\widehat{Y}_p(1)] = \mathbb{E}\left[\sum_g W_{gp}\overline{y}_{gp}(1)\right] = \frac{1}{2}\sum_g \overline{y}_{gp}(1) = \overline{y}_p(1). \tag{A5}$$

The second equality follows from Point 2 of Assumption 1. Similarly,

$$\mathbb{E}[\widehat{Y}_p(0)] = \mathbb{E}[\overline{y}_p(0)] \tag{A6}$$

$$\mathbb{E}[\widehat{Y}(d)] = \overline{y}(d), \quad \text{for } d \in \{0,1\}. \tag{A7}$$

Then, one has

$$\begin{aligned}
\widehat{\mathbb{V}}_{unit}(\widehat{\tau}) - \widehat{\mathbb{V}}_{pair}(\widehat{\tau}) &= \frac{8}{n^2}\sum_p SET_p SEU_p \\
&= \frac{8}{n^2}\sum_p \left(\sum_g W_{gp}\sum_i (y_{igp}(1) - \widehat{Y}(1))\right)\left(\sum_g (1-W_{gp})\sum_i (y_{igp}(0) - \widehat{Y}(0))\right) \\
&= \frac{8}{n^2}\sum_p \frac{n_p^2}{4}\left(\sum_g W_{gp}\sum_i \frac{y_{igp}(1)}{n_{gp}} - \widehat{Y}(1)\right)\left(\sum_g (1-W_{gp})\sum_i \frac{y_{igp}(0)}{n_{gp}} - \widehat{Y}(0)\right) \\
&= \frac{2}{P^2}\sum_p \widehat{Y}_p(1)\widehat{Y}_p(0) - \frac{2}{P}\widehat{Y}(1)\widehat{Y}(0)
\end{aligned} \tag{A8}$$

The first equality follows from Points 1 and 2 of Lemma C.1 and Assumption 2. The second equality follows from the definitions of SET_p, SEU_p, and ϵ_{igp}. The third equality follows from Point 1 of Assumption 1, and Assumption 2. The fourth equality follows from Assumption 2 and some algebra. Taking the expectation of (A8),

$$\begin{aligned}
&\mathbb{E}\left[\widehat{\mathbb{V}}_{unit}(\widehat{\tau}) - \widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right] \\
&= \frac{2}{P^2}\sum_p \left(\mathrm{Cov}(\widehat{Y}_p(1), \widehat{Y}_p(0))\right) + \frac{2}{P^2}\sum_p (\overline{y}_p(1) - \overline{y}(1))(\overline{y}_p(0) - \overline{y}(0)) - \frac{2}{P}\mathrm{Cov}(\widehat{Y}(1), \widehat{Y}(0)) \\
&= \frac{2}{P^2}\sum_p \left(\mathrm{Cov}(\widehat{Y}_p(1), \widehat{Y}_p(0))\right) + \frac{2}{P^2}\sum_p (\overline{y}_p(1) - \overline{y}(1))(\overline{y}_p(0) - \overline{y}(0)) - \frac{2}{P}\mathrm{Cov}\left(\frac{1}{P}\sum_p \widehat{Y}_p(1), \frac{1}{P}\sum_p \widehat{Y}_p(0)\right) \\
&= \frac{2(P-1)}{P^3}\sum_p \left(\mathrm{Cov}(\widehat{Y}_p(1), \widehat{Y}_p(0))\right) + \frac{2}{P^2}\sum_p (\overline{y}_p(1) - \overline{y}(1))(\overline{y}_p(0) - \overline{y}(0)).
\end{aligned}$$

The first equality follows from adding and subtracting $\frac{2}{P}\,\mathbb{E}[\widehat{Y}(1)]\,\mathbb{E}[\widehat{Y}(0)]$ and $\frac{2}{P^2}\sum_p \mathbb{E}[\widehat{Y}_p(1)]\,\mathbb{E}[\widehat{Y}_p(0)]$, and from Equations (A5), (A6) and (A7). The third equality follows from Point 3 of Assumption 1. Therefore,

$$\frac{P}{P-1}\,\mathbb{E}\left[\widehat{\mathbb{V}}_{unit}(\widehat{\tau}) - \widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right] = \frac{2}{P^2}\sum_p \left(\mathrm{Cov}(\widehat{Y}_p(1), \widehat{Y}_p(0))\right) + \frac{2}{P(P-1)}\sum_p (\overline{y}_p(0) - \overline{y}(0))(\overline{y}_p(1) - \overline{y}(1)). \tag{A9}$$

Finally,

$$\begin{aligned}
\mathrm{Cov}\left(\widehat{Y}_p(1), \widehat{Y}_p(0)\right) &= \mathbb{E}[\widehat{Y}_p(1)\widehat{Y}_p(0)] - \mathbb{E}[\widehat{Y}_p(1)]\,\mathbb{E}[\widehat{Y}_p(0)] \\
&= \left(\frac{1}{2}\overline{y}_{1p}(1)\overline{y}_{2p}(0) + \frac{1}{2}\overline{y}_{2p}(1)\overline{y}_{1p}(0)\right) - \left(\frac{1}{2}\sum_g \overline{y}_{gp}(1)\right)\left(\frac{1}{2}\sum_g \overline{y}_{gp}(0)\right) \\
&= \frac{1}{4}\overline{y}_{1p}(1)\overline{y}_{2p}(0) + \frac{1}{4}\overline{y}_{2p}(1)\overline{y}_{1p}(0) - \frac{1}{4}\overline{y}_{1p}(1)\overline{y}_{1p}(0) - \frac{1}{4}\overline{y}_{2p}(1)\overline{y}_{2p}(0) \\
&= \frac{1}{4}\left(\overline{y}_{1p}(1) - \overline{y}_{2p}(1)\right)\left(\overline{y}_{2p}(0) - \overline{y}_{1p}(0)\right) \\
&= -\frac{1}{2}\sum_g \left(\overline{y}_{gp}(0) - \overline{y}_p(0)\right)\left(\overline{y}_{gp}(1) - \overline{y}_p(1)\right) \qquad \text{(A10)}
\end{aligned}$$

The second equality follows from Points 1 and 2 of Assumption 1, and Equations (A5) and (A6). The third, fourth, and fifth equalities follow after some algebra. The result follows plugging Equation (A10) into (A9).
QED.

B. Large sample results for the pair- and unit-clustered variance estimators

In this section, we present the large sample distributions of the t-tests attached to the four variance estimators we considered in Section II. Let

$$\begin{aligned}
\sigma^2_{pair} &= \lim_{P\to+\infty} \frac{P\mathbb{V}(\widehat{\tau})}{P\mathbb{V}(\widehat{\tau}) + \frac{1}{P}\sum_p(\tau_p - \tau)^2}, \\
\Delta_{cov,P} &= \frac{1}{P}\sum_p (\overline{y}_p(0) - \overline{y}(0))(\overline{y}_p(1) - \overline{y}(1)) - \frac{1}{P}\sum_p \frac{1}{2}\sum_g \left(\overline{y}_{gp}(0) - \overline{y}_p(0)\right)\left(\overline{y}_{gp}(1) - \overline{y}_p(1)\right), \\
\text{and } \sigma^2_{unit} &= \lim_{P\to+\infty} \frac{P\mathbb{V}(\widehat{\tau})}{P\mathbb{V}(\widehat{\tau}) + \frac{1}{P}\sum_p(\tau_p - \tau)^2 + 2\Delta_{cov,P}},
\end{aligned}$$

where Assumption 3 below ensures the limits in the previous display exist.

ASSUMPTION 3:

1) *For every* d*,* g *and* p*, there is a constant* M *such that* $\left|\overline{y}_{gp}(d)\right| < M < +\infty$*.*

2) *When* $P \to +\infty$*,* $\frac{1}{P}\sum_p \tau_p$*,* $\frac{1}{P}\sum_p(\tau_p - \tau)^2$*, and* $\Delta_{cov,P}$ *converge towards finite limits, and* $P\mathbb{V}(\widehat{\tau})$ *and* $P\mathbb{V}(\widehat{\tau}) + \frac{1}{P}\sum_p(\tau_p - \tau)^2 + 2\Delta_{cov,P}$ *converge towards strictly positive finite limits.*

3) *As* $P \to +\infty$*,* $\sum_{p=1}^P \mathbb{E}[|\widehat{\tau}_p - \tau_p|^{2+\epsilon}]/S_P^{2+\epsilon} \to 0$ *for some* $\epsilon > 0$*, where* $S_P^2 \equiv P^2\mathbb{V}(\widehat{\tau})$*.*

Point 1 of Assumption 3 guarantees that we can apply the strong law of large numbers (SLLN) in Lemma 1 in Liu (1988) to the sequence $(\widehat{\tau}_p^2)_{p=1}^{+\infty}$. Point 2 ensures that $P\mathbb{V}\left(\widehat{\tau}\right)$ and $P\widehat{\mathbb{V}}_{unit}(\widehat{\tau})$ do not converge towards 0. Point 3 guarantees that we can apply the Lyapunov central limit theorem to $(\widehat{\tau}_p)_{p=1}^{+\infty}$.

THEOREM B.1: *(t-stats' asymptotic behavior) Under Assumptions 1, 2 and 3,*

1) $(\widehat{\tau} - \tau)/\sqrt{\widehat{\mathbb{V}}_{pair}(\widehat{\tau})} = (\widehat{\tau}_{fe} - \tau)/\sqrt{\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})} \xrightarrow{d} \mathcal{N}(0, \sigma^2_{pair})$*.* $\sigma^2_{pair} \leq 1$*, and if* $\tau_p = \tau$ *for every* p*,* $\sigma^2_{pair} = 1$*.*

2) $(\widehat{\tau}_{fe} - \tau)/\sqrt{\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})} \xrightarrow{d} \mathcal{N}(0, 2\sigma^2_{pair})$*.*

3) $(\widehat{\tau} - \tau)/\sqrt{\widehat{\mathbb{V}}_{unit}(\widehat{\tau})} \xrightarrow{d} \mathcal{N}(0, \sigma^2_{unit})$*.*

4) $\sigma^2_{unit} \leq \sigma^2_{pair}$ *if and only if* $\Delta_{cov,P}$ *converges towards a positive limit.*

PROOF:

See Online Appendix H.

Point 3 is related to Theorem 3.1 in Bai, Romano and Shaikh (2021), who show that when $n_{gp} = 1$, the t-test in Point 3 under-rejects. The asymptotic variance we obtain is different from theirs, because our results are derived under different assumptions. For instance, we assume a fixed population, while Bai, Romano and Shaikh (2021) assume that the experimental units are an i.i.d. sample drawn from an infinite superpopulation, and that asymptotically the expectation of the potential outcomes of two units in the same pair become equal.

C. Clustered variance estimators

LEMMA C.1 (Clustered variance estimators for $\widehat{\tau}$ and $\widehat{\tau}_{fe}$):

1) *The pair-clustered variance estimator (PCVE) of* $\widehat{\tau}$ *is* $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) = \sum_{p=1}^P \left(\frac{SET_p}{T} - \frac{SEU_p}{C}\right)^2$*.*

2) *The unit-clustered variance estimator (UCVE) of* $\widehat{\tau}$ *is* $\widehat{\mathbb{V}}_{unit}(\widehat{\tau}) = \sum_{p=1}^P \left(\frac{SET_p^2}{T^2} + \frac{SEU_p^2}{C^2}\right)$*.*

3) *The PCVE of* $\widehat{\tau}_{fe}$ *is* $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe}) = \sum_{p=1}^P \omega_p^2 \left(\widehat{\tau}_p - \widehat{\tau}_{fe}\right)^2$*.*

4) The UCVE of $\widehat{\tau}_{fe}$ is $\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe}) = \sum_{p=1}^{P} \omega_p^2 \left(\widehat{\tau}_p - \widehat{\tau}_{fe}\right)^2 \left(\left(\frac{n_{1p}}{n_p}\right)^2 + \left(\frac{n_{2p}}{n_p}\right)^2\right)$.

PROOF:

See Online Appendix H.

D. Variance estimators that rely on pairs of pairs

We also study two other estimators of $\mathbb{V}(\widehat{\tau})$. Those estimators have been proposed in the one-observation-per-unit special case, but it is straightforward to extend them to the case where all units have the same number of observations, as stated in Assumption 2.[13]

The first alternative estimator we consider is a slightly modified version of the pairs-of-pairs (POP) variance estimator (POPVE) proposed by Abadie and Imbens (2008). We only define it when the number of pairs P is even, but in our application in Subsection D.D4 below we propose a simple method to extend it to cases where the number of pairs is odd. Let $x_{g,p}$ denote the value of a predictor of the outcome in pair p's unit g. Pairs are ordered according to their value of $\frac{x_{1,p}+x_{2,p}}{2}$, the two pairs with the lowest value are matched together, the next two pairs are matched together, and so on and so forth. Let $R = \frac{P}{2}$. For any $r \in \{1, ..., R\}$ and for any $p \in \{1, 2\}$, let $\widehat{\tau}_{pr}$ denote the treatment effect estimator in pair p of POP r. Then, the POPVE is defined as

$$\widehat{\mathbb{V}}_{pop}(\widehat{\tau}) = \frac{1}{P^2} \sum_{r=1}^{R} (\widehat{\tau}_{1r} - \widehat{\tau}_{2r})^2.$$

$x_{g,p}$, the variable used to match pairs into POPs, could be the average value of the outcome at baseline in pair p's unit g. Or it could be the covariate used to form the pairs, when only one covariate is used. In our application in subsection D.D4, we use the baseline outcome to match pairs into POPs, because the covariates used to match units into pairs are unavailable in most of the data sets of the papers we revisit. Based on Lemma D.1, we will argue below that the baseline outcome should often be a good choice to match pairs into POPs. The variable one uses to form POPs should be pre-specified and not a function of the treatment assignment. Otherwise, researchers could try to find the variable minimizing the POPVE, which would lead to incorrect inference.

There are two differences between the POPVE and the variance estimator proposed in Equation (3) in Abadie and Imbens (2008). First, we match pairs with respect to a single covariate, while Abadie and Imbens (2008) consider matching with respect to a potentially multidimensional vector of covariates. This difference is not of essence: we could easily allow pairs to be matched on several covariates. We focus on the unidimensional case as that is the one we use in our

[13] Extending those variance estimators when Assumption 2 fails is left for future work.

application, where the matching is done based on the baseline outcome. Second, the estimator in Abadie and Imbens (2008) matches pairs with replacement, while $\widehat{\mathbb{V}}_{pop}(\widehat{\tau})$ matches pairs without replacement. If after ordering pairs according to their value of $\frac{x_{1,p}+x_{2,p}}{2}$, pair 2 is closer to pair 3 than pair 4, pair 2 is matched to pairs 1 and 3 in Abadie and Imbens (2008), while $\widehat{\mathbb{V}}_{pop}(\widehat{\tau})$ matches pair 1 to pair 2 and pair 3 to pair 4. Matching without replacement makes the properties of $\widehat{\mathbb{V}}_{pop}(\widehat{\tau})$ easier to analyze.

The second alternative variance estimator we consider is that proposed by Bai, Romano and Shaikh (2021) in their Equation (20) (BRSVE). Again, we define this estimator when the number of pairs P is even. With our notation, their estimator is

$$\widehat{V}_{brs}(\widehat{\tau}) = \frac{1}{P^2}\sum_{p=1}^{P}\widehat{\tau}_p^2 - \frac{1}{2}\left(\frac{2}{P^2}\sum_{r=1}^{R}\widehat{\tau}_{1r}\widehat{\tau}_{2r} + \frac{\widehat{\tau}^2}{P}\right).$$

Bai, Romano and Shaikh (2021) propose another variance estimator in their Equation (28). That estimator is less amenable to simple comparisons with the UCVE, PCVE, and POPVE, so we do not analyze its properties. However, we compute it in our applications, and find that it is typically similar to the POPVE and BRSVE.

D1. Finite-sample results

Let $\tau_{\cdot r} = \frac{1}{2}(\tau_{1r} + \tau_{2r})$ denote the average treatment effect in POP r.

LEMMA D.1: *If Assumptions 1 and 2 hold and P is even,*

1) $\mathbb{E}\left[\widehat{\mathbb{V}}_{pop}(\widehat{\tau})\right] = \mathbb{V}(\widehat{\tau}) + \frac{1}{P^2}\sum_{r=1}^{R}(\tau_{1r} - \tau_{2r})^2.$

2) $\widehat{\mathbb{V}}_{brs}(\widehat{\tau}) = \frac{1}{2}\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) + \frac{1}{2}\widehat{\mathbb{V}}_{pop}(\widehat{\tau}).$

3) If $\frac{1}{R}\sum_{r=1}^{R}\sum_{p=1,2}\frac{1}{2}(\tau_{pr} - \tau_{\cdot r})^2 \leq \frac{1}{R-1}\sum_{r=1}^{R}(\tau_{\cdot r} - \tau)^2$,

a) $\mathbb{E}\left[\widehat{\mathbb{V}}_{pop}(\widehat{\tau})\right] \leq \mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right],$

b) $\mathbb{E}\left[\widehat{\mathbb{V}}_{pop}(\widehat{\tau})\right] \leq \mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{brs}(\widehat{\tau})\right],$

c) $\mathbb{E}\left[\widehat{\mathbb{V}}_{brs}(\widehat{\tau})\right] \leq \mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right].$

PROOF:

See Online Appendix H.

Point 1 of Lemma D.1 shows that the POPVE is upward biased in general, and unbiased if the treatment effect is constant within POP. The less treatment effect heterogeneity within POP, the less upward biased the POPVE. An important practical consequence of Point 1 is that the variable used to form POPs should be a good predictor of pairs' treatment effect. The baseline value of the outcome may often be a good predictor of pairs' treatment effect. For instance, treatments sometimes produce a stronger effect on units with the lowest baseline outcome, thus leading to a catch-up mechanism (see for instance Glewwe, Park and Zhao, 2016a).

Point 1 of Lemma D.1 is related to Theorem 1 in Abadie and Imbens (2008), though there are a few differences. Abadie and Imbens (2008) assume that the experimental units are drawn from a super population, and show that once properly normalized, their estimator is consistent for the normalized conditional variance of $\widehat{\tau}$.[14] The fact that the POPVE is upward biased in Lemma D.1 and consistent in their Theorem 1 is because we do not assume that the experimental units are an i.i.d. sample from a super population. The intuition is the following. In Abadie and Imbens (2008), when the number of units grows, the covariates X_i on which pairing is based become equal to the same value x for units in the same POP: with an infinity of units, each unit can be matched to another unit with the same X_i, and each pair can be matched to another pair with the same X_i. Then, asymptotically those units are an i.i.d. sample drawn from the super-population conditional on $X_i = x$, and they all have the same expectation of their treatment effect. Treatment effect heterogeneity within POPs, the source of the POPVE's upward bias in Lemma D.1, vanishes asymptotically. On the other hand, with a convenience sample, units in the same POP may have asymptotically the same covariates, but they could still have different treatment effects, because they are not i.i.d. draws from a superpopulation.

Point 2 shows that the BRSVE is equal to the average of the PCVE and POPVE. Then, it follows from Point 1 of Lemma I and Point 1 of Lemma D.1 that $\frac{P}{P-1}\widehat{\mathbb{V}}_{brs}(\widehat{\tau})$ is upward biased. Point 2 is related to Lemma 6.4 and Theorem 3.3 in Bai, Romano and Shaikh (2021), where the authors show that $P\widehat{\mathbb{V}}_{brs}(\widehat{\tau})$ is consistent for the normalized variance of $\widehat{\tau}$. Here as well, the fact that $P\widehat{\mathbb{V}}_{brs}(\widehat{\tau})$ is upward biased in Lemma D.1 and consistent in Bai, Romano and Shaikh (2021) comes from the fact we do not assume that the experimental units are an i.i.d. sample drawn from a super population.

Finally, Point 3 shows that if the treatment effect varies less within than across POPs, the POPVE is less upward biased than the degrees-of-freedom-adjusted PCVE and BRSVE, and the BRSVE is less upward biased than the degrees-of-freedom-adjusted PCVE. A sufficient condition to have that the treatment effect varies less within than across POPs is $\frac{1}{R}\sum_{r=1}^{R}(\tau_{1r}-\tau)(\tau_{2r}-\tau) \geq 0$, meaning that the treatment effects of the two pairs in the same POP are positively correlated.

[14] In our setting, the covariates are assumed to be fixed, so the fact that we consider the unconditional variance of $\widehat{\tau}$ while they consider its conditional variance does not explain the difference between our results.

D2. Large-sample results

ASSUMPTION 4: *When $P \to +\infty$, $\frac{1}{P}\sum_r(\tau_{1r}-\tau_{2r})^2$ converges towards a finite limit.*

Let

$$\sigma^2_{pop} = \lim_{P\to+\infty} \frac{P\mathbb{V}(\widehat{\tau})}{P\mathbb{V}(\widehat{\tau}) + \frac{1}{P}\sum_r(\tau_{1r}-\tau_{2r})^2},$$
$$\sigma^2_{brs} = \lim_{P\to+\infty} \frac{P\mathbb{V}(\widehat{\tau})}{P\mathbb{V}(\widehat{\tau}) + \frac{1}{2P}\sum_r(\tau_{1r}-\tau_{2r})^2 + \frac{1}{2P}\sum_p(\tau_p-\tau)^2},$$

where Assumptions 3 and 4 ensure the limits in the previous display exist.

THEOREM D.2: *(t-stats' asymptotic behavior) Under Assumptions 1, 2, 3, and 4,*

1) $(\widehat{\tau}-\tau)/\sqrt{\widehat{\mathbb{V}}_{pop}(\widehat{\tau})} \xrightarrow{d} \mathcal{N}(0,\sigma^2_{pop})$. $\sigma^2_{pop} \leq 1$, *and if* $\tau_{1r}=\tau_{2r}$ *for every* r, $\sigma^2_{pop}=1$.

2) $(\widehat{\tau}-\tau)/\sqrt{\widehat{\mathbb{V}}_{brs}(\widehat{\tau})} \xrightarrow{d} \mathcal{N}(0,\sigma^2_{brs})$. $\sigma^2_{brs} \leq 1$, *and if* $\tau_p=\tau$ *for every* p, $\sigma^2_{brs}=1$.

3) $\sigma^2_{pair} \leq \sigma^2_{brs} \leq \sigma^2_{pop}$ *if and only if* $0 \leq \lim_{P\to+\infty}\frac{1}{R}\sum_{r=1}^{R}(\tau_{1r}-\tau)(\tau_{2r}-\tau)$.

PROOF:

See Online Appendix H.

Points 1 and 2 of Theorem D.2 show that when the number of pairs grows, the t-statistic using the POPVE and BRSVE, respectively, converges to a normal distribution with a mean equal to 0 and a variance lower than 1 in general, but equal to 1 when the treatment effect is homogenous across pairs. Therefore, those t-tests under-reject. Point 3 shows that whenever there is a positive correlation between the treatment effects of the two pairs in the same POP, the t-test using the POPVE under-rejects less than that using the BRSVE, which itself under-rejects less than that using the PCVE.

D3. Simulations

For 26 of the 82 regressions in Crépon et al. (2015a), the baseline outcome is available in the authors' data set, so for those outcomes we can simulate the POPVE and BRSVE as well. Those estimators are defined under Assumption 2, which does not hold. Therefore in those simulations, we aggregate the data at the village level. We use two samples of 80 and 20 randomly selected pairs out of the original 81 pairs, so as to have an even number of pairs. For each outcome, we simulate 3,000 vectors of treatment assignments, assigning one of the two villages to treatment in each pair. Then, we compute $\widehat{\tau}$, $\widehat{\mathbb{V}}_{pair}(\tau)$, $\widehat{\mathbb{V}}_{pop}(\tau)$, and $\widehat{\mathbb{V}}_{brs}(\tau)$, and the three corresponding 5% level t-tests.

The estimated error rate of each t-test is shown in Table D1 below. The error rate of the t-test using the PCVE is close to 5% with as few as 20 pairs. On the other hand, the error rates of the t-tests using the POPVE and BRSVE are larger than 5%, even with 80 pairs. Accordingly, we run simulations again, duplicating the random sample of 80 pairs twice to have 160 pairs. The error rate of the t-test using the BRSVE is now close to 5%, but the error rate of the t-test using the POPVE is still larger than 5%. With a sample of 320 pairs obtained by duplicating the random sample of 80 pairs four times, all tests have error rates close to 5%. With 20 and 80 pairs, we find in our simulations that the correlation between $\widehat{\mathbb{V}}_{pop}(\tau)$ and $|\widehat{\tau}|$ is much weaker than that between $\widehat{\mathbb{V}}_{pair}(\tau)$ and $|\widehat{\tau}|$. This explains why the t-test using $\widehat{\mathbb{V}}_{pop}(\tau)$ over-rejects, despite the fact $\widehat{\mathbb{V}}_{pop}(\tau)$ is unbiased: when $|\widehat{\tau}|$ is large, $\widehat{\mathbb{V}}_{pop}(\tau)$ is less likely to be large than $\widehat{\mathbb{V}}_{pair}(\tau)$, so the POPVE t-test rejects more often. With 160 and 320 pairs, this phenomenon becomes less pronounced. Overall, the asymptotic approximations in Points 1 and 2 of Theorem D.2 seem to hold only with a large number of pairs, contrary to that in Point 1 of Theorem B.1.

Table D1—: Simulations with data aggregated at village-level to compute $\widehat{\mathbb{V}}_{pop}$ and $\widehat{\mathbb{V}}_{brs}$

Variance estimator	5% level t-test error rate			
	With 20 pairs	With 80 pairs	With 160 pairs	With 320 pairs
PCVE	0.0504	0.0505	0.0506	0.0503
POPVE	0.1301	0.0818	0.0656	0.0565
BRSVE	0.0808	0.0619	0.0571	0.0530

Note: The table reports the error rates of three 5% level t-tests in Crépon et al. (2015a), aggregating data at the village level. For each of the 26 outcomes in the paper for which the baseline outcome is available, we randomly drew 3,000 simulated treatment assignments, following the paired assignment used by the authors, and computed the treatment effect estimator $\widehat{\tau}$, the pair-clustered variance estimator (PCVE), the pairs-of-pairs variance estimator (POPVE) in Abadie and Imbens (2008), the variance estimator in Bai, Romano and Shaikh (2021) (BRSVE), and the three corresponding t-tests. The error rate of each test is the percent of times, across the 78,000 regressions (26 outcomes $\times$ 3,000 replications), that the test leads the researcher to wrongly conclude that the treatment has an effect. Column 2 (resp. 3, 4, 5) shows the results using a random sample of 20 pairs (resp. a random sample of 80 pairs, the same random sample of 80 pairs duplicated twice, the same random sample of 80 pairs duplicated four times).

D4. Application

For 152 of the 294 regressions in Panel A of Table 2, the baseline outcome is available in the data set, so we can compute the POPVE and BRSVE. Those estimators are defined under Assumption 2, which does not hold in all those regressions. Therefore, we compute the POPVE and BRSVE after aggregating the data at the unit level. When the number of pairs is odd, we compute the POPVE twice, first excluding the pair with the lowest value of the baseline outcome, then excluding the pair with the highest value of the baseline outcome, and we finally take the average of the two estimators. We do the same for the BRSVE when the number of pairs is odd. We also recompute the PCVE without pair fixed effects with the aggregated data, using the exact same sample as that used to compute the POPVE and BRSVE. Across those 152 regressions, the POPVE divided by the PCVE is on average equal to 1.026. The BRSVE divided by the PCVE is on average equal to 1.014.[15] In those regressions, the POPVE and BRSVE do not lead to power gains.

E. Extension: stratified experiments with few units per strata

In this section, we perform Monte-Carlo simulations to assess how our results in Section II extend to stratified RCTs where the number of units per strata is larger than two, but still fairly small. Three main findings emerge. First, the error rates of t-tests using stratum-clustered standard errors are equal to 5%. Second, the error rates of t-tests using standard errors clustered at the unit level are larger than 5% in regressions with stratum fixed effects, but decrease as the number of units per strata increases. With 5 units per strata, and averaging across Panels A to D of Table E1 below, the error rate of a 5% level test with UCVE and stratum fixed effects is around 7.9%, while with 10 units per strata this error rate is around 6.2%. Finally, the error rates of t-tests using standard errors clustered at the unit level tend to be lower than 5% in regressions without stratum fixed effects.

We draw the potential and observed outcomes from the following data generating process (DGP),

(E1)

$$Y_{igp} = W_{gp}y_{igp}(1) + (1 - W_{gp})y_{igp}(0) + \gamma_p, \qquad i = 1, \dots, n_{gp};\ g = 1, \dots, G;\ p = 1, \dots, P,$$

where $y_{igp}(1)$ and $y_{igp}(0)$ are independent and both follow a $\mathcal{N}(0, 1)$ distribution, $\{\gamma_p\}_p \sim$ iid $\mathcal{N}(0, \sigma_\gamma^2)$, and $(y_{igp}(1), y_{igp}(0)) \perp\!\!\!\perp \gamma_p$. We either let $\sigma_\eta = 0$ or $\sigma_\eta = \sqrt{0.1}$. $\sigma_\eta = 0$ corresponds to a model with no stratum common shock, while $\sigma_\eta = \sqrt{0.1}$ corresponds to a model with a shock. We draw potential outcomes once and keep them fixed, so $y_{igp}(1)$, $y_{igp}(0)$ and γ_p do not vary across simulations.

Each stratum has G units. We vary G from two to ten. If G is even, then half of the units are randomly assigned to the control and the remaining to the treatment. If G is odd, then $(G+1)/2$ units are randomly assigned to the control.

[15] The variance estimator in Equation (28) of Bai, Romano and Shaikh (2021) is also on average higher than the PCVE.

We also set $n_{gp} = 5$ or $n_{gp} = 100$, and we let the number of strata P be equal to 100.

We compute t-tests based on unit- and stratum-clustered standard errors in regressions of the outcome on the treatment with and without stratum fixed effects. We perform 10,000 simulations for each DGP. Table E1 presents the error rates of the t-tests in each DGP.

t-tests using stratum-clustered standard errors achieve error rates close to 5% for all data configurations (as in Table 1, with $n_{gp} = 5$, the $t-$test using the PCVE with stratum fixed effects under-rejects slightly, due to the DOF-adjustment). In contrast, t-tests based on unit-clustered standard errors in regressions with stratum fixed effects overreject the true null of no treatment effect. These results are in line with Points 1 and 2 of Theorem B.1, which covered the special case where $G = 2$. t-tests based on unit-clustered standard errors in regressions with stratum fixed effects over-reject less as the number of units per strata increases from two (column 2) to ten (column 10). Interestingly, it seems that unit-clustered standard errors are approximately equal to $\sqrt{\frac{G-1}{G}}$ times the stratum-clustered standard errors. If $G = 2$, the ratio of those two standard errors is exactly equal to $\sqrt{(2-1)/2} = \sqrt{1/2}$ as shown in Lemma I, but this relationship seems to still hold in expectation for larger values of G.

In Panel A, t-tests based on unit-clustered standard errors in regressions without stratum fixed effects have error rates close to 5%. When $\sigma_\eta = 0$, there is no between and within strata heterogeneity in $\overline{y}_{gp}(0)$, so it follows from Point 3 of Theorem B.1 that in the special case where $G = 2$, t-tests based on unit-clustered standard errors in regressions without stratum fixed effects have error rates close to 5%. Our simulations suggest that this result still holds when $G > 2$. However, in Panel B, t-tests using unit-clustered standard errors in regressions without stratum fixed effects have error rates lower than 5%, because there is now between-strata heterogeneity in $\overline{y}_{gp}(0)$. We obtain similar results with five observations per unit (Panels C and D).

Table E1—: Error rates of t-test in simulated stratified RCTs with small strata

	Number of units per strata								
	2	3	4	5	6	7	8	9	10
Panel A. iid standard normal potential outcomes and $n_{gp} = 100$									
UCVE without FE	0.0415	0.0734	0.0501	0.0480	0.0501	0.0478	0.0468	0.0538	0.0479
UCVE with FE	0.1685	0.1160	0.0928	0.0806	0.0705	0.0696	0.0631	0.0662	0.0614
SCVE without FE	0.0557	0.0587	0.0529	0.0545	0.0514	0.0516	0.0495	0.0524	0.0518
SCVE with FE	0.0554	0.0582	0.0527	0.0544	0.0511	0.0515	0.0493	0.0524	0.0516
$\widehat{s.e.}_{unit}(\widehat{\tau}_{fe})/\widehat{s.e.}_{strat}(\widehat{\tau}_{fe})$	0.7053	0.8169	0.8720	0.8986	0.9188	0.9308	0.9408	0.9492	0.9551
Panel B. Stratum-level shock affecting potential outcomes and $n_{gp} = 100$									
UCVE without FE	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
UCVE with FE	0.1682	0.1105	0.0909	0.0812	0.0775	0.0671	0.0641	0.0650	0.0637
SCVE without FE	0.0518	0.0507	0.0532	0.0524	0.0566	0.0504	0.0511	0.0546	0.0528
SCVE with FE	0.0510	0.0506	0.0531	0.0520	0.0565	0.0503	0.0508	0.0545	0.0526
$\widehat{s.e.}_{unit}(\widehat{\tau}_{fe})/\widehat{s.e.}_{strat}(\widehat{\tau}_{fe})$	0.7053	0.8163	0.8700	0.8995	0.9187	0.9313	0.9420	0.9494	0.9548
Panel C. iid standard normal potential outcomes and $n_{gp} = 5$									
UCVE without FE	0.0547	0.0456	0.0474	0.0575	0.0492	0.0527	0.0512	0.0527	0.0544
UCVE with FE	0.1478	0.0981	0.0872	0.0745	0.0719	0.0687	0.0629	0.0646	0.0627
SCVE without FE	0.0515	0.0542	0.0542	0.0552	0.0557	0.0548	0.0514	0.0543	0.0552
SCVE with FE	0.0397	0.0471	0.0481	0.0502	0.0522	0.0519	0.0486	0.0515	0.0529
$\widehat{s.e.}_{unit}(\widehat{\tau}_{fe})/\widehat{s.e.}_{strat}(\widehat{\tau}_{fe})$	0.7053	0.8163	0.8695	0.8979	0.9175	0.9327	0.9404	0.9485	0.9551
Panel D. Stratum-level shock affecting potential outcomes and $n_{gp} = 5$									
UCVE without FE	0.0128	0.0152	0.0207	0.0122	0.0146	0.0173	0.0130	0.0160	0.0158
UCVE with FE	0.1539	0.1033	0.0793	0.0730	0.0687	0.0682	0.0625	0.0640	0.0671
SCVE without FE	0.0533	0.0546	0.0529	0.0520	0.0513	0.0564	0.0507	0.0536	0.0574
SCVE with FE	0.0430	0.0469	0.0476	0.0470	0.0475	0.0526	0.0482	0.0510	0.0545
$\widehat{s.e.}_{unit}(\widehat{\tau}_{fe})/\widehat{s.e.}_{strat}(\widehat{\tau}_{fe})$	0.7053	0.8171	0.8703	0.8991	0.9179	0.9315	0.9419	0.9484	0.9552

Note: The table shows the error rates of t-tests based on unit- and stratum-clustered standard errors in regressions with and without stratum fixed effects. Across simulations, we vary the number of units per strata from two to ten ($G = 2, \ldots, 10$); we vary the number of observations per unit to either $n_{gp} = 5$ or $n_{gp} = 100$; and we set the number of strata to $P = 100$. For each value of G, we simulated 10,000 samples from the following data generating processes: independent and identically distributed (iid) standard normal potential outcomes in Panels A and C, and a model with an additive stratum-level shock affecting both potential outcomes in Panel B and D. UCVE and SCVE stand for unit- and stratum-clustered variance estimators, respectively. FE stands for stratum fixed effects. $\frac{\widehat{s.e.}_{unit}(\widehat{\tau}_{fe})}{\widehat{s.e.}_{strat}(\widehat{\tau}_{fe})}$ is the average across simulations of the ratio of standard errors clustering at the unit and stratum levels in regressions with stratum fixed effects.

F. Articles in our survey of paired or small strata experiments

Table F1—: Paired RCTs and stratified RCTs with small strata

Reference	Search source
Paired RCTs	
Ashraf, Karlan and Yin (2006)	AEA registry
Banerjee et al. (2015)	*AEJ: Applied*
Crépon et al. (2015a)	*AEJ: Applied*
Beuermann et al. (2015a)	*AEJ: Applied*
Fryer Jr, Devi and Holden (2016)	AEA registry
Glewwe, Park and Zhao (2016a)	AEA registry
Bruhn et al. (2016a)	*AEJ: Applied*
Fryer Jr (2017)	AEA registry
Small-strata RCTs	
Attanasio et al. (2015)	*AEJ: Applied*
Angelucci, Karlan and Zinman (2015)	*AEJ: Applied*
Ambler, Aycinena and Yang (2015)	*AEJ: Applied*
Björkman Nyqvist, de Walque and Svensson (2017)	*AEJ: Applied*
Banerji, Berry and Shotland (2017)	*AEJ: Applied*
Lafortune, Riutort and Tessada (2018)	*AEJ: Applied*
Somville and Vandewalle (2018)	*AEJ: Applied*

Note: The table presents economics papers that have conducted clustered and paired RCTs, or clustered and stratified RCTs with ten or less units per strata. We searched the *AEJ: Applied Economics* for papers published in 2014-2018 and using the words "random" and "experiment" in the abstract, title, keywords, or main text. Four of those papers had conducted a clustered and paired RCT and seven had conducted a clustered and stratified RCT with ten units or less per strata. We also searched the AEA's registry website for RCTs (`https://www.socialscienceregistry.org`). We looked at all completed projects, whose randomization method included the word "pair" and that had either a working or a published paper. Thus, we found four more papers that had conducted a clustered and paired RCT. Beuermann et al. (2015a) use a paired design to estimate the spillover effects of the intervention they consider. Their estimation of the direct effects of that intervention relies on another type of randomization. We only include their spillover analysis in our survey and in our replication.

G. Results when the number of observations varies across units

In this section, we extend some of the results in Section II to instances where units may have different numbers of observations, as is often the case in practice.

G1. Upward bias of the pair-clustered variance estimator (PCVE)

In this subsection, we show that when units have different numbers of observations, our recommendation of using the PCVE still applies.

When units have different numbers of observations, there are several estimators of the treatment effect one may consider. $\widehat{\tau}$, the standard difference in means estimator, is such that

$$\widehat{\tau} = \frac{1}{T}\sum_{p=1}^{P}\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}} Y_{igp}W_{gp} - \frac{1}{C}\sum_{p=1}^{P}\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}} Y_{igp}(1-W_{gp}),$$

where T and C respectively denote the total number of treated and control observations. When the number of observations varies across units, T and C are stochastic. For instance, assume one has two pairs. In pair 1, units 1 and 2 both have 1 observation, but in pair 2 unit 1 has 1 observations while unit 2 has 2 observations. Then, T is equal to 2 with probability 1/2, and to 3 with probability 1/2. These stochastic denominators in $\widehat{\tau}$ make it impossible to derive a closed-form expression of its expectation and variance. One can still show that when the number of pairs goes to infinity, $\widehat{\tau}$ converges toward τ, the average treatment effect, and one could also use the delta method to show that $\widehat{\tau}$ is asymptotically normal and derive its asymptotic variance. However, throughout the paper we have focused on estimators' finite sample variances.

Therefore, instead of $\widehat{\tau}$ we consider another, closely related estimator, whose expectation and variance are straightforward to derive even when the number of observations varies across units, and which is unbiased for a causal effect that differs from τ but that is still relatively natural (see Imai, King and Nall (2009) for closely related discussions). Let $\widetilde{\tau}$ denote the coefficient of W_{gp} in the weighted OLS regression of Y_{igp} on a constant and W_{gp}, with weights $V_{gp} = n_p/n_{gp}$.[16] Let $\widetilde{\alpha}$ be the intercept in that regression. One can show that

$$\widetilde{\tau} = \frac{1}{P}\sum_{p}\frac{n_p}{\overline{n}}\sum_{g}(W_{gp}\overline{Y}_{gp} - (1-W_{gp})\overline{Y}_{gp}) = \frac{1}{P}\sum_{p}\frac{n_p}{\overline{n}}\widehat{\tau}_p, \tag{G1}$$

where $\bar{n} = n/P$. Under Assumption 2, $\widetilde{\tau} = \widehat{\tau}$. Hence, $\widetilde{\tau}$ generalizes $\widehat{\tau}$ to the case where the number of observations varies across units. $\widetilde{\tau}$ is also one of the

[16] Specifically, the intercept $\widetilde{\alpha}$ and $\widetilde{\tau}$ are such that $(\widetilde{\alpha},\widetilde{\tau}) = \text{argmin}_{\alpha,\tau}\sum_p\sum_g\sum_i V_{gp}(Y_{igp}-\alpha-\tau W_{gp})^2$.

estimators considered by Imai, King and Nall (2009), though the fact $\widetilde{\tau}$ can be obtained by weighted least squares is not noted therein.

$\widetilde{\tau}$ is generally not unbiased for τ, unless in every pair, the two units have the same number of observations, i.e., $n_{1p} = n_{2p}$ for all p (Imai, King and Nall, 2009). On the other hand, $\widetilde{\tau}$ is unbiased for

$$\tau^* = \frac{1}{P}\sum_p \frac{n_p}{\overline{n}}\left(\frac{\tau_{1p}}{2} + \frac{\tau_{2p}}{2}\right)$$

where $\tau_{gp} = \frac{1}{n_{gp}}\sum_{i=1}^{n_{gp}}[y_{igp}(1) - y_{igp}(0)]$ denotes the average treatment effect in unit g of pair p.[17] τ^* is a weighted average of the pair-specific average treatment effects $(\tau_{1p} + \tau_{2p})/2$. Those pair-specific average treatment effects give equal weight to the average treatment effect in each unit, rather than weighting them according to their number of observations like τ_p does.Imai, King and Nall (2009) show that

$$\mathbb{V}(\widetilde{\tau}) = \frac{1}{4P^2}\sum_p \frac{n_p^2}{\overline{n}^2}\left(\Delta_p(1) + \Delta_p(0)\right)^2,$$

where $\Delta_p(1) \equiv \overline{y}_{1p}(1) - \overline{y}_{2p}(1)$ and $\Delta_p(0) \equiv \overline{y}_{1p}(0) - \overline{y}_{2p}(0)$. They propose various estimators of that variance, and show that they are upward biased. Instead, we rely on the fact $\widetilde{\tau}$ can be obtained by weighted least squares to propose an estimator whose properties have not been studied in the randomization-inference framework we consider: the PCVE attached to $\widetilde{\tau}$.

First, the following lemma extends Lemma C.1 to the PCVE in a weighted OLS regression.[18]

LEMMA G.1 (Pair-clustered variance estimator for $\widetilde{\tau}$): $\widehat{\mathbb{V}}_{pair}(\widetilde{\tau}) = \frac{1}{P^2}\sum_p \frac{n_p^2}{\overline{n}^2}\left[\widehat{\tau}_p - \widetilde{\tau}\right]^2.$

PROOF:

See Online Appendix H.

Then, we study the asymptotic distribution of the t-statistic attached to $\widetilde{\tau}$ and $\widehat{\mathbb{V}}_{pair}(\widetilde{\tau})$. To do so, we make the following assumption.

ASSUMPTION 5:

1) *For all g and p, $1 \leq n_{gp} \leq N$ for some fixed $N < +\infty$.*
2) *As $P \to +\infty$, $\frac{1}{P}\sum_p\left(\frac{n_p}{\overline{n}}\right)^2$, $\frac{1}{P}\sum_p\left(\frac{n_p}{\overline{n}}\right)^2\mathbb{E}[\widehat{\tau}_p]$, and $\frac{1}{P}\sum_p\left(\frac{n_p}{\overline{n}}\right)^2(\mathbb{E}[\widehat{\tau}_p])^2$ converge to strictly positive constants, and $\tau^* = \frac{1}{P}\sum_p \frac{n_p}{\overline{n}}\left(\frac{\tau_{1p}}{2} + \frac{\tau_{2p}}{2}\right)$ converges to a constant τ^∞.*

[17] With a slight abuse of notation, τ_{1r} and τ_{2r} refer to the ATE in pairs 1 and 2 of POP r, while τ_{1p} and τ_{2p} refer to the ATE in units 1 and 2 of pair p.

[18] We follow the definition of clustered variance estimators for weighted least squares in Equation (15) of Cameron and Miller (2015).

3) As $P \to +\infty$, $\sum_{p=1}^{P} \mathbb{E}\left[|\frac{n_p}{\overline{n}}|^{2+\epsilon}|\widehat{\tau}_p - \mathbb{E}[\widehat{\tau}_p]|^{2+\epsilon}\right]/\widetilde{S}_P^{2+\epsilon} \to 0$ for some $\epsilon > 0$, where $\widetilde{S}_P^2 \equiv P^2\mathbb{V}(\widetilde{\tau})$.

Point 1 of Assumption 5 requires that the number of observations in every unit is greater than 1 and lower than some fixed N. Combined with Point 2 of Assumption 3, Point 2 of Assumption 5 ensures that $P\widehat{\mathbb{V}}_{pair}(\widetilde{\tau})$ converges towards a strictly positive limit. Point 3 guarantees that we can apply the Lyapunov central limit theorem to $\left(\frac{n_p}{\overline{n}}\widehat{\tau}_p\right)_{p=1}^{+\infty}$. Let $\sigma^2_{wls} = \lim_{P\to+\infty} \frac{P\mathbb{V}(\widetilde{\tau})}{P\mathbb{V}(\widetilde{\tau})+\frac{1}{P}\sum_p\left(\frac{n_p}{\overline{n}}\right)^2(\mathbb{E}(\widehat{\tau}_p)-\tau^\infty)^2}$.

THEOREM G.2: *If Assumptions 1 and 5, and Points 1 and 2 of Assumption 3 hold,*
$(\widetilde{\tau} - \tau^*)/\sqrt{\widehat{\mathbb{V}}_{pair}(\widetilde{\tau})} \xrightarrow{d} \mathcal{N}(0, \sigma^2_{wls})$. $\sigma^2_{wls} \leq 1$, and if $\tau_{gp} = \tau$ for every g and p, or if $n_{1p} = n_{2p}$ and $\tau_p = \tau$ for every p, then $\sigma^2_{wls} = 1$.

PROOF:
See Online Appendix H.

This theorem shows that when the number of pairs grows, the t-statistic of the weighted least squares estimator using the PCVE converges to a normal distribution with a mean equal to 0 and a variance lower than 1 in general, but equal to 1 when the treatment effect is homogenous across units, or when the treatment effect is homogenous across pairs and in every pair the two units have the same number of observations.

Theorem G.2 shows that when units have different numbers of observations, the PCVE attached to $\widetilde{\tau}$ is upward biased asymptotically. We now show that the same holds for $\widehat{\tau}_{fe}$, the pair fixed effects estimator, provided one applies some kind of degrees-of-freedom correction to its PCVE. As shown in Point 3 of Lemma C.1, the PCVE of $\widehat{\tau}_{fe}$ is $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe}) = \sum_{p=1}^{P} \omega_p^2 \left(\widehat{\tau}_p - \widehat{\tau}_{fe}\right)^2$. Let $\widetilde{\omega}_p = \omega_p \left(1 - 2\omega_p\right)^{-1/2}$.

LEMMA G.3 (The adjusted PCVE for $\widehat{\tau}_{fe}$ is upward biased): *Under Assumption 1, and if $\omega_p \leq 1/2$ for all p, $\mathbb{E}\left[\sum_p \widetilde{\omega}_p^2(\widehat{\tau}_p - \widehat{\tau}_{fe})^2\right] = \mathbb{V}(\widehat{\tau}_{fe})\left(1 + \sum_p \widetilde{\omega}_p^2\right) + \sum_p \widetilde{\omega}_p^2[\mathbb{E}(\widehat{\tau}_p - \widehat{\tau}_{fe})]^2$. PROOF:*
See Online Appendix H.

Lemma G.3 shows that the adjusted PCVE, where the ω_p are replaced by $\widetilde{\omega}_p$, is upward biased for the variance of $\widehat{\tau}_{fe}$. The adjustment in $\widetilde{\omega}_p$ is similar to a degrees-of-freedom adjustment. In fact, under Assumption 2, the adjusted PCVE is equal to $\frac{P}{P-2}\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$. The requirement that $\omega_p \leq 1/2$ for all p is mild. For instance, if $n_{1p} = n_{2p}$ for all p, this only requires that every pair has fewer observations than all other pairs combined. If there is an integer L such that $n_p \leq L$ for every p, one can show that $\liminf_{P\to+\infty} \mathbb{E}\left[P\left(\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe}) - \mathbb{V}(\widehat{\tau}_{fe})\right)\right] \geq 0$: the unadjusted PCVE is also upward biased asymptotically. When the number of observations varies across units, $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$ does not coincide with the estimator of the variance of $\widehat{\tau}_{fe}$ considered in Imai, King and Nall (2009). It seems that

Lemma G.3 above is the first result to justify the use of the PCVE attached to $\widehat{\tau}_{fe}$, in paired RCTs where the number of observations varies across units.

G2. Ratio of the UCVE and PCVE with pair fixed effects

In this subsection, we derive the ratio of the UCVE and PCVE with pair fixed effects when units have different numbers of observations.

LEMMA G.4 (Ratio of the UCVE and PCVE with pair fixed effects when units have different numbers o
$\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})/\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe}) = \sum_p \zeta_p \left(\left(\frac{n_{1p}}{n_p}\right)^2 + \left(\frac{n_{2p}}{n_p}\right)^2\right)$, *where, for all* p $\zeta_p \geq 0$ *and* $\sum_p \zeta_p = 1$. *Therefore,* $\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})/\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe}) \in \left[\frac{1}{2}, 1\right]$.

PROOF:
The formula for $\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})/\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$ follows from Points 3 and 4 of Lemma C.1, with

$$\zeta_p = \frac{\omega_p^2 \left(\widehat{\tau}_p - \widehat{\tau}_{fe}\right)^2}{\sum_{p=1}^{P} \omega_p^2 \left(\widehat{\tau}_p - \widehat{\tau}_{fe}\right)^2}.$$

$n_{1p}^2 + n_{2p}^2 \leq (n_{1p} + n_{2p})^2$, so $\left(\frac{n_{1p}}{n_p}\right)^2 + \left(\frac{n_{2p}}{n_p}\right)^2 \leq 1$. $(n_{1p} - n_{2p})^2 = n_{1p}^2 - 2n_{1p}n_{2p} + n_{2p}^2 \geq 0$, so $2n_{1p}^2 + 2n_{2p}^2 \geq (n_{1p} + n_{2p})^2$, and $\left(\frac{n_{1p}}{n_p}\right)^2 + \left(\frac{n_{2p}}{n_p}\right)^2 \geq \frac{1}{2}$. Therefore, $\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})/\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe}) \in \left[\frac{1}{2}, 1\right]$.

Lemma G.4 shows that $\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})/\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$ is a weighted average across pairs of the sum of the squared shares that each unit accounts for in the pair. The sum of these squared shares is included between a half and one, so this ratio is included between a half and one. Figure G1 plots this ratio when n_{1p}/n_{2p} is constant across pairs. $\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})/\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$ is close to 1/2 when n_{1p}/n_{2p} is included between 0.5 and 2, meaning that the first unit has between half and twice as many observations as the second one. For instance, if in every pair, one unit has twice as many observations as the other, then the ratio of the two variances is equal to 5/9. Based on Figure G1, one can also derive an upper bound for $\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})/\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$, when n_{1p}/n_{2p} varies across pairs. For instance, if in every pair, one unit has at most twice as many observations as the other, as should often be the case in practice, then the ratio of the two variances is at most equal to 5/9. Overall, Lemma G.4 shows that Point 2 of Lemma I still approximately holds when units in each pair have different numbers of observations, unless they have an extremely unbalanced number of observations.

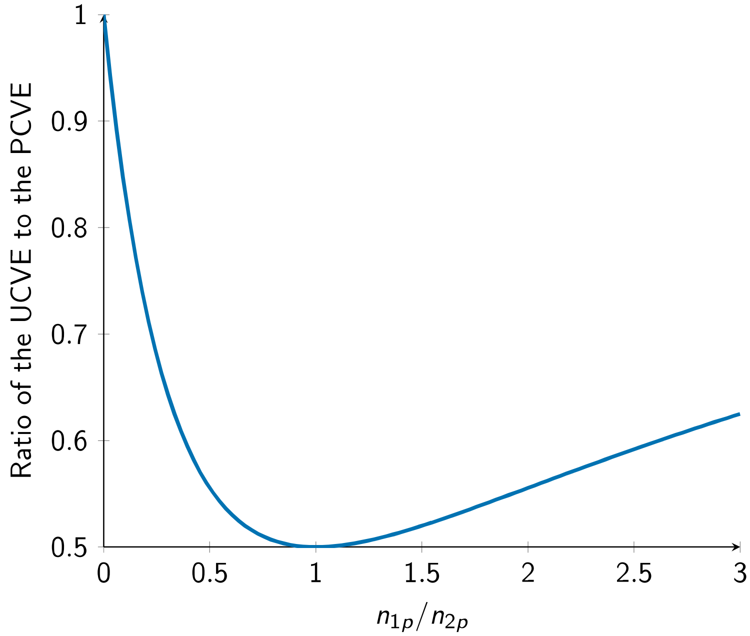

Figure G1. : Ratio of Unit-Clustered and Pair-Clustered Variance Estimators with Pair Fixed Effects

Note: UCVE and PCVE stand for unit- and pair clustered variance estimators, respectively. n_{1p} and n_{2p} are the number of observations in units 1 and 2 of pair p, respectively.

H. Proofs of the results in the Online Appendix

H1. Proof of Theorem B.1

The proof relies on Lemma H.1 and on the two equations below.

Using a similar reasoning as that used to show Equation (H40) in the proof of Lemma H.1, one can show that

$$\mathbb{E}\left[\left|\widehat{Y}_p(d)\right|^{2+\epsilon}\right] \leq M_1 < +\infty. \tag{H1}$$

for all d and p and for some $M_1 > 0$.

By Lemma H.1, Assumption 2, and Point 2 of Assumption 3,

$$\widehat{\tau} = \frac{1}{P}\sum_p \widehat{\tau}_p \xrightarrow{\mathbb{P}} \lim_{P\to+\infty} \frac{1}{P}\sum_p \mathbb{E}[\widehat{\tau}_p] = \lim_{P\to+\infty} \frac{1}{P}\sum_p \tau_p = \lim_{P\to+\infty} \tau. \tag{H2}$$

Point 1

Note that by Point 3 of Assumption 1, $\widehat{\tau} - \tau = \widehat{\tau} - \mathbb{E}[\widehat{\tau}] = \sum_p (\widehat{\tau}_p - \mathbb{E}[\widehat{\tau}_p])/P$ is a sum of independent random variables $(\widehat{\tau}_p - \mathbb{E}[\widehat{\tau}_p])_{p=1}^P$ with mean zero and with a finite variance by Equation (H40). As $\sum_{p=1}^P \mathbb{E}[|\widehat{\tau}_p - \tau_p|^{2+\epsilon}/S_P^{2+\epsilon}] \to 0$ for some $\epsilon > 0$ (by Point 3 of Assumption 3), then, by the Lyapunov central limit theorem, $(\widehat{\tau} - \tau)/(S_P/P) = \sum_p (\widehat{\tau}_p - \tau_p)/S_P \xrightarrow{d} \mathcal{N}(0,1)$ as $P \to +\infty$, where $S_P^2 = \sum_{p=1}^P \mathbb{V}(\widehat{\tau}_p) = P^2\mathbb{V}(\widehat{\tau})$. Therefore,

$$(\widehat{\tau} - \tau)/\sqrt{\mathbb{V}(\widehat{\tau})} \xrightarrow{d} \mathcal{N}(0,1). \tag{H3}$$

Then,

$$\begin{aligned}
P\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) - P\mathbb{V}(\widehat{\tau}) &= \sum_{p=1}^P \frac{\widehat{\tau}_p^2}{P} - \widehat{\tau}^2 - \sum_{p=1}^P \frac{\mathbb{V}(\widehat{\tau}_p)}{P} \\
&= \sum_{p=1}^P \frac{\widehat{\tau}_p^2}{P} - \widehat{\tau}^2 - \sum_{p=1}^P \frac{\mathbb{E}[\widehat{\tau}_p^2] - \mathbb{E}[\widehat{\tau}_p]^2}{P} \\
&= \sum_{p=1}^P \frac{\widehat{\tau}_p^2 - \mathbb{E}[\widehat{\tau}_p^2]}{P} - \widehat{\tau}^2 + \sum_{p=1}^P \frac{\tau_p^2}{P} && \text{(H4)} \\
&\xrightarrow{\mathbb{P}} \lim_{P\to+\infty} \frac{1}{P}\sum_{p=1}^P (\tau_p - \tau)^2. && \text{(H5)}
\end{aligned}$$

The first equality follows from Equations (3) and (A4). The third equality follows from $\mathbb{E}[\widehat{\tau}_p] = \tau_p$. Let's consider each of the terms in Equation (H4). $\sum_{p=1}^P \frac{\widehat{\tau}_p^2 - \mathbb{E}[\widehat{\tau}_p^2]}{P} \xrightarrow{\mathbb{P}} 0$ by Lemma H.1. Then, $\widehat{\tau}^2 \xrightarrow{\mathbb{P}} \lim_{P\to+\infty} \tau^2$ by Equation (H2) and the continuous mapping theorem (CMT). Equation (H5) follows from these facts, and from Point 2 of Assumption 3.

Given Equation (H5), Point 2 of Assumption 3, the Slutsky Lemma and the CMT, as $P \to +\infty$,

$$\frac{\widehat{\tau} - \tau}{\sqrt{\widehat{\mathbb{V}}_{pair}(\widehat{\tau})}} = \frac{\widehat{\tau} - \tau}{\sqrt{\mathbb{V}(\widehat{\tau})}}\sqrt{\frac{P\mathbb{V}(\widehat{\tau})}{P\widehat{\mathbb{V}}_{pair}(\widehat{\tau})}} \xrightarrow{d} \mathcal{N}(0, \sigma_{pair}^2). \tag{H6}$$

Finally, by Lemma I, $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) = \widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$, and by Assumption 2, $\widehat{\tau} = \widehat{\tau}_{fe}$.
QED.

Point 2

By Lemma 3, $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) = 2\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})$, so given Point 1 of this theorem, the result follows.
QED.

Point 3

$$
\begin{aligned}
& P\widehat{\mathbb{V}}_{unit}(\widehat{\tau}) - P\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) \\
& = \frac{2}{P}\sum_p \widehat{Y}_p(1)\widehat{Y}_p(0) - 2\frac{1}{P}\sum_p \widehat{Y}_p(1)\frac{1}{P}\sum_p \widehat{Y}_p(0) \\
& \xrightarrow{\mathbb{P}} 2\lim_{P\to+\infty}\left\{\frac{1}{P}\sum_p \mathbb{E}[\widehat{Y}_p(1)\widehat{Y}_p(0)] - \mathbb{E}[\widehat{Y}(1)]\,\mathbb{E}[\widehat{Y}(0)]\right\} \\
& \text{(H7)} \\
& = 2\lim_{P\to+\infty}\frac{1}{P}\sum_p\left\{\left(\overline{y}_p(0)-\overline{y}(0)\right)\left(\overline{y}_p(1)-\overline{y}(1)\right) - \frac{1}{2}\sum_g(\overline{y}_{gp}(0)-\overline{y}_p(0))(\overline{y}_{gp}(1)-\overline{y}_p(1))\right\}.
\end{aligned}
$$

The first equality follows from Equation (A8). The convergence arrow follows from the fact $\mathbb{E}\left[\left|\widehat{Y}_p(1)\widehat{Y}_p(0)\right|^{1+\epsilon/2}\right]$ is bounded uniformly in p by Equation (H1) and the Cauchy-Schwarz inequality, from the fact that $\mathbb{E}\left[\left|\widehat{Y}_p(d)\right|^{1+\epsilon/2}\right]$ is also bounded uniformly in p, from Point 3 of Assumption 1, from the SLLN in Lemma 1 in Liu (1988), from the CMT, and from Point 2 of Assumption 3. The last equality follows from the same steps as those used to prove Lemma 3. The result follows from Equations (H7), (H5), and (H3), and a reasoning similar to that used to prove Equation (H6).
QED.

H2. Proof of Lemma C.1

Point 1
First, we introduce the formulas for the PCVE and UCVE in a general linear regression. Let ϵ_{igp} be the residual from the regression of Y_{igp} on a K-vector of covariates $\boldsymbol{X}_{igp}$, and $\boldsymbol{X}$ the $(n \times K)$ matrix whose rows are $\boldsymbol{X}'_{igp}$. The PCVE of the OLS estimator, $\widehat{\boldsymbol{\beta}}$, is defined as follows (Liang and Zeger (1986), Abadie

et al. (2017))

(H8)

$$\widehat{\mathbb{V}}_{pair}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{p=1}^{P}\left(\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}}\epsilon_{igp}\boldsymbol{X}_{igp}\right)\left(\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}}\epsilon_{igp}\boldsymbol{X}_{igp}\right)'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

The UCVE of the OLS estimator, $\widehat{\boldsymbol{\beta}}$, is defined as follows

(H9)

$$\widehat{\mathbb{V}}_{unit}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{p=1}^{P}\sum_{g=1}^{2}\left(\sum_{i=1}^{n_{gp}}\epsilon_{igp}\boldsymbol{X}_{igp}\right)\left(\sum_{i=1}^{n_{gp}}\epsilon_{igp}\boldsymbol{X}_{igp}\right)'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Subtract from Equation (1) the average outcome in the population $\overline{Y} \equiv \frac{1}{n}\sum_p\sum_g\sum_i Y_{igp} = \widehat{\alpha} + \widehat{\tau}\overline{W} + \overline{\epsilon}$, where $\overline{W} \equiv \frac{1}{n}\sum_p\sum_g\sum_i W_{gp}$, and $\overline{\epsilon} \equiv \frac{1}{n}\sum_p\sum_g\sum_i \epsilon_{igp} = 0$ by construction. Then,

$$Y_{igp} - \overline{Y} = \widehat{\tau}(W_{gp} - \overline{W}) + \epsilon_{igp}. \tag{H10}$$

Apply Equation (H8) to the residuals and covariates of the regression defined by Equation (H10).[19] Then,

$$\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) = \frac{\sum_p\left[\sum_g(W_{gp} - \overline{W})\sum_i\epsilon_{igp}\right]^2}{\left[\sum_p\sum_g\sum_i(W_{gp} - \overline{W})^2\right]^2}. \tag{H11}$$

The numerator of $\widehat{\mathbb{V}}_{pair}(\widehat{\tau})$ equals

$$\begin{aligned}\sum_p\left[\sum_g(W_{gp} - \overline{W})\sum_i\epsilon_{igp}\right]^2 &= \sum_p\left[(1 - \overline{W})SET_p - \overline{W}SEU_p\right]^2\\ &= \sum_p\left[\frac{C}{n}SET_p - \frac{T}{n}SEU_p\right]^2.\end{aligned} \tag{H12}$$

The first equality follows from the definition of SET_p and SEU_p. The second equality follows from the definition of T and C.

[19] The clustered variance estimators of $\widehat{\tau}$ in the demeaned regression in Equation (H10) and in the regression with an intercept in Equation (1) are equal (Cameron and Miller, 2015).

The denominator of $\widehat{\mathbb{V}}_{pair}(\widehat{\tau})$ equals

$$\begin{aligned}
\left[\sum_p \sum_g \sum_i (W_{gp} - \overline{W})^2\right]^2 &= \left[\sum_p \sum_g (W_{gp} - \overline{W})^2 n_{gp}\right]^2 \\
&= \left[(1-\overline{W})^2 \sum_p T_p + \overline{W}^2 \sum_p C_p\right]^2 \\
&= \left[\frac{C^2}{n^2} T + \frac{T^2}{n^2} C\right]^2 \\
&= \left[\frac{CT}{n}\right]^2. \qquad \text{(H13)}
\end{aligned}$$

The first equality follows from $(W_{gp} - \overline{W})$ being constant across units. The second equality follows from the definition of T_p and C_p. The third equality follows from the definition of T and C.
Then, combining Equations (H11), (H12) and (H13),

$$\begin{aligned}
\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) &= \frac{\sum_p \left[\frac{C}{n} SET_p - \frac{T}{n} SEU_p\right]^2}{\left[\frac{CT}{n}\right]^2} \\
&= \sum_p \left[\frac{SET_p}{T} - \frac{SEU_p}{C}\right]^2.
\end{aligned}$$

QED.

Point 2

Apply Equation (H9) to the residuals and covariates of the regression defined by Equation (H10). Then,

$$\widehat{\mathbb{V}}_{unit}(\widehat{\tau}) = \frac{\sum_p \sum_g \left[(W_{gp} - \overline{W}) \sum_i \epsilon_{igp}\right]^2}{\left[\sum_p \sum_g \sum_i (W_{gp} - \overline{W})^2\right]^2}. \qquad \text{(H14)}$$

The numerator of $\widehat{\mathbb{V}}_{unit}(\widehat{\tau})$ equals

$$\sum_p \sum_g \left[(W_{gp} - \overline{W}) \sum_i \epsilon_{igp} \right]^2 = \sum_p \sum_g (W_{gp} - \overline{W})^2 \left(\sum_i \epsilon_{igp} \right)^2$$
$$= \sum_p \left[(1 - \overline{W})^2 SET_p^2 + \overline{W}^2 SEU_p^2 \right]$$
$$= \sum_p \left[\frac{C^2}{n^2} SET_p^2 + \frac{T^2}{n^2} SEU_p^2 \right]. \tag{H15}$$

The second equality follows from the definition of SET_p and SEU_p. The third equality follows from the definition of T and C. Then, combining Equations (H13), (H14) and (H15),

$$\widehat{\mathbb{V}}_{unit}(\widehat{\tau}) = \frac{\sum_p \left[\frac{C^2}{n^2} SET_p^2 + \frac{T^2}{n^2} SEU_p^2 \right]}{\left[\frac{CT}{n} \right]^2}$$
$$= \sum_p \left[\frac{SET_p^2}{T^2} + \frac{SEU_p^2}{C^2} \right].$$

QED.

Point 3

Let $SET_{p,fe} = \sum_{g=1}^{2} \sum_{i=1}^{n_{gp}} W_{gp} u_{igp}$ and $SEU_{p,fe} = \sum_{g=1}^{2} \sum_{i=1}^{n_{gp}} (1 - W_{gp}) u_{igp}$ respectively be the sum of the residuals u_{igp} for the treated and untreated observations in pair p. Averaging Equation (2) across units in pair p,

$$\overline{Y}_p = \widehat{\tau}_{fe} \overline{W}_p + \widehat{\gamma}_p + \overline{u}_p, \tag{H16}$$

where $\overline{Y}_p = \frac{1}{n_p} \sum_{g=1}^{2} \sum_{i=1}^{n_{gp}} Y_{igp}$, $\overline{W}_p = \frac{1}{n_p} \sum_{g=1}^{2} \sum_{i=1}^{n_{gp}} W_{gp} = \frac{1}{n_p} \sum_{g=1}^{2} W_{gp} n_{gp} = \frac{T_p}{n_p}$, and $\overline{u}_p = \frac{1}{n_p} \sum_{g=1}^{2} \sum_{i=1}^{n_{gp}} u_{igp}$. Substracting Equation (H16) from Equation (2),

$$Y_{igp} - \overline{Y}_p = \widehat{\tau}_{fe} (W_{gp} - \overline{W}_p) + u_{igp} - \overline{u}_p. \tag{H17}$$

$\{u_{ijp'}\}$ is orthogonal to the pair-p fixed effect indicator $\{\delta_{igp}\}$, so

$$\sum_{p'=1}^{P}\sum_{g=1}^{2}\sum_{i=1}^{n_{jp'}} u_{ijp'}\delta_{igp} = 0$$

$$\Leftrightarrow \sum_{g=1}^{2}\sum_{i=1}^{n_{gp}} u_{igp} = 0, \tag{H18}$$

where the equivalence holds because $\delta_{igp} = 1$ if and only if observation i belongs to pair p. This implies that for all p $\overline{u}_p = 0$. Equation (H17) then becomes a regression with one covariate and the same residuals as in Equation (2):

$$Y_{igp} - \overline{Y}_p = \widehat{\tau}_{fe}(W_{gp} - \overline{W}_p) + u_{igp}. \tag{H19}$$

Now, it follows from Equations (H8) and (H19) that[20]

$$\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe}) = \frac{\left[\sum_{p=1}^{P}\left(\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}} u_{igp}(W_{gp} - \overline{W}_p)\right)^2\right]}{\left(\sum_{p=1}^{P}\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}}(W_{gp} - \overline{W}_p)^2\right)^2}. \tag{H20}$$

The denominator of $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$ equals

$$\begin{aligned}
\left[\sum_p\sum_g\sum_i (W_{gp} - \overline{W}_p)^2\right]^2 &= \left[\sum_p\sum_g (W_{gp} - \overline{W}_p)^2 n_{gp}\right]^2 \\
&= \left[\sum_p [T_p(1 - \overline{W}_p)^2 + C_p\overline{W}_p^2]\right]^2 \\
&= \left[\sum_p\left(T_p\frac{C_p^2}{n_p^2} + C_p\frac{T_p^2}{n_p^2}\right)\right]^2 \\
&= \left[\sum_p \frac{T_pC_p}{n_p}\right]^2 \\
&= \left[\sum_p (n_{1p}^{-1} + n_{2p}^{-1})^{-1}\right]^2. \qquad \text{(H21)}
\end{aligned}$$

[20]The clustered variance estimators of $\widehat{\tau}_{fe}$ in the regression residualized from the pair fixed effects in Equation (H19) and in the regression with pair fixed effects in Equation (2) are equal (Cameron and Miller, 2015).

The numerator of $\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe})$ is equal to

$$
\begin{aligned}
\sum_{p=1}^{P}\left(\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}} u_{igp}(W_{gp}-\overline{W}_p)\right)^2 &= \sum_{p=1}^{P}\left(\sum_{g=1}^{2}(W_{gp}-\overline{W}_p)\sum_{i=1}^{n_{gp}} u_{igp}\right)^2 \\
&= \sum_{p=1}^{P}\left(-\overline{W}_p(SET_{p,fe}+SEU_{p,fe})+SET_{p,fe}\right)^2 \\
&= \sum_{p=1}^{P}\left(SET_{p,fe}\right)^2, \qquad \text{(H22)}
\end{aligned}
$$

where $SET_{p,fe}+SEU_{p,fe}=\sum_{g=1}^{2}\sum_{i=1}^{n_{gp}} u_{igp}=0$ from Equation (H18). Finally,

$$
\begin{aligned}
SET_{p,fe} &= \sum_{g,i} W_{gp}[Y_{igp}-\widehat{\gamma}_p-\widehat{\tau}_{fe}W_{gp}] \\
&= \sum_{g,i} W_{gp}Y_{igp}-(\widehat{\gamma}_p+\widehat{\tau}_{fe})\sum_{g,i} W_{gp} \\
&= \sum_{g,i} W_{gp}Y_{igp}-(\overline{Y}_p-\widehat{\tau}_{fe}\overline{W}_p+\widehat{\tau}_{fe})\overline{W}_p n_p \\
&= \sum_{g,i} W_{gp}Y_{igp}-\overline{W}_p\sum_{g,i} Y_{igp}-(1-\overline{W}_p)\widehat{\tau}_{fe}\overline{W}_p n_p \\
&= \sum_{g,i} W_{gp}Y_{igp}-\overline{W}_p\left[\sum_{g,i} W_{gp}Y_{igp}+\sum_{g,i}(1-W_{gp})Y_{igp}\right]-(1-\overline{W}_p)\overline{W}_p n_p\widehat{\tau}_{fe} \\
&= (1-\overline{W}_p)\sum_{g,i} W_{gp}Y_{igp}-\overline{W}_p\sum_{g,i}(1-W_{gp})Y_{igp}-(1-\overline{W}_p)\overline{W}_p n_p\widehat{\tau}_{fe} \\
&= (1-\overline{W}_p)\overline{W}_p n_p\left(\frac{\sum_{g,i} W_{gp}Y_{igp}}{\overline{W}_p n_p}-\frac{\sum_{g,i}(1-W_{gp})Y_{igp}}{(1-\overline{W}_p)n_p}-\widehat{\tau}_{fe}\right) \\
&= \frac{n_{1p}n_{2p}}{n_p^2}n_p\left(\frac{\sum_{g,i} W_{gp}Y_{igp}}{\sum_{g,i} W_{gp}}-\frac{\sum_{g,i}(1-W_{gp})Y_{igp}}{\sum_{g,i}(1-W_{gp})}-\widehat{\tau}_{fe}\right) \\
&= \frac{n_{1p}n_{2p}}{n_{1p}+n_{2p}}(\widehat{\tau}_p-\widehat{\tau}_{fe}). \qquad \text{(H23)}
\end{aligned}
$$

The first equality follows from the definition of $SET_{p,fe}$. The second equality follows from the definition of u_{igp} in Equation (2). The third equality follows from the definition of $\overline{W}_p$ and Equations (H16) and (H18). The ninth equality follows from the definition of $\widehat{\tau}_p$.

Therefore, combining Equations (H20), (H21), (H22) and (H23),

$$\widehat{\mathbb{V}}_{pair}(\widehat{\tau}_{fe}) = \sum_{p=1}^{P} \omega_p^2 (\widehat{\tau}_p - \widehat{\tau})^2$$

QED.

Point 4

Applying the definition of the UCVE from Equation (H9) to the regression in Equation (H19),

$$\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe}) = \frac{\left[\sum_{p=1}^{P} \sum_{g=1}^{2} \left(\sum_{i=1}^{n_{gp}} u_{igp}(W_{gp} - \overline{W}_p)\right)^2\right]}{\left(\sum_{p=1}^{P} \sum_{g=1}^{2} \sum_{i=1}^{n_{gp}} (W_{gp} - \overline{W}_p)^2\right)^2}. \tag{H24}$$

The numerator of $\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe})$ equals

$$\begin{aligned}
\sum_{p=1}^{P} \sum_{g=1}^{2} \left(\sum_{i=1}^{n_{gp}} u_{igp}(W_{gp} - \overline{W}_p)\right)^2 &= \sum_{p=1}^{P} \sum_{g=1}^{2} (W_{gp} - \overline{W}_p)^2 \left(\sum_{i=1}^{n_{gp}} u_{igp}\right)^2 \\
&= \sum_{p=1}^{P} \left((1 - \overline{W}_p)^2 SET_{p,fe}^2 + \overline{W}_p^2 SEU_{p,fe}^2\right) \\
&= \sum_{p=1}^{P} SET_{p,fe}^2 \left(\frac{C_p^2}{n_p^2} + \frac{T_p^2}{n_p^2}\right) \\
&= \sum_{p=1}^{P} \frac{C_p^2 T_p^2}{n_p^2} SET_{p,fe}^2 \left(\frac{1}{T_p^2} + \frac{1}{C_p^2}\right) \\
&= \sum_{p=1}^{P} (n_{1p}^{-1} + n_{2p}^{-1})^{-2} SET_{p,fe}^2 \left(\frac{1}{n_{1p}^2} + \frac{1}{n_{2p}^2}\right). \qquad \text{(H25)}
\end{aligned}$$

The second equality follows from the definitions of $SET_{p,fe}$ and $SEU_{p,fe}$. The third equality follows from Equation (H18), i.e., $SET_{p,fe} + SEU_{p,fe} = \sum_g \sum_i u_{igp} = 0$, for all p, so $SET_{p,fe}^2 = SEU_{p,fe}^2$, and the definitions of T_p and C_p. Finally, combining Equations (H21), (H23), (H24) and (H25),

$$\widehat{\mathbb{V}}_{unit}(\widehat{\tau}_{fe}) = \sum_{p=1}^{P} \omega_p^2 (\widehat{\tau}_p - \widehat{\tau})^2 \left(\left(\frac{n_{1p}}{n_p} \right)^2 + \left(\frac{n_{2p}}{n_p} \right)^2 \right).$$

QED.

H3. Proof of Lemma D.1

Point 1

$$\begin{aligned}
\widehat{\mathbb{V}}_{pop}(\widehat{\tau}) &= \frac{1}{P^2} \sum_{r=1}^{R} (\widehat{\tau}_{1r} - \widehat{\tau}_{2r})^2, \\
&= \frac{1}{P^2} \sum_{r=1}^{R} (\widehat{\tau}_{1r}^2 + \widehat{\tau}_{2r}^2 - 2\widehat{\tau}_{1r}\widehat{\tau}_{2r}).
\end{aligned}$$

Taking expected value,

$$\begin{aligned}
\mathbb{E}[\widehat{\mathbb{V}}_{pop}(\widehat{\tau})] &= \frac{1}{P^2} \sum_{r=1}^{R} \mathbb{E}(\widehat{\tau}_{1r}^2 + \widehat{\tau}_{2r}^2 - 2\widehat{\tau}_{1r}\widehat{\tau}_{2r}), \\
&= \frac{1}{P^2} \sum_{r=1}^{R} (\mathbb{V}(\widehat{\tau}_{1r}) + \mathbb{V}(\widehat{\tau}_{2r}) + \tau_{1r}^2 + \tau_{2r}^2 - 2\tau_{1r}\tau_{2r}), \\
&= \frac{1}{P^2} \sum_{p=1}^{P} \mathbb{V}(\widehat{\tau}_p) + \frac{1}{P^2} \sum_{r=1}^{R} (\tau_{1r} - \tau_{2r})^2, \\
&= \mathbb{V}(\widehat{\tau}) + \frac{1}{P^2} \sum_{r=1}^{R} (\tau_{1r} - \tau_{2r})^2. \qquad \text{(H26)}
\end{aligned}$$

The second equality follows from properties of the variance and that $\mathbb{E}[\widehat{\tau}_{1r}] = \tau_{1r}$ and $\mathbb{E}[\widehat{\tau}_{2r}] = \tau_{2r}$. The third equality follows from $P = 2R$. The fourth equality follows from Equation (3). **QED**.

Point 2

$$\begin{aligned}
\widehat{\mathbb{V}}_{brs}(\widehat{\tau}) &= \frac{1}{P^2}\sum_p \widehat{\tau}_p^2 - \frac{1}{2}\left(\frac{2}{P^2}\sum_r \widehat{\tau}_{1r}\widehat{\tau}_{2r} + \frac{\widehat{\tau}^2}{P}\right). \\
&= \frac{1}{2P^2}\sum_p (\widehat{\tau}_p - \widehat{\tau})^2 + \frac{1}{2P^2}\sum_r (\widehat{\tau}_{1r}^2 + \widehat{\tau}_{2r}^2 - 2\widehat{\tau}_{1r}\widehat{\tau}_{2r}). \\
&= \frac{1}{2}\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) + \frac{1}{2}\widehat{\mathbb{V}}_{pop}(\widehat{\tau}).
\end{aligned}$$

QED.

POINT 3

$$\mathbb{E}[\widehat{\mathbb{V}}_{pop}(\widehat{\tau})] \leq \mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right],$$

$$\Leftrightarrow (2R-1)\sum_{r=1}^{R}(\tau_{1r} - \tau_{2r})^2 \leq 2R\sum_{p=1}^{P}(\tau_p - \tau)^2,$$

$$\Leftrightarrow (2R-1)\sum_{r=1}^{R}(\tau_{1r}^2 + \tau_{2r}^2 - 2\tau_{1r}\tau_{2r}) \leq 2R\sum_{r=1}^{R}[\tau_{1r}^2 - 2\tau_{1r}\tau + \tau^2 + \tau_{2r}^2 - 2\tau_{2r}\tau + \tau^2],$$

$$\Leftrightarrow 0 \leq \sum_{r=1}^{R}(\tau_{1r} - \tau_{2r})^2 + 2R\sum_{r=1}^{R}[2\tau_{1r}\tau_{2r} - 2(\tau_{1r} + \tau_{2r})\tau + 2\tau^2],$$

$$\Leftrightarrow 0 \leq \sum_{r=1}^{R}(\tau_{1r} - \tau_{2r})^2 + 4R\sum_{r=1}^{R}(\tau_{1r} - \tau)(\tau_{2r} - \tau).$$

The second inequality follows from Points 1 and 1 of this lemma. Let $\tau_{\cdot r} = \frac{1}{2}(\tau_{1r} + \tau_{2r})$. Then,

$$\mathbb{E}[\widehat{\mathbb{V}}_{pop}(\widehat{\tau})] \leq \mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right],$$

$$\Leftrightarrow 0 \leq \sum_{r=1}^{R}\sum_{p=1,2} 2(\tau_{pr} - \tau_{\cdot r})^2 + 4R\sum_{r=1}^{R}(\tau_{1r} - \tau_{\cdot r} + \tau_{\cdot r} - \tau)(\tau_{2r} - \tau_{\cdot r} + \tau_{\cdot r} - \tau),$$

$$\Leftrightarrow 0 \leq \sum_{r=1}^{R}\sum_{p=1,2} \frac{1}{2}(\tau_{pr} - \tau_{\cdot r})^2 + R\sum_{r=1}^{R}[(\tau_{1r} - \tau_{\cdot r})(\tau_{2r} - \tau_{\cdot r}) + (\tau_{\cdot r} - \tau)^2],$$

$$\Leftrightarrow 0 \leq \sum_{r=1}^{R}\sum_{p=1,2} \frac{1}{2}(\tau_{pr} - \tau_{\cdot r})^2 + R\sum_{r=1}^{R}\left[-\sum_{p=1,2}\frac{1}{2}(\tau_{pr} - \tau_{\cdot r})^2 + (\tau_{\cdot r} - \tau)^2\right],$$

$$\Leftrightarrow \frac{1}{R}\sum_{r=1}^{R}\sum_{p=1,2} \frac{1}{2}(\tau_{pr} - \tau_{\cdot r})^2 \leq \frac{1}{R-1}\sum_{r=1}^{R}(\tau_{\cdot r} - \tau)^2.$$

This proves inequality a).

Then, if $\frac{1}{R}\sum_{r=1}^{R}\sum_{p=1,2}\frac{1}{2}(\tau_{pr}+\tau_{\cdot r})^2 \leq \frac{1}{R-1}\sum_{r=1}^{R}(\tau_{\cdot r}-\tau)^2$, it follows from Point 2 of the lemma and the previous display that

$$\begin{aligned}
\mathbb{E}\left[\widehat{\mathbb{V}}_{pop}(\widehat{\tau})\right] &\leq \frac{1}{2}\,\mathbb{E}\left[\widehat{\mathbb{V}}_{pop}(\widehat{\tau})\right] + \frac{1}{2}\,\mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right] \\
&\leq \frac{1}{2}\,\mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pop}(\widehat{\tau})\right] + \frac{1}{2}\,\mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right] \\
&= \mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{brs}(\widehat{\tau})\right],
\end{aligned}$$

which proves inequality b).

Similarly, if $\frac{1}{R}\sum_{r=1}^{R}\sum_{p=1,2}\frac{1}{2}(\tau_{pr} + \tau_{\cdot r})^2 \leq \frac{1}{R-1}\sum_{r=1}^{R}(\tau_{\cdot r} - \tau)^2$, it follows from Point 2 of the lemma and the previous display that

$$\begin{aligned}
\mathbb{E}\left[\widehat{\mathbb{V}}_{brs}(\widehat{\tau})\right] &\leq \frac{1}{2}\,\mathbb{E}\left[\widehat{\mathbb{V}}_{pop}(\widehat{\tau})\right] + \frac{1}{2}\,\mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right] \\
&\leq \frac{1}{2}\,\mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right] + \frac{1}{2}\,\mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right] \\
&= \mathbb{E}\left[\frac{P}{P-1}\widehat{\mathbb{V}}_{pair}(\widehat{\tau})\right],
\end{aligned}$$

which proves inequality c).

QED.

H4. Proof of Theorem D.2

Point 1

$$\begin{aligned}
P\widehat{\mathbb{V}}_{pop}(\widehat{\tau}) - P\mathbb{V}(\widehat{\tau}) &= \frac{1}{P}\sum_{r=1}^{R}[\widehat{\tau}_{1r}^2 - 2\widehat{\tau}_{1r}\widehat{\tau}_{2r} + \widehat{\tau}_{2r}^2] - \frac{1}{P}\sum_{p=1}^{P}\mathbb{V}(\widehat{\tau}_p) \\
&= \frac{1}{P}\sum_{p=1}^{P}\widehat{\tau}_p^2 - \frac{2}{P}\sum_{r=1}^{R}\widehat{\tau}_{1r}\widehat{\tau}_{2r} - \frac{1}{P}\sum_{p=1}^{P}[\mathbb{E}(\widehat{\tau}_p^2) - \tau_p^2] \\
&= \sum_{p=1}^{P}\frac{\widehat{\tau}_p^2 - \mathbb{E}[\widehat{\tau}_p^2]}{P} - \frac{1}{R}\sum_{r=1}^{R}\widehat{\tau}_{1r}\widehat{\tau}_{2r} + \frac{1}{P}\sum_{r=1}^{R}(\tau_{1r}^2 + \tau_{2r}^2) \\
&\xrightarrow{\mathbb{P}} \lim_{P\to+\infty}\frac{1}{P}\sum_{r=1}^{R}(\tau_{1r} - \tau_{2r})^2. \qquad \text{(H27)}
\end{aligned}$$

The second equality follows from the properties of the variance. As $P \to +\infty$, by Lemma H.1, $\sum_{p=1}^{P}\frac{\widehat{\tau}_p^2-\mathbb{E}[\widehat{\tau}_p^2]}{P} \xrightarrow{\mathbb{P}} 0$. Likewise, as $R = P/2 \to +\infty$, by Lemma 1 in Liu (1988), $\sum_{r=1}^{R}\widehat{\tau}_{1r}\widehat{\tau}_{2r}/R - \sum_{r=1}^{R}\tau_{1r}\tau_{2r}/R \xrightarrow{\mathbb{P}} 0$, because $\mathbb{E}[|\widehat{\tau}_{1r}\widehat{\tau}_{2r}|^{1+\epsilon/2}]$ is uniformly bounded in r by Equation (H40) and the Cauchy-Schwarz inequality, $(\widehat{\tau}_{1r}\widehat{\tau}_{2r})_{r=1}^{+\infty}$ is a sequence of independent random variables by Point 3 of Assumption 1, and $\mathbb{E}(\widehat{\tau}_{1r}\widehat{\tau}_{2r}) = \mathbb{E}(\widehat{\tau}_{1r})\,\mathbb{E}(\widehat{\tau}_{2r}) = \tau_{1r}\tau_{2r}$. Finally, the convergence arrow follows from Point 2 of Assumption 3 and some algebra.

The result follows from Equations (H3) and (H27) and a reasoning similar to that used to prove Equation (H6).
QED.

Point 2

$$\begin{aligned}
P\widehat{\mathbb{V}}_{bsr}(\widehat{\tau}) - P\mathbb{V}(\widehat{\tau}) &= \frac{1}{2}P(\widehat{\mathbb{V}}_{pair}(\widehat{\tau}) - \mathbb{V}(\widehat{\tau})) + \frac{1}{2}P(\widehat{\mathbb{V}}_{pop}(\widehat{\tau}) - \mathbb{V}(\widehat{\tau})) \\
&\xrightarrow{\mathbb{P}} \frac{1}{2}\lim_{P\to+\infty}\frac{1}{P}\sum_{p=1}^{P}(\tau_p - \tau)^2 + \frac{1}{2}\lim_{P\to+\infty}\frac{1}{P}\sum_{r=1}^{R}(\tau_{1r} - \tau_{2r})^2.
\end{aligned}$$

The first equality follows from Point 2 of Lemma I. The convergence arrow follows from Equations (H5) and (H27). The result follows from the previous display, Equation (H3), and a reasoning similar to that used to prove Equation (H6).
QED.

Point 3

$$
\begin{aligned}
&\sigma^2_{pair} \leq \sigma^2_{pop},\\
\Leftrightarrow &\lim_{P\to+\infty} \frac{1}{R}\sum_{r=1}^{R}(\tau_{1r}-\tau_{2r})^2 \leq \lim_{P\to+\infty} \frac{1}{R}\sum_{p=1}^{P}(\tau_p-\tau)^2,\\
\Leftrightarrow &\lim_{P\to+\infty} \frac{1}{R}\sum_{r=1}^{R}(\tau_{1r}^2+\tau_{2r}^2-2\tau_{1r}\tau_{2r}) \leq \lim_{P\to+\infty} \frac{1}{R}\sum_{r=1}^{R}[\tau_{1r}^2+\tau_{2r}^2-2(\tau_{1r}+\tau_{2r})\tau+2\tau^2],\\
\Leftrightarrow &\, 0 \leq \lim_{P\to+\infty} \frac{1}{R}\sum_{r=1}^{R}[2\tau_{1r}\tau_{2r}-2(\tau_{1r}+\tau_{2r})\tau+2\tau^2],\\
\Leftrightarrow &\, 0 \leq \lim_{P\to+\infty} \frac{1}{R}\sum_{r=1}^{R}(\tau_{1r}-\tau)(\tau_{2r}-\tau).
\end{aligned}
$$

Then, $\sigma^2_{pair} \leq \sigma^2_{bsr} \leq \sigma^2_{pop} \Leftrightarrow \sigma^2_{pair} \leq \sigma^2_{pop}$.

Point 4 is straightforward so we do not prove it.
QED.

H5. Proof of Lemma G.1

Let e_{igp} be the residual from the weighted least squares regression. One has

$$Y_{igp} = \widetilde{\alpha} + \widetilde{\tau}W_{gp} + e_{igp}.$$

Let $\widetilde{Y} = \frac{1}{n}\sum_{i,g,p} V_{gp}Y_{igp}$. The previous display implies that

$$
\begin{aligned}
\widetilde{Y} &= \widetilde{\alpha}\sum_{i,g,p}\frac{V_{gp}}{n} + \widetilde{\tau}\frac{1}{n}\sum_{i,g,p} V_{gp}W_{gp} + \frac{1}{n}\sum_{i,g,p} V_{gp}e_{igp}\\
&= 2\widetilde{\alpha} + \widetilde{\tau},
\end{aligned}
$$

where the second equality follows from $\frac{1}{n}\sum_{i,g,p} V_{gp}e_{igp} = 0$, by the first-order condition attached to $\widetilde{\alpha}$ in the weighted OLS minimization problem. Then, combining the two preceding displays implies that

$$Y_{igp} - \frac{1}{2}\widetilde{Y} = \widetilde{\tau}\left(W_{gp} - \frac{1}{2}\right) + e_{igp}. \tag{H28}$$

The next step is to compute the clustered variance estimators for the weighted least squares estimator. To do so, we apply Equation (15) in Cameron and Miller (2015) to the residuals and covariates of the regression defined by Equation (H28). This equation implies that

$$\widehat{\mathbb{V}}_{pair}(\widetilde{\tau}) = \frac{\sum_p \left[\sum_g V_{gp}(W_{gp} - \frac{1}{2})\sum_i e_{igp}\right]^2}{\left[\sum_p \sum_g \sum_i V_{gp}(W_{gp} - \frac{1}{2})^2\right]^2}. \tag{H29}$$

Let $\widehat{Y}_{igp} = \widetilde{\alpha} + W_{gp}\widetilde{\tau}$, $\widehat{Y}(0) = \widetilde{\alpha}$, and $\widehat{Y}(1) = \widetilde{\alpha} + \widetilde{\tau}$. Note that

$$\begin{aligned}\sum_{i,g} W_{gp}\frac{e_{igp}}{n_{gp}} &= \sum_{i,g} W_{gp}(Y_{igp} - \widehat{Y}_{igp})/n_{gp} \\ &= \sum_g W_{gp}\overline{y}_{gp}(1) - \widehat{Y}(1)\sum_g W_{gp} \\ &= \widehat{Y}_p(1) - \sum_{p'} \frac{n_{p'}}{n}\widehat{Y}_{p'}(1) \end{aligned} \tag{H30}$$

The second equality follows from $W_{gp}Y_{igp} = W_{gp}Y_{igp}(1)$, the definition of $\overline{y}_{gp}(1)$ and $W_{gp}\widehat{Y}_{igp} = W_{gp}\widehat{Y}(1)$. The third equality follows from the definition of $\widehat{Y}_p(1)$, Point 2 of Assumption 1, and the definition of $\widehat{Y}(1)$.

Likewise,

$$\sum_{i,g}(1 - W_{gp})\frac{e_{igp}}{n_{gp}} = \widehat{Y}_p(0) - \sum_{p'} \frac{n_{p'}}{n}\widehat{Y}_{p'}(0) \tag{H31}$$

The numerator of $\widehat{\mathbb{V}}_{pair}(\widetilde{\tau})$ equals

$$\begin{aligned}
\sum_p \left[\sum_g V_{gp}\left(W_{gp}-\frac{1}{2}\right)\sum_i e_{igp}\right]^2 &= \sum_p \left[\sum_g n_p\left(W_{gp}-\frac{1}{2}\right)(W_{gp}+1-W_{gp})\sum_i \frac{e_{igp}}{n_{gp}}\right]^2 \\
&= \sum_p n_p^2\left[\left(1-\frac{1}{2}\right)\sum_{i,g} W_{gp}\frac{e_{igp}}{n_{gp}} - \frac{1}{2}\sum_{i,g}(1-W_{gp})\frac{e_{igp}}{n_{gp}}\right]^2 \\
&= \sum_p \frac{n_p^2}{4}\left[\widehat{Y}_p(1) - \sum_{p'}\frac{n_{p'}}{n}\widehat{Y}_{p'}(1) - \widehat{Y}_p(0) + \sum_{p'}\frac{n_{p'}}{n}\widehat{Y}_{p'}(0)\right]^2 \\
&= \sum_p \frac{n_p^2}{4}\left[\widehat{\tau}_p - \widetilde{\tau}\right]^2. \qquad \text{(H32)}
\end{aligned}$$

The second equality follows from the fact that $W_{gp}-\frac{1}{2} = 1-\frac{1}{2}$ for the treated units and $W_{gp}-\frac{1}{2} = -\frac{1}{2}$ for the untreated units. The third equality follows from Equations (H30) and (H31).

The denominator of $\widehat{\mathbb{V}}_{pair}(\widetilde{\tau})$ equals

$$\begin{aligned}
\left[\sum_p\sum_g\sum_i V_{gp}\left(W_{gp}-\frac{1}{2}\right)^2\right]^2 &= \left[2n\frac{1}{4}\right]^2 \\
&= \frac{n^2}{4}. \qquad \text{(H33)}
\end{aligned}$$

Then, combining Equations (H29), (H32) and (H33),

$$\widehat{\mathbb{V}}_{pair}(\widetilde{\tau}) = \sum_p \frac{n_p^2}{n^2}\left[\widehat{\tau}_p - \widetilde{\tau}\right]^2 = \frac{1}{P^2}\sum_p \frac{n_p^2}{\overline{n}^2}\left[\widehat{\tau}_p - \widetilde{\tau}\right]^2. \qquad \text{(H34)}$$

QED.

H6. Proof of Theorem G.2

It follows from Lemma H.1 that

$$\widetilde{\tau} - \tau^* = \frac{1}{P}\sum_p \frac{n_p}{\overline{n}}\left(\widehat{\tau}_p - \mathbb{E}[\widehat{\tau}_p]\right) \xrightarrow{\mathbb{P}} 0, \qquad \text{(H35)}$$

and

$$\frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2[\widehat{\tau}_p^2-\mathbb{E}(\widehat{\tau}_p^2)]\xrightarrow{\mathbb{P}}0. \tag{H36}$$

By a similar argument to the one used in the proof of Lemma H.1, one can also show that

$$\frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2[\widehat{\tau}_p-\mathbb{E}(\widehat{\tau}_p)]\xrightarrow{\mathbb{P}}0. \tag{H37}$$

We now use Point 3 of Assumption 5 to derive the asymptotic distribution of $(\widetilde{\tau}-\tau^*)/(\widetilde{S}_P/P)$. As $\sum_{p=1}^{P}\mathbb{E}\left[|\frac{n_p}{\overline{n}}|^{2+\epsilon}|\widehat{\tau}_p-\mathbb{E}\left[\widehat{\tau}_p\right]|^{2+\epsilon}/\widetilde{S}_P^{2+\epsilon}\right]\to 0$ for some $\epsilon>0$ (by Point 3 of Assumption 5), then, by the Lyapunov central limit theorem, $(\widetilde{\tau}-\tau^*)/(\widetilde{S}_P/P)=\sum_p\frac{n_p}{\overline{n}}(\widehat{\tau}_p-\mathbb{E}[\widehat{\tau}_p])/\widetilde{S}_P\xrightarrow{d}\mathcal{N}(0,1)$ as $P\to+\infty$, as $\widetilde{S}_P^2=P^2\mathbb{V}(\widetilde{\tau})=\sum_{p=1}^{P}\mathbb{V}\left(\frac{n_p}{\overline{n}}\widehat{\tau}_p\right)$.
Therefore,

$$(\widetilde{\tau}-\tau^*)/\sqrt{\mathbb{V}(\widetilde{\tau})}\xrightarrow{d}\mathcal{N}(0,1). \tag{H38}$$

Then,

$$
\begin{aligned}
&P\widehat{\mathbb{V}}_{pair}(\widetilde{\tau}) - P\mathbb{V}(\widetilde{\tau}) \\
&= \frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2(\widehat{\tau}_p - \widetilde{\tau})^2 - \frac{1}{P}\sum_{p=1}^{P}\left(\frac{n_p}{\overline{n}}\right)^2\mathbb{V}(\widehat{\tau}_p) \\
&= \frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2(\widehat{\tau}_p - \widetilde{\tau})^2 - \frac{1}{P}\sum_{p=1}^{P}\left(\frac{n_p}{\overline{n}}\right)^2[\mathbb{E}(\widehat{\tau}_p^2) - \mathbb{E}[\widehat{\tau}_p]^2] \\
&= \frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2(\widehat{\tau}_p^2 - 2\widetilde{\tau}\widehat{\tau}_p + \widetilde{\tau}^2) - \frac{1}{P}\sum_{p=1}^{P}\left(\frac{n_p}{\overline{n}}\right)^2[\mathbb{E}(\widehat{\tau}_p^2) - \mathbb{E}[\widehat{\tau}_p]^2] \\
&= \frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2(\widehat{\tau}_p^2 - \mathbb{E}[\widehat{\tau}_p^2]) - 2\widetilde{\tau}\frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2\widehat{\tau}_p + \widetilde{\tau}^2\frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2 + \frac{1}{P}\sum_{p=1}^{P}\left(\frac{n_p}{\overline{n}}\right)^2\mathbb{E}[\widehat{\tau}_p]^2 \\
&\xrightarrow{\mathbb{P}} -2\tau^{\infty}\lim_{P\to+\infty}\frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2\mathbb{E}[\widehat{\tau}_p] + (\tau^{\infty})^2\lim_{P\to+\infty}\frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2 + \lim_{P\to+\infty}\frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2\mathbb{E}[\widehat{\tau}_p]^2 \\
&= \lim_{P\to+\infty}\frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2\left[\mathbb{E}[\widehat{\tau}_p]^2 - 2\tau^{\infty}\,\mathbb{E}[\widehat{\tau}_p] + (\tau^{\infty})^2\right]
\end{aligned}
$$

(H39)

$$
= \lim_{P\to+\infty}\frac{1}{P}\sum_{p=1}^{P}\left(\frac{n_p}{\overline{n}}\right)^2[\mathbb{E}[\widehat{\tau}_p] - \tau^{\infty}]^2.
$$

The first equality follows from Equation (H34) and the fact that the $(\widehat{\tau}_p)_{p=1}^{P}$ are independent across p by Point 3 of Assumption 1. The second equality follows from the definition of variance. The convergence in probability follows from Equations (H35) and (H36), (H37), and Point 2 of Assumption 5.
Then,

$$
\begin{aligned}
\frac{\widetilde{\tau} - \tau^*}{\sqrt{\widehat{\mathbb{V}}_{pair}(\widetilde{\tau})}} &= \frac{\widetilde{\tau} - \mathbb{E}[\widetilde{\tau}]}{\sqrt{\mathbb{V}(\widetilde{\tau})}}\sqrt{\frac{P\mathbb{V}(\widetilde{\tau})}{P\widehat{\mathbb{V}}_{pair}(\widetilde{\tau})}} \\
&\xrightarrow{d} \mathcal{N}(0, \sigma_{wls}^2).
\end{aligned}
$$

The convergence in distribution follows from Equation (H39), Equation (H38), Lemma H.2, the Slutsky Lemma, and the CMT.
QED.

H7. *Proof of Lemma G.3*

$$\begin{aligned}
\mathbb{E}\left[\sum_p \widetilde{\omega}_p^2(\widehat{\tau}_p - \widehat{\tau}_{fe})^2\right] &= \sum_p \widetilde{\omega}_p^2\, \mathbb{E}[(\widehat{\tau}_p - \widehat{\tau}_{fe})^2] \\
&= \sum_p \widetilde{\omega}_p^2 \left[\mathbb{V}(\widehat{\tau}_p - \widehat{\tau}_{fe}) + [\mathbb{E}(\widehat{\tau}_p - \widehat{\tau}_{fe})]^2\right] \\
&= \sum_p \widetilde{\omega}_p^2 \left[\mathbb{V}(\widehat{\tau}_p) + \mathbb{V}(\widehat{\tau}_{fe}) - 2\mathrm{Cov}(\widehat{\tau}_p, \widehat{\tau}_{fe}) + [\mathbb{E}(\widehat{\tau}_p - \widehat{\tau}_{fe})]^2\right] \\
&= \sum_p \widetilde{\omega}_p^2 \left[\mathbb{V}(\widehat{\tau}_p) + \mathbb{V}(\widehat{\tau}_{fe}) - 2\omega_p \mathbb{V}(\widehat{\tau}_p) + [\mathbb{E}(\widehat{\tau}_p - \widehat{\tau}_{fe})]^2\right] \\
&= \sum_p \widetilde{\omega}_p^2\, [1 - 2\omega_p]\, \mathbb{V}(\widehat{\tau}_p) + \mathbb{V}(\widehat{\tau}_{fe}) \sum_p \widetilde{\omega}_p^2 + \sum_p \widetilde{\omega}_p^2 [\mathbb{E}(\widehat{\tau}_p - \widehat{\tau}_{fe})]^2 \\
&= \sum_p \omega_p^2 \mathbb{V}(\widehat{\tau}_p) + \mathbb{V}(\widehat{\tau}_{fe}) \sum_p \widetilde{\omega}_p^2 + \sum_p \widetilde{\omega}_p^2 [\mathbb{E}(\widehat{\tau}_p - \widehat{\tau}_{fe})]^2 \\
&= \mathbb{V}(\widehat{\tau}_{fe}) \left(1 + \textstyle\sum_p \widetilde{\omega}_p^2\right) + \sum_p \widetilde{\omega}_p^2 [\mathbb{E}(\widehat{\tau}_p - \widehat{\tau}_{fe})]^2
\end{aligned}$$

The first equality follows from the linearity of the expectation and the fact that the weights ω_p are not stochastic. The fourth equality follows from Point 3 of Assumption 1. The sixth equality follows from the definition of $\widetilde{\omega}_p$. The seventh equality follows from the definition of the variance, the definition of $\widehat{\tau}_{fe}$ and Point 3 of Assumption 1.
QED.

H8. *Auxiliary Lemmas to prove Theorems B.1, D.2, and G.2*

LEMMA H.1: *Let $q \geq 1$, under Points 2 and 3 of Assumption 1, and Assumption 2 or Point 1 of Assumption 3,*

$$\frac{1}{P} \sum_p \left(\frac{n_p}{\overline{n}}\right)^q [\widehat{\tau}_p^q - \mathbb{E}(\widehat{\tau}_p^q)] \xrightarrow{\mathbb{P}} 0$$

Proof. Assumption 2 implies Point 1 of Assumption 3, so it is sufficient to show that the result holds under Points 2 and 3 of Assumption 1, and Point 1 of Assumption 3.

Note that by Point 3 of Assumption 1, $((\frac{n_p}{\overline{n}}\widehat{\tau}_p)^q - \mathbb{E}[(\frac{n_p}{\overline{n}}\widehat{\tau}_p)^q])_{p=1}^P$, $q \geq 1$, is a sequence of independent random variables with mean zero.

Note that, for all p,

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{n_p}{\overline{n}}\widehat{\tau}_p\right|^{q+\epsilon}\right]^{1/(q+\epsilon)} &= \frac{n_p}{\overline{n}}\,\mathbb{E}\left[\left|\widehat{Y}_p(1)-\widehat{Y}_p(0)\right|^{q+\epsilon}\right]^{1/(q+\epsilon)} \\
&\leq N\left(\left(\mathbb{E}\left[\left|\widehat{Y}_p(1)\right|^{q+\epsilon}\right]\right)^{1/(q+\epsilon)}+\left(\mathbb{E}\left[\left|\widehat{Y}_p(0)\right|^{q+\epsilon}\right]\right)^{1/(q+\epsilon)}\right) \\
&= N\left(\left(\mathbb{E}\left[\left|\sum_g W_{gp}\overline{y}_{gp}(1)\right|^{q+\epsilon}\right]\right)^{1/(q+\epsilon)}+\left(\mathbb{E}\left[\left|\sum_g (1-W_{gp})\overline{y}_{gp}(0)\right|^{q+\epsilon}\right]\right)^{1/(q+\epsilon)}\right) \\
&\leq N\left(\sum_g\left(\mathbb{E}\left[\left|W_{gp}\overline{y}_{gp}(1)\right|^{q+\epsilon}\right]\right)^{1/(q+\epsilon)}+\sum_g\left(\mathbb{E}\left[\left|(1-W_{gp})\overline{y}_{gp}(0)\right|^{q+\epsilon}\right]\right)^{1/(q+\epsilon)}\right) \\
&= N\left(\sum_g\left(\mathbb{E}[W_{gp}]\left|\overline{y}_{gp}(1)\right|^{q+\epsilon}\right)^{1/(q+\epsilon)}+\sum_g\left(\mathbb{E}[1-W_{gp}]\left|\overline{y}_{gp}(0)\right|^{q+\epsilon}\right)^{1/(q+\epsilon)}\right) \\
&= N\left(\sum_g\left(\frac{1}{2}\left|\overline{y}_{gp}(1)\right|^{q+\epsilon}\right)^{1/(q+\epsilon)}+\sum_g\left(\frac{1}{2}\left|\overline{y}_{gp}(0)\right|^{q+\epsilon}\right)^{1/(q+\epsilon)}\right) \\
&< N\frac{4}{2^{1/(q+\epsilon)}}M<+\infty. \qquad \text{(H40)}
\end{aligned}$$

The first equality follows from the definition of $\widehat{\tau}_p$. The first inequality follows from Minkowski's inequality, and from Point 1 of Assumption 5. The third line follows from the definitions of $\widehat{Y}_p(1)$ and $\widehat{Y}_p(0)$. The fourth line follows from Minkowski's inequality. The fifth line follows from W_{gp} being a binary variable. The sixth line follows from Point 2 of Assumption 1. The seventh line follows from Point 1 of Assumption 3.

Using the LLN in Lemma 1 in Liu (1988), the previous facts and the fact that almost sure convergence implies convergence in probability, one concludes that

$$\frac{1}{P}\sum_p\left(\frac{n_p}{\overline{n}}\right)^q\left[\widehat{\tau}_p^q-\mathbb{E}(\widehat{\tau}_p^q)\right]\xrightarrow{\mathbb{P}}0. \qquad \text{(H41)}$$

QED.

LEMMA H.2: *[Strictly positive limit for $P\mathbb{V}(\widetilde{\tau})$] Under Point 2 of Assumption 3 and Point 1 of Assumption 5,* $\lim_{P\to+\infty}P\mathbb{V}(\widetilde{\tau})>0$.

Proof. Note that

$$\begin{aligned}
\lim_{P\to+\infty} P\mathbb{V}(\widetilde{\tau}) &= \lim_{P\to+\infty} \frac{1}{P}\sum_{p}\left(\frac{n_p}{\overline{n}}\right)^2 \mathbb{V}(\widehat{\tau}_p) \\
&\geq \frac{1}{N^2} \lim_{P\to+\infty} \frac{1}{P}\sum_{p} \mathbb{V}(\widehat{\tau}_p) \\
&= \frac{1}{N^2} \lim_{P\to+\infty} P\mathbb{V}(\widehat{\tau}) \\
&> 0.
\end{aligned}$$

The first equality follows from the definition of $\widetilde{\tau}$ and Point 3 of Assumption 1. The first inequality follows from the fact that $0 < \frac{1}{N} \leq \frac{n_p}{\overline{n}} \leq N$ (which follows from Point 1 of Assumption 5). The second equality follows from the definition of $\mathbb{V}(\widehat{\tau})$. The second inequality follows from Point 2 of Assumption 3.
QED.