On-Line Appendix

"Matching on the Estimated Propensity Score" (Abadie and Imbens, 2015)

Alberto Abadie and Guido W. Imbens
Current version: August 10, 2015

The first part of this appendix contains additional proofs. The second part reports the results of a Monte Carlo study that confirms the theoretical properties of the propensity score matching estimators derived in the article.

I. Additional Proofs:

We first state and prove a number of preliminary results. For real numbers a, $\lfloor a \rfloor$ is the largest integer less than or equal to a, and $\lceil a \rceil$ is the smallest integer greater than or equal to a. If a is an integer, then $\lceil a \rceil = \lfloor a \rfloor$, otherwise $\lceil a \rceil = \lfloor a \rfloor + 1$.

Lemma A.1 *Consider two independent samples of sizes n_0 and n_1 from continuous distributions F_0 and F_1 with common support: $X_{0,1}, \ldots, X_{0,n_0} \sim$ i.i.d. F_0 and $X_{1,1}, \ldots, X_{1,n_1} \sim$ i.i.d. F_1. Let $N = n_0 + n_1$. Assume that the support of F_0 and F_1 is an interval inside $[0,1]$. Let f_0 and f_1 be the densities of F_0 and F_1, respectively. Suppose that for any x in the supports of F_0 and F_1, $f_1(x)/f_0(x) \leq \bar{r}$. For $1 \leq i \leq n_1$ and $1 \leq m \leq M \leq n_0$, let $|U_{n_0,n_1,i}|_{(m)}$ be the m-th order statistic of $\{|X_{1,i} - X_{0,1}|, \ldots, |X_{1,i} - X_{0,n_0}|\}$. Then, for $n_0 \geq 3$:*

$$E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{n_1}\frac{1}{M}\sum_{m=1}^{M}|U_{n_0,n_1,i}|_{(m)}\right] \leq \bar{r}\,\frac{n_1}{N^{1/2}\lfloor n_0^{3/4}\rfloor} + M\frac{n_1}{N^{1/2}}\,n_0^{M-1/4}\exp(-n_0^{1/4}).$$

Proof of Lemma A.1: Consider N balls assigned at random among n bins of equal probability. It is known that the mean of the number of bins with exactly m balls is equal to

$$n\begin{pmatrix} N \\ m \end{pmatrix}\left(\frac{1}{n}\right)^m\left(1-\frac{1}{n}\right)^{N-m}$$

(see Johnson and Kotz, 1977). Because $f_1(x)/f_0(x) \leq \bar{r}$, for any measurable set A:

$$\Pr(X_{1,i} \in A) = \int_A f_1(x)\,dx = \int_A \left(\frac{f_1(x)}{f_0(x)}\right) f_0(x)\,dx \leq \bar{r}\Pr(X_{0,i} \in A).$$

Divide the support of F_0 and F_1 in $\lfloor n_0^{3/4}\rfloor$ cells of equal probability, $1/\lfloor n_0^{3/4}\rfloor$, under F_0. Let Z_{M,n_0} be the number of such cells that are not occupied by at least M observations from the sample: $X_{0,1}, \ldots, X_{0,n_0}$. Let $\mu_{M,n_0} = E[Z_{M,n_0}]$. Notice that $n_0 \geq 3$ implies $\lfloor n_0^{3/4}\rfloor \geq 2$. Then,

$$\begin{aligned}
\mu_{M,n_0} &= \sum_{m=0}^{M-1}\lfloor n_0^{3/4}\rfloor\begin{pmatrix} n_0 \\ m \end{pmatrix}\left(\frac{1}{\lfloor n_0^{3/4}\rfloor}\right)^m\left(1-\frac{1}{\lfloor n_0^{3/4}\rfloor}\right)^{n_0-m} \\
&\leq \sum_{m=0}^{M-1}\lfloor n_0^{3/4}\rfloor\frac{n_0^m}{m!}\left(\frac{1}{\lfloor n_0^{3/4}\rfloor}\right)^m\left(1-\frac{1}{\lfloor n_0^{3/4}\rfloor}\right)^{n_0-m} \\
&\leq Mn_0^{M-1/4}\left(1-\frac{1}{n_0^{3/4}}\right)^{n_0}.
\end{aligned}$$

Using Markov's inequality,

$$\Pr(Z_{M,n_0} > 0) = \Pr(Z_{M,n_0} \geq 1) \leq \mu_{M,n_0} \leq M n_0^{M-1/4} \left(1 - \frac{1}{n_0^{3/4}}\right)^{n_0}.$$

Notice that for any positive a, we have that $a - 1 \geq \log(a)$. Therefore, for any $b < N$, we have that $\log(1 - b/N) \leq -b/N$ and $(1 - b/N)^N \leq \exp(-b)$. As a result, we obtain:

$$\left(1 - \frac{1}{n_0^{3/4}}\right)^{n_0} = \left(1 - \frac{n_0^{1/4}}{n_0}\right)^{n_0} \leq \exp(-n_0^{1/4}).$$

Putting together the last two displayed equations, we obtain the following exponential bound for $\Pr(Z_{M,n_0} > 0)$:

$$\Pr(Z_{M,n_0} > 0) \leq M n_0^{M-1/4} \exp(-n_0^{1/4}).$$

Notice that $|U_{n_0,n_1,i}|_{(m)} \leq 1$. For $0 \leq n \leq \lfloor n_0^{3/4} \rfloor$, let $c_{n_0,n}$ be the point in the support of F_0 such that $c_{n_0,n} = F_0^{-1}(n/\lfloor n_0^{3/4} \rfloor)$, then

$$\begin{aligned}
E\left[\sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \,\Big|\, Z_{M,n_0} = 0\right] &\leq \sum_{n=1}^{\lfloor n_0^{3/4} \rfloor} M\big(c_{n_0,n} - c_{n_0,n-1}\big) \Pr\big(c_{n_0,n-1} \leq X_{1,i} \leq c_{n_0,n} \big| Z_{M,n_0} = 0\big) \\
&\leq \frac{M\bar{r}}{\lfloor n_0^{3/4} \rfloor} \sum_{n=1}^{\lfloor n_0^{3/4} \rfloor} \big(c_{n_0,n} - c_{n_0,n-1}\big) \\
&\leq \frac{M\bar{r}}{\lfloor n_0^{3/4} \rfloor}.
\end{aligned}$$

Now,

$$\begin{aligned}
E\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} \frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)}\right] &= E\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} \frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \Big| Z_{M,n_0} = 0\right] \Pr(Z_{M,n_0} = 0) \\
&+ E\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} \frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \Big| Z_{M,n_0} > 0\right] \Pr(Z_{M,n_0} > 0) \\
&\leq E\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} \frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \Big| Z_{M,n_0} = 0\right] \\
&+ \frac{n_1}{N^{1/2}} \Pr(Z_{M,n_0} > 0) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} E\left[\frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \Big| Z_{M,n_0} = 0\right] \\
&+ \frac{n_1}{N^{1/2}} \Pr(Z_{M,n_0} > 0) \\
&\leq \bar{r} \frac{n_1}{N^{1/2} \lfloor n_0^{3/4} \rfloor} + M \frac{n_1}{N^{1/2}} n_0^{M-1/4} \exp(-n_0^{1/4}).
\end{aligned}$$

□

LEMMA A.2 *Suppose that the propensity score $p(X) = \Pr(W = 1|X)$ is continuously distributed, with continuous density, $f(p)$, and interval support, $[\underline{p}, \overline{p}]$, with $\underline{p} > 0$ and $\overline{p} < 1$. Let $f_w(p)$ be* *the density of the propensity score conditional on $W = w$, where $w \in \{0, 1\}$. Then, the densities $f_1(p)$ and $f_0(p)$ are continuous and share a common support. Moreover, the ratio $f_1(p)/f_0(p)$ is continuous, uniformly bounded, and uniformly bounded away from zero for all p such that $f(p) > 0$.*

PROOF OF LEMMA A.2: Applying Bayes' Theorem, for all $0 \leq p \leq 1$,

$$\begin{aligned} f_w(p) &= \frac{f(p)\Pr(W = w|p(X) = p)}{\Pr(W = w)} \\ &= \frac{p}{\Pr(W = w)} f(p). \end{aligned}$$

Therefore, the functions $f_1(p)$ and $f_0(p)$ are continuous, and the supports of $f_1(p)$ and $f_0(p)$ are equal to the support of $f(p)$. Now, for any p such that $f(p) > 0$, we obtain,

$$\frac{f_1(p)}{f_0(p)} = \frac{p}{1-p}\frac{\Pr(W = 0)}{\Pr(W = 1)} < \frac{\overline{p}}{1-\overline{p}}\frac{1-\underline{p}}{\underline{p}}.$$

Similarly,

$$\frac{f_0(p)}{f_1(p)} < \frac{1-\underline{p}}{\underline{p}}\frac{\overline{p}}{1-\overline{p}}.$$

Therefore, we obtain $\overline{\eta} = (\overline{p}\,(1-\underline{p}))/((1-\overline{p})\,\underline{p}) > 1$, and

$$\frac{1}{\overline{\eta}} < \frac{f_1(p)}{f_0(p)} < \overline{\eta}.$$

$\square$

LEMMA A.3 *(Inverse Moments of the Doubly Truncated Binomial Distribution) Let N_0 be a Binomial variable with parameters $(N, (1-p))$ that is left-truncated for values smaller than M and right-truncated for values greater than $N - M$, where $M < N/2$. Let $N_1 = N - N_0$. Then, for any $r > 0$, there exist a constant C_r, such that*

$$E\left[\left(\frac{N}{N_0}\right)^r\right] \leq C_r \quad \textit{and} \quad E\left[\left(\frac{N}{N_1}\right)^r\right] \leq C_r$$

for all $N > 2M$.

PROOF OF LEMMA A.3: Here, we prove the first assertion of the lemma. The proof of the second assertion is analogous. Let $N_1 = N - N_0$. Let $\bar{q}$ be a scalar greater than one. Then,

$$\begin{aligned} E\left[\left(\frac{N}{N_0}\right)^r\right] &= E\left[\left(\frac{N}{N_0}\right)^r 1\left\{\frac{N}{N_0} > \bar{q}\right\}\right] + E\left[\left(\frac{N}{N_0}\right)^r 1\left\{\frac{N}{N_0} \leq \bar{q}\right\}\right] \\ &\leq \left(\frac{N}{M}\right)^r \Pr\left(\frac{N}{N_0} > \bar{q}\right) + \bar{q}^r \\ &= \left(\frac{N}{M}\right)^r \Pr\left(N_1 > \left(1 - \frac{1}{\bar{q}}\right)N\right) + \bar{q}^r. \end{aligned}$$

Notice that:

$$\Pr\left(N_1 > \left(1 - \frac{1}{\bar{q}}\right)N\right) = \frac{\displaystyle\sum_{x>(1-1/\bar{q})N,\, x\geq M}^{x\leq N-M} \binom{N}{x} p^x(1-p)^{N-x}}{\displaystyle\sum_{x\geq M}^{x\leq N-M} \binom{N}{x} p^x(1-p)^{N-x}}.$$

For $N > 2M$ the denominator can be bounded away from zero. Therefore, for some positive constant C, and $\bar{q} > 1/(1-p)$,

$$\begin{aligned}
\Pr\left(N_1 > \left(1 - \frac{1}{\bar{q}}\right) N\right) &\leq C \sum_{x > (1-1/\bar{q})N, x \geq M}^{x \leq N-M} \binom{N}{x} p^x (1-p)^{N-x} \\
&\leq C \sum_{x > (1-1/\bar{q})N} \binom{N}{x} p^x (1-p)^{N-x} \\
&\leq C \exp\left\{-2(1 - 1/\bar{q} - p)^2 N\right\},
\end{aligned}$$

by Hoeffding's Inequality (e.g. van der Vaart and Wellner, 1996, p. 459). Therefore $E[(N/N_0)^r]$ is uniformly bounded for $N > 2M$. $\square$

For a sample of scalars $X_1, \ldots, X_N$ from a cumulative distribution function $F : [a, b] \mapsto [0, 1]$, let $\widehat{F}$ be the empirical cumulative distribution function:

$$\widehat{F}(x) = \frac{1}{N} \sum_{i=1}^{N} 1_{(-\infty, x]}(X_i).$$

Let $\widehat{F}^{-1}$ be the empirical inverse cumulative distribution function:

$$\widehat{F}^{-1}(q) = \inf_{a \leq x \leq b} \left\{x : \widehat{F}(x) \geq q\right\}.$$

Let $\xi_{1:N}, \ldots, \xi_{N:N}$ be the order statistics for a random sample of size N from the uniform distribution. Let $\widehat{G}$ and $\widehat{G}^{-1}$ be the empirical cumulative distribution function for that sample and its inverse.

LEMMA A.4 *Suppose $F : [a, b] \mapsto [0, 1]$ is a continuous and strictly increasing cumulative distribution function with $F(a) = 0$. Then,*

(i)

$$\sup_{0 \leq q \leq 1} \left|\widehat{F}^{-1}(q) - F^{-1}(q)\right| \xrightarrow{a.s.} 0,$$

(ii)

$$\max_{i=1:N} \left|F^{-1}(\xi_{i:N}) - F^{-1}(i/N)\right| \xrightarrow{a.s.} 0,$$

(iii) for any two integers J and M:

$$\max_{i=J+1:N-M} \left|F^{-1}(\xi_{i+M:N}) - F^{-1}(\xi_{i-J:N})\right| \xrightarrow{a.s.} 0.$$

PROOF OF LEMMA A.4: First, notice that F^{-1} is uniformly continuous, because it is a continuous function defined on a compact set (e.g., Rudin (1976), theorems 4.17 and 4.19). Because $\widehat{F}^{-1}$ has the same distribution as $F^{-1} \circ \widehat{G}^{-1}$ and

$$\sup_{0 \leq q \leq 1} |\widehat{G}^{-1}(q) - q| \xrightarrow{a.s.} 0$$

(see Shorack and Wellner, 2009, page 95), the result in part *(i)* is implied by uniform continuity of F^{-1} as follows. Fix $\delta > 0$. Because if F^{-1} is uniformly continuous, then for each $\delta > 0$, there exists a $\varepsilon > 0$ such that if $|p-q| < \varepsilon$, then $|F^{-1}(p) - F^{-1}(q)| < \delta$. With probability one, for each $\varepsilon > 0$ there exists $N(\varepsilon)$ such that for $N \geq N(\varepsilon)$,

$$\sup_{0 \leq q \leq 1} \left| \widehat{G}^{-1}(q) - q \right| < \varepsilon.$$

Therefore, with probability one, there exists $N(\varepsilon)$ such that for $N \geq N(\varepsilon)$,

$$\sup_{0 \leq q \leq 1} \left| F^{-1}(\widehat{G}^{-1}(q)) - F^{-1}(q) \right| < \delta,$$

which proves *(i)*. Part *(ii)* is directly implied by *(i)* and by the fact that $\widehat{F}^{-1}$ has the same distribution as $F^{-1} \circ \widehat{G}^{-1}$. To prove *(iii)* notice that

$$\begin{aligned}\max_{i=J+1:N-M} \left| F^{-1}(\xi_{i+M:N}) - F^{-1}(\xi_{i-J:N}) \right| \leq & \max_{i=J+1:N-M} \left| F^{-1}(\xi_{i+M:N}) - F^{-1}((i+M)/N) \right| \\ & + \max_{i=J+1:N-M} \left| F^{-1}(\xi_{i-J:N}) - F^{-1}((i-J)/N) \right| \\ & + \max_{i=J+1:N-M} \left| F^{-1}((i+M)/N) - F^{-1}((i-J)/N) \right|.\end{aligned}$$

Result *(ii)* implies that the first two terms on the right-hand-side of last equation converge almost surely to zero. Then, uniform continuity of F^{-1} implies that the last term on the right-hand-side of last equation converges to zero, which proves *(iii)*. □

LEMMA A.5 *Let $Y_1, \ldots Y_{N+1}$ be independent and distributed as standard exponential (equivalently, $\Gamma(1,1)$, where Γ denotes the Gamma distribution with parameters $(1,1)$). Let $S_j = \sum_{i=1}^{j} Y_i$ and $S_{N+1,j} = \sum_{i=1}^{j} Y_i / \sum_{i=1}^{N+1} Y_i$, for $1 \leq j \leq N+1$. Let k denote a positive integer. Then:*

(i) S_j has a Gamma distribution with parameters $(j,1)$ and moments

$$E[(S_j)^k] = \frac{(j+k-1)!}{(j-1)!}$$

(iii) $S_{N+1,j}$ has a Beta distribution with parameters $(j, N-j+1)$ and moments:

$$E[(S_{N+1,j})^k] = \frac{(j+k-1)!N!}{(j-1)!(N+k)!}.$$

PROOF OF LEMMA A.5: See, e.g., Poirier (1995), pages 98-106. □

LEMMA A.6 *Suppose $F : [a,b] \mapsto [0,1]$ is a strictly increasing and absolutely continuous cumulative distribution function with $F(a) = 0$ and derivative $f(x)$. Suppose $m : [a,b] \mapsto \mathbb{R}$ is non-negative and continuous (hence, bounded on $[a,b]$). Then, for any nonnegative integer, M:*

$$\sum_{i=M+1}^{N-M} m\left(F^{-1}\left(\xi_{i:N}\right)\right)\left(\xi_{i+M:N} - \xi_{i-M:N}\right) \xrightarrow{p} 2M \int_a^b m(s) f(s) ds,$$

and

$$\sum_{i=M+1}^{N-M} m\left(F^{-1}\left(\xi_{i:N}\right)\right) N \left(\xi_{i+M:N} - \xi_{i-M:N}\right)^2 \xrightarrow{p} 2M(2M+1) \int_a^b m(s) f(s) ds.$$

Proof of Lemma A.6: We first prove three results,

$$\sum_{i=M+1}^{N-M}\left(m\left(F^{-1}\left(\xi_{i:N}\right)\right)-m\left(F^{-1}\left(i/N\right)\right)\right)\left(\xi_{i+M:N}-\xi_{i-M:N}\right)=o_p(1), \tag{A.1}$$

$$\sum_{i=M+1}^{N-M} m\left(F^{-1}\left(i/N\right)\right)\left(\xi_{i+M:N}-\xi_{i-M:N}\right)-\frac{2M}{N}\sum_{i=M+1}^{N-M} m\left(F^{-1}\left(i/N\right)\right)=o_p(1), \tag{A.2}$$

and

$$\frac{1}{N}\sum_{i=M+1}^{N-M} m\left(F^{-1}\left(i/N\right)\right)-\int_a^b m(s)f(s)ds=o(1), \tag{A.3}$$

which together imply the first result in the lemma.

First consider (A.1). For N large enough, we have $M+1\leq N-M$. Let k be an integer such that $-M\leq k\leq M-1$. It is enough to prove

$$\sum_{i=M+1}^{N-M}\left(m\left(F^{-1}\left(\xi_{i:N}\right)\right)-m\left(F^{-1}\left(i/N\right)\right)\right)\left(\xi_{i+k+1:N}-\xi_{i+k:N}\right)=o_p(1).$$

Then, equation (A.1) follows from summation from $k=-M$ to $k=M-1$. Because m is uniformly continuous on $[a,b]$, then for any $\varepsilon>0$, there exists a $\delta>0$ such that if x_1 and x_2 belong to $[a,b]$ and $|x_2-x_1|<\delta$, then $|m(x_2)-m(x_1)|<\varepsilon$. Consider $\varepsilon>0$. Then, there exists a $\delta>0$ such that, applying Lemma A.4(ii) we obtain:

$$\begin{aligned}&\Pr\left(\max_{i=1,\dots,N}\left|m\left(F^{-1}\left(\xi_{i:N}\right)\right)-m\left(F^{-1}\left(i/N\right)\right)\right|<\varepsilon\right)\\&\qquad\geq\Pr\left(\max_{i=1,\dots,N}\left|F^{-1}\left(\xi_{i:N}\right)-F^{-1}\left(i/N\right)\right|<\delta\right)\longrightarrow 1.\end{aligned}$$

Because this derivation holds for any $\varepsilon>0$, we obtain:

$$\max_{i=1,\dots,N}\left|m\left(F^{-1}\left(\xi_{i:N}\right)\right)-m\left(F^{-1}\left(i/N\right)\right)\right|\xrightarrow{p}0. \tag{A.4}$$

Therefore

$$\begin{aligned}&\left|\sum_{i=M+1}^{N-M}\left(m\left(F^{-1}\left(\xi_{i:N}\right)\right)-m\left(F^{-1}\left(i/N\right)\right)\right)\left(\xi_{i+k+1:N}-\xi_{i+k:N}\right)\right|\\&\qquad<\max_{i=1,\dots,N}\left|m\left(F^{-1}\left(\xi_{i:N}\right)\right)-m\left(F^{-1}\left(i/N\right)\right)\right|\left|\sum_{i=M+1}^{N-M}\left(\xi_{i+k+1:N}-\xi_{i+k:N}\right)\right|\xrightarrow{p}0,\end{aligned}$$

which proves (A.1). Now consider (A.2). As before, it is enough to prove:

$$\sum_{i=M+1}^{N-M} m\left(F^{-1}\left(i/N\right)\right)\left(\xi_{i+k+1:N}-\xi_{i+k:N}\right)-\frac{1}{N}\sum_{i=M+1}^{N-M} m\left(F^{-1}\left(i/N\right)\right)=o_p(1),$$

for $-M\leq k\leq M-1$. For $i=1,\dots,N+1$, define the spacings $\delta_{Ni}=\xi_{i:N}-\xi_{i-1:N}$ (with $\xi_{0:N}=0$ and $\xi_{N+1:N}=1$). The left-hand-side of last equation can be expressed as:

$$\sum_{i=M+1}^{N-M} m\left(F^{-1}\left(i/N\right)\right)\left(\delta_{Ni+k+1}-\frac{1}{N}\right).$$

We will show that the last equation converges to zero in probability by showing that its mean and variance converge to zero. Using Lemma A.5, the representation of uniform spacings via exponential variables (see, e.g., Shorack and Wellner, 2009, page 721), we obtain:

$$E[\delta_{Ni+k+1}] = \frac{1}{N+1}, \qquad \mathrm{var}(\delta_{Ni+k+1}) = \frac{N}{(N+1)^2(N+2)} < \frac{1}{N^2}. \tag{A.5}$$

Notice also that for any $k, l \in \{1, \cdots, N\}$ with $k < l$, $x \in [0,1]$ and $y \in [0, 1-x]$, we have:

$$\Pr\left(\sum_{j=k+1}^{l} \delta_{Nj} \leq y \,\middle|\, \sum_{j=1}^{k} \delta_{Nj} = x \right) = \Pr(\xi_{Nl} - \xi_{Nk} \leq y | \xi_{Nk} = x).$$

The distribution of $\xi_{Nl} - \xi_{Nk}$ conditional on $\xi_{Nk} = x$ is the same as the distribution of $(l-k)$-th order statistic in a sample of size $N-k$ from a uniform distribution on $[0, 1-x]$. As a result, the probability of the event $\xi_{Nl} - \xi_{Nk} \leq y$ conditional on $\xi_{Nk} = x$ is non-decreasing in x. Therefore, $\sum_{j=k+1}^{l} \delta_{Nj}$ is negatively regression dependent on $\sum_{j=1}^{k} \delta_{Nj}$ (see, e.g., Lehmann, 1966). By exchangeability, this result extends to any two sums S_N, Q_N of disjoint sets of uniform spacings. Then, negative regression dependence implies that for any non-decreasing function, $e()$:

$$\mathrm{cov}(e(S_N), e(Q_N)) < 0. \tag{A.6}$$

Because m is bounded on $[a, b]$,

$$E\left[\sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right) \left(\delta_{i+k+1:N} - \frac{1}{N} \right) \right] = \frac{-1}{N(N+1)} \sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right) \to 0.$$

Because m is non-negative,

$$\begin{aligned} \mathrm{var}\left(\sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right) \left(\delta_{i+k+1:N} - \frac{1}{N} \right) \right) &\leq \sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right)^2 \mathrm{var}\left(\delta_{i+k+1:N} - \frac{1}{N} \right) \\ &\leq \frac{1}{N^2} \sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right)^2 \to 0. \end{aligned}$$

Therefore (A.2) holds. Finally, consider (A.3):

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right) = \int_0^1 m(F^{-1}(y))dy = \int_a^b m(s)f(s)d(s).$$

Next we will prove,

$$\sum_{i=M+1}^{N-M} \left(m\left(F^{-1}(\xi_{i:N})\right) - m\left(F^{-1}(i/N)\right) \right) N \left(\xi_{i+M:N} - \xi_{i-M:N}\right)^2 = o_p(1), \tag{A.7}$$

$$\sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right) N \left(\xi_{i+M:N} - \xi_{i-M:N}\right)^2 - \frac{2M(2M+1)}{N} \sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right) = o_p(1), \tag{A.8}$$

which along with (A.3) imply the second result in the Lemma. Because of equation (A.4), to prove equation (A.7) it is sufficient to prove that

$$\sum_{i=M+1}^{N-M} N \left(\xi_{i+M:N} - \xi_{i-M:N}\right)^2$$

is bounded in probability. By Lemma A.5, for $N \geq 2M$:

$$E\left[\sum_{i=M+1}^{N-M} N\left(\xi_{i+M:N} - \xi_{i-M:N}\right)^2\right] = N\frac{2M(2M+1)}{(N+1)(N+2)}(N-2M) \longrightarrow 2M(2M+1).$$

Also, because of equation (A.6), exchangeability, and the Cauchy-Schwarz inequality:

$$\begin{aligned}
&\operatorname{var}\left(\sum_{i=M+1}^{N-M} N\left(\xi_{i+M:N} - \xi_{i-M:N}\right)^2\right) \\
&\quad = \operatorname{var}\left(\sum_{i=M+1}^{N-M} N\left(\delta_{Ni+M} + \cdots + \delta_{Ni-M+1}\right)^2\right) \\
&\quad \leq \sum_{i=M+1}^{N-M} \operatorname{var}\left(N\left(\delta_{Ni+M} + \cdots + \delta_{Ni-M+1}\right)^2\right) \\
&\quad + \sum_{i=M+1}^{N-M} \sum_{\substack{0 \vee i-2M+1 \\ j \neq i}}^{N+1 \wedge i+2M-1} \operatorname{cov}(N\left(\delta_{Ni+M} + \cdots + \delta_{Ni-M+1}\right)^2, N\left(\delta_{Nj+M} + \cdots + \delta_{Nj-M+1}\right)^2) \\
&\quad \leq \sum_{i=M+1}^{N-M} \operatorname{var}\left(N\left(\delta_{Ni-M+1} + \cdots + \delta_{Ni+M}\right)^2\right) \\
&\quad + (4M-2)\sum_{i=M+1}^{N-M} \operatorname{var}\left(N\left(\delta_{Ni-M+1} + \cdots + \delta_{Ni+M}\right)^2\right).
\end{aligned}$$

Lemma A.5 implies:

$$\begin{aligned}
\operatorname{var}\left(N\left(\delta_{Ni-M+1} + \cdots + \delta_{Ni+M}\right)^2\right) &< N^2 E\left[\left(\delta_{Ni-M+1} + \cdots + \delta_{Ni+M}\right)^4\right] \\
&= N^2\frac{(2M+3)!N!}{(2M-1)!(N+4)!}.
\end{aligned}$$

Therefore,

$$\operatorname{var}\left(\sum_{i=M+1}^{N-M} N\left(\xi_{i+M:N} - \xi_{i-M:N}\right)^2\right) \longrightarrow 0,$$

which proves (A.7). Using Lemma A.5, notice that the expectation of left hand side of equation (A.8) is:

$$\begin{aligned}
&\sum_{i=M+1}^{N-M} E\left[m\left(F^{-1}(i/N)\right)\left(N\left(\xi_{i+M:N} - \xi_{i-M:N}\right)^2 - \frac{2M(2M+1)}{N}\right)\right] \\
&\qquad = \left(\frac{2M(2M+1)N}{(N+1)(N+2)} - \frac{2M(2M+1)}{N}\right)\sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right) \to 0. \qquad \text{(A.9)}
\end{aligned}$$

Using equation (A.6), Cauchy-Schwarz inequality, exchangeability of the uniform spacings, and boundedness of m, we obtain that for some constant C_m:

$$
\begin{aligned}
&\operatorname{var}\left(\sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right)\left(N\left(\xi_{i+M:N}-\xi_{i-M:N}\right)^2\right)\right)\\
&\quad\leq C_m^2\sum_{i=M+1}^{N-M}\operatorname{var}\left(N\left(\delta_{Ni+M}+\cdots+\delta_{Ni-M+1}\right)^2\right)\\
&\quad+\sum_{i=M+1}^{N-M} m\left(F^{-1}(i/N)\right)\sum_{\substack{0\vee i-2M+1\\ j\neq i}}^{N+1\wedge i+2M-1} m\left(F^{-1}(j/N)\right)\operatorname{cov}(N\left(\delta_{Ni+M}+\cdots+\delta_{Ni-M+1}\right)^2, N\left(\delta_{Nj+M}+\cdots+\delta_{Nj-M+1}\right)^2)\\
&\quad\leq C_m^2\sum_{i=M+1}^{N-M}\operatorname{var}\left(N\left(\delta_{Ni-M+1}+\cdots+\delta_{Ni+M}\right)^2\right)\\
&\quad+(4M-2)C_m^2\sum_{i=M+1}^{N-M}\operatorname{var}\left(N\left(\delta_{Ni-M+1}+\cdots+\delta_{Ni+M}\right)^2\right)\longrightarrow 0,
\end{aligned}
$$

which proves (A.8). □

Lemma A.7 *Let X be a scalar random variable with interval support $[a,b]$, distribution function G, and density function g that is continuous on $[a,b]$. For a random sample $X_1,\ldots,X_N$, let $X_{j:N}$ be the j-th order statistic. We will adopt the convention $X_{j:N}=a$ if $j<1$ and $X_{j:N}=b$ if $j>N$. Let V_{Nk} be the rank of observation k in the sample. Let P_{Nk} be the probability that observation k will be a match for an out-of sample observation with continuous density f:*

$$
\begin{aligned}
P_{Nk}=&\int_a^{(X_{V_{Nk}:N}+X_{V_{Nk}+M:N})/2} f(x)\,dx\,1\{V_{Nk}\leq M\}\\
&+\int_{(X_{V_{Nk}:N}+X_{V_{Nk}-M:N})/2}^{(X_{V_{Nk}:N}+X_{V_{Nk}+M:N})/2} f(x)\,dx\,1\{M+1\leq V_{Nk}\leq N-M\}\\
&+\int_{(X_{V_{Nk}:N}+X_{V_{Nk}-M:N})/2}^{b} f(x)\,dx\,1\{N-M+1\leq V_{Nk}\}.
\end{aligned}\tag{A.10}
$$

Assume that f/g is bounded on $[a,b]$. Let σ^2 be a bounded function with domain $[a,b]$. Then,

$$
\sum_{k=1}^{N}\sigma^2(X_k)\left(P_{Nk}-\frac{f(X_k)}{g(X_k)}\frac{G(X_{V_{Nk}+M:N})-G(X_{V_{Nk}-M:N})}{2}\right)\xrightarrow{p}0,\tag{A.11}
$$

and

$$
\sum_{k=1}^{N}\sigma^2(X_k)N\left(P_{Nk}^2-\left(\frac{f(X_k)}{g(X_k)}\frac{G(X_{V_{Nk}+M:N})-G(X_{V_{Nk}-M:N})}{2}\right)^2\right)\xrightarrow{p}0.\tag{A.12}
$$

Proof of Lemma A.7: Fix $k\in\{1,2,\ldots,N\}$ (e.g., $k=1$). Notice that,

$$
NP_{Nk}-N\int_{(X_{V_{Nk}:N}+X_{V_{Nk}-M:N})/2}^{(X_{V_{Nk}:N}+X_{V_{Nk}+M:N})/2} f(x)\,dx=o_p(1).
$$

Let

$$
Z_{Nk}^{(1)}=\sigma^2(X_k)N\left(P_{Nk}-\frac{f(X_k)}{g(X_k)}\frac{G(X_{V_{Nk}+M:N})-G(X_{V_{Nk}-M:N})}{2}\right).
$$

Given that f and g are continuous, there are mean values $\bar{X}_{f,k,N,M}$ and $\bar{X}_{g,k,N,M}$ (in $((X_{V_{Nk}:N} + X_{V_{Nk}-M:N})/2, (X_{V_{Nk}:N} + X_{V_{Nk}+M:N})/2)$ and $(X_{V_{Nk}-M:N}, X_{V_{Nk}+M:N})$, respectively) such that:

$$\begin{aligned}
0 &= \sigma^2(X_k)N\left(P_{Nk} - \frac{f(\bar{X}_{f,k,N,M})}{g(\bar{X}_{g,k,N,M})}\frac{G(X_{V_{Nk}+M:N}) - G(X_{V_{Nk}-M:N})}{2}\right) + o_p(1) \\
&= Z^{(1)}_{Nk} + \sigma^2(X_k)\left(\frac{f(X_k)}{g(X_k)} - \frac{f(\bar{X}_{f,k,N,M})}{g(\bar{X}_{g,k,N,M})}\right)\frac{N(G(X_{V_{Nk}+M:N}) - G(X_{V_{Nk}-M:N}))}{2} + o_p(1) \\
&= Z^{(1)}_{Nk} + o_p(1)O_p(1) + o_p(1).
\end{aligned}$$

The last equality follows from equation (A.5) and the continuous mapping theorem. Therefore $Z^{(1)}_{Nk} \xrightarrow{p} 0$. Next, we will show that $E[|Z^{(1)}_{Nk}|]$ is bounded uniformly in N. Let P^b_{Nk}, P^m_{Nk}, and P^t_{Nk} be the three terms on the right hand side of equation (A.10), arranged in the same order as in the equation. By Cauchy's generalization of the Mean Value Theorem, there is $\bar{X}_{N,M}$ in $(0, X_{2M:N})$, such that:

$$\begin{aligned}
NP^b_{Nk} \le NF(X_{2M:N}) &= \frac{F(X_{2M:N})}{G(X_{2M:N})}NG(X_{2M:N}) \\
&\le \frac{f(\bar{X}_{N,M})}{g(\bar{X}_{N,M})}NG(X_{2M:N}).
\end{aligned}$$

Because f/g is bounded and because for any positive integer r,

$$E\left[(NG(X_{2M:N}))^r\right] = N^r\,\frac{(2M+r-1)!N!}{(2M-1)!(N+r)!} < \frac{(2M+r-1)!}{(2M-1)!},$$

we obtain that $E[|NP^b_{Nk}|^r]$ is bounded by a constant that does not depend on N or k. Similarly, by Cauchy's generalization of the Mean Value Theorem, there is $\bar{X}_{k,N,M}$ in $(X_{V_{Nk}-M:N}, X_{V_{Nk}+M:N})$, such that:

$$\begin{aligned}
NP^m_{Nk} &\le N\int_{(X_{V_{Nk}:N}+X_{V_{Nk}-M:N})/2}^{X_{V_{Nk}:N}+X_{V_{Nk}+M:N})/2} f(x)\,dx \\
&\le N\int_{X_{V_{Nk}-M:N}}^{X_{V_{Nk}+M:N}} f(x)\,dx \\
&= \frac{F(X_{V_{Nk}+M:N}) - F(X_{V_{Nk}-M:N})}{G(X_{V_{Nk}+M:N}) - G(X_{V_{Nk}-M:N})}N(G(X_{V_{Nk}+M:N}) - G(X_{V_{Nk}-M:N})) \\
&= \frac{f(\bar{X}_{k,N,M})}{g(\bar{X}_{k,N,M})}N(G(X_{V_{Nk}+M:N}) - G(X_{V_{Nk}-M:N})).
\end{aligned}$$

Because f/g is bounded and because for any positive integer r,

$$E\left[(N(G(X_{V_{Nk}+M:N}) - G(X_{V_{Nk}-M:N}))^r\right] \le N^r\,\frac{(2M+r-1)!N!}{(2M-1)!(N+r)!} < \frac{(2M+r-1)!}{(2M-1)!},$$

we obtain that $E[|NP^m_{Nk}|^r]$ is bounded by a constant that does not depend on N. Using an analogous argument it can be shown that for any positive integer, r, $E[|NP^t_{Nk}|^r]$ is bounded by a constant that does not depend on N. As a result, $E[|NP^m_{Nk}|^r]$ is bounded by a constant that does not depend on N. Because σ^2 is bounded, applying Minkowski inequality we obtain that, for $r > 0$, $E[|Z^{(1)}_{Nk}|^r]$ is uniformly bounded in N. This, in combination with the fact that $Z^{(1)}_{Nk} \xrightarrow{p} 0$, implies $E[|Z^{(1)}_{Nk}|] \to 0$. Now, exchangeability of $Z^{(1)}_{Nk}$ for any given N implies that

$E[|Z^{(1)}_{N1}|] = E[|Z^{(1)}_{N2}|] = \cdots = E[|Z^{(1)}_{NN}|]$. Using Markov's Inequality:

$$\varepsilon \Pr\left(\left|\frac{1}{N}\sum_{k=1}^{N} Z^{(1)}_{Nk}\right| > \varepsilon\right) \le E\left[\left|\frac{1}{N}\sum_{k=1}^{N} Z^{(1)}_{Nk}\right|\right] \le \frac{1}{N}\sum_{k=1}^{N} E[|Z^{(1)}_{Nk}|] = E[|Z^{(1)}_{N1}|] \longrightarrow 0,$$

which proves equation (A.11). The proof of equation (A.12) is analogous. □

Lemma A.8 *Assume that (i)* $(W_1, X_1), \ldots, (W_N, X_N)$ *are independent and identically distributed, (ii)* X *has a continuous distribution, and* W *has a binary distribution with* $\Pr(W = 1) \in (0, 1)$*; (iii) for* $w = 0, 1$*, the cumulative distribution functions of* X *given* $W = w$*,* F_w*, is absolutely continuous with derivative* f_w*; (iv) there exists a constant* C_h *such that for* $h = f_1/f_0$ *we have* $(1/C_h) \le h \le C_h$*. Let* M *be a positive integer not greater than the minimum of* $N_1 = \sum_{i=1}^{N} W_i$ *and* $N_0 = N - N_1$*. Let* $K_M(i)$ *be the number of times observation* i *is used as a match when each unit is matched with replacement to the closest* M *units in the opposite treatment group, and* k *be any positive integer. Then,* $E[K_M(i)^k|W_i = w]$ *is uniformly bounded in* N *by a finite constant.*

Proof of Lemma A.8: Let V_i denote the rank of X_i among the N_{W_i} units with $W_j = W_i$, so that $V_i \in \{1, \ldots, \max(N_0, N_1)\}$. Also, let $X_{w,(j)}$ be the j-th order statistic of X-values among the N_w units with $W_i = w$. Let $\boldsymbol{X}_w$ be the N_w vector with the stacked covariate values for units with $W_i = w$. For $1 \le v \le N_w$, define the probability that unit i is a match for unit j, conditional on $W_i = w$, $W_j = 1 - w$, $\boldsymbol{X}_w$, $V_i = v$, $N_0 = n_0$ and $N_1 = n_1$:

$$p(M, n_0, n_1, w, v, \boldsymbol{X}_w) = \Pr(i \in \mathcal{J}_M(j)|N_0 = n_0, N_1 = n_1, W_i = w, W_j = 1 - w, V_i = v, \boldsymbol{X}_w).$$

First we prove that for all $n_0 \ge M$, $n_1 \ge M$, $w \in \{0, 1\}$, and $v \in \{1, \ldots, n_w\}$,

$$E\left[\left(n_w p(M, n_0, n_1, w, v, \boldsymbol{X}_w)\right)^k\right] < C_h^k \frac{(2M + k - 1)!}{(2M - 1)!}, \tag{A.13}$$

where the expectation on the left-hand-side integrates over the distribution of $\boldsymbol{X}_w$, which has length equal to n_w. Let us first focus on the case with $w = 1$ and $M + 1 \le v \le n_1 - M$. Then

$$\begin{aligned} p(M, n_0, n_1, 1, v, \boldsymbol{X}_w) &= \Pr\left(\frac{X_{1,(v-M)} + X_{1,(v)}}{2} \le X_j \le \frac{X_{1,(v+M)} + X_{1,(v)}}{2} \middle| \boldsymbol{X}_1, W_j = 0\right) \\ &\le \Pr\left(X_{1,(v-M)} \le X_j \le X_{1,(v+M)} \,|\boldsymbol{X}_1, W_j = 0\right) \\ &= F_0\left(X_{1,(v+M)}|\boldsymbol{X}_1\right) - F_0\left(X_{1,(v-M)}|\boldsymbol{X}_1\right). \end{aligned}$$

By continuity of F_1 and F_0, boundedness of h, Cauchy's version of the Mean Value Theorem, exchangeability of uniform spacings, and Lemma A.5 we obtain:

$$
\begin{aligned}
E[(n_1 p(M, n_0, n_1, 1, v, \boldsymbol{X}_1))^k] &\leq E\left[\left(n_1\left(F_0\left(X_{1,(v+M)}\right) - F_0\left(X_{1,(v-M)}\right)\right)\right)^k\right] \\
&= n_1^k E\left[\left(\frac{F_0\left(X_{1,(v+M)}\right) - F_0\left(X_{1,(v-M)}\right)}{F_1\left(X_{1,(v+M)}\right) - F_1\left(X_{1,(v-M)}\right)}\left(F_1\left(X_{1,(v+M)}\right) - F_1\left(X_{1,(v-M)}\right)\right)\right)^k\right] \\
&\leq C_h^k n_1^k E\left[\left(\xi_{v+M:n_1} - \xi_{v-M:n_1}\right)^k\right] \\
&= n_1^k C_h^k \frac{(2M+k-1)!}{(2M-1)!} \frac{n_1!}{(n_1+k)!} \\
&< C_h^k \frac{(2M+k-1)!}{(2M-1)!}.
\end{aligned}
$$

For the case with $w = 1$ and $v \leq M$ we have:

$$
\begin{aligned}
p(M, n_0, n_1, w, v, \boldsymbol{X}_w) &= \Pr\left(X_j \leq \frac{X_{1,(v)} + X_{1,(v+M)}}{2} \,\middle|\, \boldsymbol{X}_1, W_j = 0\right) \\
&\leq \Pr\left(X_j \leq X_{1,(v+M)} \,\middle|\, \boldsymbol{X}_1, W_j = 0\right) \\
&\leq F_0\left(X_{1,(2M)} | \boldsymbol{X}_1\right).
\end{aligned}
$$

Using the same argument as before, the expectation of $(n_1 p(M, n_0, n_1, 1, v, \boldsymbol{X}_w))^k$ can be bounded by $C_h^k (2M+k-1)!/(2M-1)!$. The argument for the case with $v + M \geq N + 1$ is similar and is omitted. This proves (A.13). Conditional on $\boldsymbol{X}_1, N_0 = n_0, N_1 = n_1, W_i = 1$, and $V_i = v$, the random variable $K_M(i)$ follows a Binomial distribution with parameters $(n_0, p(M, n_0, n_1, 1, v, \boldsymbol{X}_1))$. Therefore,

$$
\begin{aligned}
E\left[K_M(i)^k \,\middle|\, \begin{array}{l} \boldsymbol{X}_1, N_0 = n_0, N_1 = n_1, \\ W_i = 1, V_i = v \end{array}\right] &= \sum_{r=0}^{k} \frac{S(k,r) n_0! p(M, n_0, n_1, 1, v, \boldsymbol{X}_1))^r}{(n_0 - r)!} \\
&\leq \sum_{r=0}^{k} S(k,r) \left(\frac{n_0}{n_1}\right)^r (n_1 p(M, n_0, n_1, 1, v, \boldsymbol{X}_1))^r,
\end{aligned} \tag{A.14}
$$

where $S(k, r)$ is a Stirling number of the second kind (see Johnson, Kotz, and Kemp, 1993). Taking expectations over $\boldsymbol{X}_1$, we obtain:

$$
E\left[K_M(i)^k \,\middle|\, \begin{array}{l} N_0 = n_0, N_1 = n_1, \\ W_i = 1, V_i = v \end{array}\right] \leq \sum_{r=0}^{k} S(k,r) \left(\frac{n_0}{n_1}\right)^r C_h^r \frac{(2M+r-1)!}{(2M-1)!}. \tag{A.15}
$$

Because the bound does not depend on v, it applies also to $E[K_M(i)^k | N_0 = n_0, N_1 = n_1, W_i = 1]$. Now, from Lemma A.3, it follows that:

$$
E\left[K_M(i)^k | W_i = 1\right] \leq \sum_{r=0}^{k} S(k,r) C_r C_h^r \frac{(2M+r-1)!}{(2M-1)!}.
$$

The same argument applies to $E[K_M(i)^k | W_i = 0]$. □

LEMMA A.9 $(W_1, X_1), \ldots, (W_N, X_N)$ *are independent and identically distributed, where* X *has a continuous distribution on* $[a, b]$, *and* W *has a binary distribution with* $\Pr(W = 1) \in (0, 1)$. *Let* P_{Nk} *be the probability that observation* k *is used as a match for any particular observation in* *the opposite treatment arm, conditional on* $\boldsymbol{W}$ *and* $\boldsymbol{X}_{W_k}$. *Assume that* f_1/f_0 *is bounded and bounded away from zero, then for all* $\delta > 0$:

$$\max_{k=1,\ldots,N} P_{Nk} = o_p(N^{-1+\delta}).$$

Proof of Lemma A.9: For the proof, we focus on the case when $W_k = 0$. The derivations for the treated observations are analogous. Let $\{P_{0,(1)}, \ldots, P_{0,(N_0)}\}$ be the catchment probabilities for $\{X_{0,(1)}, \ldots, X_{0,(N_0)}\}$. If k is such that $k \geq M+1$ and $k \leq N_0 - M$, then

$$P_{0,(k)} = \int_{(X_{0,(k-M)}+X_{0,(k)})/2}^{(X_{0,(k+M)}+X_{0,(k)})/2} f_1(x)dx.$$

Now apply the change of variables $z = F_0(x)$ to obtain:

$$\begin{aligned} P_{0,(k)} &= \int_{F_0((X_{0,(k-M)}+X_{0,(k)})/2)}^{F_0((X_{0,(k+M)}+X_{0,(k)})/2)} \frac{f_1(F_0^{-1}(z))}{f_0(F_0^{-1}(z))}\, dz \\ &\leq \int_{F_0(X_{0,(k-M)})}^{F_0(X_{0,(k+M)})} \frac{f_1(F_0^{-1}(z))}{f_0(F_0^{-1}(z))}\, dz. \end{aligned}$$

Using the assumption that f_1/f_0 is bounded by some constant C, we obtain:

$$P_{0,(k)} \leq C(F_0(X_{0,(k+M)}) - F_0(X_{0,(k-M)})).$$

The derivation $k < M+1$ and $k > N_0 - M$ is similar, so it is omitted. Now the result follows from the fact that the maximal uniform spacing is $o_p(N^{-1+\delta})$ for all $\delta > 0$ (see, e.g., Theorem 1 in Shorack and Wellner (2009), page 726). □

Lemma A.10 *Assume that the conditions of Lemma A.9 hold. In addition, assume that* $\sigma^2(w,x) = \mathrm{var}(Y|W=w, X=x)$ *is uniformly bounded. Then, (i)*

$$\frac{1}{N_w}\sum_{i:W_i=w}^{N} \sigma^2(w,X_i)K_M(i) - \frac{1}{N_w}\sum_{i:W_i=w}^{N} \sigma^2(w,X_i)N_{1-w}P_{Ni} = o_p(1),$$

and (ii)

$$\frac{1}{N_w}\sum_{i:W_i=w}^{N} \sigma^2(w,X_i)K_M(i)^2 - \frac{1}{N_w}\sum_{i:W_i=w}^{N} \sigma^2(w,X_i)\left(N_{1-w}^2P_{Ni}^2 + N_{1-w}P_{Ni}(1-P_{Ni})\right) = o_p(1).$$

Proof of Lemma A.10: Here we prove *(ii)* only. The proof of *(i)* is analogous but slightly less involved. In the proof we adopt $w = 0$. The proof for $w = 1$ is identical after switching treatment subscripts. Re-order the sample units so those units with $W_i = 0$ come first, $W_1 = \cdots = W_{N_0} = 0$. For $1 \leq k \leq N_0$, $1 \leq l \leq N_0$, and $k \neq l$, let $\mu_{r,k} = E[K_M^r(k)|\boldsymbol{W}, \boldsymbol{X}_0]$ and $\mu_{r,s,k,l} = E[K_M^r(k)K_M^s(l)|\boldsymbol{W}, \boldsymbol{X}_0]$. We first prove the results for the single match case, with $M = 1$. Notice that, if $M = 1$, then conditional on $\boldsymbol{W}$ and $\boldsymbol{X}_0$, the vector $(K_1(1), \ldots, K_1(N_0))$ has a multinomial distribution with parameters N_1 and $P_{0,1}, \ldots, P_{0,N_0}$. The moment generating function of this distribution is:

$$M(t_1, \ldots, t_{N_0}) = \left(\sum_{k=1}^{N_0} P_{0,k}e^{t_k}\right)^{N_1}.$$

Then,

$$\begin{aligned}\mu_{2,k} &= N_1 P_{0,k} + N_1(N_1-1)P_{0,k}^2,\\ \mu_{2,2,k,l} &= N_1(N_1-1)P_{0,k}P_{0,l} + N_1(N_1-1)(N_1-2)(P_{0,k}^2 P_{0,l} + P_{0,k}P_{0,l}^2)\\ &\quad + N_1(N_1-1)(N_1-2)(N_1-3)P_{0,k}^2 P_{0,l}^2.\end{aligned}$$

Therefore,

$$\begin{aligned}\mu_{2,2,k,l} - \mu_{2,k}\mu_{2,l} =& - N_0 P_{0,k}P_{0,l} - 2N_0(N_0-1)\big(P_{0,k}^2 P_{0,l} + P_{0,k}P_{0,l}^2\big)\\ & - 2N_0(N_0-1)(2N_0-3)P_{0,k}^2 P_{0,l}^2 \leq 0.\end{aligned}$$

Now,

$$\begin{aligned}&E\left[\left(\frac{1}{N_0}\sum_{i=1}^{N_0}\sigma^2(0,X_i)(K_1^2(i)-\mu_{2,i})\right)^2\Big|\boldsymbol{W},\boldsymbol{X}_0\right]\\ &\quad = \frac{1}{N_0^2}\sum_{i=1}^{N_0}\sum_{j=1}^{N_0}\sigma^2(0,X_i)\sigma^2(0,X_j)E\Big[(K_1^2(i)-\mu_{2,i})(K_1^2(j)-\mu_{2,j})\Big|\boldsymbol{W},\boldsymbol{X}_0\Big]\\ &\quad = \frac{1}{N_0^2}\sum_{i=1}^{N_0}\sigma^4(0,X_i)(\mu_{4,i}-\mu_{2,i}^2) + \frac{1}{N_0^2}\sum_{i=1}^{N_0}\sum_{j\neq i}\sigma^2(0,X_i)\sigma^2(0,X_j)(\mu_{2,2,i,j}-\mu_{2,i}\mu_{2,j})\\ &\quad \leq \frac{1}{N_0^2}\sum_{i=1}^{N_0}\sigma^4(0,X_i)\mu_{4,i}.\end{aligned}$$

Therefore, from equation (A.15), we obtain

$$\begin{aligned}E\left[\left(\frac{1}{N_0}\sum_{i=1}^{N_0}\sigma^2(0,X_i)(K_1^2(i)-\mu_{2,i})\right)^2\Big|\boldsymbol{W}\right] &\leq C_{\sigma^4}\frac{1}{N_0}E[\mu_{4,i}|\boldsymbol{W}]\\ &\leq \frac{C_{\sigma^4}}{N}\left(1+\frac{N_1}{N_0}\right)\sum_{r=0}^{4}S(4,r)\left(\frac{N_0}{N_1}\right)^r C_h^r(r+1)!\end{aligned}$$

where C_{σ^4} is a bound on $\sigma^4(w,x)$ for $w\in\{0,1\}$ and x in the support of X. Now, Lemma A.3 implies

$$E\left[\left(\frac{1}{N_0}\sum_{i=1}^{N_0}\sigma^2(0,X_i)(K_1^2(i)-\mu_{2,i})\right)^2\right]\longrightarrow 0,$$

which yields *(ii)* for the case of $w=0$ and $M=1$.

Now consider the case with $M>1$,

$$\begin{aligned}&E\left[\left(\frac{1}{N_0}\sum_{i=1}^{N_0}\sigma^2(0,X_i)(K_M^2(i)-\mu_{2,i})\right)^2\Big|\boldsymbol{W},\boldsymbol{X}_0\right]\\ &\quad = \frac{1}{N_0^2}\sum_{i=1}^{N_0}\sigma^4(0,X_i)(\mu_{4,i}-\mu_{2,i}^2) + \frac{1}{N_0^2}\sum_{i=1}^{N_0}\sum_{j\neq i}\sigma^2(0,X_i)\sigma^2(0,X_j)(\mu_{2,2,i,j}-\mu_{2,i}\mu_{2,j}). \quad \text{(A.16)}\end{aligned}$$

The proof that the expectation of the first term on the left-hand-side of equation (A.16) converges to zero is the same as for the case of $M=1$. However, if $M>1$, then it is not the case that all elements of the second term on the left-hand-side of equation (A.16) are negative. If $M>1$, an observation i, with $W_i = 0$ and $M + 1 \leq V_i = v \leq N_0 - M$ is a match for all observations in the opposite treatment arm that have covariate value in the catchment interval

$$A_M(i) = \left(\frac{X_{0,(v+M)} + X_{0,(v)}}{2}, \frac{X_{0,(v-M)} + X_{0,(v)}}{2} \right),$$

which overlaps with the catchment intervals of other $2(M-1)$ untreated observations. The catchment intervals of an untreated observation with $V_i \leq M$ or $V_i > N_0 - M$ overlaps with less than $2(M-1)$ catchment intervals of other untreated observations. As a result, relative to the proof for $M = 1$, overlapping catchment intervals create terms, $\mu_{2,2,i,j} - \mu_{2,i}\mu_{2,j}$, that are not necessarily negative. The number of such potentially non-negative terms is smaller than $2(M-1)N_0$. Let I_{ij} be an indicator function that takes value equal to one if the catchment intervals of observations i and j overlap, and value zero otherwise. Let $K_M(i,j)$ be the number of treated observations with covarite values in $A_M(i) \cup A_M(j)$. Then,

$$\begin{aligned}
\frac{1}{N_0^2} \sum_{i=1}^{N_0} \sum_{j \neq i} \sigma^2(0, X_i)\sigma^2(0, X_j) E\Big[(K_M^2(i) - \mu_{2,i})(K_M^2(j) - \mu_{2,j}) \big| \boldsymbol{W}, \boldsymbol{X}_0\Big] \\
\leq \frac{1}{N_0^2} \sum_{i=1}^{N_0} \sum_{j \neq i} \sigma^2(0, X_i)\sigma^2(0, X_j) \Big(E\Big[K_M^2(i) K_1^2(j) \big| \boldsymbol{W}, \boldsymbol{X}_0\Big] - \mu_{2,i}\mu_{2,j} \Big) I_{ij} \\
\leq \frac{1}{N_0^2} \sum_{i=1}^{N_0} \sum_{j \neq i} \sigma^2(0, X_i)\sigma^2(0, X_j) E\Big[K_M^2(i) K_1^2(j) \big| \boldsymbol{W}, \boldsymbol{X}_0\Big] I_{ij} \\
\leq \frac{C_{\sigma^4}}{N_0^2} \sum_{i=1}^{N_0} \sum_{j \neq i} E\Big[K_M^4(i,j) \big| \boldsymbol{W}, \boldsymbol{X}_0\Big] I_{ij}.
\end{aligned}$$

Using the same argument as in Lemma A.8, it can be seen

$$E\Big[K_M^4(i,j) \big| \boldsymbol{W}, I_{ij} = 1\Big] \leq \sum_{k=0}^{4} c_k \left(\frac{N_1}{N_0} \right)^k,$$

for some positive constants, c_k. Now, Lemma A.3 yields the result. □

Lemma A.11 *Suppose that the assumptions of Lemma A.10 hold. Assume also that the density of X is continuous on $[a, b]$. Finally, assume that, for $w = 0, 1$, $\sigma^2(w, x)$ is continuous on $[a, b]$. Let $p^* = \Pr(W = 1)$, then, for $w = 0, 1$, (i)*

$$\frac{1}{N_w} \sum_{i:W_i=w} \sigma^2(w, X_i) K_M(i) \xrightarrow{p} M E \left[\sigma^2(w, X_i) \left(\frac{p^*}{1-p^*} \right)^{1-2w} \frac{f_{1-w}(X_i)}{f_w(X_i)} \middle| W_i = w \right]$$

and (ii)

$$\begin{aligned}
\frac{1}{N_w} \sum_{i:W_i=w} \sigma^2(w, X_i) K_M(i)^2 \xrightarrow{p} M E \left[\sigma^2(w, X_i) \left(\frac{p^*}{1-p^*} \right)^{1-2w} \frac{f_{1-w}(X_i)}{f_w(X_i)} \middle| W_i = w \right] \\
+ \frac{M(2M+1)}{2} E \left[\sigma^2(w, X_i) \left(\left(\frac{p^*}{1-p^*} \right)^{1-2w} \frac{f_{1-w}(X_i)}{f_w(X_i)} \right)^2 \middle| W_i = w \right].
\end{aligned}$$

Proof of Lemma A.11: The result follows from lemmas A.6, A.7, and A.10. □

Proof of Proposition 1: Let

$$
\begin{aligned}
D_N &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\bar{\mu}(1, p(X_i)) - \bar{\mu}(0, p(X_i)) - \tau\right) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (2W_i - 1) \left(1 + \frac{K_M(i)}{M}\right) \left(Y_i - \bar{\mu}(W_i, p(X_i))\right),
\end{aligned}
$$

and

$$
R_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (2W_i - 1) \left(\bar{\mu}(1 - W_i, p(X_i)) - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} \bar{\mu}(1 - W_i, p(X_j)) \right).
$$

Notice that

$$
\sqrt{N}(\widehat{\tau}^*_N - \tau) = D_N + R_N.
$$

We will first show that $D_N \stackrel{d}{\to} N(0, \sigma^2)$. Notice that:

$$
D_N = \sum_{k=1}^{2N} \xi_{N,k},
$$

where

$$
\xi_{N,k} = \frac{1}{\sqrt{N}} \left(\bar{\mu}(1, p(X_k)) - \bar{\mu}(0, p(X_k)) - \tau\right)
$$

for $1 \leq k \leq N$, and

$$
\xi_{N,k} = \frac{1}{\sqrt{N}} (2W_{k-N} - 1) \left(1 + \frac{K_M(k-N)}{M}\right) \left(Y_{k-N} - \bar{\mu}(W_{k-N}, p(X_{k-N}))\right)
$$

for $N+1 \leq k \leq 2N$. Consider the σ-fields $\mathcal{F}_{N,k} = \sigma\{W_1, \ldots, W_k, p(X_1), \ldots, p(X_k)\}$ for $1 \leq k \leq N$ and $\mathcal{F}_{N,k} = \sigma\{W_1, \ldots, W_N, p(X_1), \ldots, p(X_N), Y_1, \ldots, Y_{k-N}\}$ for $N+1 \leq k \leq 2N$. Then for each $N \geq 1$,

$$
\left\{ \sum_{j=1}^{i} \xi_{N,j}, \mathcal{F}_{N,i}, 1 \leq i \leq 2N \right\}
$$

is a martingale. To obtain the result of the proposition, we apply the Central Limit Theorem for martingale arrays (e.g., Billingsley, 1995). The following three conditions are sufficient:

$$
\sum_{k=1}^{2N} E\left[|\xi_{N,k}|^{2+\delta}\right] \to 0 \quad \text{for some } \delta > 0, \tag{A.17}
$$

$$
\sum_{k=1}^{N} E[\xi^2_{N,k} | \mathcal{F}_{N,k-1}] \stackrel{p}{\to} E\left[\left(\bar{\mu}(1, p(X)) - \bar{\mu}(0, p(X)) - \tau\right)^2\right], \tag{A.18}
$$

and

$$
\begin{aligned}
\sum_{k=N+1}^{2N} E[\xi^2_{N,k} | \mathcal{F}_{N,k-1}] \stackrel{p}{\to} E&\left[\bar{\sigma}^2(1, p(X)) \left(\frac{1}{p(X)} + \frac{1}{2M}\left(\frac{1}{p(X)} - p(X)\right)\right)\right] \\
&+ E\left[\bar{\sigma}^2(0, p(X)) \left(\frac{1}{1-p(X)} + \frac{1}{2M}\left(\frac{1}{1-p(X)} - (1 - p(X))\right)\right)\right].
\end{aligned} \tag{A.19}
$$

Equation (A.17) is the Lyapounov condition (which is sufficient for the usual Lindeberg condition to hold). Because the functions $\bar{\mu}(w,p)$ are continuous on a compact support, these functions are also bounded. Let $C_{\bar{\mu}}$ be a bound on $\max_{w,p}\bar{\mu}(w,p)$ for $w \in \{0,1\}$ and $p \in [\underline{p},\overline{p}]$. We obtain

$$\sum_{k=1}^{N} E\left[|\xi_{N,k}|^{2+\delta}\right] \leq \frac{(2C_{\bar{\mu}}+|\tau|)^{2+\delta}}{N^{\delta/2}} \to 0.$$

Now, let $C_{\bar{\sigma}^{2+\delta}}$ be a bound on $E[|Y_i - \bar{\mu}(W_i,P(X_i))|^{2+\delta}|W_i,P(X_i)]$. Now, using the Law of Iterated Expectation and the fact that $K_M(i)$ has bounded moments (Lemma A.8), we obtain:

$$\sum_{k=N+1}^{2N} E\left[|\xi_{N,k}|^{2+\delta}\right] \leq \frac{C_{\bar{\sigma}^{2+\delta}} E\left[\left(1+\frac{K_M(i)}{M}\right)^{2+\delta}\right]}{N^{\delta/2}} \to 0.$$

This proves equation (A.17). Equation (A.18) is easy to prove because the data are *i.i.d.* To prove equation (A.19) notice that:

$$\sum_{k=N+1}^{2N} E[\xi_{N,k}^2|\mathcal{F}_{N,k-1}] = \frac{1}{N}\sum_{i=1}^{N}\left(1+\frac{K_M(i)}{M}\right)^2 \bar{\sigma}^2(W_i,p(X_i)).$$

Now, Lemma A.11 implies the result in equation (A.19) after some algebra. To finish the proof of part (i), we will show that $R_N \xrightarrow{p} 0$. We can write $R_N = R_{N,0} + R_{N,1}$, where

$$R_{N,w} = \frac{1}{\sqrt{N}}\sum_{i:W_i=w}^{N}\left(\bar{\mu}(1-W_i,p(X_i)) - \frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}\bar{\mu}(1-W_i,p(X_j))\right).$$

We prove $R_{N,1} = o_p(1)$. The proof for $R_{N,0} = o_p(1)$ is analogous, so we omit it. By Markov's Inequality, it is enough to show that $E[|R_{N,1}|] \to 0$. Without loss of generality and to simplify notation, we reorder the observations in the sample, so that the observations with $W_i = 1$ come first. Because the function $\bar{\mu}$ has a derivative bounded by $C_{\bar{\mu}}$, we obtain that for $N_0 = n_0$ (with $N_0 = n_0, M \leq n_0 \leq N - M$) and $n_1 = N - n_0$:

$$\begin{aligned}
E\big[|R_{N,1}|\big|N_0=n_0\big] &\leq \frac{1}{\sqrt{N}}\sum_{i=1}^{n_1}\frac{1}{M}\sum_{j\in\mathcal{J}_M(i)} E\big[\,|\bar{\mu}(0,p(X_i)) - \bar{\mu}(0,p(X_j))|\,\big|\,N_0=n_0\big] \\
&\leq \frac{C_{\bar{\mu}}}{\sqrt{N}}\sum_{i=1}^{n_1}\frac{1}{M}\sum_{j\in\mathcal{J}_M(i)} E\big[\,|p(X_i)-p(X_j)|\,\big|\,N_0=n_0\big].
\end{aligned}$$

Now, lemmas A.1 and A.2 imply that there is some $\bar{r} > 0$ such that:

$$E\big[|R_{N,1}|\big|N_0=n_0\big] \leq C_{\bar{\mu}}\bar{r}\,\frac{n_1}{N^{1/2}\lfloor n_0^{3/4}\rfloor} + M\frac{n_1}{N^{1/2}}\,n_0^{M-1/4}\exp(-n_0^{1/4}).$$

Because $n_0^{3/4}/\lfloor n_0^{3/4}\rfloor$ and $n_0^{M+1/2}\exp(-n_0^{1/4})$ are bounded for all $n_0 \geq 1$, there exists some constant C such that:

$$E\big[|R_{N,1}|\big|N_0=n_0\big] \leq \frac{C}{N^{1/4}}\left(\frac{n_1}{N}\frac{N^{3/4}}{n_0^{3/4}}\right).$$

Therefore,

$$E\big[|R_{N,1}|\big] \leq \frac{C}{N^{1/4}}\,E\left[\left(\frac{N}{N_0}\right)^{3/4}\,\middle|\,M \leq N_0 \leq N-M\right],$$

which converges to zero because of the result in Lemma A.3. This completes the proof of part (i). To prove part (ii), first notice that $\sqrt{N}(\widehat{\tau}^{*}_{t,N}-\tau_t)=D_{t,N}+R_{t,N}$, where

$$\begin{aligned}D_{t,N} &= \frac{\sqrt{N}}{N_1}\sum_{i=1}^{N} W_i\left(\bar{\mu}(1,p(X_i))-\bar{\mu}(0,p(X_i))-\tau_t\right)\\ &+\frac{\sqrt{N}}{N_1}\sum_{i=1}^{N}\left(W_i-(1-W_i)\frac{K_M(i)}{M}\right)\left(Y_i-\bar{\mu}(W_i,p(X_i))\right)\end{aligned}$$

and

$$R_{t,N}=\frac{\sqrt{N}}{N_1}\sum_{i=1}^{N} W_i\left(\bar{\mu}(0,p(X_i))-\frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}\bar{\mu}(0,p(X_j))\right).$$

Because $R_{t,N}=(N/N_1)R_{N,1}$, $N/N_1\xrightarrow{p}1/E[p(X)]$ and $R_{N,1}\xrightarrow{p}0$, we obtain: $R_{t,N}\xrightarrow{p}0$. Therefore, we need to show that $D_{t,N}\xrightarrow{d}N(0,\sigma_t^2)$. Notice that:

$$D_{t,N}=\sum_{k=1}^{2N}\xi_{t,N,k},$$

where

$$\xi_{t,N,k}=\frac{\sqrt{N}}{N_1}W_k\left(\bar{\mu}(1,p(X_k))-\bar{\mu}(0,p(X_k))-\tau_t\right)$$

for $1\le k\le N$, and

$$\xi_{t,N,k}=\frac{\sqrt{N}}{N_1}\left(W_{k-N}-(1-W_{k-N})\frac{K_M(k-N)}{M}\right)\left(Y_{k-N}-\bar{\mu}(W_{k-N},p(X_{k-N}))\right)$$

for $N+1\le k\le 2N$. Consider the σ-fields $\mathcal{F}_{t,N,k}=\sigma\{W_1,\ldots,W_N,p(X_1),\ldots,p(X_k)\}$ for $1\le k\le N$ and $\mathcal{F}_{t,N,k}=\sigma\{W_1,\ldots,W_N,p(X_1),\ldots,p(X_N),Y_1,\ldots,Y_{k-N}\}$ for $N+1\le k\le 2N$. Then for each $N\ge 1$,

$$\left\{\sum_{j=1}^{i}\xi_{t,N,j},\mathcal{F}_{t,N,i},1\le i\le 2N\right\}$$

is a martingale. To obtain the result of the proposition, we apply the Martingale Central Limit Theorem. The following three conditions are sufficient:

$$\sum_{k=1}^{2N}E\left[|\xi_{t,N,k}|^{2+\delta}\right]\to 0\quad\text{for some }\delta>0,\tag{A.20}$$

$$\sum_{k=1}^{N}E[\xi^2_{t,N,k}|\mathcal{F}_{t,N,k-1}]\xrightarrow{p}\frac{1}{E[p(X)]^2}E\left[p(X)\left(\bar{\mu}(1,p(X))-\bar{\mu}(0,p(X))-\tau_t\right)^2\right],\tag{A.21}$$

and

$$\begin{aligned}\sum_{k=N+1}^{2N}E[\xi^2_{t,N,k}|\mathcal{F}_{t,N,k-1}] &\xrightarrow{p}\frac{1}{E[p(X)]^2}E\left[p(X)\bar{\sigma}^2(1,p(X))\right]\\ &+\frac{1}{E[p(X)]^2}E\left[\bar{\sigma}^2(0,p(X))\left(\frac{p^2(X)}{1-p(X)}+\frac{1}{M}p(X)+\frac{1}{2M}\frac{p^2(X)}{1-p(X)}\right)\right].\end{aligned}\tag{A.22}$$

Equation (A.20) follows from the same arguments employed for equation (A.17) and from the fact that the moments of N/N_1 are bounded (Lemma A.3). In particular:

$$\sum_{k=1}^{N} E\left[|\xi_{t,N,k}|^{2+\delta}\right] \leq \sum_{k=1}^{N} E\left[\frac{N^{(2+\delta)/2}}{N_1^{2+\delta}}(2C_{\bar{\mu}}+|\tau_t|)^{2+\delta}\right]$$
$$= \frac{(2C_{\bar{\mu}}+|\tau_t|)^{2+\delta}}{N^{\delta/2}} E\left[\left(\frac{N}{N_1}\right)^{2+\delta}\right] \xrightarrow{p} 0.$$

Using the Law of Iterated Expectation and the fact that $K_M(i)$ has bounded moments, and the Cauchy-Schwartz inequality we obtain:

$$\sum_{k=N+1}^{2N} E\left[|\xi_{t,N,k}|^{2+\delta}\right] \leq \frac{C_{\bar{\sigma}^{2+\delta}} E\left[\left(\frac{N}{N_1}\right)^{2+\delta}\left(1+\frac{K_M(i)}{M}\right)^{2+\delta}\right]}{N^{\delta/2}} \to 0.$$

It is easy to show that

$$\sum_{k=N+1}^{2N} E\left[\frac{N}{N_1^2} W_{k-N}(Y_{k-N}-\bar{\mu}(1,p(X_{k-N})))^2 \Big| \mathcal{F}_{t,N,k-1}\right] = \left(\frac{N}{N_1}\right)^2 \frac{1}{N}\sum_{i=1}^{N} W_i \bar{\sigma}^2(1,p(X_i))$$
$$\xrightarrow{p} \frac{1}{E[p(X)]^2} E\left[p(X)\bar{\sigma}^2(1,p(X))\right].$$

In addition:

$$\sum_{k=N+1}^{2N} E\left[\frac{N}{N_1^2}(1-W_{k-N})\left(\frac{K_M(k-N)}{M}\right)^2 (Y_{k-N}-\bar{\mu}(W_{k-N},p(X_{k-N})))^2 \Big| \mathcal{F}_{t,N,k-1}\right]$$
$$= \left(\frac{N}{N_1}\right)^2 \frac{1}{N}\sum_{i=1}^{N}(1-W_i)\left(\frac{K_M(i)}{M}\right)^2 \bar{\sigma}^2(0,p(X_i)).$$

Now, Lemma A.11 implies the result in equation (A.22). □

PROOF OF LEMMA 1: Assumption 4*(i)* implies that there exist finite constants C_L and C_U such that $C_L \leq x'\theta \leq C_U$ for any $\theta \in \text{int}(\Theta)$ and x in the support of X. Because f is continuous and everywhere positive, it follows that it is bounded away from zero on the compact interval $[C_L, C_U]$ (see, e.g., Rudin (1976, Theorem 4.16)). Similarly $F(1-F)$ is bounded and bounded away from zero on $[C_L, C_U]$. Therefore, there exist positive constants c_U and c_L such that $c_L \leq f^2(x'\theta)/(F(x'\theta)(1-F(x'\theta))) \leq c_U$, for any $\theta \in \text{int}(\Theta)$ and x in the support of X. Then, for $\theta \in \text{int}(\Theta)$ and any non-zero $v \in \mathbb{R}^k$, we obtain

$$E\left[\|X\|^2 \frac{f^2(X'\theta)}{F(X'\theta)(1-F(X'\theta))}\right] < \infty,$$

and

$$v'E\left[X\frac{f^2(X'\theta)}{F(X'\theta)(1-F(X'\theta))}X'\right]v \geq c_L v'E[XX']v > 0.$$

As a result, I_θ is finite and positive definite for all $\theta \in \text{int}(\Theta)$. By Assumption 4 *(i)* and *(ii)*, $f^2/(F(1-F))$ is continuous and bounded in $[C_L, C_U]$ and $\|X\|$ is bounded. By the Dominated Convergence Theorem, we obtain that I_θ is continuous on $\text{int}(\Theta)$. This, along with continuous differentiability of F, implies that the parametrization $\theta \mapsto P^\theta$ is regular on $\text{int}(\Theta)$ (see Proposition 2.1.1 in Bickel, *et al.* 1998).

Now, Proposition 2.1.2 in Bickel, *et al.* (1998) implies that, under P^{θ_N},

$$\Lambda_N(\theta^*|\theta_N) = -h'\Delta_N(\theta_N) - \frac{1}{2}h'I_{\theta_N}h + o_p(1).$$

Continuity of I_θ implies the result in equation (1). Also for regular parametric models, equation (2) is derived in the proof of Proposition 2.1.2 in Bickel et al. (1998).

Suppose that $\widehat{\theta}_N$ is asymptotically linear. That is, under P^{θ^*}

$$\sqrt{N}(\widehat{\theta}_N - \theta^*) = I_{\theta^*}^{-1}\Delta_N(\theta^*) + o_p(1).$$

By contiguity (see, e.g., Proposition 2.1.3 in Bickel et al., 1998), we obtain that the previous equation holds also under P^{θ_N}. Therefore, under P^{θ_N},

$$\sqrt{N}(\widehat{\theta}_N - \theta_N) = -h + I_{\theta^*}^{-1}\Delta_N(\theta^*) + o_p(1).$$

Now, equation (3) follows from equation (15) in Proposition 2.1.2 in Bickel et al. (1998). Therefore, if $\widehat{\theta}_N$ is asymptotically linear then equation (3) holds. To prove asymptotic linearity of $\widehat{\theta}_N$, notice first that, given that $F(x'\theta)$ is continuously differentiable and bounded away from zero and one on int(Θ) for all x in the support of X, and that I_θ exists and is continuous, Lemma 7.6 in van der Vaart (1998) implies that $\{P^\theta : \theta \in \text{int}(\Theta)\}$ is differentiable in quadratic mean. Moreover, given that $f(x'\theta)/F(x'\theta)$ and $f(x'\theta)/(1-F(x'\theta))$ are uniformly bounded by a constant for all $\theta \in \text{int}(\Theta)$ and x in the support of X, and that I_{θ^*} is nonsingular, then by Theorem 5.39 in var der Vaart (1998) consistency of $\widehat{\theta}_N$ implies that equation (3) holds. Now, given that $E[XX']$ is nonsingular, $F(x'\theta)$ is continuous in θ for all x in the support of X and bounded away from zero and one, and Θ is compact, Theorem 2.5 in Newey and McFadden (1998) implies $\widehat{\theta}_N \xrightarrow{p} \theta^*$.
□

In the setting of section III, the following lemma provides primitive conditions for Assumption 5.

Lemma A.12 *Let X_0 be the first coordinate of X and let X_1 be the sub-vector of the last $k-1$ coordinates of X. Assume that the distribution of X_0 conditional on (X_1, Y_1) (respectively, (X_1, Y_0)) admits a density, $f_{X_0|X_1,Y_1}(x_0, x_1, y_1)$ (respectively, $f_{X_0|X_1,Y_0}(x_0, x_1, y_0)$), with respect to the Borel measure, λ. Assume that $f_{X_0|X_1,Y_1}(x_0, x_1, y_1)$ and $f_{X_0|X_1,Y_0}(x_0, x_1, y_0)$ are bounded and continuous functions of x_0. For $N = 1, 2, \ldots$, let $X'\theta_N = X_0\theta_{0N} + X_1'\theta_{1N}$. Assume that $\theta_N \to \theta^*$ such that $\theta_0^* \neq 0$. Let $r(y, w, x)$ be a bounded function from $\mathbb{R}^{k+2}$ to $\mathbb{R}$, continuous in the first coordinate of x. Then*

$$E_{\theta_N}[r(Y, W, X)|W, F(X'\theta_N)] \to E[r(Y, W, X)|W, F(X'\theta^*)],$$

almost surely.

Proof of Lemma A.12: Because F is strictly increasing, we have

$$E_\theta[r(Y, W, X)|W, F(X'\theta)] = E_\theta[r(Y, W, X)|W, X'\theta].$$

Therefore, it is enough to prove

$$E_{\theta_N}[r(Y, W, X)|W, X'\theta_N] \to E[r(Y, W, X)|W, X'\theta^*],$$

almost surely. Notice also that

$$E_\theta[r(Y,W,X)|W,F(X'\theta)] = WE_\theta[r(Y,W,X)|W=1,F(X'\theta)] \\ + (1-W)E_\theta[r(Y,W,X)|W=0,F(X'\theta)].$$

It is, therefore, enough to prove convergence of $E_{\theta_N}[r(Y,W,X)|W=1,X'\theta_N]$ and $E_{\theta_N}[r(Y,W,X)|W=0,X'\theta_N]$. We will prove convergence for $E_{\theta_N}[r(Y,W,X)|W=1,X'\theta_N]$; the proof for $E_{\theta_N}[r(Y,W,X)|W=0,X'\theta_N]$ is analogous. Let $r_1(y,x)=r(y,1,x)$. Given that W is independent of (Y_1,X) conditional on the propensity score, we obtain

$$E_{\theta_N}[r(Y,W,X)|W=1,X'\theta_N] = E_{\theta_N}[r_1(Y_1,X)|W=1,X'\theta_N] = E_{\theta_N}[r_1(Y_1,X)|X'\theta_N] \\ = E[r_1(Y_1,X)|X'\theta_N].$$

Similarly, $E[r(Y,W,X)|W=1,X'\theta^*] = E[r_1(Y_1,X)|X'\theta^*]$. Hence, we aim to prove

$$E[r_1(Y_1,X)|X'\theta_N] \to E[r_1(Y_1,X)|X'\theta^*]$$

almost surely. More precisely, if we make $g_\theta(X)=E[r_1(Y_1,X)|X'\theta]$, we aim to prove $g_{\theta_N}(X)\to g_{\theta^*}(X)$ as $\theta_N\to\theta^*$, almost surely.

Let $f_{Y_1,X_1}(y,x_1)$ be the density of (Y_1,X_1) with respect to some σ-finite measure, μ. ((Y_1,X_1) may include variables that are not continuously distributed.) Other densities are denoted analogously. By the change of variables formula, if $\theta_0\neq 0$,

$$f_{X'\theta,Y_1,X_1}(z,y,x_1) = \frac{1}{|\theta_0|} f_{Y_1,X_0,X_1}\left(y, \frac{z-x_1'\theta_1}{\theta_0}, x_1\right) \\ = \frac{1}{|\theta_0|} f_{X_0|X_1,Y_1}\left(\frac{z-x_1'\theta_1}{\theta_0}, x_1, y\right) f_{Y_1,X_1}(y,x_1)$$

is the density of $(X'\theta,Y_1,X_1')'$ with respect to $\lambda\times\mu$. Because $r_1(y,x)=r(y,1,x)$, the function r_1 is bounded and continuous in the first coordinate of x. With a slight notational abuse we will write $r_1(Y,X_0,X_1)=r_1(Y,X)$. Then, by continuity and boundedness of r_1 and $f_{X_0|X_1,Y_1}$, and the Dominated Convergence Theorem, we obtain that for all x in the support of X

$$\frac{\displaystyle\int r_1\left(y, \frac{x'\theta_N - x_1'\theta_{1N}}{\theta_{0N}}, x_1\right) f_{X_0|X_1,Y_1}\left(\frac{x'\theta_N - x_1'\theta_{1N}}{\theta_{0N}}, x_1, y\right) f_{Y_1,X_1}(y,x_1)d\mu}{\displaystyle\int f_{X_0|X_1,Y_1}\left(\frac{x'\theta_N - x_1'\theta_{1N}}{\theta_{0N}}, x_1, y\right) f_{Y_1,X_1}(y,x_1)d\mu},$$

which is a version of $g_{\theta_N}(x)$, converges to

$$\frac{\displaystyle\int r_1\left(y, \frac{x'\theta^* - x_1'\theta_1^*}{\theta_0^*}, x_1\right) f_{X_0|X_1,Y_1}\left(\frac{x'\theta^* - x_1'\theta_1^*}{\theta_0^*}, x_1, y\right) f_{Y_1,X_1}(y,x_1)d\mu}{\displaystyle\int f_{X_0|X_1,Y_1}\left(\frac{x'\theta^* - x_1'\theta_1^*}{\theta_0^*}, x_1, y\right) f_{Y_1,X_1}(y,x_1)d\mu},$$

which is a version of $g_{\theta^*}(x)$. □

II. Monte Carlo Evidence:

In this section we report the results of a simulation exercise designed to investigate the sampling distribution of propensity score matching estimators and the quality of the approximation to that distribution that is proposed in this article. The Monte Carlo results in this section illustrate the effect of adjusting standard errors and confidence intervals for the estimation error in the propensity score and confirm our theoretical results.

We report results for five designs and for the four type of estimators considered in section II. In all cases we use $N = 5000$ observations, two covariates, and 2000 Monte Carlo replications. For each design and each estimand (ATE or ATET) we calculate two estimators. The first estimator is based on matching on the true propensity score, $\widehat{\tau}^*_N = \widehat{\tau}_N(\theta^*)$ for ATE and $\widehat{\tau}^*_{t,N} = \widehat{\tau}_{t,N}(\theta^*)$ for ATET. The second estimator is based on matching on the estimated propensity score, $\widehat{\tau}_N = \widehat{\tau}_N(\widehat{\theta}_N)$ for ATE and $\widehat{\tau}_{t,N} = \widehat{\tau}_{t,N}(\widehat{\theta}_N)$ for ATET.

We estimate standard errors in three different ways. For estimators that match on the true propensity score, we construct standard errors using the formulas derived in Abadie and Imbens (2006) for the case when matching is done directly on covariates. Those formulas are valid because in this case matching is done using a covariate, which is the true propensity score. For the case when matching is done on the estimated propensity score, we first estimate standard errors without adjusting for estimation of the propensity score. That is, these are the standard errors that are obtained when the estimated propensity score is used for matching and for the estimation of the standard errors, but the fact that the propensity score is estimated is ignored in the calculation of the standard errors. These standard errors correspond to $\widehat{\sigma}/\sqrt{N}$ for ATE and $\widehat{\sigma}_t/\sqrt{N}$ for ATET, where $\widehat{\sigma}$ and $\widehat{\sigma}_t$ are given in section IV. For the case of matching on the estimated propensity score we also calculate the standard errors $\widehat{\sigma}_{\mathrm{adj}}/\sqrt{N}$ and $\widehat{\sigma}_{\mathrm{adj},t}/\sqrt{N}$, which adjust for estimation of the propensity score. For each estimator/standard error pair we evaluate the performance of (nominally) asymptotic 95 percent confidence intervals constructed by adding and subtracting 1.96 times the standard error to the estimator.

Design I: Two covariates, X_1 and X_2, are both uniformly distributed on $[-1/2, 1/2]$ and independent of each other. The potential outcomes are generated by $Y(0) = 3X_1 - 3X_2 + U_0$ and $Y(1) = 5 + 5X_1 + X_2 + U_1$, and U_0 and U_1 are independent standard Normal random variables, independent of (W, X_1, X_2). The treatment variable, W, is related to (X_1, X_2) through the propensity score, which is logistic

$$\Pr(W = 1|X_1 = x_1, X_2 = x_2) = \frac{\exp(x_1 + 2x_2)}{1 + \exp(x_1 + 2x_2)}.$$

ATE and ATET estimators use one match ($M = 1$).

Design II: X_1 and X_2 are distributed as in Design I. Potential outcomes are generated by $Y(0) = 10X_1 + U_0$ and $Y(1) = 5 - 10X_1 + U_1$, where U_0 and U_1 are standard Normal random variables independent of each other and of (W, X_1, X_2). The treatment variable, W, is related to (X_1, X_2) through the propensity score, which is logistic

$$\Pr(W = 1|X_1 = x_1, X_2 = x_2) = \frac{\exp(2x_2)}{1 + \exp(2x_2)}.$$

In this design, average treatment effect varies widely as a function of X, so $\partial\tau_t(\theta^*)/\partial\theta$ is large. In addition, in this design each value of the propensity score is associated with a unique value for the covariates, and therefore both c and c_t are equal to zero. ATE and ATET estimators use one match ($M = 1$).

Design III: In this design we modify Design I by using four matches ($M = 4$) on the estimated or true propensity score rather than a single match. The remainder of the design, including the potential outcome distributions and the assignment mechanism, is identical to that in Design I.

Design IV: This design is identical to Design II, except the estimated propensity score is misspecified as:

$$\frac{\exp((x_2 + 1/2)^2)}{1 + \exp((x_2 + 1/2)^2)}.$$

Notice that this specification is still a valid balancing score (Rosenbaum and Rubin, 1983) because the log odds ratio is a monotone function of the covariates.

Design V: In this design we change the distribution of the covariates. With probability 0.7, X_1 has a uniform distribution on the interval $[-1/2, 0]$, and with probability 0.3 it has a uniform distribution on the interval $[0, 1/2]$. With probability 0.6, X_2 has a truncated exponential distribution on the interval $[-1/2, 0]$, and with probability 0.4 it has a truncated exponential distribution on the interval $[0, 1/2]$. Hence the distribution of the covariates is discontinuous at zero.

In Table 1 we present Monte Carlo results for the ATE and in Table 2 we present Monte Carlo results for the ATET. In Design I the standard deviation of the estimator based on matching on the true propensity score is equal to 0.055 for ATE and 0.064 for ATET (row 1). The average the standard error across simulations are 0.055 for ATE and 0.065 for ATET, very close to the standard deviations. Asymptotic 95 percent confidence intervals provide coverage close to nominal (row 3). As predicted by the theoretical results, matching on the estimated propensity score leads to a smaller standard deviation for ATE, 0.057, (row 4) than matching on the true propensity score. For this design, the same is true for ATET: estimation of the ATET parameter matching of the estimated propensity score is more precise than matching on the true propensity score. Row 5 reports average standard errors when the fact that the propensity score was estimated is ignored in the construction of the standard errors. Row 6 reports average standard errors that adjust for the estimation of the propensity score. Standard errors that do not account for the estimation of the propensity score are severely biased as they approximate the standard deviation of the estimator for the case when the propensity score is known (in row 1). In contrast, the adjusted standard errors closely approximate the standard deviation of the estimators that match on the estimated propensity score (in row 4). Accordingly, using unadjusted standard errors to construct confidence intervals leads to over-coverage, while confidence intervals constructed using adjusted standard errors produce coverage rates that are close to nominal.

Our theoretical results predict that in Design II both the adjusted and unadjusted standard errors for ATE should perform well. The reason is that $c = 0$, so adjusting the ATE standard errors is not necessary. Moreover, our theoretical results predict that, in this design, the adjustment to the ATET standard errors for first step estimation of the propensity score is positive. The reason is that, in this design, $c_t = 0$ but $\partial\tau_t(\theta^*)/\partial\theta \neq 0$, which implies that matching on the estimated propensity score produces estimators with higher variance than matching on the true propensity score. Accordingly, confidence intervals for ATET constructed using matching on the estimated propensity score and unadjusted standard errors lead to under-coverage. The simulation results are consistent with these predictions.

In Design III with the number of matches equal to four instead of one, the precision of the estimator improves. There remains a bias in the standard errors that are based on ignoring the estimation of the propensity score. In Design IV the propensity score is misspecified. However, the misspecified propensity score is still a valid balancing score, and thus matching on it continues to remove the bias from all covariates. Coverage rates of the confidence intervals are still close to nominal levels. In the fifth and last design the distributions of the covariates include discontinuity points. This appears to have little effect on the performance of the confidence intervals.

Additional References

Abadie, A. and Imbens, G.W. (2006b) Additional Proofs for "Large Sample Properties of Matching Estimators for Average Treatment Effects", web appendix.

Johnson, N. and Kotz, S. (1977) *Urn Models and Their Applications*, John Wiley & Sons, New York.

Johnson, N., Kotz, S., and Kemp, A. (1993) *Univariate Discrete Distributions*, John Wiley & Sons, New York.

Lehmann, E.L. (1966) "Some Concepts of Dependence," *Annals of Mathematical Statistics*, vol. 37, no. 5, 1137-1153.

Poirier, D.J. (1995) *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press, Cambridge.

Rudin, W. (1976) *Principles of Mathematical Analysis*, McGraw-Hill, New York.

Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics.* Wiley, New York.

van der Vaart, A.W. and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.

Table I – Monte Carlo Results for Average Treatment Effect
($N = 5000$, Number of Replications = 2000)

Design	I	II	III	IV	V
Panel A: Matching on the True Propensity Score (PS)					
Standard Deviation	0.055	0.106	0.052	0.054	0.053
Average Standard Error	0.055	0.106	0.052	0.054	0.053
Coverage Rate of 95% Confidence Interval	0.947	0.950	0.943	0.946	0.943
Panel B: Matching on the Estimated PS					
Standard Deviation	0.045	0.105	0.039	0.045	0.044
Average Standard Error:					
Ignoring Estimation of the PS	0.055	0.106	0.046	0.054	0.053
Accounting for Estimation of the PS	0.045	0.106	0.039	0.044	0.044
Coverage Rate of 95% Confidence Interval:					
Ignoring Estimation of PS	0.985	0.950	0.981	0.986	0.981
Accounting for Estimation of the PS	0.951	0.950	0.947	0.949	0.946

Table II – Monte Carlo Results for Average Treatment Effect on the Treated
($N = 5000$, Number of Replications = 2000)

Design	I	II	III	IV	V
Panel A: Matching on the True Propensity Score (PS)					
Standard Deviation	0.064	0.123	0.058	0.059	0.063
Average Standard Error	0.065	0.123	0.057	0.059	0.063
Coverage Rate of 95% Confidence Interval	0.950	0.942	0.946	0.948	0.949
Panel B: Matching on the Estimated PS					
Standard Deviation	0.057	0.143	0.048	0.051	0.055
Average Standard Error:					
Ignoring Estimation of the PS	0.065	0.122	0.057	0.059	0.063
Accounting for Estimation of the PS	0.057	0.145	0.049	0.050	0.055
Coverage Rate of 95% Confidence Interval:					
Ignoring Estimation of PS	0.972	0.899	0.970	0.978	0.972
Accounting for Estimation of the PS	0.953	0.952	0.949	0.945	0.952