A Appendix

A.1 Proof of Theorem 3.1

First we argue that $\hat{\theta}_n \xrightarrow{P} \theta^{\text{obs}}$. Note that

$$\hat{\theta}_n = \frac{\sum_{1 \le i \le n} Y_i(1) R_i(1) D_i}{\sum_{1 \le i \le n} R_i(1) D_i} - \frac{\sum_{1 \le i \le n} Y_i(0) R_i(0)(1 - D_i)}{\sum_{1 \le i \le n} R_i(0)(1 - D_i)} ~.$$

By Lemma S.1.5 in Bai et al. (2021),

$$\begin{aligned} \frac{1}{n/2} \sum_{1 \le i \le n} Y_i(1) R_i(1) D_i &\xrightarrow{P} E[Y_i(1) R_i(1)] ~, \\ \frac{1}{n/2} \sum_{1 \le i \le n} R_i(1) D_i &\xrightarrow{P} E[R_i(1)] ~, \\ \frac{1}{n/2} \sum_{1 \le i \le n} Y_i(0) R_i(0)(1 - D_i) &\xrightarrow{P} E[Y_i(0) R_i(0)] ~, \\ \frac{1}{n/2} \sum_{1 \le i \le n} R_i(0) D_i &\xrightarrow{P} E[R_i(0)] ~. \end{aligned}$$

Hence the result follows by the continuous mapping theorem. Next we argue that $\hat{\theta}_n^{\text{drop}} \xrightarrow{P} \theta^{\text{drop}}$. To begin, recall that $\hat{\theta}_n^{\text{drop}} = \mathbb{B}_n^{-1} \mathbb{C}_n$, where

$$\begin{aligned} \mathbb{B}_n &= \frac{1}{n/2} \sum_{1 \le j \le n/2} R_{\pi(2j-1)} R_{\pi(2j)} \\ \mathbb{C}_n &= \frac{1}{n/2} \sum_{1 \le j \le n/2} R_{\pi(2j-1)} R_{\pi(2j)} (Y_{\pi(2j-1)} - Y_{\pi(2j)})(D_{\pi(2j-1)} - D_{\pi(2j)}) \end{aligned}$$

For $\mathbb{B}_n$, it follows Assumptions 3.1, 3.2(a), 3.3, the fact that $R_i(d) \in \{0, 1\}$ for $d \in \{0, 1\}$ and therefore has finite second moments, and similar arguments to those in the proof of Lemma S.1.6 of Bai et al. (2021) that as $n \to \infty$,

$$\mathbb{B}_n \xrightarrow{P} E[E[R_i(1) | X_i] E[R_i(0) | X_i]] ~. \tag{9}$$

Next, we turn to $\mathbb{C}_n$. Note

$$\begin{aligned} \mathbb{C}_n = \frac{1}{n/2} \sum_{1 \le j \le n/2} \Big(& R_{\pi(2j-1)}(1) R_{\pi(2j)}(0) (Y_{\pi(2j-1)}(1) - Y_{\pi(2j)}(0)) D_{\pi(2j-1)} \\ & + R_{\pi(2j-1)}(0) R_{\pi(2j)}(1) (Y_{\pi(2j)}(1) - Y_{\pi(2j-1)}(0))(1 - D_{\pi(2j-1)}) \Big) ~. \end{aligned}$$

It follows from Assumption 3.1 and $Q_n = Q^n$ that

$$E[\mathbb{C}_n|X^{(n)}] = \frac{1}{n}\sum_{1\le j\le n/2}\Big(E[Y_{\pi(2j-1)}(1)R_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]E[R_{\pi(2j)}(0)|X_{\pi(2j)}]$$
$$- E[Y_{\pi(2j)}(0)R_{\pi(2j)}(0)|X_{\pi(2j)}]E[R_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]$$
$$+ E[Y_{\pi(2j)}(1)R_{\pi(2j)}(1)|X_{\pi(2j)}]E[R_{\pi(2j-1)}(0)|X_{\pi(2j-1)}]$$
$$- E[Y_{\pi(2j-1)}(0)R_{\pi(2j-1)}(0)|X_{\pi(2j-1)}]E[R_{\pi(2j)}(1)|X_{\pi(2j)}]\Big)$$

Next, it follows from Assumptions 2.1(a), 3.2(b), 3.1, 3.3, and similar arguments to those in the proof of Lemma S.1.6 of Bai et al. (2021) that as $n \to \infty$,

$$\frac{1}{n}\sum_{1\le j\le n/2}\Big(E[Y_{\pi(2j-1)}(1)R_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]E[R_{\pi(2j)}(0)|X_{\pi(2j)}]$$
$$+ E[Y_{\pi(2j)}(1)R_{\pi(2j)}(1)|X_{\pi(2j)}]E[R_{\pi(2j-1)}(0)|X_{\pi(2j-1)}]\Big) \xrightarrow{P} E[E[Y_i(1)R_i(1)|X_i]E[R_i(0)|X_i]] \quad (10)$$

and

$$\frac{1}{n}\sum_{1\le j\le n/2}\Big(E[Y_{\pi(2j-1)}(0)R_{\pi(2j-1)}(0)|X_{\pi(2j-1)}]E[R_{\pi(2j)}(1)|X_{\pi(2j)}]$$
$$+ E[Y_{\pi(2j)}(0)R_{\pi(2j)}(0)|X_{\pi(2j)}]E[R_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]\Big) \xrightarrow{P} E[E[Y_i(0)R_i(0)|X_i]E[R_i(1)|X_i]]\ . \quad (11)$$

Moreover, it can be shown using similar arguments to those in the proof of Lemma S.1.6 of Bai et al. (2021) that

$$\Big|\mathbb{C}_n - E[\mathbb{C}_n|X^{(n)}]\Big| \xrightarrow{P} 0\ , \quad (12)$$

and hence by combining (10)-(12) we obtain that

$$\mathbb{C}_n \xrightarrow{P} [E[Y_i(1)R_i(1)|X_i]E[R_i(0)|X_i]] + E[E[Y_i(0)R_i(0)|X_i]E[R_i(1)|X_i]]\ . \quad (13)$$

The conclusion then follows from (9), (13), as well as the continuous mapping theorem. ■

A.2 Proof of Theorem 3.2

First we argue that $\hat{\theta}_n \xrightarrow{P} \theta^{\text{obs}}$. Note that

$$\hat{\theta}_n = \frac{\sum_{1\le i\le n} Y_i(1)R_i(1)D_i}{\sum_{1\le i\le n} R_i(1)D_i} - \frac{\sum_{1\le i\le n} Y_i(0)R_i(0)(1-D_i)}{\sum_{1\le i\le n} R_i(0)(1-D_i)}\ .$$

By Lemma B.3 in Bugni et al. (2018),

$$\frac{1}{n}\sum_{1\le i\le n} Y_i(1)R_i(1)D_i \xrightarrow{P} \nu E[Y_i(1)R_i(1)]~,$$

where we note that an inspection of their proof shows that Assumption 3.4(b) is sufficient to establish their result. Similarly,

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le i\le n} R_i(1)D_i \xrightarrow{P} \nu E[R_i(1)]~,\\
&\frac{1}{n}\sum_{1\le i\le n} Y_i(0)R_i(0)(1-D_i) \xrightarrow{P} (1-\nu)E[Y_i(0)R_i(0)]~,\\
&\frac{1}{n}\sum_{1\le i\le n} R_i(0)D_i \xrightarrow{P} (1-\nu)E[R_i(0)]~.
\end{aligned}$$

Hence the result follows by the continuous mapping theorem. Next we argue that $\hat\theta_n^{\rm sfe} \xrightarrow{P} \theta^{\rm sfe}$. To that end, write $\hat\theta_n^{\rm sfe}$ as

$$\hat\theta_n^{\rm sfe} = \frac{\sum_{1\le i\le n} R_i\tilde D_i Y_i}{\sum_{1\le i\le n} R_i \tilde D_i^2}~,$$

where $\tilde D_i$ is the projection of D_i on the strata indicators, i.e., $\tilde D_i = D_i - n_1(S_i)/n(S_i)$, and

$$\frac{n_1(S_i)}{n(S_i)} = \sum_{s\in\mathcal S} I\{S_i=s\}\frac{n_1(s)}{n(s)}~,$$

for

$$n_1(s) = \sum_{1\le i\le n} R_iD_iI\{S_i=s\}, \quad n(s) = \sum_{1\le i\le n} R_i I\{S_i=s\}~.$$

By Lemma B.3 in Bugni et al. (2018) and the continuous mapping theorem, we have

$$\begin{aligned}
\frac{n_1(s)}{n(s)} = \frac{\frac{1}{n}\sum_{1\le i\le n} R_iD_iI\{S_i=s\}}{\frac{1}{n}\sum_{1\le i\le n} R_iI\{S_i=s\}} &\xrightarrow{P} \frac{\nu E[R_i(1)I\{S_i=s\}]}{\nu E[R_i(1)I\{S_i=s\}] + (1-\nu)E[R_i(0)I\{S_i=s\}]}\\
&= \frac{\nu E[R_i(1)|S_i=s]}{\nu E[R_i(1)|S_i=s] + (1-\nu)E[R_i(0)|S_i=s]}~.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\frac{1}{n}\sum_{1\le i\le n} R_i\tilde D_i Y_i\\
&= \frac{1}{n}\sum_{1\le i\le n} R_iD_iY_i - \sum_{s\in\mathcal S}\frac{1}{n}\sum_{1\le i\le n} I\{S_i=s\}R_iY_i\frac{\nu E[R_i(1)|S_i=s]}{\nu E[R_i(1)|S_i=s] + (1-\nu)E[R_i(0)|S_i=s]}\\
&\qquad + \sum_{s\in\mathcal S}\frac{1}{n}\sum_{1\le i\le n} R_iY_iI\{S_i=s\}\left(\frac{\nu E[R_i(1)|S_i=s]}{\nu E[R_i(1)|S_i=s] + (1-\nu)E[R_i(0)|S_i=s]} - \frac{n_1(s)}{n(s)}\right)
\end{aligned}$$

$$\xrightarrow{P} \nu E[R_i(1)Y_i(1)] - \sum_{s\in\mathcal{S}} (\nu E[R_i(1)Y_i(1)I\{S_i = s\}] + (1-\nu)E[R_i(0)Y_i(0)I\{S_i = s\}])$$
$$\times \frac{\nu E[R_i(1)|S_i = s]}{\nu E[R_i(1)|S_i = s] + (1-\nu)E[R_i(0)|S_i = s]}$$
$$= \nu E[R_i(1)Y_i(1)] - \sum_{s\in\mathcal{S}} p(s)(\nu E[R_i(1)Y_i(1)|S_i = s] + (1-\nu)E[R_i(0)Y_i(0)|S_i = s])$$
$$\times \frac{\nu E[R_i(1)|S_i = s]}{\nu E[R_i(1)|S_i = s] + (1-\nu)E[R_i(0)|S_i = s]}$$
$$= \nu E[R_i(1)Y_i(1)]$$
$$- E\left[(\nu E[R_i(1)Y_i(1)|S_i] + (1-\nu)E[R_i(0)Y_i(0)|S_i])\frac{\nu E[R_i(1)|S_i]}{\nu E[R_i(1)|S_i] + (1-\nu)E[R_i(0)|S_i]}\right]$$
$$= \nu(1-\nu)E\left[\frac{E[R_i(1)Y_i(1)|S_i]E[R_i(0)|S_i] - E[R_i(0)Y_i(0)|S_i]E[R_i(1)|S_i]}{\nu E[R_i(1)|S_i] + (1-\nu)E[R_i(0)|S_i]}\right] ,$$

where in the last equality we used the fact that $E[R_i(1)Y_i(1)] = E[E[R_i(1)Y_i(1)|S_i]]$. Also note that

$$\frac{1}{n}\sum_{1\le i\le n} R_i\tilde{D}_i^2$$
$$= \frac{1}{n}\sum_{1\le i\le n} R_i\tilde{D}_i\left(D_i - \frac{n_1(S_i)}{n(S_i)}\right)$$
$$= \frac{1}{n}\sum_{1\le i\le n} R_i\left(1 - \frac{n_1(S_i)}{n(S_i)}\right)D_i$$
$$= \frac{1}{n}\sum_{1\le i\le n} R_iD_i - \sum_{s\in\mathcal{S}}\frac{n_1(s)}{n(s)}\frac{1}{n}\sum_{1\le i\le n} R_iD_iI\{S_i = s\}$$
$$\xrightarrow{P} \nu E[R_i(1)] - \sum_{s\in\mathcal{S}} p(s)\nu E[R_i(1)|S_i = s]\frac{\nu E[R_i(1)|S_i = s]}{\nu E[R_i(1)|S_i = s] + (1-\nu)E[R_i(0)|S_i = s]}$$
$$= \nu E[R_i(1)] - \nu E\left[E[R_i(1)|S_i]\frac{\nu E[R_i(1)|S_i]}{\nu E[R_i(1)|S_i] + (1-\nu)E[R_i(0)|S_i]}\right]$$
$$= \nu(1-\nu)E\left[\frac{E[R_i(1)|S_i]E[R_i(0)|S_i]}{\nu E[R_i(1)|S_i] + (1-\nu)E[R_i(0)|S_i]}\right] ,$$

where the second equality follows from $\sum_{1\le i\le n} R_i\tilde{D}_i\frac{n_1(S_i)}{n(S_i)} = 0$, which is derived as follows:

$$\sum_{1\le i\le n} R_i\tilde{D}_i\frac{n_1(S_i)}{n(S_i)} = \sum_{1\le i\le n} R_i\tilde{D}_i\sum_{s\in\mathcal{S}} I\{S_i = s\}\frac{n_1(s)}{n(s)} = \sum_{s\in\mathcal{S}}\frac{n_1(s)}{n(s)}\sum_{1\le i\le n} R_i\tilde{D}_iI\{S_i = s\}$$
$$= \sum_{s\in\mathcal{S}}\frac{n_1(s)}{n(s)}\sum_{1\le i\le n} R_iD_iI\{S_i = s\} - \sum_{s\in\mathcal{S}}\frac{n_1(s)}{n(s)}\sum_{1\le i\le n} R_iI\{S_i = s\}\frac{n_1(S_i)}{n(S_i)}$$
$$= \sum_{s\in\mathcal{S}}\frac{n_1(s)}{n(s)}n_1(s) - \sum_{s\in\mathcal{S}}\frac{n_1(s)}{n(s)}\sum_{1\le i\le n} R_iI\{S_i = s\}\sum_{k\in\mathcal{S}} I\{S_i = k\}\frac{n_1(k)}{n(k)}$$
$$= \sum_{s\in\mathcal{S}}\frac{n_1(s)^2}{n(s)} - \sum_{s\in\mathcal{S}}\frac{n_1(s)}{n(s)}\sum_{1\le i\le n} R_iI\{S_i = s\}\frac{n_1(s)}{n(s)}$$

$$= \sum_{s \in \mathcal{S}} \frac{n_1(s)^2}{n(s)} - \sum_{s \in \mathcal{S}} \frac{n_1(s)}{n(s)} n(s) \frac{n_1(s)}{n(s)} = 0 \ .$$

The conclusion then follows from the continuous mapping theorem. ■

A.3 The Limiting Distribution of $\hat{\theta}_n$

Theorem A.1. *Suppose Q satisfies Assumption 2.1 (as well as $E[Y_i^2(d)] < \infty$) and Assumption 3.2 (as well as $E[Y_i^2(d)R_i(d)|X_i = x]$ is Lipschitz for $d \in \{0,1\}$), and the treatment assignment mechanism satisfies Assumptions 3.1, 3.3 as well as*

$$\frac{1}{n} \sum_{1 \leq j \leq n} ||X_{\pi(2j-1)} - X_{\pi(2j)}||^2 \xrightarrow{P} 0 \ .$$

Then, as $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta(Q)) \xrightarrow{d} N(0, \varsigma_{\text{mp}}^2) \ ,$$

where

$$\varsigma_{\text{mp}}^2 = \text{Var}[\tilde{Y}_i(1)] + \text{Var}[\tilde{Y}_i(0)] - \frac{1}{2} E[E[\tilde{Y}_i(1) + \tilde{Y}_i(0)|X_i]^2]$$

and

$$\tilde{Y}_i(d) = \frac{R_i(d)}{E[R_i(d)]} \left(Y_i(d) - \frac{E[Y_i(d)R_i(d)]}{E[R_i(d)]} \right)$$

for $d \in \{0,1\}$.

Remark A.1. Following arguments similar to those in Bai et al. (2023), we can construct a consistent estimator of ς_{mp}^2. To that end, consider the observed adjusted outcome defined as:

$$\hat{Y}_i = \frac{R_i}{\frac{1}{n} \sum_{1 \leq j \leq 2n} R_j I\{D_j = D_i\}} \left(Y_i - \frac{\frac{1}{n} \sum_{1 \leq j \leq 2n} Y_j I\{D_j = D_i\} R_j}{\frac{1}{n} \sum_{1 \leq j \leq 2n} I\{D_j = D_i\} R_j} \right) \ ,$$

We then propose the following variance estimator:

$$\hat{\upsilon}_n^2 = \hat{\tau}_n^2 - \frac{1}{2} \hat{\lambda}_n^2 \ , \tag{14}$$

where

$$\hat{\tau}_n^2 = \frac{1}{n} \sum_{1 \leq j \leq n} \left(\hat{Y}_{\pi(2j)} - \hat{Y}_{\pi(2j-1)} \right)^2$$

$$\hat{\lambda}_n^2 = \frac{2}{n} \sum_{1 \leq j \leq \lfloor n/2 \rfloor} \left(\hat{Y}_{\pi(4j-3)} - \hat{Y}_{\pi(4j-2)} \right) \left(\hat{Y}_{\pi(4j-1)} - \hat{Y}_{\pi(4j)} \right) (D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)}) \ .$$

It follows from similar arguments to those used in Bai et al. (2023) that under appropriate assumptions $\hat{\upsilon}_n^2 \xrightarrow{P} \varsigma_{\text{mp}}^2$. ■

PROOF OF THEOREM A.1. To begin, note

$$\hat{\theta}_n = \frac{\frac{1}{n}\sum_{1\le i\le 2n} Y_i(1)R_i(1)D_i}{\frac{1}{n}\sum_{1\le i\le 2n} R_i(1)D_i} - \frac{\frac{1}{n}\sum_{1\le i\le 2n} Y_i(0)R_i(0)(1-D_i)}{\frac{1}{n}\sum_{1\le i\le 2n} R_i(0)(1-D_i)} ~.$$

Next, note by Assumption 3.1 that

$$\sqrt{n}\Big(\frac{1}{n}\sum_{1\le i\le 2n} Y_i(1)R_i(1)D_i - E[Y_i(1)R_i(1)]\Big) = \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n}(Y_i(1)R_i(1)D_i - E[Y_i(1)R_i(1)]D_i)$$

and similarly for the other three terms. The desired conclusion then follows from Lemma A.1 together with an application of the delta method. In particular, for $g(x,y,z,w) = \frac{x}{y} - \frac{z}{w}$, observe that

$$\hat{\theta}_n = g\left(\frac{1}{n}\sum_{1\le i\le 2n} Y_i(1)R_i(1)D_i, \frac{1}{n}\sum_{1\le i\le 2n} R_i(1)D_i, \frac{1}{n}\sum_{1\le i\le 2n} Y_i(0)R_i(0)(1-D_i), \frac{1}{n}\sum_{1\le i\le 2n} R_i(0)(1-D_i)\right)$$

and the Jacobian is

$$D_g(x,y,z,w) = \Big(\frac{1}{y}, -\frac{x}{y^2}, -\frac{1}{w}, \frac{z}{w^2}\Big) ~.$$

Note by the laws of total variance and total covariance that $\mathbb{V}$ in Lemma A.1 is symmetric with entries

$$\begin{aligned}
\mathbb{V}_{11} &= \text{Var}[Y_i(1)R_i(1)] - \frac{1}{2}\text{Var}[E[Y_i(1)R_i(1)|X_i]] \\
\mathbb{V}_{12} &= \text{Cov}[Y_i(1)R_i(1), R_i(1)] - \frac{1}{2}\text{Cov}[E[Y_i(1)R_i(1)|X_i], E[R_i(1)|X_i]] \\
\mathbb{V}_{13} &= \frac{1}{2}\text{Cov}[E[Y_i(1)R_i(1)|X_i], E[Y_i(0)R_i(0)|X_i]] \\
\mathbb{V}_{14} &= \frac{1}{2}\text{Cov}[E[Y_i(1)R_i(1)|X_i], E[R_i(0)|X_i]] \\
\mathbb{V}_{22} &= \text{Var}[R_i(1)] - \frac{1}{2}\text{Var}[E[R_i(1)|X_i]] \\
\mathbb{V}_{23} &= \frac{1}{2}\text{Cov}[E[R_i(1)|X_i], E[Y_i(0)R_i(0)|X_i]] \\
\mathbb{V}_{24} &= \frac{1}{2}\text{Cov}[E[R_i(1)|X_i], E[R_i(0)|X_i]] \\
\mathbb{V}_{33} &= \text{Var}[Y_i(0)R_i(0)] - \frac{1}{2}\text{Var}[E[Y_i(0)R_i(0)|X_i]] \\
\mathbb{V}_{34} &= \text{Cov}[Y_i(0)R_i(0), R_i(0)] - \frac{1}{2}\text{Cov}[E[Y_i(0)R_i(0)|X_i], E[R_i(0)|X_i]] \\
\mathbb{V}_{44} &= \text{Var}[R_i(0)] - \frac{1}{2}\text{Var}[E[R_i(0)|X_i]] ~.
\end{aligned}$$

The conclusion of the theorem then follows from direct calculation. ∎

Lemma A.1. *Suppose Q satisfies Assumption 2.1 (as well as $E[Y_i^2(d)] < \infty$) and Assumption 3.2 (as well as $E[Y_i^2(d)R_i(d)|X_i = x]$ is Lipschitz for $d \in \{0,1\}$), and the treatment assignment mechanism* *satisfies Assumptions 3.1, 3.3 as well as*

$$\frac{1}{n} \sum_{1 \le j \le n} \|X_{\pi(2j-1)} - X_{\pi(2j)}\|^2 \xrightarrow{P} 0 \ . \tag{15}$$

Define

$$\begin{aligned}
\mathbb{L}_n^{\mathrm{YA1}} &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} (Y_i(1) R_i(1) D_i - E[Y_i(1) R_i(1)] D_i) \\
\mathbb{L}_n^{\mathrm{A1}} &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} (R_i(1) D_i - E[R_i(1)] D_i) \\
\mathbb{L}_n^{\mathrm{YA0}} &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} (Y_i(0) R_i(0)(1 - D_i) - E[Y_i(0) R_i(0)](1 - D_i)) \\
\mathbb{L}_n^{\mathrm{A0}} &= \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} (R_i(0)(1 - D_i) - E[R_i(0)](1 - D_i)) \ .
\end{aligned}$$

Then, as $n \to \infty$,

$$(\mathbb{L}_n^{\mathrm{YA1}}, \mathbb{L}_n^{\mathrm{A1}}, \mathbb{L}_n^{\mathrm{YA0}}, \mathbb{L}_n^{\mathrm{A0}})' \xrightarrow{d} N(0, \mathbb{V}) \ ,$$

where

$$\mathbb{V} = \mathbb{V}_1 + \mathbb{V}_2$$

for

$$\mathbb{V}_1 = \begin{pmatrix} \mathbb{V}_1^1 & 0 \\ 0 & \mathbb{V}_1^0 \end{pmatrix}$$

$$\begin{aligned}
\mathbb{V}_1^1 &= \begin{pmatrix} E[\mathrm{Var}[Y_i(1) R_i(1) | X_i]] & E[\mathrm{Cov}[Y_i(1) R_i(1), R_i(1) | X_i]] \\ E[\mathrm{Cov}[Y_i(1) R_i(1), R_i(1) | X_i]] & E[\mathrm{Var}[R_i(1) | X_i]] \end{pmatrix} \\
\mathbb{V}_1^0 &= \begin{pmatrix} E[\mathrm{Var}[Y_i(0) R_i(0) | X_i]] & E[\mathrm{Cov}[Y_i(0) R_i(0), R_i(0) | X_i]] \\ E[\mathrm{Cov}[Y_i(0) R_i(0), R_i(0) | X_i]] & E[\mathrm{Var}[R_i(0) | X_i]] \end{pmatrix}
\end{aligned}$$

$$\mathbb{V}_2 = \frac{1}{2} \mathrm{Var}[(E[Y_i(1) R_i(1) | X_i], E[R_i(1) | X_i], E[Y_i(0) R_i(0) | X_i], E[R_i(0) | X_i])'] \ .$$

Proof of Lemma A.1. Note

$$(\mathbb{L}_n^{\mathrm{YA1}}, \mathbb{L}_n^{\mathrm{A1}}, \mathbb{L}_n^{\mathrm{YA0}}, \mathbb{L}_n^{\mathrm{A0}}) = (\mathbb{L}_{1,n}^{\mathrm{YA1}}, \mathbb{L}_{1,n}^{\mathrm{A1}}, \mathbb{L}_{1,n}^{\mathrm{YA0}}, \mathbb{L}_{1,n}^{\mathrm{A0}}) + (\mathbb{L}_{2,n}^{\mathrm{YA1}}, \mathbb{L}_{2,n}^{\mathrm{A1}}, \mathbb{L}_{2,n}^{\mathrm{YA0}}, \mathbb{L}_{2,n}^{\mathrm{A0}}) \ ,$$

where

$$\mathbb{L}_{1,n}^{\mathrm{YA1}} = \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} (Y_i(1) R_i(1) D_i - E[Y_i(1) R_i(1) D_i | X^{(n)}, D^{(n)}])$$

$$\mathbb{L}_{2,n}^{\text{YA1}} = \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} (E[Y_i(1)R_i(1)D_i|X^{(n)}, D^{(n)}] - E[Y_i(1)R_i(1)]D_i)$$

and similarly for the rest. Next, note $(\mathbb{L}_{1,n}^{\text{YA1}}, \mathbb{L}_{1,n}^{\text{A1}}, \mathbb{L}_{1,n}^{\text{YA0}}, \mathbb{L}_{1,n}^{\text{A0}}), n \ge 1$ is a triangular array of normalized sums of random vectors. We will apply the Lindeberg central limit theorem for random vectors, i.e., Proposition 2.27 of van der Vaart (1998), to this triangular array. Conditional on $X^{(n)}, D^{(n)}$, $(\mathbb{L}_{1,n}^{\text{YA1}}, \mathbb{L}_{1,n}^{\text{A1}}) \perp\!\!\!\perp (\mathbb{L}_{1,n}^{\text{YA0}}, \mathbb{L}_{1,n}^{\text{A0}})$. Moreover, it follows from $Q_n = Q^{2n}$ and Assumption 3.1 that

$$\begin{aligned} &\text{Var}\left[\begin{pmatrix} \mathbb{L}_{1,n}^{\text{YA1}} \\ \mathbb{L}_{1,n}^{\text{A1}} \end{pmatrix} \middle| X^{(n)}, D^{(n)} \right] \\ &\quad = \begin{pmatrix} \frac{1}{n}\sum_{1 \le i \le 2n} \text{Var}[Y_i(1)R_i(1)|X_i]D_i & \frac{1}{n}\sum_{1 \le i \le 2n} \text{Cov}[Y_i(1)R_i(1), R_i(1)|X_i]D_i \\ \frac{1}{n}\sum_{1 \le i \le 2n} \text{Cov}[Y_i(1)R_i(1), R_i(1)|X_i]D_i & \frac{1}{n}\sum_{1 \le i \le 2n} \text{Var}[R_i(1)|X_i]D_i \end{pmatrix} . \end{aligned}$$

For the upper left component, we have

$$\frac{1}{n} \sum_{1 \le i \le 2n} \text{Var}[Y_i(1)R_i(1)|X_i]D_i = \frac{1}{n} \sum_{1 \le i \le 2n} E[Y_i^2(1)R_i(1)|X_i]D_i - \frac{1}{n} \sum_{1 \le i \le 2n} E[Y_i(1)R_i(1)|X_i]^2 D_i . \tag{16}$$

Note

$$\begin{aligned} &\frac{1}{n} \sum_{1 \le i \le 2n} E[Y_i^2(1)R_i(1)|X_i]D_i \\ &= \frac{1}{2n} \sum_{1 \le i \le 2n} E[Y_i^2(1)R_i(1)|X_i] + \frac{1}{2}\Big(\frac{1}{n} \sum_{1 \le i \le 2n : D_i = 1} E[Y_i^2(1)R_i(1)|X_i] \\ &\qquad - \frac{1}{n} \sum_{1 \le i \le 2n : D_i = 0} E[Y_i^2(1)R_i(1)|X_i]\Big) . \end{aligned}$$

It follows from the weak law of large numbers, the application of which is permitted by $E[Y_i^2(1)] < \infty$ and the fact that $R_i(1) \in \{0, 1\}$, that

$$\frac{1}{2n} \sum_{1 \le i \le 2n} E[Y_i^2(1)R_i(1)|X_i] \xrightarrow{P} E[Y_i^2(1)R_i(1)] .$$

On the other hand, it follows from Assumption 3.2 and 3.3 that

$$\begin{aligned} &\Big|\frac{1}{n} \sum_{1 \le i \le 2n : D_i = 1} E[Y_i^2(1)R_i(1)|X_i] - \frac{1}{n} \sum_{1 \le i \le 2n : D_i = 0} E[Y_i^2(1)R_i(1)|X_i]\Big| \\ &\le \frac{1}{n} \sum_{1 \le j \le n} |E[Y_{\pi(2j-1)}^2(1)A_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] - E[Y_{\pi(2j)}^2(1)A_{\pi(2j)}(1)|X_{\pi(2j)}]| \\ &\lesssim \frac{1}{n} \sum_{1 \le j \le n} \|X_{\pi(2j-1)} - X_{\pi(2j)}\| = o_P(1) . \end{aligned}$$

Therefore,

$$\frac{1}{n}\sum_{1\le i\le 2n} E[Y_i^2(1)R_i(1)|X_i]D_i \xrightarrow{P} E[Y_i^2(1)R_i(1)] \ .$$

Meanwhile,

$$\begin{aligned}&\frac{1}{n}\sum_{1\le i\le 2n} E[Y_i(1)R_i(1)|X_i]^2 D_i \\ &= \frac{1}{2n}\sum_{1\le i\le 2n} E[Y_i(1)R_i(1)|X_i]^2 + \frac{1}{2}\Big(\frac{1}{n}\sum_{1\le i\le 2n: D_i=1} E[Y_i(1)R_i(1)|X_i]^2 \\ &\qquad - \frac{1}{n}\sum_{1\le i\le 2n: D_i=0} E[Y_i(1)R_i(1)|X_i]^2\Big) \ .\end{aligned}$$

Jensen's inequality implies $E[E[Y_i(1)R_i(1)|X_i]^2] \le E[Y_i^2(1)R_i(1)] < E[Y_i^2(1)] < \infty$, so it follows from the weak law of large numbers as above that

$$\frac{1}{2n}\sum_{1\le i\le 2n} E[Y_i(1)R_i(1)|X_i]^2 \xrightarrow{P} E[E[Y_i(1)R_i(1)|X_i]^2] \ .$$

Next,

$$\begin{aligned}&\Big|\frac{1}{n}\sum_{1\le i\le 2n: D_i=1} E[Y_i(1)R_i(1)|X_i]^2 - \frac{1}{n}\sum_{1\le i\le 2n: D_i=0} E[Y_i(1)R_i(1)|X_i]^2\Big| \\ &\le \frac{1}{n}\sum_{1\le j\le n} |E[Y_{\pi(2j-1)}(1)A_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] - E[Y_{\pi(2j)}(1)A_{\pi(2j)}(1)|X_{\pi(2j)}]| \\ &\qquad \times |E[Y_{\pi(2j-1)}(1)A_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] + E[Y_{\pi(2j)}(1)A_{\pi(2j)}(1)|X_{\pi(2j)}]| \\ &\lesssim \Big(\frac{1}{n}\sum_{1\le j\le n} \|X_{\pi(2j-1)} - X_{\pi(2j)}\|^2\Big)^{1/2} \\ &\qquad \times \Big(\frac{1}{n}\sum_{1\le j\le n} |E[Y_{\pi(2j-1)}(1)A_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] + E[Y_{\pi(2j)}(1)A_{\pi(2j)}(1)|X_{\pi(2j)}]|^2\Big)^{1/2} \\ &\lesssim \Big(\frac{1}{n}\sum_{1\le j\le n} \|X_{\pi(2j-1)} - X_{\pi(2j)}\|^2\Big)^{1/2} \\ &\qquad \times \Big(\frac{1}{n}\sum_{1\le j\le n} (|E[Y_{\pi(2j-1)}(1)A_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]|^2 + |E[Y_{\pi(2j)}(1)A_{\pi(2j)}(1)|X_{\pi(2j)}]|^2)\Big)^{1/2} \\ &\le \Big(\frac{1}{n}\sum_{1\le j\le n} \|X_{\pi(2j-1)} - X_{\pi(2j)}\|^2\Big)^{1/2}\Big(\frac{1}{n}\sum_{1\le i\le 2n} E[Y_i(1)R_i(1)|X_i]^2\Big)^{1/2} = o_P(1) \ ,\end{aligned}$$

where the first inequality follows by inspection, the second follows from Assumption 3.2 and the Cauchy-Schwarz inequality, the third follows from $(a+b)^2 \le 2a^2 + 2b^2$, the last follows by inspection again, and the convergence in probability follows from (15). Therefore, it follows from (16) that

$$\frac{1}{n}\sum_{1\le i\le 2n} \mathrm{Var}[Y_i(1)R_i(1)|X_i]D_i \xrightarrow{P} E[\mathrm{Var}[Y_i(1)R_i(1)|X_i]] \ .$$

Similar arguments imply that

$$\text{Var}\left[\begin{pmatrix}\mathbb{L}_{1,n}^{\text{YA1}} \\ \mathbb{L}_{1,n}^{\text{A1}}\end{pmatrix}\middle| X^{(n)}, D^{(n)}\right] \xrightarrow{P} \mathbb{V}_1^1 \ .$$

Similarly,

$$\text{Var}\left[\begin{pmatrix}\mathbb{L}_{1,n}^{\text{YA0}} \\ \mathbb{L}_{1,n}^{\text{A0}}\end{pmatrix}\middle| X^{(n)}, D^{(n)}\right] \xrightarrow{P} \mathbb{V}_1^0 \ .$$

If $E[\text{Var}[Y_i(1)R_i(1)|X_i]] = E[\text{Var}[R_i(1)|X_i]] = E[\text{Var}[Y_i(0)R_i(0)|X_i]] = E[\text{Var}[R_i(0)|X_i]] = 0$, then it follows from Markov's inequality conditional on $X^{(n)}$ and $D^{(n)}$, and the fact that probabilities are bounded and hence uniformly integrable, that $(\mathbb{L}_{1,n}^{\text{YA1}}, \mathbb{L}_{1,n}^{\text{A1}}, \mathbb{L}_{1,n}^{\text{YA0}}, \mathbb{L}_{1,n}^{\text{A0}}) \xrightarrow{P} 0$. Otherwise, it follows from similar arguments to those in the proof of Lemma S.1.5 of Bai et al. (2021) that

$$\rho(\mathcal{L}((\mathbb{L}_{1,n}^{\text{YA1}}, \mathbb{L}_{1,n}^{\text{A1}}, \mathbb{L}_{1,n}^{\text{YA0}}, \mathbb{L}_{1,n}^{\text{A0}})'|X^{(n)}, D^{(n)}), N(0, \mathbb{V}_1)) \xrightarrow{P} 0 \ , \tag{17}$$

where $\mathcal{L}$ denotes the distribution and ρ is any metric that metrizes weak convergence.

Next, we study $(\mathbb{L}_{2,n}^{\text{YA1}}, \mathbb{L}_{2,n}^{\text{A1}}, \mathbb{L}_{2,n}^{\text{YA0}}, \mathbb{L}_{2,n}^{\text{A0}})$. It follows from $Q_n = Q^{2n}$ and Assumption 3.1 that

$$\begin{pmatrix}\mathbb{L}_{2,n}^{\text{YA1}} \\ \mathbb{L}_{2,n}^{\text{A1}} \\ \mathbb{L}_{2,n}^{\text{YA0}} \\ \mathbb{L}_{2,n}^{\text{A0}}\end{pmatrix} = \begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{1\le i\le 2n} D_i(E[Y_i(1)R_i(1)|X_i] - E[Y_i(1)R_i(1)]) \\ \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n} D_i(E[R_i(1)|X_i] - E[R_i(1)]) \\ \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n}(1-D_i)(E[Y_i(0)R_i(0)|X_i] - E[Y_i(0)R_i(0)]) \\ \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n}(1-D_i)(E[R_i(0)|X_i] - E[R_i(0)])\end{pmatrix} .$$

For $\mathbb{L}_{2,n}^{\text{YA1}}$, note it follows from Assumptions 3.1, 3.2 and (15) that

$$\begin{aligned}\text{Var}[\mathbb{L}_{2,n}^{\text{YA1}}|X^{(n)}] &= \frac{1}{4n}\sum_{1\le j\le n}(E[Y_{\pi(2j-1)}(1)A_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] - E[Y_{\pi(2j)}(1)A_{\pi(2j)}(1)|X_{\pi(2j)}])^2 \\ &\lesssim \frac{1}{n}\sum_{1\le j\le n}\|X_{\pi(2j-1)} - X_{\pi(2j)}\|^2 \xrightarrow{P} 0 \ .\end{aligned}$$

Therefore, it follows from Markov's inequality conditional on $X^{(n)}$ and $D^{(n)}$, and the fact that probabilities are bounded and hence uniformly integrable, that

$$\mathbb{L}_{2,n}^{\text{YA1}} = E[\mathbb{L}_{2,n}^{\text{YA1}}|X^{(n)}] + o_P(1) \ .$$

Similarly,

$$\begin{pmatrix} \mathbb{L}_{2,n}^{\mathrm{YA1}} \\ \mathbb{L}_{2,n}^{\mathrm{A1}} \\ \mathbb{L}_{2,n}^{\mathrm{YA0}} \\ \mathbb{L}_{2,n}^{\mathrm{A0}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2\sqrt{n}} \sum_{1 \le i \le 2n} (E[Y_i(1)R_i(1)|X_i] - E[Y_i(1)R_i(1)]) \\ \frac{1}{2\sqrt{n}} \sum_{1 \le i \le 2n} (E[R_i(1)|X_i] - E[R_i(1)]) \\ \frac{1}{2\sqrt{n}} \sum_{1 \le i \le 2n} (E[Y_i(0)R_i(0)|X_i] - E[Y_i(0)R_i(0)]) \\ \frac{1}{2\sqrt{n}} \sum_{1 \le i \le 2n} (E[R_i(0)|X_i] - E[R_i(0)]) \end{pmatrix} + o_P(1) \ .$$

It then follows from Assumption 2.1and the central limit theorem that

$$(\mathbb{L}_{2,n}^{\mathrm{YA1}}, \mathbb{L}_{2,n}^{\mathrm{A1}}, \mathbb{L}_{2,n}^{\mathrm{YA0}}, \mathbb{L}_{2,n}^{\mathrm{A0}})' \xrightarrow{d} N(0, \mathbb{V}_2) \ .$$

Because (17) holds and $(\mathbb{L}_{2,n}^{\mathrm{YA1}}, \mathbb{L}_{2,n}^{\mathrm{A1}}, \mathbb{L}_{2,n}^{\mathrm{YA0}}, \mathbb{L}_{2,n}^{\mathrm{A0}})$ is deterministic conditional on $X^{(n)}, D^{(n)}$, the conclusion of the theorem follows from Lemma S.1.3 in Bai et al. (2021). ■

A.4 A Numerical Example

Let $X \sim N(0,1)$ and $\epsilon = (\epsilon_Y(1), \epsilon_Y(0), \epsilon_R(1), \epsilon_R(0))' \sim N(0, \Sigma)$, where the diagonal elements of Σ are 1 and all off-diagonal elements are -0.3. Suppose for $d \in \{0, 1\}$,

$$\begin{aligned} Y(d) &= \mu_d(X) + \epsilon_Y(d) \\ R(d) &= I\{\epsilon_R(d) \le \nu_d(X)\} \ , \end{aligned}$$

with $\mu_d(x)$ and $\nu_d(x)$ specified below. In the following two examples, the values of θ can be calculated by hand, and the values of θ^{obs} and θ^{drop} are computed via simulation with $n = 10^6$ random draws.

1. $\mu_1(x) = 2x$, $\mu_0(x) = x^3$, $\nu_1(x) = x$, $\nu_0(x) = x^2$. In this example, $\theta = 0$, $\theta^{\mathrm{obs}} \approx 1.17$, $\theta^{\mathrm{drop}} \approx -0.50$.

2. $\mu_1(x) = 2x$, $\mu_0(x) = x$, $\nu_1(x) = x$, $\nu_0(x) = x$. In this example, $\theta = 0$, $\theta^{\mathrm{obs}} \approx 0.56$, $\theta^{\mathrm{drop}} \approx 0.86$.

A.5 Additional Details for Empirical Survey in Section 4.2

Table 3: Additional notes about each paper used in Figure 1

Paper	Table Replicated	Additional Notes
Dhar et al. (2022)	Table 2: (1), (2) and (3)	Original specification features controls. Original estimates do not include strata fixed-effects.
Carter et al. (2021)	Figure 2: left panel ("Direct impact on treatment group")	Original specification features controls. Original estimates include strata fixed-effects. We reported both "During" and "After" estimates.
Casaburi and Reed (2022)	Table 2: (1)	Original specification does not feature controls. Original estimate includes strata fixed-effects.
Abebe et al. (2021)	Table 2, Table 3 (Column 1)	Original specification does not feature controls. Original estimates include strata fixed-effects.
Hjort et al. (2021)	Online Appendix Table A.11: (1)	Original specification does not feature controls. Original estimate does not include strata fixed-effects. This is an intent-to-treat specification.
Romero et al. (2020)	Table 3: (4)	Original specification does not feature controls. Original estimates include pair fixed-effects. These are intent-to-treat specifications.
Attanasio et al. (2020)	Table 4: Second Column	Original specification features controls. Original estimate does not include strata fixed-effects. The first column of Table 4 is estimated using a probit regression and thus is not reproduced.

Notes: For each paper considered in Section 4.2, we list the corresponding table/figure and specification(s) replicated in the second column. We include relevant notes for each application in the third column.

A.6 Details for Equation (6)

Let $\tilde{\theta}_n^{\text{drop}}$ denote the OLS estimator of θ^{drop} in (6) using only observations with $R_i = 1$. By construction, the jth entry of the OLS estimator of the projection coefficient of D_i on the pair fixed effects is given by

$$\left(\sum_{1\leq i\leq n:R_i=1} I\{i\in\{\pi(2j-1),\pi(2j)\}\right)^{-1}\sum_{1\leq i\leq n:R_i=1} D_i I\{i\in\{\pi(2j-1),\pi(2j)\}\}~. \tag{18}$$

Let $\tilde{D}_i$ denote the residual from the projection of D_i on the pair fixed effects. Fix $1\leq j\leq n$. If $R_{\pi(2j-1)}=R_{\pi(2j)}=1$, then it follows from (18) that

$$\tilde{D}_{\pi(2j)}=\frac{1}{2}\left(D_{\pi(2j)}-D_{\pi(2j-1)}\right)~,$$

$$\tilde{D}_{\pi(2j-1)}=\frac{1}{2}\left(D_{\pi(2j-1)}-D_{\pi(2j)}\right)~.$$

Next suppose the jth pair contains only one attrited unit. Without loss of generality, assume $R_{\pi(2j-1)}=0$ and $R_{\pi(2j)}=1$. It then follows from (18) that

$$\tilde{D}_{\pi(2j)}=D_{\pi(2j)}-D_{\pi(2j)}=0~.$$

By an application of the Frisch-Waugh-Lovell theorem we can thus conclude that $\tilde{\theta}_n^{\rm drop}=\hat{\theta}_n^{\rm drop}$, as desired.

A.7 Relevant Excerpts from Referenced Sources

Donner and Klar (2000) chapter 3, page 40:

"A final disadvantage of the matched pair design is that the loss to follow-up of a single cluster in a pair implies that both clusters in that pair must effectively be discarded from the trial, at least with respect to testing the effect of intervention. This problem [...] clearly does not arise if there is some replication of clusters within each combination of intervention and stratum."

King et al. (2007) page 490:

"The key additional advantage of the matched pair design from our perspective is that it enables us to protect ourselves, to a degree, from selection bias that could otherwise occur with the loss of clusters. In particular, if we lose a cluster for a reason related to one or more of the variables we matched on [...] then no bias would be induced for the remaining clusters. That is, whether we delete or impute the remaining member of the pair that suffered a loss of a cluster under these circumstances, the set of all remaining pairs in the study would still be as balanced—matched on observed background characteristics and randomized within pairs—as the original full data set. Thus, any variable we can measure and match on when creating pairs removes a potential for selection bias if later on we lose a cluster due to a reason related to that variable. [...] Classical randomization, which does not match on any variables, lacks this protective property."

Bruhn and McKenzie (2009) page 209:

"King et al. (2007) emphasize one additional advantage in the context of social science experiments when the matched pairs occur at the level of a community, village, or school, which is that it provides partial protection against political interference or drop-out. If a unit drops out of the study [...] its pair unit can also be dropped from the study, while the set of remaining pairs will still be as balanced as the original dataset. In contrast, in a pure randomized experiment, if even one unit drops out, it is no longer guaranteed that the treatment and control groups are balanced, on average."

Glennerster and Takavarasha (2013) chapter 4, page 159:

"In paired matching, for example, if we lose one of the units in the pair [...] and we include a dummy for the stratum, essentially we have to drop the other unit in the pair from the analysis. [...] Some evaluators have mistakenly seen this as an advantage of pairing [...] But in fact if we drop the pair we have just introduced even more attrition bias. [...] Our suggestion is that if there is a risk of attrition [...] use strata that have at least four units rather than pairwise randomization."