Supplemental to "Inference in Experiments with Matched Pairs"*

Yuehao Bai
Department of Economics
University of Michigan
yuehaob@umich.edu

Joseph P. Romano
Departments of Economics and Statistics
Stanford University
romano@stanford.edu

Azeem M. Shaikh
Department of Economics
University of Chicago
amshaikh@uchicago.edu

January 15, 2021

Abstract

This document provides proofs for all results for the authors' paper "Inference in Experiments with Matched Pairs," details for Remark 3.8, as well as some additional simulations.

KEYWORDS: Experiment, matched pairs, matched pairs t-test, permutation test, randomized controlled trial, treatment assignment, two-sample t-test

JEL classification codes: C12, C14

*The research of the third author is supported by NSF Grant SES-1530661.

S.1 Appendix

Please note that in what follows we will use the notation $a \lesssim b$ to denote $a \leq cb$ for some constant c.

S.1.1 Additional details for Remark 3.8

To streamline our exposition, it is helpful to introduce the following shorthand notation:

$$\begin{aligned}
\Delta_{Y,j} &= (Y_{\pi(2j)} - Y_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)}) \\
\Delta_{X,j} &= (X_{\pi(2j)} - X_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)}) \\
\hat{\gamma}_n &= \frac{1}{n} \sum_{1 \leq j \leq n} (\Delta_{Y,j} - \hat{\Delta}_n) \Delta_{X,j} \\
V_n &= \frac{1}{n} \sum_{1 \leq j \leq n} \Delta_{X,j} \Delta_{X,j}' \\
\mu_n &= \frac{1}{n} \sum_{1 \leq j \leq n} \Delta_{X,j} \ .
\end{aligned}$$

Using this notation,

$$\hat{\alpha}_n = \hat{\Delta}_n - \mu_n' \hat{\beta}_n \ ,$$

where $\hat{\beta}_n = \Omega_n^{-1} \hat{\gamma}_n$ with $\Omega_n = V_n - \mu_n \mu_n'$. Hence,

$$\sqrt{n}(\hat{\alpha}_n - \Delta(Q)) = \sqrt{n}(\hat{\Delta}_n - \Delta(Q)) - \sqrt{n} \mu_n' \hat{\beta}_n \ .$$

In order to establish (18), it therefore suffices to show that

$$\sqrt{n} \mu_n' \hat{\beta}_n = o_P(1) \ .$$

Using the Sherman-Morrison-Woodbury formula, we have that

$$\Omega_n^{-1} = V_n^{-1} + \frac{V_n^{-1} \mu_n \mu_n' V_n^{-1}}{1 - \mu_n' V_n^{-1} \mu_n} \ .$$

It follows that

$$\begin{aligned}
\sqrt{n} \mu_n' \hat{\beta}_n &= \sqrt{n} \mu_n' V_n^{-1} \hat{\gamma}_n + \frac{\sqrt{n} \mu_n' V_n^{-1} \mu_n \mu_n' V_n^{-1} \hat{\gamma}_n}{1 - \mu_n' V_n^{-1} \mu_n} \\
&= \sqrt{n} (V_n^{-\frac{1}{2}} \mu_n)' V_n^{-\frac{1}{2}} \hat{\gamma}_n + \frac{\sqrt{n} (V_n^{-\frac{1}{2}} \mu_n)' (V_n^{-\frac{1}{2}} \mu_n)(V_n^{-\frac{1}{2}} \mu_n)' (V_n^{-\frac{1}{2}} \hat{\gamma}_n)}{1 - (V_n^{-\frac{1}{2}} \mu_n)'(V_n^{-\frac{1}{2}} \mu_n)} \ .
\end{aligned}$$

It therefore suffices to show that

$$\begin{aligned}
\sqrt{n} V_n^{-\frac{1}{2}} \mu_n &= O_P(1) && \text{(S.1)} \\
V_n^{-\frac{1}{2}} \hat{\gamma}_n &= o_P(1) \ . && \text{(S.2)}
\end{aligned}$$

In order to establish (S.1), note $E[\sqrt{n} V_n^{-\frac{1}{2}} \mu_n | X^{(n)}] = 0$ and $\mathrm{Var}[\sqrt{n} V_n^{-\frac{1}{2}} \mu_n | X^{(n)}] = I_k$, where I_k is the identity matrix and $k = \dim(X_i)$. Now (S.1) follows immediately. In order to establish (S.2), note that

$$\begin{aligned}
V_n^{-\frac{1}{2}} \hat{\gamma}_n &= V_n^{-\frac{1}{2}} \frac{1}{n} \sum_{1 \leq j \leq n} \Delta_{Y,j} \Delta_{X,j} + \hat{\Delta}_n V_n^{-\frac{1}{2}} \mu_n \\
&= V_n^{-\frac{1}{2}} \frac{1}{n} \sum_{1 \leq j \leq n} \Delta_{Y,j} \Delta_{X,j} + o_P(1) \ ,
\end{aligned}$$

where in the final equality we make use of Lemma S.1.5 in the Supplemental Appendix and (S.1). To argue that

$$V_n^{-\frac{1}{2}} \frac{1}{n} \sum_{1 \leq j \leq n} \Delta_{Y,j} \Delta_{X,j} = o_P(1) \ ,$$

first use Assumption 2.1(c) and the Cauchy-Schwartz inequality to obtain

$$\left| E\left[V_n^{-\frac{1}{2}} \frac{1}{n} \sum_{1 \le j \le n} \Delta_{Y,j} \Delta_{X,j} \middle| X^{(n)} \right] \right| \lesssim \left(\frac{1}{n} \sum_{1 \le j \le n} |\Delta_{X,j}|^2 \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{1 \le j \le n} \Delta_{X,j}' V_n^{-1} \Delta_{X,j} \right)^{\frac{1}{2}} .$$

Using properties of the trace, it follows that

$$\frac{1}{n} \sum_{1 \le j \le n} \Delta_{X,j}' V_n^{-1} \Delta_{X,j} = k . \tag{S.3}$$

It therefore follows from Assumption 2.3 that

$$E\left[V_n^{-\frac{1}{2}} \frac{1}{n} \sum_{1 \le j \le n} \Delta_{Y,j} \Delta_{X,j} \middle| X^{(n)} \right] \xrightarrow{P} 0 .$$

To complete the establishment of (S.2), it is helpful to assume further that $\text{Var}[\Delta_{Y,j} | X^{(n)}]$ is bounded uniformly in $1 \le j \le n$ (as would be the case if, for example, the support of the potential outcomes were bounded). With this assumption, we have that

$$\left| \text{Var}\left[V_n^{-\frac{1}{2}} \frac{1}{n} \sum_{1 \le j \le n} \Delta_{Y,j} \Delta_{X,j} \middle| X^{(n)} \right] \right| \lesssim \frac{1}{n^2} \sum_{1 \le j \le n} \Delta_{X,j}' V_n^{-1} \Delta_{X,j} ,$$

which tends to zero from (S.3).

S.1.2 Proof of Theorem 3.1

The theorem follows immediately upon noting that (9) follows from Lemmas S.1.4–S.1.5 below. ■

S.1.3 Proof of Theorem 3.2

The theorem follows immediately upon noting that (14) follows from Lemmas S.1.4–S.1.5 and S.1.6 below. ■

S.1.4 Proof of Theorem 3.3

From Lemma S.1.4, we see that it suffices to show that $\hat{\nu}_n^2$ defined in (20) tends in probability to (S.16). Since

$$\begin{aligned} & E[\text{Var}[Y_i(1)|X_i]] + E[\text{Var}[Y_i(0)|X_i]] + \frac{1}{2} E\left[((E[Y_i(1)|X_i] - E[Y_i(1)]) - (E[Y_i(0)|X_i] - E[Y_i(0)]))^2 \right] \\ = \; & E[\text{Var}[Y_i(1)|X_i]] + E[\text{Var}[Y_i(0)|X_i]] + \frac{1}{2} \left(E\left[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2 \right] - (E[Y_i(1)] - E[Y_i(0)])^2 \right) . \end{aligned}$$

the desired conclusion follows immediately from Lemmas S.1.5–S.1.7. ■

S.1.5 Proof of Theorem 3.4

Let Q satisfying (31) be given. For such a Q, we first argue that

$$gZ^{(n)} | X^{(n)} \stackrel{d}{=} Z^{(n)} | X^{(n)} . \tag{S.4}$$

Since $\pi = \pi_n(X^{(n)})$, we have from Assumption 2.2 that

$$gD^{(n)} | X^{(n)} \stackrel{d}{=} D^{(n)} | X^{(n)} . \tag{S.5}$$

Furthermore,

$$Y^{(n)} \perp\!\!\!\perp D^{(n)} | X^{(n)} . \tag{S.6}$$

To see this, note for any set A and any d and d' in the support of $D^{(n)}|X^{(n)}$ that

$$\begin{aligned}
P\{Y^{(n)} \in A|D^{(n)} = (d_1, \dots d_{2n}), X^{(n)}\} &= P\{(Y_1(d_1), \dots Y_{2n}(d_{2n})) \in A|D^{(n)} = (d_1, \dots d_{2n}), X^{(n)}\} \\
&= P\{(Y_1(d_1), \dots Y_{2n}(d_{2n})) \in A|X^{(n)}\} \\
&= P\{(Y_1(d_1'), \dots Y_{2n}(d_{2n}')) \in A|X^{(n)}\} \\
&= P\{(Y_1(d_1'), \dots Y_{2n}(d_{2n}')) \in A|D^{(n)} = (d_1', \dots d_{2n}'), X^{(n)}\} \\
&= P\{Y^{(n)} \in A|D^{(n)} = (d_1', \dots d_{2n}'), X^{(n)}\} ,
\end{aligned}$$

where the first and fifth equalities follow from (1), the second and fourth equalities follow from $(Y^{(n)}(1), Y^{(n)}(0)) \perp\!\!\!\perp D^{(n)} \big| X^{(n)}$, the third follows from the fact that Q satisfies (31). It now follows from (S.5) and (S.6) that (S.4) holds.

Next, observe that

$$\begin{aligned}
E\left[\sum_{g \in \mathbf{G}_n(\pi)} \phi_n^{\text{rand}}(gZ^{(n)})\right] &= E\left[E\left[\sum_{g \in \mathbf{G}_n(\pi)} \phi_n^{\text{rand}}(gZ^{(n)}) \Big| X^{(n)}\right]\right] \\
&= E\left[\sum_{g \in \mathbf{G}_n(\pi)} E\left[\phi_n^{\text{rand}}(Z^{(n)}) \Big| X^{(n)}\right]\right] \\
&= E\left[2^n E\left[\phi_n^{\text{rand}}(Z^{(n)}) \Big| X^{(n)}\right]\right] \\
&= 2^n E\left[\phi_n^{\text{rand}}(Z^{(n)})\right] ,
\end{aligned} \tag{S.7}$$

where the first and final equalities follow from the law of iterated expectations, the second follows from (S.4), and the third exploits the fact that $|\mathbf{G}_n(\pi)| = 2^n$. Using the fact that $\mathbf{G}_n(\pi)$ is a group, we have with probability one that

$$\sum_{g \in \mathbf{G}_n(\pi)} \phi_n^{\text{rand}}(gZ^{(n)}) \leq 2^n \alpha .$$

Hence,

$$E\left[\sum_{g \in \mathbf{G}_n(\pi)} \phi_n^{\text{rand}}(gZ^{(n)})\right] \leq 2^n \alpha . \tag{S.8}$$

Combining (S.7) and (S.8), we see that (32) holds, as desired. ■

S.1.6 Proof of Theorem 3.5

Note that

$$\hat{\Delta}_n = \frac{1}{n} \sum_{1 \leq j \leq n} (Y_{\pi(2j)} - Y_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)}) .$$

This observation, together with the definition of $\hat{\nu}_n$ in (20), implies that

$$\hat{R}_n(t) = P\left\{ \frac{\left|\frac{1}{\sqrt{n}} \sum_{1 \leq j \leq n} \epsilon_j (Y_{\pi(2j)} - Y_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)})\right|}{\check{\nu}_n(\epsilon_1, \dots, \epsilon_n)} \leq t \Big| W^{(n)} \right\} ,$$

where, independently of $W^{(n)}$, $\epsilon_j, j = 1, \dots, n$ are i.i.d. Rademacher random variables and $\check{\nu}_n^2$ is defined as in (S.41). Note further that

$$\hat{R}_n(t) = \check{R}_n(t) - \check{R}_n(-t) ,$$

where

$$\check{R}_n(t) = P\left\{ \frac{\frac{1}{\sqrt{n}} \sum_{1 \leq j \leq n} \epsilon_j (Y_{\pi(2j)} - Y_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)})}{\check{\nu}_n(\epsilon_1, \dots, \epsilon_n)} \leq t \Big| W^{(n)} \right\} .$$

The desired conclusion now follows immediately from Lemmas S.1.8–S.1.9 together with Theorem 5.2 of Chung and Romano (2013). ■

S.1.7 Proof of Theorem 4.1

For $1 \leq i \leq 2n$, let $U_i = |X_i|$ and write $U_{(1)} \leq \cdots \leq U_{(2n)}$. Note that

$$\begin{aligned}
\frac{1}{n} \sum_{1 \leq j \leq n} |X_{\pi(2j)} - X_{\pi(2j-1)}| &= \frac{1}{n} \sum_{1 \leq j \leq n} (X_{\pi(2j)} - X_{\pi(2j-1)}) \\
&\leq \frac{1}{n}(X_{\pi(2n)} - X_{\pi(1)}) \\
&\leq \frac{1}{n} 2U_{(2n)} \\
&\xrightarrow{P} 0 ,
\end{aligned}$$

where the equality exploits the fact that $X_{\pi(2j-1)} \leq X_{\pi(2j)}$, the two inequalities follow by inspection, and the convergence in probability to zero follows from Lemma S.1.1. Similarly,

$$\begin{aligned}
\frac{1}{n} \sum_{1 \leq j \leq n} |X_{\pi(2j)} - X_{\pi(2j-1)}|^2 &\leq |X_{\pi(2n)} - X_{\pi(1)}| \left(\frac{1}{n} \sum_{1 \leq j \leq n} |X_{\pi(2j)} - X_{\pi(2j-1)}| \right) \\
&\leq \left(\frac{U_{(2n)}}{\sqrt{n}} \right)^2 \\
&\xrightarrow{P} 0 ,
\end{aligned}$$

where the first inequality follows by inspection, the second follows by arguing as before, and the convergence in probability to zero again follows from Lemma S.1.1. Finally, for any $k \in \{2, 3\}$ and $\ell \in \{0, 1\}$, we have that

$$\begin{aligned}
\frac{2}{n} \sum_{1 \leq j \leq \frac{n}{2}} |X_{\pi(4j-k)} - X_{\pi(4j-\ell)}|^2 &\leq \frac{2}{n} \sum_{1 \leq j \leq \frac{n}{2}} |X_{\pi(4j-3)} - X_{\pi(4j)}|^2 \\
&\leq |X_{\pi(2n)} - X_{\pi(1)}| \left(\frac{2}{n} \sum_{1 \leq j \leq \frac{n}{2}} |X_{\pi(4j-3)} - X_{\pi(4j)}| \right) \\
&\leq \left(\frac{U_{(2n)}}{\sqrt{n}} \right)^2 \\
&\xrightarrow{P} 0 ,
\end{aligned}$$

where the first and second inequalities follow by inspection, the third follows by arguing as before, and the convergence in probability to zero again follows from Lemma S.1.1. It thus follows that Assumptions 2.3–2.4 hold.

S.1.8 Proof of Theorem 4.2

We describe an algorithm that leads to a pairing that does not minimize the right-hand side of (35) exactly, but which leads to the desired bound, from which the result follows.

In order to describe the algorithm, it is useful to introduce some further notation. For an integer $m > 1$, divide $[0, 1]^k$ into m^k hypercubes with sides of length m^{-1}. We index these cubes by k-tuples of the form $(i_1, \ldots, i_k)$ with $1 \leq i_j \leq m$ for all $1 \leq j \leq k$. Specifically, the k-tuple $(i_1, \ldots, i_k)$ corresponds to the (closed) cube with vertices

$$\left\{ \frac{1}{m}(i_1 - 1 + \delta_1, \ldots, i_k - 1 + \delta_k) : \delta_j \in \{0, 1\} \text{ for all } 1 \leq j \leq k \right\} .$$

We further order these cubes in a "contiguous" way. We do so by defining an algorithm f_k that takes as an input a k-dimensional hypercube of the form $(i_1, \ldots, i_k)$ with $i_j \in \{1, m\}$ for all $1 \leq j \leq k$ and returns a "path" starting from $(i_1, \ldots, i_k)$ and ending at $(i'_1, \ldots, i'_k)$ with $i'_j \in \{1, m\}$ for all $1 \leq j \leq k$ that traverses all m^k of the possible k-dimensional hypercubes. We define f_1 so that

$$f_1((i_1)) = \begin{cases} (1) \mapsto (2) \mapsto \cdots \mapsto (m-1) \mapsto (m) & \text{if} \quad (i_1) = (1) \\ (m) \mapsto (m-1) \mapsto \cdots \mapsto (2) \mapsto (1) & \text{if} \quad (i_1) = (m) . \end{cases} \tag{S.9}$$

Given f_{k-1}, we define $f_k((i_1^0, \ldots, i_k^0))$ as follows. If $i_k^0 = 1$, then $f_k((i_1^0, \ldots, i_k^0))$ equals

$$\begin{aligned}
& (i_1^0, \ldots, i_{k-1}^0, 1) \mapsto \cdots \mapsto (i_1^1, \ldots, i_{k-1}^1, 1) \\
\mapsto \quad & (i_1^1, \ldots, i_{k-1}^1, 2) \mapsto \cdots \mapsto (i_1^2, \ldots, i_{k-1}^2, 2) \\
& \vdots \\
\mapsto \quad & (i_1^{j-1}, \ldots, i_{k-1}^{j-1}, j) \mapsto \cdots \mapsto (i_1^j, \ldots, i_{k-1}^j, j) \\
& \vdots \\
\mapsto \quad & (i_1^{m-1}, \ldots, i_{k-1}^{m-1}, m) \mapsto \cdots \mapsto (i_1^m, \ldots, i_{k-1}^m, m) \ ,
\end{aligned}$$

where in the preceding display it is understood that the "path" for a fixed "row," i.e.,

$$(i_1^{j-1}, \ldots, i_{k-1}^{j-1}, j) \mapsto \cdots \mapsto (i_1^j, \ldots, i_{k-1}^j, j) \ , \tag{S.10}$$

is given by applying f_{k-1} first to obtain a "path" starting from $(i_1^{j-1}, \ldots, i_{k-1}^{j-1})$ and ending at $(i_1^j, \ldots, i_{k-1}^j)$ and then "appending" j to obtain a "path" of the form (S.10). If, on the other hand, $i_k^0 = m$, then $f_k((i_1^0, \ldots, i_k^0))$ equals

$$\begin{aligned}
& (i_1^0, \ldots, i_{k-1}^0, m) \mapsto \cdots \mapsto (i_1^1, \ldots, i_{k-1}^1, m) \\
\mapsto \quad & (i_1^1, \ldots, i_{k-1}^1, m-1) \mapsto \cdots \mapsto (i_1^2, \ldots, i_{k-1}^2, m-1) \\
& \vdots \\
\mapsto \quad & (i_1^{j-1}, \ldots, i_{k-1}^{j-1}, m-j+1) \mapsto \cdots \mapsto (i_1^j, \ldots, i_{k-1}^j, m-j+1) \\
& \vdots \\
\mapsto \quad & (i_1^{m-1}, \ldots, i_{k-1}^{m-1}, 1) \mapsto \cdots \mapsto (i_1^m, \ldots, i_{k-1}^m, 1) \ ,
\end{aligned}$$

where, as before, in the preceding display it is understood that the "path" for a fixed "row," i.e.,

$$(i_1^{j-1}, \ldots, i_{k-1}^{j-1}, m-j+1) \mapsto \cdots \mapsto (i_1^j, \ldots, i_{k-1}^j, m-j+1) \ , \tag{S.11}$$

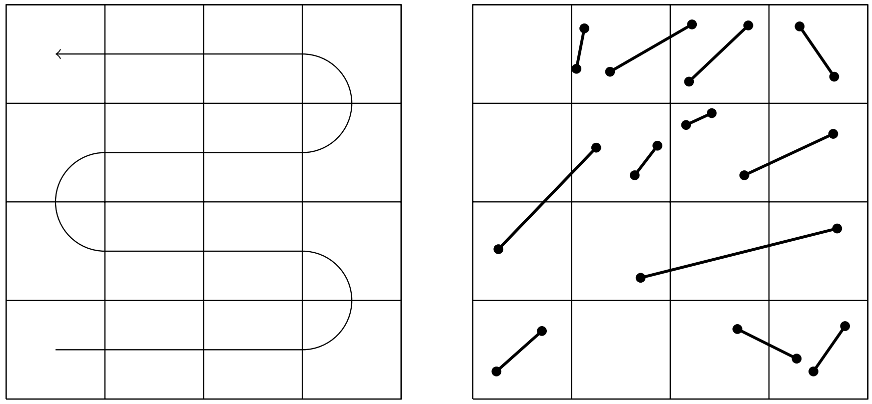
is given by applying f_{k-1} first to obtain a "path" starting from $(i_1^{j-1}, \ldots, i_{k-1}^{j-1})$ and ending at $(i_1^j, \ldots, i_{k-1}^j)$ and then "appending" $m-j+1$ to obtain a "path" of the form (S.10).

With f_k so defined, we may obtain a "path" starting with $(1, \ldots, 1)$. Figure 1(a) above illustrates the "path" obtained in this way for the case of $k = 2$ and $m = 4$. Using this "path," we are now prepared to describe our algorithm for pairing units below. We emphasize that the algorithm depends on the choice of m. For clarity, we also note that when we say in our description of the algorithm that a unit i belongs to a hypercube, we mean that X_i belongs to the hypercube. To avoid any ambiguity, whenever a unit belongs to more than one hypercube, we assign it the hypercube that appears earliest along the "path."

Algorithm S.1.1.

> Begin with the first nonempty hypercube along the "path." If there are an even number of units in that hypercube, pair them together in any fashion; if there are an odd number of units in that hypercube, pair as many as possible together. Now proceed to the "next" nonempty hypercube along the "path." If in the previous hypercube there was an unpaired unit, pair one of the units in the present hypercube with the remaining unit from the previous hypercube. If, after doing so, there are an even number of unpaired units in the hypercube, pair them in any fashion; if, after doing so, there are an odd number of unpaired units in the hypercube, pair as many as possible together. Proceed to the next nonempty hypercube along the "path." Continue in this fashion until there are no more nonempty hypercubes.

Figure 1(b) above illustrates a pairing obtained by applying Algorithm S.1.1 with $k = 2$, $n = 12$ and $m = 4$.



(a) (b)

Figure S.1: (a) Illustration of the "path" obtained by applying f_k with $k = 2$ and $m = 4$; (b) Illustration of a pairing obtained by applying Algorithm S.1.1 with $k = 2$, $n = 12$ and $m = 4$. Note that the endpoints of the line segments correspond to units and the pairs correspond to units connected by a line segments.

We now argue that Algorithm S.1.1 leads to a pairing satisfying the desired bound. To this end, first note that the maximum distance between any two points in the a k-dimensional hypercube with sides of length $\frac{1}{m}$ is $\frac{\sqrt{k}}{m}$. Note further that the maximum distance between two points in two such cubes that are contiguous (as understood according to ordering described in Section 4) is $\frac{2\sqrt{k}}{m}$. Using these facts, the bound in (36) now easily follows. Indeed, simply note that the sum that appears on the left-hand side of (36) may contain at most n terms corresponding to pairs of points within hypercubes and at most m^k terms corresponding to pairs of points in contiguous hypercubes. The desired conclusion now follows immediately. ■

S.1.9 Proof of Theorem 4.3

We prove the result for $k = 3$ and $\ell = 0$; the other values of k and ℓ can be handled similarly.

By arguing as in the proof of Theorem 4.2 and using (35), we see that

$$\frac{2}{n} \sum_{1 \le j \le \lfloor \frac{n}{2} \rfloor} |\bar{X}_{\tilde{\pi}(2j)} - \bar{X}_{\tilde{\pi}(2j-1)}|^2 \xrightarrow{P} 0 \, . \tag{S.12}$$

Note that

$$\begin{aligned}
& \frac{1}{n} \sum_{1 \le j \le \frac{n}{2}} |X_{\tilde{\pi}(4j-3)} - X_{\tilde{\pi}(4j)}|^2 \\
= \; & \frac{1}{n} \sum_{1 \le j \le \frac{n}{2}} |X_{\tilde{\pi}(4j-3)} - \bar{X}_{\tilde{\pi}(2j-1)} + \bar{X}_{\tilde{\pi}(2j-1)} - \bar{X}_{\tilde{\pi}(2j)} + \bar{X}_{\tilde{\pi}(2j)} - X_{\tilde{\pi}(4j)}|^2 \\
\lesssim \; & \frac{1}{n} \sum_{1 \le j \le \frac{n}{2}} |X_{\tilde{\pi}(4j-3)} - \bar{X}_{\tilde{\pi}(2j-1)}|^2 + |\bar{X}_{\tilde{\pi}(2j-1)} - \bar{X}_{\tilde{\pi}(2j)}|^2 + |\bar{X}_{\tilde{\pi}(2j)} - X_{\tilde{\pi}(4j)}|^2 \\
\lesssim \; & \frac{1}{n} \sum_{1 \le j \le \frac{n}{2}} |X_{\tilde{\pi}(4j-3)} - X_{\tilde{\pi}(4j-2)}|^2 + |\bar{X}_{\tilde{\pi}(2j-1)} - \bar{X}_{\tilde{\pi}(2j)}|^2 + |X_{\tilde{\pi}(4j-1)} - X_{\tilde{\pi}(4j)}|^2 \\
\lesssim \; & \frac{1}{n} \sum_{1 \le j \le n} |X_{\pi(2j)} - X_{\pi(2j-1)}|^2 + \frac{1}{n} \sum_{1 \le j \le \frac{n}{2}} |\bar{X}_{\tilde{\pi}(2j-1)} - \bar{X}_{\tilde{\pi}(2j)}|^2 \\
\xrightarrow{P} \; & 0 \, ,
\end{aligned}$$

where the first equality follows by inspection, the first inequality follows using the fact that $|a + b|^2 \le 2(|a|^2 + |b|^2)$ for any real vectors a and b, the second inequality follows from (38) and (39), the second equality follows again from (39), and the convergence to zero in probability follows from the assumption that π satisfies Assumption 2.3 and (S.12). ■

S.1.10 Auxiliary Results

Lemma S.1.1. *Let $U_i, i = 1, \ldots, n$ an i.i.d. sequence of random vectors such that $E[|U_i|^r] < \infty$. Then,*

$$n^{-\frac{1}{r}} \max_{1 \le i \le n} |U_i| \xrightarrow{P} 0$$

as $n \to \infty$.

Proof: Let $\epsilon > 0$ be given. Note that

$$\begin{aligned} P\left\{ n^{-\frac{1}{r}} \max_{1 \le i \le n} |U_i| > \epsilon \right\} &\le \sum_{1 \le i \le n} P\{|U_i|^r > \epsilon^r n\} \\ &\le \frac{1}{n\epsilon^r} \sum_{1 \le i \le n} E[|U_i|^r I\{|U_i|^r > \epsilon^r n\}] \\ &= \frac{1}{\epsilon^r} E[|U_i|^r I\{|U_i|^r > \epsilon^r n\}] \\ &\to 0 \end{aligned}$$

as $n \to \infty$, where the first inequality follows from Bonferonni's inequality, the second inequality follows from Markov's inequality, the final equality follows from the i.i.d. assumption, and the convergence to zero follows from the assumption that $E[|U_i|^r] < \infty$. ■

Lemma S.1.2. *For $n \ge 1$, let U_n and V_n be real-valued random variables and $\mathcal{F}_n$ a σ-field. Suppose*

$$P\{U_n \le u | \mathcal{F}_n\} \to \Phi(u/\tau_1) \text{ a.s.} , \tag{S.13}$$

where $\Phi(\cdot)$ is the standard normal c.d.f. Further assume V_n is $\mathcal{F}_n$-measurable and

$$V_n \xrightarrow{d} N(0, \tau_2^2) .$$

Then,

$$U_n + V_n \xrightarrow{d} N(0, \tau_1^2 + \tau_2^2) .$$

Proof: The proof is omitted, but follows easily using characteristic functions. ■

Lemma S.1.3. *Let $(U_{n,1}, \ldots, U_{n,n}) \sim G_n^* = \bigotimes_{1 \le i \le n} G_{n,i}$ with $\mu(G_{n,i}) = 0$ for all $1 \le i \le n$. Define*

$$\bar{G}_n = \frac{1}{n} \sum_{1 \le i \le n} G_{n,i} .$$

If

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} E_{\bar{G}_n}[|U| I\{|U| > \lambda\}] = 0 , \tag{S.14}$$

then $\bar{U}_n \xrightarrow{G_n^*} 0$.

Proof: The proof is omitted, but is a straightforward generalization of Lemma 11.4.2 in Lehmann and Romano (2005), where $G_{n,i} = G_n$. ■

Lemma S.1.4. *If Assumptions 2.1–2.3 hold, then*

$$\sqrt{n}(\hat{\Delta}_n - \Delta(Q)) \xrightarrow{d} N(0, \nu^2) , \tag{S.15}$$

where

$$\nu^2 = E[Var[Y_i(1)|X_i]] + E[Var[Y_i(0)|X_i]]$$

$$
\begin{aligned}
& +\frac{1}{2}E\left[((E[Y_i(1)|X_i]-E[Y_i(1)])-(E[Y_i(0)|X_i]-E[Y_i(0)]))^2\right] \\
= \quad & Var[Y_i(1)]+Var[Y_i(0)]-\frac{1}{2}E\left[((E[Y_i(1)|X_i]-E[Y_i(1)])+(E[Y_i(0)|X_i]-E[Y_i(0)]))^2\right]
\end{aligned} \qquad \text{(S.16)}
$$

as $n\to\infty$.

PROOF: Note that

$$
\begin{aligned}
\frac{1}{n}\sum_{1\le i\le 2n:D_i=1} Y_i &= \frac{1}{n}\sum_{1\le i\le 2n} Y_i(1)D_i \\
\frac{1}{n}\sum_{1\le i\le 2n:D_i=0} Y_i &= \frac{1}{n}\sum_{1\le i\le 2n} Y_i(0)(1-D_i)\ .
\end{aligned}
$$

Hence, we may write

$$
\sqrt{n}(\hat\Delta_n-\Delta(Q)) = A_n - B_n + C_n - D_n\ ,
$$

where

$$
\begin{aligned}
A_n &= \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n}\Big(Y_i(1)D_i - E[Y_i(1)D_i|X^{(n)},D^{(n)}]\Big) \\
B_n &= \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n}\Big(Y_i(0)(1-D_i) - E[Y_i(0)(1-D_i)|X^{(n)},D^{(n)}]\Big) \\
C_n &= \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n}\Big(E[Y_i(1)D_i|X^{(n)},D^{(n)}] - D_iE[Y_i(1)]\Big) \\
D_n &= \frac{1}{\sqrt{n}}\sum_{1\le i\le 2n}\Big(E[Y_i(0)(1-D_i)|X^{(n)},D^{(n)}] - (1-D_i)E[Y_i(0)]\Big)\ .
\end{aligned}
$$

Note that, conditional on $X^{(n)}$ and $D^{(n)}$, A_n and B_n are independent and C_n and D_n are constant.

We first analyze the limiting behavior of A_n. Conditional on $X^{(n)}$ and $D^{(n)}$, the terms in this sum are independent, but not identically distributed. We proceed by verifying that the condition in Linderberg's Central Limit Theorem holds in probability conditional on $X^{(n)}$ and $D^{(n)}$. To that end, define

$$
s_n^2 = s_n^2(X^{(n)},D^{(n)}) = \sum_{1\le i\le 2n}\mathrm{Var}[Y_i(1)D_i|X^{(n)},D^{(n)}]
$$

and note that

$$
\begin{aligned}
s_n^2 &= \sum_{1\le i\le 2n:D_i=1}\mathrm{Var}[Y_i(1)|X^{(n)},D^{(n)}] \\
&= \sum_{1\le i\le 2n:D_i=1}\mathrm{Var}[Y_i(1)|X^{(n)}] \\
&= \sum_{1\le i\le 2n:D_i=1}\mathrm{Var}[Y_i(1)|X_i]\ ,
\end{aligned}
$$

where the first equality follows from Assumption 2.2 and the second follows from the fact that $Q_n=Q^n$. It follows that

$$
\frac{s_n^2}{n} = \frac{1}{2n}\sum_{1\le i\le 2n}\mathrm{Var}[Y_i(1)|X_i] + \left(\frac{1}{2n}\sum_{1\le i\le 2n:D_i=1}\mathrm{Var}[Y_i(1)|X_i] - \frac{1}{2n}\sum_{1\le i\le 2n:D_i=0}\mathrm{Var}[Y_i(1)|X_i]\right)\ .
$$

Using Assumption 2.1(b), we have that

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

Note further that

$$
\left|\frac{1}{2n}\sum_{1\le i\le 2n:D_i=1}\mathrm{Var}[Y_i(1)|X_i] - \frac{1}{2n}\sum_{1\le i\le 2n:D_i=0}\mathrm{Var}[Y_i(1)|X_i]\right|
$$

$$\leq \frac{1}{2n} \sum_{1 \leq j \leq n} \left| \text{Var}[Y_{\pi(2j)}(1)|X_{\pi(2j)}] - \text{Var}[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] \right|$$
$$\lesssim \frac{1}{n} \sum_{1 \leq j \leq n} |X_{\pi(2j)} - X_{\pi(2j-1)}| \xrightarrow{P} 0 ,$$

where the first inequality follows by inspection, the second inequality exploits Assumption 2.1(c) and the convergence to zero follows from Assumption 2.3. Hence,

$$\frac{s_n^2}{n} \xrightarrow{P} E[\text{Var}[Y_i(1)|X_i]] > 0 , \tag{S.17}$$

where the final inequality exploits Assumption 2.1(a). Next, we argue for any $\epsilon > 0$ that

$$\frac{1}{s_n^2} \sum_{1 \leq i \leq 2n} E[|Y_i(1)D_i - E[Y_i(1)D_i|X^{(n)}, D^{(n)}]|^2 I\{|Y_i(1)D_i - E[Y_i(1)D_i|X^{(n)}, D^{(n)}]| > \epsilon s_n\}|X^{(n)}, D^{(n)}] \xrightarrow{P} 0 .$$

To this end, first note that for any $m > 0$ we have that

$$P\{\epsilon s_n > m\} \to 1 . \tag{S.18}$$

Note further that Assumption 2.2 implies that

$$E[Y_i(1)D_i|X^{(n)}, D^{(n)}] = D_i E[Y_i(1)|X_i] , \tag{S.19}$$

so the lefthand-side of the preceding display may be written as

$$\begin{aligned}
&\frac{1}{s_n^2} \sum_{1 \leq i \leq 2n : D_i = 1} E[|Y_i(1) - E[Y_i(1)|X_i]|^2 I\{|Y_i(1) - E[Y_i(1)|X_i]| > \epsilon s_n\}|X^{(n)}, D^{(n)}] \\
\leq\ & \left(\frac{s_n^2}{n}\right)^{-1} \frac{1}{n} \sum_{1 \leq i \leq 2n} E[|Y_i(1) - E[Y_i(1)|X_i]|^2 I\{|Y_i(1) - E[Y_i(1)|X_i]| > \epsilon s_n\}|X^{(n)}, D^{(n)}] \\
\leq\ & \left(\frac{s_n^2}{n}\right)^{-1} \frac{1}{n} \sum_{1 \leq i \leq 2n} E[|Y_i(1) - E[Y_i(1)|X_i]|^2 I\{|Y_i(1) - E[Y_i(1)|X_i]| > m\}|X^{(n)}, D^{(n)}] + o_P(1) \\
=\ & \left(\frac{s_n^2}{n}\right)^{-1} \frac{1}{n} \sum_{1 \leq i \leq 2n} E[|Y_i(1) - E[Y_i(1)|X_i]|^2 I\{|Y_i(1) - E[Y_i(1)|X_i]| > m\}|X_i] + o_P(1) \\
\xrightarrow{P}\ & (E[\text{Var}[Y_i(1)|X_i]])^{-1} E[|Y_i(1) - E[Y_i(1)|X_i]|^2 I\{|Y_i(1) - E[Y_i(1)|X_i]| > m\}] ,
\end{aligned}$$

where the first inequality follows by inspection, the second inequality exploits (S.17)–(S.18), the equality follows from Assumption 2.2 and the fact that $Q_n = Q^n$, and the convergence in probability follows from (S.17) and the fact that Assumption 2.1(b) implies

$$E[|Y_i(1) - E[Y_i(1)|X_i]|^2] = E[\text{Var}[Y_i(1)|X_i]] \leq E[Y_i^2(1)] < \infty . \tag{S.20}$$

Note further that (S.20) implies that

$$\lim_{m \to \infty} E[|Y_i(1) - E[Y_i(1)|X_i]|^2 I\{|Y_i(1) - E[Y_i(1)|X_i]| > m\}] = 0 .$$

The condition in Lindeberg's Central Limit Theorem therefore holds in probability. It follows by a subsequencing argument similar to that used in the proof of Lemma S.1.5 below that

$$\sup_{t \in \mathbf{R}} \left| P\{A_n \leq t | X^{(n)}, D^{(n)}\} - \Phi(t/\sqrt{E[\text{Var}[Y_i(1)|X_i]]}) \right| \xrightarrow{P} 0 .$$

A similar argument establishes that

$$\sup_{t \in \mathbf{R}} \left| P\{B_n \leq t | X^{(n)}, D^{(n)}\} - \Phi(t/\sqrt{E[\text{Var}[Y_i(0)|X_i]]}) \right| \xrightarrow{P} 0 .$$

Since A_n and B_n are independent conditional on $X^{(n)}$ and $D^{(n)}$, it follows by another subsequencing argument that

$$\sup_{t \in \mathbf{R}} \left| P\{A_n - B_n \leq t | X^{(n)}, D^{(n)}\} - \Phi(t/\sqrt{E[\text{Var}[Y_i(0)|X_i]] + E[\text{Var}[Y_i(0)|X_i]]}) \right| \xrightarrow{P} 0 . \tag{S.21}$$

To analyze C_n, first note that (S.19) implies that

$$C_n = \frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} D_i \left(E[Y_i(1)|X_i] - E[Y_i(1)] \right) \ , \tag{S.22}$$

so

$$E[C_n|X^{(n)}] = \frac{1}{2\sqrt{n}} \sum_{1 \le i \le 2n} \left(E[Y_i(1)|X_i] - E[Y_i(1)] \right) \ . \tag{S.23}$$

Furthermore,

$$\begin{aligned}
\mathrm{Var}[C_n|X^{(n)}] &= \mathrm{Var}[C_n - E[C_n|X^{(n)}]|X^{(n)}] \\
&= \mathrm{Var}\left[\frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} \left(D_i - \frac{1}{2} \right) \left(E[Y_i(1)|X_i] - E[Y_i(1)] \right) \middle| X^{(n)} \right] \\
&= \mathrm{Var}\left[\frac{1}{\sqrt{n}} \sum_{1 \le i \le 2n} \left(D_i - \frac{1}{2} \right) E[Y_i(1)|X_i] \middle| X^{(n)} \right] \\
&= \frac{1}{4n} \sum_{1 \le j \le n} \left(E[Y_{\pi(2j)}(1)|X_{\pi(2j)}] - E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}] \right)^2 \\
&\lesssim \frac{1}{n} \sum_{1 \le j \le n} \left(X_{\pi(2j)} - X_{\pi(2j-1)} \right)^2 \xrightarrow{P} 0 \ ,
\end{aligned}$$

where the first equality exploits properties of conditional variances, the second follows from (S.22)–(S.23), the third exploits the fact that $\sum_{1 \le i \le 2n} D_i = n$, the fourth exploits the distribution of $D^{(n)}|X^{(n)}$, the inequality follows from Assumption 2.1(c), and the convergence in probability follows from Assumption 2.3. For any $\epsilon > 0$, it thus follows from Chebychev's inequality that

$$P\{|C_n - E[C_n|X^{(n)}]| > \epsilon | X^{(n)}\} \le \frac{\mathrm{Var}[C_n|X^{(n)}]}{\epsilon^2} \xrightarrow{P} 0 \ .$$

Since probabilities are bounded, we have further that

$$P\{|C_n - E[C_n|X^{(n)}]| > \epsilon\} \xrightarrow{P} 0 \ .$$

Hence,

$$C_n = \frac{1}{2\sqrt{n}} \sum_{1 \le i \le 2n} \left(E[Y_i(1)|X_i] - E[Y_i(1)] \right) + o_P(1) \ . \tag{S.24}$$

A similar argument establishes that

$$D_n = \frac{1}{2\sqrt{n}} \sum_{1 \le i \le 2n} \left(E[Y_i(0)|X_i] - E[Y_i(0)] \right) + o_P(1) \ . \tag{S.25}$$

Hence,

$$\begin{aligned}
C_n - D_n &= \frac{1}{2\sqrt{n}} \sum_{1 \le i \le 2n} \left(\left(E[Y_i(1)|X_i] - E[Y_i(1)] \right) - \left(E[Y_i(0)|X_i] - E[Y_i(0)] \right) \right) + o_P(1) \\
&= \frac{\sqrt{2}}{2} \frac{1}{\sqrt{2n}} \sum_{1 \le i \le 2n} \left(\left(E[Y_i(1)|X_i] - E[Y_i(1)] \right) - \left(E[Y_i(0)|X_i] - E[Y_i(0)] \right) \right) + o_P(1) \\
&\xrightarrow{d} N\left(0, \frac{1}{2} E\left[\left(\left(E[Y_i(1)|X_i] - E[Y_i(1)] \right) - \left(E[Y_i(0)|X_i] - E[Y_i(0)] \right) \right)^2 \right] \right) \ ,
\end{aligned}$$

where the first equality follows from (S.24)–(S.25), the second equality follows by inspection, and the convergence in distribution follows from Slutsky's theorem and the Central Limit Theorem.

The desired conclusion (S.15) now follows by a subsequencing argument. To see this, suppose by way of contradiction that (S.15) fails. This implies that there exists $\delta > 0$ and a subsequence n_k along which

$$\sup_{t \in \mathbf{R}} |P\{\sqrt{n_k}(\hat{\Delta}_{n_k} - \Delta(Q)) \le t\} - \Phi(t/\nu)| \to \delta \ . \tag{S.26}$$

By considering a further subsequence if necessary, which, by an abuse of notation, we continue to denote by n_k, it follows from

(S.21) that

$$A_{n_k} - B_{n_k} \xrightarrow{d} N(0, E[\text{Var}[Y_i(0)|X_i]] + E[\text{Var}[Y_i(0)|X_i]]) \text{ w.p.1 (conditional on } X^{(n_k)} \text{ and } D^{(n_k)}) \ .$$

Since $C_{n_k} - D_{n_k}$ is constant conditional on $X^{(n_k)}$ and $D^{(n_k)}$, Lemma S.1.2 establishes that

$$\sqrt{n_k}(\hat{\Delta}_{n_k} - \Delta) = A_{n_k} - B_{n_k} + C_{n_k} - D_{n_k} \xrightarrow{d} N(0, \nu^2) \ ,$$

which, by Polya's Theorem, implies a contradiction to (S.26).

Finally, in order to complete the proof, note that

$$\begin{aligned}
& E[\text{Var}[Y_i(1)|X_i]] + E[\text{Var}[Y_i(0)|X_i]] + \frac{1}{2}E\left[((E[Y_i(1)|X_i] - E[Y_i(1)]) - (E[Y_i(0)|X_i] - E[Y_i(0)]))^2\right] \\
= \ & \text{Var}[Y_i(1)] + \text{Var}[Y_i(0)] - \text{Var}[E[Y_i(1)|X_i]] - \text{Var}[E[Y_i(0)|X_i]] \\
& \qquad + \frac{1}{2}E\left[((E[Y_i(1)|X_i] - E[Y_i(1)]) - (E[Y_i(0)|X_i] - E[Y_i(0)]))^2\right] \\
= \ & \text{Var}[Y_i(1)] + \text{Var}[Y_i(0)] - \frac{1}{2}\text{Var}[E[Y_i(1)|X_i]] - \frac{1}{2}\text{Var}[E[Y_i(0)|X_i]] \\
& \qquad - E[(E[Y_i(1)|X_i] - E[Y_i(1)])(E[Y_i(0)|X_i] - E[Y_i(0)])] \\
= \ & \text{Var}[Y_i(1)] + \text{Var}[Y_i(0)] - \frac{1}{2}E\left[((E[Y_i(1)|X_i] - E[Y_i(1)]) + (E[Y_i(0)|X_i] - E[Y_i(0)]))^2\right] \ ,
\end{aligned}$$

which establishes that the two expressions for ν^2 in the statement of the theorem are in fact equivalent. ■

Lemma S.1.5. *If Assumptions 2.1–2.3 hold, then $\hat{\mu}_n(d) \xrightarrow{P} E[Y_i(d)]$ and $\hat{\sigma}_n^2(d) \xrightarrow{P} Var[Y_i(d)]$, where $\hat{\mu}_n(d)$ and $\hat{\sigma}_n^2(d)$ are defined in (4) and (5), respectively.*

PROOF: Note that

$$\begin{aligned}
\hat{\mu}_n(d) &= \frac{1}{n}\sum_{1\le i\le 2n} Y_i(d)I\{D_i = d\} \\
\hat{\sigma}_n^2(d) &= \frac{1}{n}\sum_{1\le i\le 2n} (Y_i - \hat{\mu}_n(d))^2 I\{D_i = d\} \\
&= \frac{1}{n}\sum_{1\le i\le 2n} Y_i^2(d)I\{D_i = d\} - \hat{\mu}_n^2(d) \ .
\end{aligned}$$

It therefore suffices to show that

$$\frac{1}{n}\sum_{1\le i\le 2n} Y_i^r(d)I\{D_i = d\} \xrightarrow{P} E[Y_i^r(d)]$$

for $r \in \{1, 2\}$. We prove this result only for $r = 1$ and $d = 1$; the other cases can be proven similarly. To this end, write

$$\begin{aligned}
\frac{1}{n}\sum_{1\le i\le 2n} Y_i(1)I\{D_i = 1\} &= \frac{1}{n}\sum_{1\le i\le 2n} Y_i(1)D_i \\
&= \frac{1}{n}\sum_{1\le i\le 2n}\left(Y_i(1)D_i - E[Y_i(1)D_i|X^{(n)}, D^{(n)}]\right) + \frac{1}{n}\sum_{1\le i\le 2n} E[Y_i(1)D_i|X^{(n)}, D^{(n)}] \ .
\end{aligned}$$

Next, note that

$$\begin{aligned}
& \frac{1}{n}\sum_{1\le i\le 2n} E[Y_i(1)D_i|X^{(n)}, D^{(n)}] \\
= \ & \frac{1}{n}\sum_{1\le i\le 2n} D_i E[Y_i(1)|X_i] \\
= \ & \frac{1}{n}\sum_{1\le i\le 2n: D_i = 1} E[Y_i(1)|X_i] \\
= \ & \frac{1}{2n}\sum_{1\le i\le 2n} E[Y_i(1)|X_i] + \left(\frac{1}{2n}\sum_{1\le i\le 2n: D_i=1} E[Y_i(1)|X_i] - \frac{1}{2n}\sum_{1\le i\le 2n: D_i=0} E[Y_i(1)|X_i]\right) \ ,
\end{aligned}$$

where the first equality exploits (S.19) and the second and third equalities follow by inspection. Note further that

$$\begin{aligned}
&\left|\frac{1}{2n}\sum_{1\le i\le 2n: D_i=1} E[Y_i(1)|X_i] - \frac{1}{2n}\sum_{1\le i\le 2n: D_i=0} E[Y_i(1)|X_i]\right| \\
\le\ & \frac{1}{2n}\sum_{1\le j\le n} |E[Y_{\pi(2j)}(1)|X_{\pi(2j)}] - E[Y_{\pi(2j-1)}(1)|X_{\pi(2j-1)}]| \\
\lesssim\ & \frac{1}{n}\sum_{1\le j\le n} |X_{\pi(2j)} - X_{\pi(2j-1)}| \xrightarrow{P} 0 ,
\end{aligned}$$

where the first inequality follows by inspection, the second exploits Assumption 2.1(c) and the convergence in probability follows from Assumption 2.3. Since Assumption 2.1(b) implies that $E[|E[Y_i(1)|X_i]|] \le E[|Y_i(1)|] < \infty$, it follows that

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

To complete the argument, we argue that

$$\frac{1}{n}\sum_{1\le i\le 2n}\Big(Y_i(1)D_i - E[Y_i(1)D_i|X^{(n)}, D^{(n)}]\Big) \xrightarrow{P} 0 . \tag{S.27}$$

For this purpose, we proceed by verifying that (S.14) in Lemma S.1.3 holds in probability conditional on $X^{(n)}$ and $D^{(n)}$. To that end, note for any $m > 0$ that

$$\begin{aligned}
& \frac{1}{2n}\sum_{1\le i\le 2n} E[|Y_i(1)D_i - E[Y_i(1)D_i|X^{(n)}, D^{(n)}]| I\{|Y_i(1)D_i - E[Y_i(1)D_i|X^{(n)}, D^{(n)}]| > m\}|X^{(n)}, D^{(n)}] \\
=\ & \frac{1}{2n}\sum_{1\le i\le 2n} E[|Y_i(1)D_i - D_iE[Y_i(1)|X_i]| I\{|Y_i(1)D_i - D_iE[Y_i(1)|X_i]| > m\}|X^{(n)}, D^{(n)}] \\
\le\ & \frac{1}{2n}\sum_{1\le i\le 2n} E[|Y_i(1) - E[Y_i(1)|X_i]| I\{|Y_i(1) - E[Y_i(1)|X_i]| > m\}|X^{(n)}, D^{(n)}] \\
=\ & \frac{1}{2n}\sum_{1\le i\le 2n} E[|Y_i(1) - E[Y_i(1)|X_i]| I\{|Y_i(1) - E[Y_i(1)|X_i]| > m\}|X_i] \\
\xrightarrow{P}\ & E[|Y_i(1) - E[Y_i(1)|X_i]| I\{|Y_i(1) - E[Y_i(1)|X_i]| > m\}] ,
\end{aligned} \tag{S.28}$$

where the first and fourth equalities follow from (S.19), the inequality follows by inspection, and the convergence in probability follows from (S.20). The desired conclusion (S.27) now follows by a subsequencing argument. To see this, suppose by way of contradiction that (S.27) fails. This implies that there exists $\epsilon > 0$, $\delta > 0$ and a subsequence n_k along which

$$P\left\{\left|\frac{1}{n_k}\sum_{1\le i\le 2n_k}\Big(Y_i(1)D_i - E[Y_i(1)D_i|X^{(n_k)}, D^{(n_k)}]\Big)\right| > \epsilon\right\} \to \delta . \tag{S.29}$$

By considering a further subsequence if necessary, which, by an abuse of notation, we continue to denote by n_k, it follows from (S.19), (S.20) and (S.28) that

$$\begin{aligned}
\lim_{m\to\infty}\limsup_{k\to\infty} \frac{1}{2n}\sum_{1\le i\le 2n}\Big(&E[|Y_i(1)D_i - E[Y_i(1)D_i|X^{(n)}, D^{(n)}]| \\
&\times I\{|Y_i(1)D_i - E[Y_i(1)D_i|X^{(n)}, D^{(n)}]| > m\}|X^{(n)}, D^{(n)}]\Big) = 0
\end{aligned}$$

w.p.1 (conditional on $X^{(n_k)}$ and $D^{(n_k)}$). Lemma S.1.3 implies, however, that

$$\frac{1}{n_k}\sum_{1\le i\le 2n_k}\Big(Y_i(1)D_i - E[Y_i(1)D_i|X^{(n_k)}, D^{(n_k)}]\Big) \to 0 \text{ w.p.1 (conditional on } X^{(n_k)} \text{ and } D^{(n_k)}) ,$$

which implies a contradiction to (S.29). ■

Lemma S.1.6. *If Assumptions 2.1–2.3 hold, then*

$$\hat\tau_n^2 \xrightarrow{P} E[Var[Y_i(1)|X_i]] + E[Var[Y_i(0)|X_i]] + E\left[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2\right] ,$$

where $\hat{\tau}_n^2$ is defined in (21).

PROOF: Note that

$$\hat{\tau}_n^2 = \frac{1}{n}\sum_{1\le j\le n}(Y_{\pi(2j)} - Y_{\pi(2j-1)})^2 = \frac{1}{n}\sum_{1\le i\le 2n} Y_i^2 - \frac{2}{n}\sum_{1\le j\le n} Y_{\pi(2j)}Y_{\pi(2j-1)} ~.$$

Since

$$\frac{1}{n}\sum_{1\le i\le 2n} Y_i^2 = \hat{\sigma}_n^2(1) - \hat{\mu}_n^2(1) + \hat{\sigma}_n^2(0) - \hat{\mu}_n^2(0) ~,$$

it follows from Lemma S.1.5 that

$$\frac{1}{n}\sum_{1\le i\le 2n} Y_i^2 \xrightarrow{P} E[Y_i^2(1)] + E[Y_i^2(0)] ~.$$

Next, we argue that

$$\frac{2}{n}\sum_{1\le j\le n} Y_{\pi(2j)}Y_{\pi(2j-1)} \xrightarrow{P} 2E[\mu_1(X_i)\mu_0(X_i)] ~,$$

where we use the notation $\mu_d(X_i)$ to denote $E[Y_i(d)|X_i]$. To this end, first note that

$$E\left[Y_{\pi(2j)}Y_{\pi(2j-1)}\Big|X^{(n)}\right] = \frac{1}{2}\mu_1(X_{\pi(2j)})\mu_0(X_{\pi(2j-1)}) + \frac{1}{2}\mu_0(X_{\pi(2j)})\mu_1(X_{\pi(2j-1)}) ~, \tag{S.30}$$

so

$$\begin{aligned}
& E\left[\frac{2}{n}\sum_{1\le j\le n} Y_{\pi(2j)}Y_{\pi(2j-1)}\Big|X^{(n)}\right] \\
= & \frac{2}{n}\sum_{1\le j\le n} E\left[Y_{\pi(2j)}Y_{\pi(2j-1)}\Big|X^{(n)}\right] \\
= & \frac{1}{n}\sum_{1\le j\le n} \mu_1(X_{\pi(2j)})\mu_0(X_{\pi(2j-1)}) + \mu_0(X_{\pi(2j)})\mu_1(X_{\pi(2j-1)}) \\
= & \frac{1}{n}\sum_{1\le j\le n} \Big(\mu_1(X_{\pi(2j)})(\mu_0(X_{\pi(2j-1)}) - \mu_0(X_{\pi(2j)})) + \mu_1(X_{\pi(2j)})\mu_0(X_{\pi(2j)}) \\
& \qquad + \mu_1(X_{\pi(2j-1)})(\mu_0(X_{\pi(2j)}) - \mu_0(X_{\pi(2j-1)})) + \mu_1(X_{\pi(2j-1)})\mu_0(X_{\pi(2j-1)})\Big) \\
= & \frac{1}{n}\sum_{1\le i\le 2n}\mu_1(X_i)\mu_0(X_i) + \frac{1}{n}\sum_{1\le j\le n}(\mu_1(X_{\pi(2j-1)}) - \mu_1(X_{\pi(2j)}))(\mu_0(X_{\pi(2j)}) - \mu_0(X_{\pi(2j-1)})) ~,
\end{aligned}$$

where the second equality follows from (S.30) and the other equalities follow by inspection. Assumption 2.3 implies that

$$\left|\frac{1}{n}\sum_{1\le j\le n}(\mu_1(X_{\pi(2j-1)}) - \mu_1(X_{\pi(2j)}))(\mu_0(X_{\pi(2j)}) - \mu_0(X_{\pi(2j-1)}))\right| \lesssim \frac{1}{n}\sum_{1\le j\le n}|X_{\pi(2j-1)} - X_{\pi(2j)})|^2 \xrightarrow{P} 0 ~.$$

Furthermore, since

$$E[|\mu_1(X_i)\mu_0(X_i)|] \lesssim E[\mu_1^2(X_i)] + E[\mu_1^2(X_i)] \le E[Y_i^2(1)] + E[Y_i^2(0)] < \infty ~,$$

we have that

$$E\left[\frac{2}{n}\sum_{1\le j\le n} Y_{\pi(2j)}Y_{\pi(2j-1)}\Big|X^{(n)}\right] \xrightarrow{P} 2E[\mu_1(X_i)\mu_0(X_i)] ~.$$

To complete the argument, we show that

$$\frac{1}{n}\sum_{1\le j\le n}\left(Y_{\pi(2j)}Y_{\pi(2j-1)} - E\left[Y_{\pi(2j)}Y_{\pi(2j-1)}\Big|X^{(n)}\right]\right) \xrightarrow{P} 0 ~. \tag{S.31}$$

For this purpose, we proceed by verifying that (S.14) in Lemma S.1.3 holds in probability conditional on $X^{(n)}$. In what follows, we make repeated use of the following facts for any real numbers a and b and $\lambda > 0$:

$$|a+b|I\{|a+b| > \lambda\} \le 2|a|I\left\{|a| > \frac{\lambda}{2}\right\} + 2|b|I\left\{|b| > \frac{\lambda}{2}\right\} \tag{S.32}$$

$$|ab|I\{|ab| > \lambda\} \le a^2 I\{|a| > \sqrt{\lambda}\} + b^2 I\{|b| > \sqrt{\lambda}\} ~. \tag{S.33}$$

Note that the second of these facts follows from the first together with the inequality $2|ab| \leq a^2 + b^2$. Next, note that

$$
\begin{aligned}
& \frac{1}{n} \sum_{1 \leq j \leq n} E\left[\left|Y_{\pi(2j)} Y_{\pi(2j-1)} - E\left[Y_{\pi(2j)} Y_{\pi(2j-1)} \middle| X^{(n)}\right]\right| I\left\{\left|Y_{\pi(2j)} Y_{\pi(2j-1)} - E\left[Y_{\pi(2j)} Y_{\pi(2j-1)} \middle| X^{(n)}\right]\right| > \lambda\right\} \middle| X^{(n)}\right] \\
\lesssim \; & \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j)} I\left\{\left|Y_{\pi(2j)}\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] + \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j-1)} I\left\{\left|Y_{\pi(2j-1)}\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] \\
& \quad + \frac{1}{n} \sum_{1 \leq j \leq n} \left|E\left[Y_{\pi(2j)} Y_{\pi(2j-1)} \middle| X^{(n)}\right]\right| I\left\{\left|E\left[Y_{\pi(2j)} Y_{\pi(2j-1)} \middle| X^{(n)}\right]\right| > \frac{\lambda}{2}\right\} \\
\lesssim \; & \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j)} I\left\{\left|Y_{\pi(2j)}\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] + \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j-1)} I\left\{\left|Y_{\pi(2j-1)}\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] \\
& \quad + \frac{1}{n} \sum_{1 \leq j \leq n} \left|\mu_1(X_{\pi(2j)}) \mu_0(X_{\pi(2j-1)})\right| I\left\{\left|\mu_1(X_{\pi(2j)}) \mu_0(X_{\pi(2j-1)})\right| > \frac{\lambda}{2}\right\} \\
& \quad + \frac{1}{n} \sum_{1 \leq j \leq n} \left|\mu_0(X_{\pi(2j)}) \mu_1(X_{\pi(2j-1)})\right| I\left\{\left|\mu_0(X_{\pi(2j)}) \mu_1(X_{\pi(2j-1)})\right| > \frac{\lambda}{2}\right\} \\
\lesssim \; & \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j)} I\left\{\left|Y_{\pi(2j)}\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] + \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j-1)} I\left\{\left|Y_{\pi(2j-1)}\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] \\
& \quad + \frac{1}{n} \sum_{1 \leq j \leq n} \mu_1^2(X_{\pi(2j)}) I\left\{\left|\mu_1(X_{\pi(2j)})\right| > \sqrt{\frac{\lambda}{2}}\right\} + \frac{1}{n} \sum_{1 \leq j \leq n} \mu_0^2(X_{\pi(2j-1)}) I\left\{\left|\mu_0(X_{\pi(2j-1)})\right| > \sqrt{\frac{\lambda}{2}}\right\} \\
& \quad + \frac{1}{n} \sum_{1 \leq j \leq n} \mu_0^2(X_{\pi(2j)}) I\left\{\left|\mu_0(X_{\pi(2j)})\right| > \sqrt{\frac{\lambda}{2}}\right\} + \frac{1}{n} \sum_{1 \leq j \leq n} \mu_1^2(X_{\pi(2j-1)}) I\left\{\left|\mu_1(X_{\pi(2j-1)})\right| > \sqrt{\frac{\lambda}{2}}\right\} \\
\lesssim \; & \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j)}(1) I\left\{\left|Y_{\pi(2j)}(1)\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] + \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j)}(0) I\left\{\left|Y_{\pi(2j)}(0)\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] \\
& \quad + \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j-1)}(1) I\left\{\left|Y_{\pi(2j-1)}(1)\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] + \frac{1}{n} \sum_{1 \leq j \leq n} E\left[Y^2_{\pi(2j-1)}(0) I\left\{\left|Y_{\pi(2j-1)}(0)\right| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X^{(n)}\right] \\
& \quad + \frac{1}{n} \sum_{1 \leq j \leq n} \mu_1^2(X_{\pi(2j)}) I\left\{\left|\mu_1(X_{\pi(2j)})\right| > \sqrt{\frac{\lambda}{2}}\right\} + \frac{1}{n} \sum_{1 \leq j \leq n} \mu_0^2(X_{\pi(2j-1)}) I\left\{\left|\mu_0(X_{\pi(2j-1)})\right| > \sqrt{\frac{\lambda}{2}}\right\} \\
& \quad + \frac{1}{n} \sum_{1 \leq j \leq n} \mu_0^2(X_{\pi(2j)}) I\left\{\left|\mu_0(X_{\pi(2j)})\right| > \sqrt{\frac{\lambda}{2}}\right\} + \frac{1}{n} \sum_{1 \leq j \leq n} \mu_1^2(X_{\pi(2j-1)}) I\left\{\left|\mu_1(X_{\pi(2j-1)})\right| > \sqrt{\frac{\lambda}{2}}\right\} \\
\lesssim \; & \frac{1}{n} \sum_{1 \leq i \leq 2n} E\left[Y_i^2(1) I\left\{|Y_i(1)| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X_i\right] + \frac{1}{n} \sum_{1 \leq i \leq 2n} E\left[Y_i^2(0) I\left\{|Y_i(0)| > \sqrt{\frac{\lambda}{2}}\right\} \middle| X_i\right] \\
& \quad + \frac{1}{n} \sum_{1 \leq i \leq 2n} \mu_0^2(X_i) I\left\{|\mu_0(X_i)| > \sqrt{\frac{\lambda}{2}}\right\} + \frac{1}{n} \sum_{1 \leq i \leq 2n} \mu_1^2(X_i) I\left\{|\mu_1(X_i)| > \sqrt{\frac{\lambda}{2}}\right\} \\
\xrightarrow{P} \; & E\left[Y_i^2(1) I\left\{|Y_i(1)| > \sqrt{\frac{\lambda}{2}}\right\}\right] + E\left[Y_i^2(0) I\left\{|Y_i(0)| > \sqrt{\frac{\lambda}{2}}\right\}\right] \\
& \quad + E\left[\mu_1^2(X_i) I\left\{|\mu_1(X_i)| > \sqrt{\frac{\lambda}{2}}\right\}\right] + E\left[\mu_0^2(X_i) I\left\{|\mu_0(X_i)| > \sqrt{\frac{\lambda}{2}}\right\}\right]
\end{aligned}
$$

where the second inequality exploits (S.30). Since $E[Y_i^2(d)] < \infty$ and $E[\mu_d^2(X_i)] \leq E[Y_i^2(d)]$, we have that

$$
\lim_{\lambda \to \infty} E\left[\mu_d^2(X_i) I\left\{|\mu_d(X_i)| > \sqrt{\frac{\lambda}{2}}\right\}\right] = 0
$$

$$
\lim_{\lambda \to \infty} E\left[Y_i^2(1) I\left\{|Y_i(1)| > \sqrt{\frac{\lambda}{2}}\right\}\right] = 0 \; .
$$

It now follows from a subsequencing argument as in the proof of Lemma S.1.5 that (S.31) holds. Hence,

$$
\begin{aligned}
\hat{\tau}_n^2 \quad & \xrightarrow{P} \quad E[Y_i^2(1)] + E[Y_i^2(0)] - 2E[\mu_1(X_i)\mu_0(X_i)] \\
& = \quad E[\mathrm{Var}[Y_i(1)|X_i]] + E[\mathrm{Var}[Y_i(0)|X_i]] + E\left[(\mu_1(X_i) - \mu_0(X_i))^2\right]
\end{aligned}
$$

$$= \quad E[\mathrm{Var}[Y_i(1)|X_i]] + E[\mathrm{Var}[Y_i(0)|X_i]] + E\left[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2\right] \ ,$$

as desired. ■

Lemma S.1.7. *If Assumptions 2.1–2.4 hold, then*

$$\hat\lambda_n^2 \xrightarrow{P} E[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2] \ , \tag{S.34}$$

where $\hat\lambda_n$ is defined in (22).

PROOF: Let $\mu_d(X_i)$ denote $E[Y_i(d)|X_i]$ and note that

$$\begin{aligned}
& E\left[(Y_{\pi(4j-3)} - Y_{\pi(4j-2)})(Y_{\pi(4j-1)} - Y_{\pi(4j)})(D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)})\Big|X^{(n)}\right] \\
= \ & \frac{1}{4}(\mu_1(X_{\pi(4j-3)}) - \mu_0(X_{\pi(4j-2)}))(\mu_1(X_{\pi(4j-1)}) - \mu_0(X_{\pi(4j)})) \\
& -\frac{1}{4}(\mu_0(X_{\pi(4j-3)}) - \mu_1(X_{\pi(4j-2)}))(\mu_1(X_{\pi(4j-1)}) - \mu_0(X_{\pi(4j)})) \\
& -\frac{1}{4}(\mu_1(X_{\pi(4j-3)}) - \mu_0(X_{\pi(4j-2)}))(\mu_0(X_{\pi(4j-1)}) - \mu_1(X_{\pi(4j)})) \\
& +\frac{1}{4}(\mu_0(X_{\pi(4j-3)}) - \mu_1(X_{\pi(4j-2)}))(\mu_0(X_{\pi(4j-1)}) - \mu_1(X_{\pi(4j)})) \\
= \ & \frac{1}{4}\sum_{k\in\{2,3\},\ell\in\{0,1\}} \mu_1(X_{\pi(4j-k)})\mu_1(X_{\pi(4j-\ell)}) + \frac{1}{4}\sum_{k\in\{2,3\},\ell\in\{0,1\}} \mu_0(X_{\pi(4j-k)})\mu_0(X_{\pi(4j-\ell)}) \\
& -\frac{1}{4}\sum_{k\in\{2,3\},\ell\in\{0,1\}} \left(\mu_0(X_{\pi(4j-k)})\mu_1(X_{\pi(4j-\ell)}) + \mu_1(X_{\pi(4j-k)})\mu_0(X_{\pi(4j-\ell)})\right) \ .
\end{aligned}$$

Hence, in order to show that

$$E[\hat\lambda_n^2|X^{(n)}] \xrightarrow{P} E[(\mu_1(X_i) - \mu_0(X_i))^2] \ , \tag{S.35}$$

it suffices to show that

$$\frac{1}{2n}\sum_{1\le j\le\frac{n}{2}}\sum_{k\in\{2,3\},\ell\in\{0,1\}} \mu_1(X_{\pi(4j-k)})\mu_1(X_{\pi(4j-\ell)}) \xrightarrow{P} E[\mu_1^2(X_i)] \tag{S.36}$$

$$\frac{1}{2n}\sum_{1\le j\le\frac{n}{2}}\sum_{k\in\{2,3\},\ell\in\{0,1\}} \mu_0(X_{\pi(4j-k)})\mu_0(X_{\pi(4j-\ell)}) \xrightarrow{P} E[\mu_0^2(X_i)] \tag{S.37}$$

$$\begin{aligned}
& \frac{1}{2n}\sum_{1\le j\le\frac{n}{2}}\sum_{k\in\{2,3\},\ell\in\{0,1\}} \left(\mu_0(X_{\pi(4j-k)})\mu_1(X_{\pi(4j-\ell)}) + \mu_1(X_{\pi(4j-k)})\mu_0(X_{\pi(4j-\ell)})\right) \\
& \qquad \xrightarrow{P} 2E[\mu_1(X_i)\mu_0(X_i)] \ .
\end{aligned} \tag{S.38}$$

We first prove (S.36). To see this, note that

$$\begin{aligned}
\mu_1(X_{\pi(4j-k)})\mu_1(X_{\pi(4j-\ell)}) &= \mu_1^2(X_{\pi(4j-k)}) + \mu_1(X_{\pi(4j-k)})(\mu_1(X_{\pi(4j-\ell)}) - \mu_1(X_{\pi(4j-k)})) \\
\mu_1(X_{\pi(4j-k)})\mu_1(X_{\pi(4j-\ell)}) &= \mu_1^2(X_{\pi(4j-\ell)}) - \mu_1(X_{\pi(4j-\ell)})(\mu_1(X_{\pi(4j-\ell)}) - \mu_1(X_{\pi(4j-k)})) \ ,
\end{aligned}$$

so

$$\mu_1(X_{\pi(4j-k)})\mu_1(X_{\pi(4j-\ell)}) = \frac{1}{2}\mu_1^2(X_{\pi(4j-k)}) + \frac{1}{2}\mu_1^2(X_{\pi(4j-\ell)}) - \frac{1}{2}(\mu_1(X_{\pi(4j-\ell)}) - \mu_1(X_{\pi(4j-k)}))^2 \ .$$

It follows that

$$\begin{aligned}
& \frac{1}{2n}\sum_{1\le j\le\frac{n}{2}}\sum_{k\in\{2,3\},\ell\in\{0,1\}} \mu_1(X_{\pi(4j-k)})\mu_1(X_{\pi(4j-\ell)}) \\
= \ & \frac{1}{2n}\sum_{1\le i\le 2n}\mu_1^2(X_i) - \frac{1}{4n}\sum_{k\in\{2,3\},\ell\in\{0,1\}}\sum_{1\le j\le\frac{n}{2}} (\mu_1(X_{\pi(4j-\ell)}) - \mu_1(X_{\pi(4j-k)}))^2 \ .
\end{aligned}$$

But, Assumption 2.1 implies that

$$\frac{1}{4n}\sum_{k\in\{2,3\},\ell\in\{0,1\}}\sum_{1\le j\le\frac{n}{2}} (\mu_1(X_{\pi(4j-\ell)}) - \mu_1(X_{\pi(4j-k)}))^2 \lesssim \frac{1}{n}\sum_{1\le j\le\frac{n}{2}} |X_{\pi(4j-k)} - X_{\pi(4j-\ell)}|^2 \xrightarrow{P} 0 \ ,$$

where the convergence in probability to zero follows from Assumption 2.4. Since $E[\mu_1^2(X_i)] \leq E[Y_i^2(1)]$, we have that

$$\frac{1}{2n} \sum_{1 \leq i \leq 2n} \mu_1^2(X_i) \xrightarrow{P} E[\mu_1^2(X_i)] \ .$$

It thus follows that (S.36) holds. Similar arguments may be used to establish (S.37)-(S.38), from which (S.35) follows.

To complete the proof, it remains only to show that

$$\hat{\lambda}_n^2 - E[\hat{\lambda}_n^2 | X^{(n)}] \xrightarrow{P} 0 \ .$$

This fact may be established by verifying that (S.14) in Lemma S.1.3 holds in probability conditionally on $X^{(n)}$, which may be accomplished by repeated application of (S.32) and (S.33), as in the proof of Lemma S.1.6. ■

Lemma S.1.8. *Let*

$$\tilde{R}_n(t) = P\left\{ \frac{1}{\sqrt{n}} \sum_{1 \leq j \leq n} \epsilon_j (Y_{\pi(2j)} - Y_{\pi(2j-1)})(D_{\pi(2j)} - D_{\pi(2j-1)}) \leq t \middle| W^{(n)} \right\} \ ,$$

where, independently of $W^{(n)}$, $\epsilon_j, j = 1, \ldots, n$ are i.i.d. Rademacher random variables. If Assumptions 2.1–2.3 hold, then

$$\sup_{t \in \mathbf{R}} \left| \tilde{R}_n(t) - \Phi(t/\tau) \right| \xrightarrow{P} 0 \ ,$$

where

$$\tau^2 = E[\mathit{Var}[Y_i(1)|X_i]] + E[\mathit{Var}[Y_i(0)|X_i]] + E\left[(E[Y_i(1)|X_i] - E[Y_i(0)|X_i])^2 \right] \ . \tag{S.39}$$

Proof: Using the fact that $\epsilon_j, j = 1, \ldots, n$ and $\epsilon_j(D_{\pi(2j)} - D_{\pi(2j-1)}), j = 1, \ldots, n$ have the same distribution conditional on $W^{(n)}$, we have that

$$\tilde{R}_n(t) = P\left\{ \frac{1}{\sqrt{n}} \sum_{1 \leq j \leq n} \epsilon_j (Y_{\pi(2j)} - Y_{\pi(2j-1)}) \leq t \middle| W^{(n)} \right\} \ .$$

We now proceed by applying part (ii) of Lemma 11.3.3 in Lehmann and Romano (2005) with $C_{n,j} = (Y_{\pi(2j)} - Y_{\pi(2j-1)})$, which requires

$$\frac{\max_{1 \leq j \leq n} C_{n,j}^2}{\sum_{1 \leq j \leq n} C_{n,j}^2} \xrightarrow{P} 0 \ . \tag{S.40}$$

From Lemma S.1.6, we see that $\frac{1}{n} \sum_{1 \leq j \leq n} C_{n,j}^2 = \hat{\tau}_n^2 \xrightarrow{P} \tau^2 > 0$, where the inequality exploits Assumption 2.1(a). Furthermore,

$$\begin{aligned} \frac{\max_{1 \leq j \leq n} C_{n,j}^2}{n} &\lesssim \frac{\max_{1 \leq j \leq n} (Y_{\pi(2j-1)}^2 + Y_{\pi(2j)}^2)}{n} \\ &\lesssim \frac{\max_{1 \leq i \leq 2n} Y_i^2}{n} \\ &\lesssim \frac{\max_{1 \leq i \leq 2n} (Y_i(1)^2 + Y_i(0)^2)}{n} \\ &\xrightarrow{P} 0 \ , \end{aligned}$$

where the first inequality follows by exploiting the fact that $|a - b|^2 \leq 2(a^2 + b^2)$ for any real numbers a and b, the second and third inequalities follow by inspection, and the convergence in probability to zero follows from Lemma S.1.1 and Assumption 2.1(b). Hence, (S.40) holds, from which the desired conclusion now follows easily by appealing to the aforementioned lemma and Polya's theorem. ■

Lemma S.1.9. *Let*

$$\check{\nu}_n^2(\epsilon_1, \ldots, \epsilon_n) = \hat{\tau}_n^2 - \frac{1}{2}(\check{\lambda}_n^2(\epsilon_1, \ldots, \epsilon_n) + \check{\Delta}_n^2(\epsilon_1, \ldots, \epsilon_n)) \ , \tag{S.41}$$

where $\hat{\tau}_n^2$ is defined in (21),

$$\begin{aligned} &\check{\lambda}_n^2(\epsilon_1, \ldots, \epsilon_n) \\ &\quad = \frac{2}{n} \sum_{1 \leq j \leq \lfloor \frac{n}{2} \rfloor} \epsilon_{2j-1}\epsilon_{2j} (Y_{\pi(4j-3)} - Y_{\pi(4j-2)})(Y_{\pi(4j-1)} - Y_{\pi(4j)})(D_{\pi(4j-3)} - D_{\pi(4j-2)})(D_{\pi(4j-1)} - D_{\pi(4j)}) \end{aligned}$$

$$\check{\Delta}_n(\epsilon_1,\ldots,\epsilon_n) = \frac{1}{n}\sum_{1\le j\le n}\epsilon_j(Y_{\pi(2j)}-Y_{\pi(2j-1)})(D_{\pi(2j)}-D_{\pi(2j-1)}) \ ,$$

and, independently of $W^{(n)}$, $\epsilon_j, j=1,\ldots,n$ are i.i.d. Rademacher random variables. If Assumptions 2.1–2.3 hold, then

$$\check{\nu}_n^2(\epsilon_1,\ldots,\epsilon_n) \xrightarrow{P} \tau^2 \ ,$$

where τ^2 is defined in (S.39).

PROOF: From Lemma S.1.6, we see that $\hat{\tau}_n^2 \xrightarrow{P} \tau^2$. From Lemma S.1.8, we have further that $\check{\Delta}_n(\epsilon_1,\ldots,\epsilon_n) \xrightarrow{P} 0$. It therefore suffices to show that $\check{\lambda}_n^2(\epsilon_1,\ldots,\epsilon_n) \xrightarrow{P} 0$. In order to do so, note that $\check{\lambda}_n^2(\epsilon_1,\ldots,\epsilon_n)$ may be decomposed into sums of the form

$$\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}\epsilon_{2j-1}\epsilon_{2j}Y_{\pi(4j-k)}Y_{\pi(4j-\ell)}D_{\pi(4j-k')}D_{\pi(4j-\ell')} \ . \tag{S.42}$$

where $(k,k')\in\{2,3\}^2$ and $(\ell,\ell')\in\{0,1\}^2$. Furthermore, conditional on $W^{(n)}$, the terms in any such sum are independent with mean zero. We may therefore argue that any such sum tends to zero in probability by verifying that (S.14) in Lemma S.1.3 holds in probability conditional on $W^{(n)}$. To this end, note that

$$\begin{aligned}
& \frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}E\Big[|\epsilon_{2j-1}\epsilon_{2j}Y_{\pi(4j-k)}Y_{\pi(4j-\ell)}D_{\pi(4j-k')}D_{\pi(4j-\ell')}| \\
&\qquad \times I\left\{|\epsilon_{2j-1}\epsilon_{2j}Y_{\pi(4j-k)}Y_{\pi(4j-\ell)}D_{\pi(4j-k')}D_{\pi(4j-\ell')}|>\lambda\right\}\Big|W^{(n)}\Big] \\
\le\ & \frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}E\left[|Y_{\pi(4j-k)}Y_{\pi(4j-\ell)}|I\left\{|Y_{\pi(4j-k)}Y_{\pi(4j-\ell)}|>\lambda\right\}\Big|W^{(n)}\right] \\
\le\ & \frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}|Y_{\pi(4j-k)}Y_{\pi(4j-\ell)}|I\left\{|Y_{\pi(4j-k)}Y_{\pi(4j-\ell)}|>\lambda\right\} \\
\le\ & \frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}Y^2_{\pi(4j-k)}|I\left\{|Y_{\pi(4j-k)}|>\sqrt{\lambda}\right\}+\frac{2}{n}\sum_{1\le j\le\lfloor\frac{n}{2}\rfloor}Y^2_{\pi(4j-\ell)}|I\left\{|Y_{\pi(4j-\ell)}|>\sqrt{\lambda}\right\} \\
\lesssim\ & \frac{1}{n}\sum_{1\le i\le 2n}Y_i^2 I\left\{|Y_i|>\sqrt{\lambda}\right\} \\
\le\ & \frac{1}{n}\sum_{1\le i\le 2n}(Y_i^2(1)+Y_i^2(0))|I\left\{(Y_i^2(1)+Y_i^2(0))^{\frac{1}{2}}>\sqrt{\lambda}\right\} \\
\xrightarrow{P}\ & E\left[(Y_i^2(1)+Y_i^2(0))|I\left\{(Y_i^2(1)+Y_i^2(0))^{\frac{1}{2}}>\sqrt{\lambda}\right\}\right] \ ,
\end{aligned}$$

where the first inequality follows from the fact that $|\epsilon_j|=1$ for all $1\le j\le n$ and $|D_i|\le 1$ for all $1\le i\le 2n$, the second inequality exploits the fact that $\pi=\pi_n(X^{(n)})$ and both $Y^{(n)}$ and $X^{(n)}$ are contained in $W^{(n)}$, the third inequality follows from (S.33) used in the proof of Lemma S.1.6, the fourth inequality follows by inspection, the fifth inequality uses the fact that $Y_i^2\le Y_i^2(1)+Y_i^2(0)$, and the convergence in probability follows from Assumption 2.1(b). Since $E[Y_i^2(d)]<\infty$, we have that

$$\lim_{\lambda\to\infty}E\left[(Y_i^2(1)+Y_i^2(0))|I\left\{(Y_i^2(1)+Y_i^2(0))^{\frac{1}{2}}>\sqrt{\lambda}\right\}\right]=0 \ .$$

It now follows from a subsequencing argument as in the proof of Lemma S.1.5 that (S.42) tends to zero in probability. The desired result thus follows. ■

S.1.11 Additional Simulations

S.1.11.1 Power Curves

In the following figure, we present power curves for the different tests discussed in Section 5 using the specification given by Model 4 in Table 3 with $\gamma=1$.

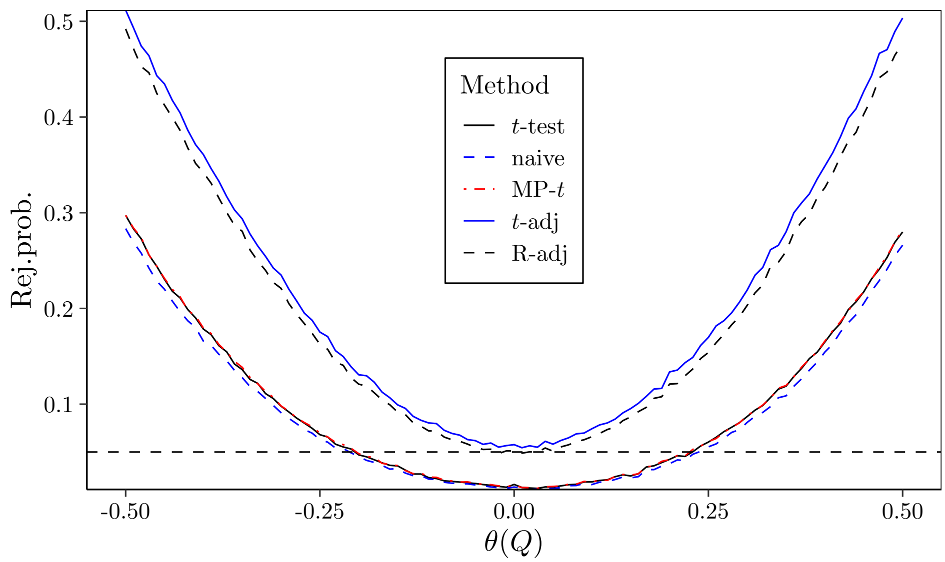


Figure S.2: Power curves for Model 4 in Table 3 with $\gamma = 1$.

S.1.11.2 $\epsilon_{d,i} \sim t_4$

In the following tables, we repeat the analysis presented in Tables 1–3, but with $\epsilon_{d,i} \sim t_4$ instead of standard normal.

	Under H_0 — $\Delta = 0$					Under H_1 — $\Delta = 1/4$				
Model	***t*-test**	**näive**	**MP-*t***	***t*-adj**	**R-adj**	***t*-test**	**näive**	**MP-*t***	***t*-adj**	**R-adj**
1	4.70	5.13	5.41	5.54	5.22	23.45	23.91	24.94	24.79	23.91
2	4.57	4.83	5.08	5.23	5.04	23.70	24.26	24.95	24.74	23.83
3	4.09	4.82	5.13	5.24	4.92	22.52	24.21	25.01	25.38	24.26
4	1.91	1.78	1.95	5.14	4.60	6.07	5.46	6.15	13.84	13.01
5	5.30	0.84	1.01	5.28	4.64	9.22	2.14	2.39	9.17	8.35
6	0.93	0.93	1.00	5.44	4.87	4.94	4.31	4.74	17.46	15.35
7	4.23	5.17	5.57	5.70	5.40	22.35	24.89	25.74	25.81	25.23
8	1.14	1.07	1.14	4.60	4.29	0.95	0.97	0.96	5.15	4.74
9	5.08	0.72	0.81	4.18	3.94	7.12	1.42	1.64	5.78	5.35
10	2.02	4.26	4.58	4.79	4.66	13.19	20.10	20.99	20.81	20.52
11	0.72	0.79	0.86	3.37	3.23	0.65	0.76	0.85	3.86	3.63
12	4.31	0.66	0.70	3.55	3.35	4.41	0.63	0.68	3.72	3.57
13	0.00	4.27	4.76	4.72	4.59	0.00	6.09	6.37	6.47	6.40
14	0.49	0.49	0.54	3.84	3.58	0.62	0.67	0.74	3.74	3.50
15	4.22	0.55	0.62	4.02	3.84	4.06	0.52	0.55	3.96	3.73

Table S.1: Rej. prob. for Models 1–15 with $\gamma = \iota_{\dim(X)}$ for Models 1–15, $\sigma_1 = 1$, $\rho = 0.2$. Errors are distributed as t_4.

	Under H_0 — $\Delta = 0$					Under H_1 — $\Delta = 1/4$				
Model	**t-test**	**näive**	**MP-t**	**t-adj**	**R-adj**	**t-test**	**näive**	**MP-t**	**t-adj**	**R-adj**
1	5.29	5.36	5.80	5.70	5.40	17.85	17.84	18.59	18.81	17.94
2	5.19	5.08	5.47	5.50	4.92	17.71	17.94	18.73	18.73	18.02
3	4.34	4.66	4.97	5.11	4.79	17.37	17.82	18.89	18.94	18.35
4	2.15	2.07	2.19	5.33	4.77	5.72	5.30	5.81	12.29	11.35
5	5.52	0.88	1.06	5.55	4.73	9.05	2.36	2.59	8.99	8.14
6	1.10	0.89	1.03	5.42	4.65	4.50	3.82	4.24	16.80	14.94
7	1.38	4.75	5.14	4.98	5.01	7.34	18.34	19.13	19.00	18.70
8	1.81	1.78	1.90	6.00	5.72	0.95	1.01	1.04	4.94	4.66
9	5.67	1.22	1.32	5.20	4.68	7.82	2.20	2.19	6.98	6.54
10	0.98	3.13	3.41	3.55	3.48	6.16	12.99	13.45	13.42	13.26
11	0.57	0.52	0.66	3.60	3.42	0.47	0.49	0.57	3.73	3.51
12	3.70	0.47	0.49	3.43	3.23	3.86	0.45	0.50	3.80	3.51
13	0.00	3.52	3.81	3.81	3.63	0.00	5.98	6.37	6.53	6.21
14	0.46	0.50	0.54	4.14	3.79	0.35	0.44	0.47	4.02	3.72
15	3.98	0.51	0.55	4.17	3.89	3.88	0.34	0.37	3.92	3.70

Table S.2: Rej. prob. for Models 1–15 with $\gamma = 1$ for Models 1–6, $\gamma' = (1, 4)$ for Models 7–9, $\gamma = \iota_{\dim(X)}$ for Models 10–15, $\sigma_1 = 2$, $\rho = 0.7$. Errors are distributed as t_4.

	Under H_0 — $\Delta = 0$					Under H_1 — $\Delta = 1/4$				
Model	**t-test**	**näive**	**MP-t**	**t-adj**	**R-adj**	**t-test**	**näive**	**MP-t**	**t-adj**	**R-adj**
1	4.49	4.96	5.15	5.06	4.85	23.13	23.74	24.66	24.63	23.95
2	4.42	5.09	5.31	5.35	5.00	24.44	24.75	25.62	25.56	24.69
3	3.79	4.70	4.84	5.05	4.78	22.90	24.89	25.88	26.22	25.14
4	1.59	1.55	1.72	4.94	4.37	5.77	5.30	5.96	13.52	12.38
5	5.27	0.80	0.90	5.20	4.67	9.10	2.11	2.28	9.11	8.36
6	0.72	0.62	0.71	4.77	4.24	4.79	4.07	4.58	18.41	16.77
7	1.03	5.14	5.53	5.46	5.04	10.23	24.15	24.84	24.98	24.43
8	0.90	0.85	0.92	4.39	4.12	1.26	1.28	1.38	5.85	5.47
9	5.67	0.62	0.73	4.68	4.28	7.15	1.17	1.40	5.67	5.29
10	2.42	3.78	3.95	4.08	3.95	14.38	17.60	18.39	18.28	17.98
11	0.73	0.80	0.88	3.30	3.01	0.81	0.82	0.91	3.42	3.16
12	3.95	0.62	0.64	2.66	2.44	3.69	0.63	0.69	2.72	2.58
13	2.12	3.56	3.82	3.93	3.73	4.47	6.52	6.94	7.21	6.95
14	2.86	2.69	2.98	3.42	3.28	3.08	2.92	3.17	3.46	3.35
15	3.56	1.92	2.08	2.92	2.62	3.83	2.06	2.37	3.05	2.90

Table S.3: Rej. prob. for Models 1–12 with $\gamma = 1$ for Models 1–6, $\gamma' = (4, 1)$ for Models 7–9, $\gamma = \iota_{\dim(X)}$ for Models 10–15, $\sigma_1 = 1$, $\rho = 0$. Errors are distributed as t_4.

S.1.11.3 $n = 40$

In the following table, we repeat the analysis presented in Table 1, but with $n = 40$ instead of $n = 100$.

	Under H_0 — $\Delta = 0$					Under H_1 — $\Delta = 1/4$				
Model	***t*-test**	**näive**	**MP-*t***	***t*-adj**	**R-adj**	***t*-test**	**näive**	**MP-*t***	***t*-adj**	**R-adj**
1	4.90	5.16	6.07	6.24	5.24	19.65	20.29	22.65	22.72	20.17
2	4.65	4.80	5.92	5.93	4.90	19.24	19.16	21.60	21.44	19.04
3	4.03	5.04	5.85	6.17	5.14	16.60	18.71	21.05	21.32	19.00
4	1.57	1.27	1.73	5.61	4.28	2.60	2.18	2.79	9.65	7.16
5	5.41	0.76	1.08	5.39	4.03	7.64	1.10	1.43	7.50	5.85
6	1.14	0.80	1.23	6.02	4.44	1.88	1.27	1.72	9.74	7.14
7	3.46	4.81	5.65	5.98	5.36	15.92	18.66	20.90	20.77	19.35
8	1.48	1.28	1.53	4.58	3.93	1.09	0.83	1.09	4.17	3.59
9	5.71	0.92	1.16	4.14	3.62	7.01	1.30	1.62	4.95	4.40
10	2.12	4.51	4.79	4.69	4.68	13.51	20.35	21.11	20.99	20.54
11	0.67	0.69	0.75	3.76	3.59	0.71	0.80	0.92	3.82	3.62
12	4.32	0.55	0.64	3.69	3.39	4.39	0.61	0.73	3.66	3.37
13	0.00	4.50	4.66	4.79	4.56	0.00	10.38	10.98	11.17	10.80
14	0.72	0.77	0.86	4.13	3.88	0.65	0.71	0.77	3.88	3.55
15	4.09	0.53	0.59	3.80	3.50	4.18	0.64	0.71	3.96	3.60

Table S.4: Rej. prob. for Models 1–15 with $\gamma = \iota_{\dim(X)}$ for Models 1–15, $\sigma_1 = 1$, $\rho = 0.2$. Sample size is $2n = 80$.

References

Chung, E. and Romano, J. P. (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics*, **41** 484–507.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. 3rd ed. Springer, New York.