# REMATCHING ESTIMATORS FOR AVERAGE TREATMENT EFFECTS

Lam Lam Hui and Kin Wai Chan

*Department of Statistics, The Chinese University of Hong Kong*

**Supplementary Material**

## S1. Assumptions

We need the following regularity conditions for our asymptotic theory.

**Assumption S1.** $X$ is continuously distributed on a compact and convex support $\mathcal{X} \subset \mathbb{R}^k$. The density of $X$ is bounded and bounded away from zero on $\mathcal{X}$.

**Assumption S2.**

1. (i) $D$ is independent of $\{Y(0), Y(1)\}$ conditional on $X = x$ for almost every $x$.

   (ii) There exists $c > 0$ such that $c < \text{pr}(D = 1 \mid X = x) < 1 - c$ for almost every $x$.

2. $\{Y_i, D_i, X_i\}_{i=1}^{N}$ is an independent and identically distributed sample of $(Y, D, X)$.

**Assumption S3.**

1. (i) $D$ is independent of $Y(0)$ conditional on $X = x$ for almost every $x$.

   (ii) There exists $c > 0$ such that $\text{pr}(D = 1 \mid X = x) < 1 - c$ for almost every $x$.

2. Conditional on $D_i = d$, $\{(Y_i, X_i)\}_{i=1}^{N_d}$ is an independent and identically distributed sample from the distribution of $(Y, X \mid D = d)$ for each $d \in \{0, 1\}$. For some $r \geq 1$, $N_1^r / N_0 \to \theta$ with $\theta \in \mathbb{R}^+$.

**Assumption S4.** For each $d \in \{0, 1\}$,

1. $x \mapsto \mu(d, x)$ and $x \mapsto \sigma^2(d, x)$ are Lipschitz continuous in $\mathcal{X}$,
2. $x \mapsto E(Y^4 \mid D = d, X = x)$ is bounded uniformly on $\mathcal{X}$, and
3. $x \mapsto \sigma^2(d, x)$ is bounded away from zero on $\mathcal{X}$.

As is prevalent in this literature, analyses are based on Assumption S2.1 that (i) assignment to treatment is independent of potential outcomes conditional on observed pretreatment variables, and (ii) there is sufficient overlap in the support of the conditional distribution of $X_i$ given $D_i = 0$ and that of the conditional distribution of $X_i$ given $D_i = 1$. The combination of the unconfoundedness assumption in S2.1.(i) and the overlap assumption in S2.1.(ii) is referred to as strong ignorability (Rosenbaum and Rubin, 1983). These two assumptions are strong and may not be satisfied in some cases. However, according to Imbens and Wooldridge (2009), there is no general approach to estimating treatment effects without unconfoundedness, and they describe several methods for assessing its plausibility. The assumption of compactness and convexity of the support of the covariates in Assumption S1 are convenient regularity conditions. Assumption S2.2 assumes that the sampling is random. And Assumption S4 requires weak smoothness restrictions on the conditional distribution of $Y$ given $X$.

Let $\lambda = (\lambda_1, \ldots, \lambda_k)^{\mathrm{T}}$ be a $k$-dimensional vector of non-negative integers. The following assumptions are needed for deriving asymptotic properties of $\hat{\tau}_{\mathrm{bc}}$ and $\hat{\tau}^t_{\mathrm{bc}}$

**Assumption S5.** Let $|\lambda| = \sum_{\ell=1}^k \lambda_\ell$ and $x^\lambda = \prod_{\ell=1}^k x_\ell^{\lambda_\ell}$, where $x_\ell$ is the $\ell$th element of $x$. Define a series $\{\lambda(q)\}_{q=1}^\infty$ that contains all distinct vectors such that $|\lambda(q)|$ is non-decreasing. Let $Q$ be the series length and $p^Q(x) = (x^{\lambda(1)}, \ldots, x^{\lambda(Q)})^{\mathrm{T}}$. And let $N \mapsto Q(N)$ be an increasing function of $N$. A non-parametric series regression estimator $\hat{\mu}(d,x)$ is given by

$$\hat{\mu}(d,x) = p^{Q(N)}(x)^{\mathrm{T}} \left( \sum_{i:D_i=d} p^{Q(N)}(X_i) p^{Q(N)}(X_i)^{\mathrm{T}} \right)^{-} \sum_{i:D_i=d} p^{Q(N)}(X_i) Y_i,$$

where $(\cdot)^{-}$ is a generalized inverse.

**Assumption S6.**

1. The support of $X$, $\mathcal{X} \subset \mathbb{R}^k$, is a Cartesian product of compact intervals.
2. $Q(N) = N^{\nu}$, with $0 < \nu < \min\{2/(4k+3), 2/(4k^2 - k)\}$.
3. There is a $C$ such that for each multi-index $\lambda$ the $\lambda$th partial derivative of $\mu(d,x)$ exists for $d = 0, 1$ and is bounded by $C^{|\lambda|}$.

Here, Assumption S5 summarises the conditions for the series estimator $\hat{\mu}(d,x)$ employed by Abadie and Imbens (2011), and Assumption S6 is used to establish the asymptotic behaviour of bias-corrected rematching estimators $\hat{\tau}_{\mathrm{bc}}$ and $\hat{\tau}^{t}_{\mathrm{bc}}$.

## S2. Proof

Proposition 1 follows directly from Equation (3.6), and its proof is therefore omitted.

## S2.1 Proof of Proposition 2

First, it can be proven that

$$\begin{aligned}
\hat{\tau} &= \frac{1}{N}\sum_{i=1}^{N}\tilde{Y}_i(1)-\tilde{Y}_i(0) \\
&= \frac{1}{N}\left\{\sum_{\substack{1\le i\le N\\ D_i=1}}\left(Y_i-\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}Y_j\right)+\sum_{\substack{1\le i\le N\\ D_i=0}}\left(\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}Y_j-Y_i\right)\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N}(2D_i-1)Y_i-\sum_{\substack{1\le i\le N\\ D_i=1}}\sum_{\substack{1\le j\le N\\ D_j=0}}\frac{1}{M_0+M_{\mathrm{re}}(i)}\mathbb{1}\{j\in\mathcal{J}(i)\}Y_j \\
&\quad+\sum_{\substack{1\le i\le N\\ D_i=0}}\sum_{\substack{1\le j\le N\\ D_j=1}}\frac{1}{M_0+M_{\mathrm{re}}(i)}\mathbb{1}\{j\in\mathcal{J}(i)\}Y_j.
\end{aligned}$$

Note that

$$\begin{aligned}
&\sum_{\substack{1\le i\le N\\ D_i=0}}\sum_{\substack{1\le j\le N\\ D_j=1}}\frac{1}{M_0+M_{\mathrm{re}}(i)}\mathbb{1}\{j\in\mathcal{J}(i)\}Y_j \\
&\qquad=\sum_{\substack{1\le i\le N\\ D_i=0}}\sum_{\substack{1\le j\le N\\ D_j=1}}\frac{1}{M_0+M_{\mathrm{re}}(i)}[\mathbb{1}\{j\in\mathcal{J}_{M_0}(i)\}+\mathbb{1}\{j\in\mathcal{L}_{\mathrm{re}}(i)\}]Y_j \\
&\qquad=\sum_{\substack{1\le j\le N\\ D_j=1}}\left[\sum_{\substack{1\le i\le N\\ D_i=0}}\frac{\mathbb{1}\{j\in\mathcal{J}_{M_0}(i)\}}{M_0+M_{\mathrm{re}}(i)}+\sum_{\substack{1\le i\le N\\ D_i=0}}\frac{\mathbb{1}\{j\in\mathcal{L}_{\mathrm{re}}(i)\}}{M_0+M_{\mathrm{re}}(i)}\right]Y_j \\
&\qquad=\sum_{\substack{1\le j\le N\\ D_j=1}}\left[\sum_{\substack{1\le i\le N\\ D_i=0}}\frac{\mathbb{1}\{j\in\mathcal{J}_{M_0}(i)\}}{M_0+M_{\mathrm{re}}(i)}+\sum_{\substack{1\le i\le N\\ D_i=0}}\frac{\mathbb{1}\{i\in\mathcal{J}_{\mathrm{re}}(j)\}\mathbb{1}\{K_{M_0}(j)=0\}}{M_0+M_{\mathrm{re}}(\mathcal{J}_{\mathrm{re}}(j))}\right]Y_j \\
&\qquad=\sum_{\substack{1\le j\le N\\ D_j=1}}\left[\sum_{\substack{1\le i\le N\\ D_i=0}}\frac{\mathbb{1}\{j\in\mathcal{J}_{M_0}(i)\}}{M_0+M_{\mathrm{re}}(i)}+\frac{\mathbb{1}\{K_{M_0}(j)=0\}}{M_0+M_{\mathrm{re}}(\mathcal{J}_{\mathrm{re}}(j))}\right]Y_j \\
&\qquad=\sum_{\substack{1\le j\le N\\ D_j=1}}\frac{K(j)}{M_0}Y_j,
\end{aligned}$$

and similarly, we have

$$\sum_{\substack{1\leq i\leq N\\ D_i=1}}\sum_{\substack{1\leq j\leq N\\ D_j=0}}\frac{1}{M_0+M_{\mathrm{re}}(i)}\mathbb{1}\{j\in\mathcal{J}(i)\}Y_j=\sum_{\substack{1\leq j\leq N\\ D_j=0}}\frac{K(j)}{M_0}Y_j.$$

Hence,

$$\begin{aligned}\hat{\tau}&=\frac{1}{N}\sum_{i=1}^{N}(2D_i-1)Y_i-\sum_{\substack{1\leq j\leq N\\ D_j=0}}\frac{K(j)}{M_0}Y_j+\sum_{\substack{1\leq j\leq N\\ D_j=1}}\frac{K(j)}{M_0}Y_j\\ &=\frac{1}{N}\sum_{i=1}^{N}(2D_i-1)\left\{1+\frac{K(i)}{M_0}\right\}Y_i.\end{aligned}$$

Similarly, we can prove that

$$\hat{\tau}^t=\frac{1}{N_1}\sum_{i=1}^{N}\left\{D_i-(1-D_i)\frac{K(i)}{M_0}\right\}Y_i,$$

which is the desired result. □

### S2.2 Proof of Theorem 1

The proof below follows a similar argument as in the proof of Theorem 1(i) in Abadie and Imbens (2006). Recall from Section 4.1 that the bias term is given by

$$\begin{aligned}B&=\frac{1}{N}\sum_{i=1}^{N}(2D_i-1)\left[\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}\{\mu(1-D_i,X_i)-\mu\left(1-D_i,X_j\right)\}\right]\\ &=\frac{1}{N}\sum_{i=1}^{N}\sum_{j\in\mathcal{J}(i)}\frac{B_{j,i}}{M(i)},\end{aligned}$$

where $M(i) = |\mathcal{J}(i)|$, and

$$B_{j,i} = (2D_i - 1)\{\mu(1 - D_i, X_i) - \mu(1 - D_i, X_j)\}$$

is the unit-level bias for the match with index $j$ of the unit $i$. Let $U_{j,i} = X_j - X_i$ be the unit-level matching discrepancy. Then $B_{j,i}$ can be rewritten as follows:

$$\begin{aligned} B_{j,i} &= D_i\{\mu(0, X_i) - \mu(0, X_j)\} - (1 - D_i)\{\mu(1, X_i) - \mu(1, X_j)\} \\ &= D_i\{\mu(0, X_i) - \mu(0, X_i + U_{j,i})\} - (1 - D_i)\{\mu(1, X_i) - \mu(1, X_i + U_{j,i})\}, \end{aligned}$$

where $j \in \mathcal{J}(i)$. Next, we prove that $M(i)$ is bounded in probability. Recall that $M(i) = M_0 + M_{\mathrm{re}}(i)$, where $M_{\mathrm{re}}(i) = \sum_{j=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{\mathrm{re}}(j)\}$. By the definition of $\mathcal{J}_{\mathrm{re}}(i)$, we know that $\mathcal{J}_{\mathrm{re}}(j) \subseteq \mathcal{J}_1(j)$ for all $j = 1, \ldots, N$. So, we have $\mathbb{1}\{i \in \mathcal{J}_{\mathrm{re}}(j)\} \leq \mathbb{1}\{i \in \mathcal{J}_1(j)\}$ for all $i, j = 1, \ldots, N$, implying that

$$M_{\mathrm{re}}(i) \leq \sum_{j=1}^{N} \mathbb{1}\{i \in \mathcal{J}_1(j)\} \equiv K_1(i),$$

where the inequality essentially means that the number of times unit $i$ is used as the nearest match among *unmatched* units must be bounded above by the number of times unit $i$ is used as the nearest match among *all* units. According to Lemma 3(i) of Abadie and Imbens (2006), we have $K_1(i) = O_p(1)$. Hence, $M(i) \leq M_0 + K_1(i) = O_p(1)$.

Let $J(i) = \arg\max_{j\in\mathcal{J}(i)} B_{j,i},\ i = 1,\dots,N$, denoting the index of the farthest match in $\mathcal{J}(i)$. We obtain

$$\begin{aligned}
E(N^{2/k}B^2) =& N^{2/k}E\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j\in\mathcal{J}(i)}\frac{B_{j,i}}{M(i)}\right)^2 \\
\leq& N^{2/k}E\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j\in\mathcal{J}(i)}\frac{|B_{J(i),i}|}{M(i)}\right)^2 \\
=& N^{2/k}E\left(\frac{1}{N}\sum_{i=1}^{N}|B_{J(i),i}|\right)^2 .
\end{aligned}$$

Under the Lipschitz continuity assumption on $\mu(d, X)$ for $d \in \{0, 1\}$, we have $|B_{j,i}| \leq C_1\|U_{j,i}\|$, where $C_1$ is some positive constant that is finite by assumption. Hence, we get

$$\begin{aligned}
E(N^{2/k}B^2) \leq& N^{2/k}E\left(\frac{1}{N}\sum_{i=1}^{N}C_1\|U_{J(i),i}\|\right)^2 \\
=& N^{2/k}C_1^2\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E(\|U_{J(i),i}\|\|U_{J(j),j}\|) \\
\leq& N^{2/k}C_1^2\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(\frac{\|U_{J(i),i}\|^2+\|U_{J(j),j}\|^2}{2}\right) \\
=& N^{2/k}C_1^2\frac{2}{N^2}E\left(\sum_{i=1}^{N}\frac{N}{2}\|U_{J(i),i}\|^2\right) \\
=& N^{2/k-1}C_1^2\sum_{i=1}^{N}E\|U_{J(i),i}\|^2 \\
=& N^{2/k-1}C_1^2E\left\{\frac{1}{N_0^{2/k}}\sum_{i:D_i=1}E(N_0^{2/k}\left\|U_{J(i),i}\right\|^2 \mid D_1,\dots,D_N,X_i)\right. \\
&\left.+\frac{1}{N_1^{2/k}}\sum_{i:D_i=0}E(N_1^{2/k}\left\|U_{J(i),i}\right\|^2 \mid D_1,\dots,D_N,X_i)\right\}, \quad \text{(S2.1)}
\end{aligned}$$

where the last equality holds using the tower property.

According to Lemma 2 of Abadie and Imbens (2006), if Assumption S1 holds, then the normalized moments of the matching discrepancies, i.e., $N^{1/k}\|U_{j,i}\|$ for $j \in \mathcal{J}_m(i)$ and $m < \infty$, are bounded for all points in the support. However, for our proposed rematching estimator, the matched index set for the $i$th unit, i.e., $\mathcal{J}(i)$, is not equal to any $\mathcal{J}_m(i)$ with a fixed $m$. To derive an upper bound for $\|U_{J(i),i}\| \leq \sup_{j\in\mathcal{J}(i)} \|X_i - X_j\|$, we will bound the matching discrepancies and rematching discrepancies separately. By definition, $\mathcal{J}(i) = \mathcal{J}_{M_0}(i) \cup \mathcal{L}_{\text{re}}(i)$. So, it holds that

$$\sup_{j\in\mathcal{J}(i)} \|X_i - X_j\| \leq \sup_{j\in\mathcal{J}_{M_0}(i)} \|X_i - X_j\| + \sup_{j\in\mathcal{L}_{\text{re}}(i)} \|X_i - X_j\|, \qquad \text{(S2.2)}$$

where $\sup_{j\in\emptyset} \|X_i - X_j\| := 0$. Since $\mathcal{L}_{\text{re}}(i) = \{\ell \in \{1, \dots, N\} : i \in \mathcal{J}_{\text{re}}(\ell)\}$, we know that $j \in \mathcal{L}_{\text{re}}(i)$ implies $i \in \mathcal{J}_{\text{re}}(j)$. Also, note that

$$\begin{aligned} \mathcal{J}_{\text{re}}(j) &= \begin{cases} \mathcal{J}_1(j) & \text{if } K_{M_0}(j) = 0; \\ \emptyset & \text{if } K_{M_0}(j) > 0, \end{cases} \\ &\subseteq \mathcal{J}_1(j). \end{aligned}$$

Hence, $j \in \mathcal{L}_{\text{re}}(i)$ further implies $i \in \mathcal{J}_1(j)$. Then, let $\mathcal{S}(i) = \{j \ : \ i \in \mathcal{J}_1(j)\}$. We can bound (S2.2) as follows:

$$\sup_{j\in\mathcal{J}(i)} \|X_i - X_j\| \leq \sup_{j\in\mathcal{J}_{M_0}(i)} \|X_i - X_j\| + \sup_{j\in\mathcal{S}(i)} \|X_i - X_j\|. \qquad \text{(S2.3)}$$

Now let $\kappa(k)$ be the kissing number in dimension $k$, i.e., the maximum number of non-overlapping unit spheres that can touch a common unit sphere. By considering the ratio of volumes of $k$-dimensional spheres with radii 1 and 3, it is not hard to see that $\kappa(k) \leq 3^k$ is bounded for any finite $k$; see, e.g., Cohn and Zhao (2014) for a sharper bound. Note that a particular point $X_i$ can only accommodate a limited number of the points regarding $X_i$ as their nearest neighbor due to spatial constraints. The value of $|\mathcal{S}(i)|$ is maximized when we arrange the maximum number of unit spheres around a unit sphere centered at $X_i$ in a way that each sphere touches the central sphere without overlapping. By definition, this maximum number is the kissing number; see, e.g., Kozakova et al. (2006) and Zong (1998) for more details. In other words, the set $\mathcal{S}(i)$ contains at most $\kappa(k)$ points, i.e., $|\mathcal{S}(i)| \leq \kappa(k) \leq 3^k$.

Hence, $j \in \mathcal{S}(i)$ implies $j \in \mathcal{J}_{3^k}(i)$. Consequently, (S2.3) can be further bounded as follows:

$$\sup_{j \in \mathcal{J}(i)} \|X_i - X_j\| \leq \sup_{j \in \mathcal{J}_{M_0}(i)} \|X_i - X_j\| + \sup_{j \in \mathcal{J}_{3^k}(i)} \|X_i - X_j\|. \tag{S2.4}$$

Each term on the right-hand side of (S2.4) is bounded in view of Lemma 2 of Abadie and Imbens (2006). Moreover, the conditional expectations $E(N_0^{2/k} \left\|U_{J(i),i}\right\|^2 \mid D_1, \ldots, D_N, X_i)$ and $E(N_1^{2/k} \left\|U_{J(i),i}\right\|^2 \mid D_1, \ldots, D_N, X_i)$ are both bounded by a constant $C$ that does not depend on $i$ according to

Lemma 2 of Abadie and Imbens (2006). Thus, using this uniform bound, we have

$$\sum_{i:D_i=1} E(N_0^{2/k} \left\|U_{J(i),i}\right\|^2 \mid D_1, \ldots, D_N, X_i) \leq CN_1,$$

$$\sum_{i:D_i=0} E(N_1^{2/k} \left\|U_{J(i),i}\right\|^2 \mid D_1, \ldots, D_N, X_i) \leq CN_0.$$

Therefore, (S2.1) can be bounded as follows:

$$E(N^{2/k}B^2) \leq C_2 E\left\{\left(\frac{N}{N_0}\right)^{2/k}\frac{N_1}{N} + \left(\frac{N}{N_1}\right)^{2/k}\frac{N_0}{N}\right\}$$

for some positive constant $C_2$. By Chernoff's inequality, any moment of $N/N_1$ or $N/N_0$ is uniformly bounded in $N$, as in Abadie and Imbens (2006). Part 1 of the theorem follows from Markov's inequality.

The proof of part 2 is similar to that of part 1 and is omitted. □

### S2.3 Proof of Lemma 1

By Lemma 3(i) of Abadie and Imbens (2006), $K_{M_0}(i) = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\} = O_p(1)$. Consider

$$K(i) = \left[\sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\}\frac{M_0}{M_{\mathrm{re}}(\ell) + M_0}\right] + \mathbb{1}\{K_{M_0}(i) = 0\}\frac{M_0}{M_{\mathrm{re}}(\mathcal{J}_{\mathrm{re}}(i)) + M_0}$$
$$= \begin{cases} \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\}M_0/\{M_{\mathrm{re}}(\ell) + M_0\} & \text{if } K_{M_0}(i) \neq 0; \\ M_0/\{M_{\mathrm{re}}(\mathcal{J}_{\mathrm{re}}(i)) + M_0\} & \text{if } K_{M_0}(i) = 0, \end{cases}$$

where $M_0/\{M_{\mathrm{re}}(i)+M_0\} \in (0,1]$ is the adjusted weight for unit $i$ after rematching and $M_0/\{M_{\mathrm{re}}(\mathcal{J}_{\mathrm{re}}(i))+M_0\}$ is the weight given, after rematching, to unit $i$ that is unmatched in the simple matching. When $K_{M_0}(i) \neq 0$, we have

$$K(i) = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\} \frac{M_0}{M_{\mathrm{re}}(\ell) + M_0} \leq K_{M_0}(i) = O_p(1).$$

And when $K_{M_0}(i) = 0$,

$$K(i) = \frac{M_0}{M_{\mathrm{re}}(\mathcal{J}_{\mathrm{re}}(i)) + M_0} \in (0,1).$$

Therefore, $K(i) = O_p(1)$.

According to Lemma 3 of Abadie and Imbens (2006), the $q$th moment of $K_{M_0}(i)$ conditional on $X_i = x, D_{1:N}$, which is $E\{K_{M_0}(i)^q \mid X_i = x, D_{1:N}\}$, is uniformly bounded in $N$. Since $K(i)$ is either $\leq K_{M_0}(i)$ or $\in (0,1)$, $E\{K(i)^q \mid X_i = x, D_{1:N}\}$ will also be bounded in $N$, so the first part of the lemma follows.

Part 2 of Lemma 1 can be proven using the same argument as for $E\{K(i)^q\}$.

For part 3 of Lemma 1, because the moments of $K(i)$ are bounded uniformly in $N$, and because the variance $\sigma^2(d,x)$ is bounded by $\bar{\sigma}^2 = \sup_{x,d} \sigma^2(d,x)$, which is finite by Assumption S4, $E[\{1+K(i)/M_0\}^2\sigma^2(d,x)]$ is bounded by $\bar{\sigma}^2 E[\{1+K(i)/M_0\}^2]$ and is finite. Hence $E(V^R) = O(1)$. $\square$

### S2.4 Proof of Theorem 2

Before proving Theorem 2, we need some preliminary results given by Abadie and Imbens (2002). First, let $\bar{r}_d(x)$ be the number of units with $D_i = d$ and $X_i \geq x$. Define $X_{(i,k)} = X_j$ if $\bar{r}_{D_i}(X_i) - \bar{r}_{D_i}(X_j) = k$, and $\bar{r}_{D_i}(X_i) - \lim_{x \uparrow X_j} \bar{r}_{D_i}(x) = k - 1$. When $X$ is a scalar, $A_{M_0}(i) = (X_i/2 + X_{(i,-M_0)}/2, X_i/2 + X_{(i,M_0)}/2)$. Then, the exact conditional distribution of $K_{M_0}(i)$ is,

$$K_{M_0}(i) \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1 \sim \text{Bin}\left(N_0, \int_{A_{M_0}(i)} f_0(z)dz\right),$$

$$K_{M_0}(i) \mid \mathbf{D}, \{X_j\}_{D_j=0}, D_i = 0 \sim \text{Bin}\left(N_1, \int_{A_{M_0}(i)} f_1(z)dz\right).$$

Next, let Gamma($\alpha$, $\beta$) denote the gamma distribution with a shape parameter $\alpha$ and a scale parameter $\beta$. Given $X_i = x$, and $D_i = 1$,

$$2N_1 \cdot \frac{f_1(x)}{f_0(x)} \cdot \int_{A_{M_0}(i)} f_0(z)dz \to \text{Gamma}(2M_0, 1)$$

in distribution, and given $X_i = x$ and $D_i = 0$,

$$2N_0 \cdot \frac{f_0(x)}{f_1(x)} \cdot \int_{A_{M_0}(i)} f_1(z)dz \to \text{Gamma}(2M_0, 1)$$

in distribution.

Note that

$$N\text{var}(\hat{\tau}) = E(V^R) + V^{\tau(X)}$$

$$\begin{aligned}
&= E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(D_i, X_i)\right] + V^{\tau(X)} \\
&= E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(1, X_i) \mid D_i = 1\right] p \\
&\quad + E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(0, X_i) \mid D_i = 0\right] (1-p) + V^{\tau(X)},
\end{aligned}$$

where $p$ is the probability of treatment assignment. Similarly,

$$\begin{aligned}
N\text{var}(\hat{\tau}_0) &= E(V_0^R) + V^{\tau(X)} \\
&= E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(1, X_i) \mid D_i = 1\right] p \\
&\quad + E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(0, X_i) \mid D_i = 0\right] (1-p) + V^{\tau(X)}.
\end{aligned}$$

The number of times a unit is used as a match is calculated as

$$K_{M_0}(i) = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\}$$

when simple matching estimators are used. Recall that $M(\ell) = M_0 + M_{\text{re}}(\ell)$.

By Proposition 1,

$$K(i) = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}(\ell)\} \frac{M_0}{M(\ell)},$$

which is not comparable with $K_{M_0}(i)$ as it is a weighted form. To obtain a comparable form, we let

$$H(i) = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}(\ell)\}.$$

It is the number of times a unit is used as a match when rematching estimators are used. Equivalently,

$$H(i) = \max\{K_{M_0}(i), 1\},$$

since if a unit remains unmatched after the first matching, it will be matched to its nearest neighbour in the rematching process.

The conditional expectation of $H(i)$ is

$$\begin{aligned}
&E[H(i) \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] \\
&= E[\max\{K_{M_0}(i), 1\} \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] \\
&= E[K_{M_0}(i)\mathbb{1}\{K_{M_0}(i) \geq 1\} \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] \\
&\qquad + E[\mathbb{1}\{K_{M_0}(i) < 1\} \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] \\
&= E[K_{M_0}(i) \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] \\
&\qquad + \mathrm{pr}[K_{M_0}(i) < 1 \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1],
\end{aligned}$$

and the conditional expectation of $H^2(i)$ is

$$\begin{aligned}
&E[H^2(i) \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] \\
&= E[K_{M_0}^2(i) \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] \\
&\qquad + \mathrm{pr}[K_{M_0}(i) < 1 \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1]
\end{aligned}$$

Define $P_{M_0}(i) = D_i \cdot \int_{A_{M_0}(i)} f_0(z)dz + (1-D_i) \cdot \int_{A_{M_0}(i)} f_1(z)dz$. According to Abadie and Imbens (2002),

$$E[K_{M_0}(i) \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] = N_0 P_{M_0}(i),$$

$$E[K^2_{M_0}(i) \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] = N_0 P_{M_0}(i) + N_0(N_0 - 1) P_{M_0}(i)^2,$$

and therefore,

$$\begin{aligned} & E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(1, X_i) \middle| D_i = 1\right] \\ & = E\left(\left[1 + \frac{1}{M^2}\{N_0 P_{M_0}(i) + N_0(N_0-1)P_{M_0}(i)^2\}\right.\right. \\ & \qquad \left.\left. + \ \frac{2}{M_0} N_0 P_{M_0}(i)\right] \sigma^2(1, X_i) \middle| \ D_i = 1\right) \\ & = E\left(\left[1 + \frac{1}{M_0}\left\{\frac{(1-p)}{p}\frac{f_0(X_i)}{f_1(X_i)} + \frac{(1-p)^2}{2p^2}\frac{f_0(X_i)^2}{f_1(X_i)^2}(2M+1)\right\}\right.\right. \\ & \qquad \left.\left. + \ \frac{2(1-p)}{p}\frac{f_0(X_i)}{f_1(X_i)}\right] \sigma^2(1, X_i) \middle| \ D_i = 1\right) + o(1). \end{aligned}$$

Similarly, we derive that

$$E[H(i) \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] = N_0 P_{M_0}(i) + \{1 - P_{M_0}(i)\}^{N_0},$$

$$\begin{aligned} & E[H^2(i) \mid \mathbf{D}, \{X_j\}_{D_j=1}, D_i = 1] \\ & \qquad = N_0 P_{M_0}(i) + N_0(N_0-1)P_{M_0}(i)^2 + \{1 - P_{M_0}(i)\}^{N_0}. \end{aligned}$$

And

$$\begin{aligned} & E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(1, X_i) \middle| D_i = 1\right] \\ & \leq \ E\left[\left\{1 + \frac{H(i)}{M_0}\right\}^2 \sigma^2(1, X_i) \middle| D_i = 1\right] \end{aligned}$$

$$= E\left\{\left(1 + \frac{1}{M^2}[N_0 P_{M_0}(i) + N_0(N_0-1)P_{M_0}(i)^2 + \{1 - P_{M_0}(i)\}^{N_0}] + \frac{2}{M_0}\{N_0 P_{M_0}(i) + (1 - P_{M_0}(i))^{N_0}\}\right)\sigma^2(1, X_i)\,\middle|\, D_i = 1\right\}.$$

The first inequality holds because $K(i)/M_0 = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}(\ell)\}/M(\ell)$, $H(i)/M_0 = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}(\ell)\}/M_0$, and $M(\ell) \geq M_0$ for all $\ell$. Since $\{1 - P_{M_0}(i)\} \in [0, 1)$ when $M < \min(N_0, N_1)$, we know that $\{1 - P_{M_0}(i)\}^{N_0} = o(1)$. So, similarly, we know that

$$\begin{aligned}
&E\left[\left(1 + \frac{H(i)}{M_0}\right)^2 \sigma^2(1, X_i) \mid D_i = 1\right] \\
&= E\left[\left(1 + \frac{1}{M_0}\left\{\frac{(1-p)}{p}\frac{f_0(X_i)}{f_1(X_i)} + \frac{(1-p)^2}{2p^2}\frac{f_0(X_i)^2}{f_1(X_i)^2}(2M+1)\right\} + \frac{2(1-p)}{p}\frac{f_0(X_i)}{f_1(X_i)}\right)\sigma^2(1, X_i)\,\middle|\, D_i = 1\right] + o(1) \\
&= E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(1, X_i)\,\middle|\, D_i = 1\right].
\end{aligned}$$

That is,

$$\begin{aligned}
&E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(1, X_i)\,\middle|\, D_i = 1\right] \\
&\qquad \leq E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(1, X_i)\,\middle|\, D_i = 1\right].
\end{aligned}$$

Similarly,

$$\begin{aligned}
&E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(0, X_i) \mid D_i = 0\right] \\
&\qquad \leq E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(0, X_i) \mid D_i = 0\right].
\end{aligned}$$

Then, it is obvious that $N\text{var}(\hat{\tau}) \leq N\text{var}(\hat{\tau}_0)$. □

### S2.5 Proof of Theorem 3

Note that

$$\begin{aligned}
&N\text{var}(\hat{\tau}) \\
&= E(V^R) + V^{\tau(X)} \\
&= E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(D_i, X_i)\right] + V^{\tau(X)} \\
&= E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(1, X_i) \mid D_i = 1\right] p \\
&\quad + E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(0, X_i) \mid D_i = 0\right] (1-p) + V^{\tau(X)},
\end{aligned}$$

where $p$ is the probability of treatment assignment. Similarly,

$$\begin{aligned}
N\text{var}(\hat{\tau}_0) &= E(V_0^R) + V^{\tau(X)} \\
&= E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(1, X_i) \mid D_i = 1\right] p \\
&\quad + E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(0, X_i) \mid D_i = 0\right] (1-p) + V^{\tau(X)}.
\end{aligned}$$

It is known that

$$K_{M_0}(i) = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\}$$

and

$$\frac{K(i)}{M_0} = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\} \frac{1}{M_{\text{re}}(\ell) + M_0} + \mathbb{1}\{K_{M_0}(i) = 0\} \frac{1}{M_{\text{re}}(\mathcal{J}_{\text{re}}(i)) + M_0}$$

$$= \begin{cases} \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\} 1/\{M_{\mathrm{re}}(\ell) + M_0\} & \text{if } K_{M_0}(i) \neq 0; \\ 1/\{M_{\mathrm{re}}(\mathcal{J}_{\mathrm{re}}(i)) + M_0\} & \text{if } K_{M_0}(i) = 0. \end{cases}$$

Then we state the results below given the values of $N_0$ and $N_1$. Under the assumptions that $k = 1$, $f_0(x) = f_1(x)$, and $M_0 N_1 < N_0$, we have, for $D_i = 1$,

$$K(i) = K_{M_0}(i) \sim \mathrm{Bin}(M_0 N_0, 1/N_1).$$

For $D_i = 0$, given that $f_0(x) = f_1(x)$, the term $K_{M_0}(i)$ can only take values in $\{0, 1\}$ when there exist unmatched points. When $K_{M_0}(i) = 1$ for the control unit $i$, exactly one of $\mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\}$ $(\ell = 1, \dots, N)$ is 1. Since the number of matched control is $M_0 N_1$, we have $K_{M_0}(i) \sim \mathrm{Bern}(M_0 N_1 / N_0)$, and

$$\frac{K(i)}{M_0} = \begin{cases} \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\} 1/\{M_{\mathrm{re}}(\ell) + M_0\} & \text{if } K_{M_0}(i) = 1; \\ 1/\{M_{\mathrm{re}}(\mathcal{J}_{\mathrm{re}}(i)) + M_0\} & \text{if } K_{M_0}(i) = 0. \end{cases}$$

And also,

$$[K_{M_0}(i) \mid D = 1] \sim \mathrm{Bin}(M_0 N_0, 1/N_1),$$

$$[\mathbb{1}\{K_{M_0}(i) = 1\} \mid D = 0] \sim \mathrm{Bern}(M_0 N_1 / N_0),$$

$$[M_{\mathrm{re}}(\ell) \mid K_{M_0}(i) = 1, D = 0] \sim \mathrm{Bin}(N_0 - M_0 N_1, 1/N_1),$$

$$[\{M_{\mathrm{re}}(\mathcal{J}_{re}(i)) - 1\} \mid K_{M_0}(i) = 0, D = 0] \sim \mathrm{Bin}(N_0 - M_0 N_1 - 1, 1/N_1).$$

Denoting $[K_{M_0}(i) \mid D = 1]$ as $K$, $[\mathbb{1}\{K_{M_0}(i) = 1\} \mid D = 0]$ as $I$, $[M_{\text{re}}(\ell) \mid K_{M_0}(i) = 1, D = 0]$ and $[M_{\text{re}}(\mathcal{J}_{re}(i)) - 1 \mid K_{M_0}(i) = 0, D = 0]$ as $[H \mid I]$, we have

$$
\begin{aligned}
K &\sim \text{Bin}(M_0 N_0, 1/N_1), \\
I &\sim \text{Bern}(M_0 N_1/N_0), \\
[H \mid I] &\sim \text{Bin}(N_0 - M_0 N_1 - \mathbb{1}(I = 0), 1/N_1).
\end{aligned}
$$

Now, using the assumption that $\sigma^2(d, x) = \sigma_d^2$ is a constant as a function of $x$ for each $d \in \{0, 1\}$, we have that

$$
\begin{aligned}
N\text{var}(\hat{\tau}) \to\ & E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(1, X_i) \mid D_i = 1\right] p \\
& + E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \sigma^2(0, X_i) \mid D_i = 0\right] (1 - p) + V^{\tau(X)} \\
=\ & \sigma_0^2 (1 - p) E\left[\left\{1 + \sum_{\ell=1}^{N} \frac{\mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\}}{M_{\text{re}}(\ell) + M_0} + \frac{\mathbb{1}\{K_{M_0}(i) = 0\}}{M_{\text{re}}(\mathcal{J}_{\text{re}}(i)) + M_0}\right\}^2 \mid D_i = 0\right] \\
& + \sigma_1^2 p E\left[\left\{1 + \frac{K(i)}{M_0}\right\}^2 \mid D_i = 1\right] + V^{\tau(X)} \\
=\ & \sigma_1^2 p E\left(1 + \frac{K}{M_0}\right)^2 + \sigma_0^2 (1 - p) E\left(1 + \frac{I}{H + M_0} + \frac{1 - I}{H + 1 + M_0}\right)^2 + V^{\tau(X)}.
\end{aligned}
$$

For comparison, $N\text{var}(\hat{\tau}_0)$ can be written as

$$
\begin{aligned}
N\text{var}(\hat{\tau}_0) \to\ & E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(1, X_i) \mid D_i = 1\right] p \\
& + E\left[\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\}^2 \sigma^2(0, X_i) \mid D_i = 0\right] (1 - p) + V^{\tau(X)} \\
=\ & \sigma_1^2 p E\left(1 + \frac{K}{M_0}\right)^2 + \sigma_0^2 (1 - p) E\left(1 + \frac{I}{M_0}\right)^2 + V^{\tau(X)}.
\end{aligned}
$$

Similarly, we have

$$N_1 \text{var}(\hat{\tau}^t) \to \sigma_1^2 p + \sigma_0^2 (1-p) E\left(\frac{I}{H+M_0} + \frac{1-I}{H+1+M_0}\right)^2 + V^{\tau(X),t},$$
$$N_1 \text{var}(\hat{\tau}_0^t) \to \sigma_1^2 p + \sigma_0^2 (1-p) E\left(\frac{I}{M_0}\right)^2 + V^{\tau(X),t}.$$

Then the desired results follow. □

## S2.6 Proof of Proposition 3

By Theorem 3, we have

$$N\{\text{var}(\hat{\tau}_0^t) - \text{var}(\hat{\tau}^t)\} \to \sigma_0^2(1-p)\left\{E\left(\frac{I}{M_0}\right)^2 - E\left(\frac{I}{H+M_0} + \frac{1-I}{H+1+M_0}\right)^2\right\}.$$

Note that

$$E\left(\frac{I}{M_0}\right)^2 = \frac{1}{M_0^2}\frac{M_0 N_1}{N_0} = \frac{N_1}{N_0 M_0},$$

and

$$\begin{aligned}
&E\left(\frac{I}{H+M_0} + \frac{1-I}{H+1+M_0}\right)^2 \\
&= E\left\{\left(\frac{1}{H+M_0}\right)^2 \mid I=1\right\}\text{pr}(I=1) + E\left\{\left(\frac{1}{H+1+M_0}\right)^2 \mid I=0\right\}\text{pr}(I=0) \\
&= E\left(\frac{1}{H_1+M_0}\right)^2 \frac{M_0 N_1}{N_0} + E\left(\frac{1}{H_0+1+M_0}\right)^2\left(1 - \frac{M_0 N_1}{N_0}\right) \\
&= E\left(\frac{M_0}{H_1+M_0}\right)^2 \frac{N_1}{N_0 M_0} + E\left(\frac{M_0}{H_0+1+M_0}\right)^2\left(\frac{1}{M_0^2} - \frac{N_1}{N_0 M_0}\right).
\end{aligned}$$

We have

$N\{\text{var}(\hat{\tau}_0^t) - \text{var}(\hat{\tau}^t)\}$

$$= \sigma_0^2(1-p)\left[\frac{N_1}{N_0 M_0}\left\{1 - E\left(\frac{M_0}{H_1+M_0}\right)^2 + E\left(\frac{M_0}{H_0+1+M_0}\right)^2\left(\frac{N_0}{N_1 M_0}-1\right)\right\}\right].$$

It is obvious that $E\left(\frac{M_0}{H_1+M_0}\right)^2 \leq 1$ , and we have $N_0 > N_1 M_0$ by assumption. Hence,

$$N\{\text{var}(\hat{\tau}_0^t) - \text{var}(\hat{\tau}^t)\} > 0.$$

Then the result follows. □

## S2.7 Proof of Theorem 4

We consider each of the terms in (4.9) separately. For the first term, by the law of large numbers $\bar{\tau}(X) \to \tau$ in probability. For the weighted average of the residual $R$, we have $NE\{(R)^2\} = E[\{1 + K(i)/M_0\}^2\sigma^2(d,x)] = O(1)$, which is finite, hence, $R = O_p(N^{-1/2}) = o_p(1)$. And for the bias term, by Theorem 1, $B = O_p(N^{-1/k}) = o_p(1)$. We thus prove the first part of the theorem. The proof of the second part follows the same argument and is therefore omitted. □

## S2.8 Proof of Theorem 5

Since the second part of the theorem follows the same argument as the first part, we only prove part (i). First, $(V^{\tau(X)})^{-1/2}N^{1/2}\{\bar{\tau}(X) - \tau\} \to \mathrm{N}(0,1)$ in distribution by central limit theorem. Second, since $E[\{1 + K(i)/M_0\}^4]$

is uniformly bounded, we then obtain

$$\frac{N^{1/2}\sum_{i=1}^{N} R_i}{[\sum_{i=1}^{N}\{1+K(i)/M_0\}^2\sigma^2(D_i,X_i)]^{1/2}} = (V^R)^{-1/2}N^{1/2}R \to \mathrm{N}(0,1)$$

in distribution in a similar manner to Proof of Theorem 4 of Abadie and Imbens (2006) using a Lindeberg-Feller central limit theorem. In addition, $(V^R)^{-1/2}N^{1/2}R$ and $(V^{\tau(X)})^{-1/2}N^{1/2}\{\bar{\tau}(X)-\tau\}$ are asymptotically independent. Therefore, the convergence to $\mathrm{N}(0,1)$ of both $(V^R)^{-1/2}N^{1/2}R$ and $(V^{\tau(X)})^{-1/2}N^{1/2}\{\bar{\tau}(X)-\tau\}$, boundedness of $V^R$ and $V^{\tau(X)}$, and boundedness of $V^R$ away from zero imply that

$$\{V^R + V^{\tau(X)}\}^{-1/2}N^{1/2}(\hat{\tau} - B - \tau) \to \mathrm{N}(0,1)$$

in distribution. Then the desired results follow. □

### S2.9 Proof of Theorem 6

We only prove the first part since part (ii) of the theorem follows the same argument as part (i). Since it is proven in Section S2.2 that $\sup_{j\in\mathcal{J}(i)}\|X_i - X_j\|$ is bounded, it follows from Lemma A.2 of Abadie and Imbens (2011) that

$$\max_{i=1,\ldots,N}|\hat{\mu}(D_i,X_i)-\hat{\mu}(D_i,X_j)-\{\mu(D_i,X_i)-\mu(D_i,X_j)\}| = o_p(N^{-1/2}),$$

where $j\in\mathcal{J}(i)$. The difference $|\hat{B}-B|$ can be written as

$|\hat{B}-B|$

$$
\begin{aligned}
&\leq \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}|D_i\{\hat{\mu}(0,X_i)-\hat{\mu}(0,X_j)\}-(1-D_i)\{\hat{\mu}(1,X_i)-\hat{\mu}(1,X_j)\}\\
&\quad -[D_i\{\mu(0,X_i)-\mu(0,X_j)\}-(1-D_i)\{\mu(1,X_i)-\mu(1,X_j)\}]|\\
&\leq \max_{\substack{i=1,\ldots,N\\ j\in\mathcal{J}(i)}}\sum_{d=0,1}|\hat{\mu}(d,X_i)-\hat{\mu}(d,X_j)-\{\mu(d,X_i)-\mu(d,X_j)\}|\\
&=o_p(N^{-1/2}).
\end{aligned}
$$

Then the desired results follow. □

## S2.10 Proof of Theorem 7

First notice that

$$
\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0}\right\}^2\{\sigma^2(D_i,X_i)-\hat{\sigma}^2(D_i,X_i)\}\right|\\
&\qquad\leq\left|\frac{1}{N}\sum_{i=1}^{N}\left(1+\frac{K(i)}{M_0}\right)^2[\sigma^2(D_i,X_i)-E\{\hat{\sigma}^2(D_i,X_i)\mid X_{1:N},D_{1:N}\}]\right|\\
&\qquad\quad+\left|\frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0}\right\}^2[E\{\hat{\sigma}^2(D_i,X_i)\mid X_{1:N},D_{1:N}\}-\hat{\sigma}^2(D_i,X_i)]\right|.
\end{aligned}
$$

Given Lemma 1.1 that all moments of $K(i)$ are bounded uniformly in $N$, it follows from Proof of Theorem 6 of Abadie and Imbens (2002) that both terms on the right-hand side of the above inequality converge to zero. And thus

$$
\frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0}\right\}^2\hat{\sigma}^2(D_i,X_i)\to V^R \tag{S2.5}
$$

in probability. Now, consider

$$\frac{1}{N}\sum_{i=1}^{N}\{\tilde{Y}_i(1)-\tilde{Y}_i(0)-\hat{\tau}\}^2 = \frac{1}{N}\sum_{i=1}^{N}\{\tilde{Y}_i(1)-\tilde{Y}_i(0)-\tau\}^2-(\hat{\tau}-\tau)^2$$
$$= \frac{1}{N}\sum_{i=1}^{N}\{\tilde{Y}_i(1)-\tilde{Y}_i(0)-\tau\}^2+o_p(1).$$

We decompose

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\tilde{Y}_i(1)-\tilde{Y}_i(0)-\tau\right\}^2$$
$$=\frac{1}{N}\sum_{i=1}^{N}\left[(2D_i-1)\left\{\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}\mu(D_i,X_i)-\mu(1-D_i,X_j)\right\}-\tau\right]^2+\frac{1}{N}\sum_{i=1}^{N}\left\{\epsilon_i-\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}\epsilon_j\right\}^2$$
$$+\frac{2}{N}\sum_{i=1}^{N}\left[(2D_i-1)\left\{\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}\mu(D_i,X_i)-\mu(1-D_i,X_j)\right\}-\tau\right](2D_i-1)\left\{\epsilon_i-\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}\epsilon_j\right\}.$$

Since the sample maximum of the norms of the matching discrepancies $\|X_i - X_j\|$ is $o_p(1)$, and the regression functions $\mu(d,x)$ are Lipschitz, for $d = 0, 1$, we follow a similar argument to the proof of Theorem 7 of Abadie and Imbens (2006) to obtain

$$\frac{1}{N}\sum_{i=1}^{N}\left[(2D_i-1)\left\{\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}\mu(D_i,X_i)-\mu(1-D_i,X_j)\right\}-\tau\right]^2$$
$$=\frac{1}{N}\sum_{i=1}^{N}\{\mu(1,X_i)-\mu(0,X_i)-\tau\}^2+o_p(1)$$

for the first term on the right-hand side of the above equation. Also similar to Proof of Theorem 7 of Abadie and Imbens (2006), we have

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\epsilon_i-\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}\epsilon_j\right\}^2-\frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0^2}\right\}\sigma^2(D_i,X_i)=o_p(1)$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\{\tilde{Y}_i(1)-\tilde{Y}_i(0)-\hat{\tau}\}^2 = \frac{1}{N}\sum_{i=1}^{N}\{\mu_1(X_i)-\mu_0(X_i)-\tau\}^2 + \frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0^2}\right\}\sigma^2(D_i,X_i)+o_p(1)$$

for the second and the last terms on the right-hand side of the equation, respectively. Similar to the convergence (S2.5), it can be obtained that

$$\left|\frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0^2}\right\}\sigma^2(D_i,X_i)-\frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0^2}\right\}\hat{\sigma}^2(D_i,X_i)\right| = o_p(1).$$

Therefore,

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\tilde{Y}_i(1)-\tilde{Y}_i(0)-\hat{\tau}\right\}^2 - \frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0^2}\right\}\hat{\sigma}^2(D_i,X_i)\to V^{\tau(X)} \tag{S2.6}$$

in probability. Since

$$\begin{aligned}\hat{V} &= \frac{1}{N}\sum_{i=1}^{N}\{\tilde{Y}_i(1)-\tilde{Y}_i(0)-\hat{\tau}\}^2 \\ &\quad + \frac{1}{N}\sum_{i=1}^{N}\left[\left\{\frac{K(i)}{M_0}\right\}^2+\left(\frac{2M-1}{M_0}\right)\left\{\frac{K(i)}{M_0}\right\}\right]\hat{\sigma}^2(D_i,X_i) \\ &= \frac{1}{N}\sum_{i=1}^{N}\{\tilde{Y}_i(1)-\tilde{Y}_i(0)-\hat{\tau}\}^2+\frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0}\right\}^2\hat{\sigma}^2(D_i,X_i) \\ &\quad - \frac{1}{N}\sum_{i=1}^{N}\left\{1+\frac{K(i)}{M_0^2}\right\}\hat{\sigma}^2(D_i,X_i),\end{aligned}$$

the convergences (S2.5) and (S2.6) imply that

$$\hat{V}\to V^R+V^{\tau(X)}$$

in probability, which finishes the proof for the first part of the theorem. The proof for the second part is similar to that of part one and is omitted. □

## S3. Simulation results

In our simulation designs, the average treatment effects in this simulation design are set as $\tau = \tau^t = 1$. The parameter $c$ controls the ratio of the treated. Since matching estimators have been used when the interest is in $\tau^t$, and there are more controls than the treated, we adjust $c$ to fix the ratio of the treated roughly at 0.3 for all cases. For example, letting $c = 0.4$ will lead to a treated ratio of 0.3 for all single covariate cases. For multivariate cases, we fix $c$ at 0.45 for Cases 2(a), 2(b), 2(d), and 2(e) and at 0.55 for Case 2(c); see Remark S1 for more discussions. For each case, we set the sample size as $N = 100$, the number of fixed matches as $M_0 = 1, 2, 3, 4, 5$, and the number of Monte Carlo replications as $2^{12}$.

Five cases with a single covariate are considered. In particular, we let $X \sim \text{Beta}(1.2, 1.2)$, where $\text{Beta}(\alpha, \beta)$ denotes the beta distribution with parameters $\alpha, \beta > 0$. For the mean function $m(x)$, we consider five non-linear curves presented in Table S1. We set $P(x) = 0.15 + 0.7x$ which assigns a unit to the treated group with higher probability if its corresponding covariate is of a greater value. The simulation results are in Figure S1.

Table S1: Simulation designs for $m(x)$.

| Cases | $m(x)$ |
|---|---|
| 1(a) | $0.1 + z/2 + \exp(-200(z-0.7)^2)/2$ |
| 1(b) | $0.8 - 2(z-0.9)^2 - 5(z-0.7)^3 - 10(z-0.6)^{10}$ |
| 1(c) | $0.2 + \sqrt{1-z} - 0.6(0.9-z)^2$ |
| 1(d) | $0.2 + \sqrt{1-z} - 0.6(0.9-z)^2 - 0.1z\cos(30z)$ |
| 1(e) | $0.4 + 0.25\sin(8z-5) + 0.4\exp(-16(4z-2.5)^2)$ |

**Remark S1.** In a preliminary simulation study, we consider different treated ratio and find that results are similar overall. For example, if we choose a smaller $c$, the treated ratio will be larger. But as long as we have more treated units than control units, our conclusion will not change with the choice of $c$. Figure S2 shows the results with the same simulation setting except that $c = 0.3$ for all cases. The same conclusions are obtained.

We also consider cases with multiple covariates $X_i^{\mathrm{T}} = (X_{i1}, \ldots, X_{i6})$. In particular, we set the number of covariates at 6 with one covariate being discretely distributed, i.e., $k = 5$. For the first three covariates, $(X_{i1}, X_{i2}, X_{i3})^{\mathrm{T}}$ is generated from a multivariate normal distribution with mean $\mu_x \in \mathbb{R}^3$, $\mathrm{var}(X_{i1}) = 2$, $\mathrm{var}(X_{i2}) = \mathrm{var}(X_{i3}) = \mathrm{cov}(X_{i1}, X_{i2}) = 1$, $\mathrm{cov}(X_{i1}, X_{i3}) = -1$, and $\mathrm{cov}(X_{i2}, X_{i3}) = -0.5$. For the remaining three covariates, we generate $X_{i4} \sim \mathrm{Unif}(-3, 3)$, $X_{i5} \sim \chi_1^2$, and $X_{i6} \sim \mathrm{Bern}(0.5)$, where $\chi_1^2$ denotes the chi-squared distribution with one degree of freedom. Suppose that the mean function admits the form $m(x) = \gamma^{\mathrm{T}} g(x)$, where $x = (x_1, x_2, \ldots, x_6)^{\mathrm{T}}$, $\gamma \in \mathbb{R}^h$ is a $h$-vector of coefficients, $g : \mathbb{R}^k \to \mathbb{R}^h$ is some link function, and $h \in \mathbb{N}$. The propensity score $P(x) = x^{\mathrm{T}}\eta$, where $\eta \in \mathbb{R}^k$ is a vector of coefficient. Five cases for $\mu_x$, $\eta$, $g$ and $\gamma$ are considered; see Table S2.

Since the dimension of covariate is $k > 1$, the bias term is non-negligible and a bias adjustment is needed. The bias can be corrected by the ordinary least-squares estimator via a linear regression model, which is a special case of the power series estimator. As it is implemented by Abadie et al. (2004), we use the weighted linear regression for correcting the bias of our proposed estimators, i.e., $\hat{\mu}(d, x) = \hat{\alpha}_d + x^{\mathrm{T}}\hat{\beta}_d$ for $d = 0, 1$, where $(\hat{\alpha}_d, \hat{\beta}_d) = \arg\min_{(\alpha_d, \beta_d)} \sum_{i:D_i=d} K(i)/M_0(Y_i - \alpha_d - X_i^{\mathrm{T}}\beta_d)^2$. The weight $K(i)/M_0 = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}(\ell)\}/M(\ell)$ is the number of times unit $i$ is used as a match for all units $\ell$, each time weighted by the total number of matches for unit $\ell$. Notice that here we do the matching by Mahalanobis distance which is the vector norm $\|x\|_A = (x'Ax)^{1/2}$, where the $k \times k$ matrix $A$ is chosen to be the inverse of the sample covariance matrix of the covariates, corresponding to the Mahalanobis metric:

$$A = \left\{ \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})(X_i - \bar{X})^{\mathrm{T}} \right\}^{-1},$$

where $\bar{X} = \sum_{i=1}^{N} X_i/N$. We also include in the comparison the genetic matching estimator (Diamond and Sekhon, 2013), which uses a state-of-art

Table S2: The values of parameters used in the simulation experiments.

| Cases | Parameters | Values |
|---|---|---|
| 2(a) | $\mu_x$ | $(0,0,0)^{\mathrm{T}}$ |
| | $\eta$ | $0.7 \times (0,1,1,1,-2,1,0.5)^{\mathrm{T}}$ |
| | $g(x)$ | $(1, x_{1:2}^{\mathrm{T}}, (x_3/2), \exp(-200(x_4-0.7)^2), x_{5:6}^{\mathrm{T}})^{\mathrm{T}}$ |
| | $\gamma$ | $(-3,2,1,1,2,2,2)^{\mathrm{T}}$ |
| 2(b) | $\mu_x$ | $(0,0,0)^{\mathrm{T}}$ |
| | $\eta$ | $0.7 \times (0,1,1,1,-2,1,0.5)^{\mathrm{T}}$, |
| | $g(x)$ | $(1, x_1, x_2^4, x_3, \sqrt{3-x_4}, x_{5:6}^{\mathrm{T}})^{\mathrm{T}}$, |
| | $\gamma$ | $(-3,2,1,1,2,-1,1)^{\mathrm{T}}$ |
| 2(c) | $\mu_x$ | $(-0.5,0,0)^{\mathrm{T}}$ |
| | $\eta$ | $0.7 \times (0,-1,1.5,1,-0.5,1,0.5)^{\mathrm{T}}$ |
| | $g(x)$ | $(1, x_{1:6}^{\mathrm{T}})^{\mathrm{T}}$ |
| | $\gamma$ | $(-3,2,1,1,2,2,2)^{\mathrm{T}}$ |
| 2(d) | $\mu_x$ | $(0,0,0)^{\mathrm{T}}$ |
| | $\eta$ | $0.7 \times (0,1,1,1,-2,1,0.5)^{\mathrm{T}}$, |
| | $g(x)$ | $(1, x_1, (0.2-x_2)^3, x_{3:6}^{\mathrm{T}}$, |
| | $\gamma$ | $(-3,2,1,1,2,2,2)^{\mathrm{T}}$ |
| 2(e) | $\mu_x$ | $(0,0,0)^{\mathrm{T}}$ |
| | $\eta$ | $0.7 \times (0,1,1,1,-2,1,0.5)^{\mathrm{T}}$, |
| | $g(x)$ | $(1, x_1, x_2^4, (1.5-x_3)^3, \sqrt{3-x_4}, (x_5-0.9)^2, x_6)^{\mathrm{T}}$, |
| | $\gamma$ | $(-3,2,1,1,2,2,2)^{\mathrm{T}}$ |

iterative algorithm to maximize a criterion related to covariate balance and perform the nearest neighbour matching using the scaled generalized Mahalanobis distance. However, the genetic matching has a few tuning parameters that slow the estimation for better results. For instance, a higher value of the tuning parameter called ‘pop.size’ of the genetic matching slower the estimation for better results. For fairness, we set ‘pop.size=20’ to make the computational time roughly the same for all estimators. If we allow the computational time to be longer for the genetic matching estimator, it is likely that it will perform better in terms of the covariate balance, but at the cost of efficiency.

Figure S3 is the performance comparison when $X$ is multi-dimensional. It is found that the mean squared errors of $\hat{\tau}^t_{\text{bc}}$ and $\hat{\tau}^t_{0,\text{bc}}$ are much lower compared with others. Hence, we extract the results for $\hat{\tau}^t_{\text{bc}}$ and $\hat{\tau}^t_{0,\text{bc}}$ in Figure S4 for better comparison. Interpreting the results from multi-dimensional cases is difficult due to the bias estimation. Using more data points allows a better fit but meanwhile introduces more bias if they are not good matches.

Intuitively, a complicated mean function requires more data for a better bias estimation, which is the case for Case 2(d) in Figure S4. Therefore, we may discover a decrease in bias with increasing $M_0$ without rematching. However, the rematching estimator performs better in terms of bias across $M_0$ since rematching can provide extra information for the underlying mean function. In fact, the rematching estimator uses more information with

Table S3: Coverage rates of the 95% and the 90% confidence intervals.

| | | $M_0 = 1$ | | $M_0 = 2$ | | $M_0 = 3$ | | $M_0 = 4$ | | $M_0 = 5$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CI | Cases | 95% | 90% | 95% | 90% | 95% | 90% | 95% | 90% | 95% | 90% |
| $\hat{C}^t$ | 1(a) | 0.939 | 0.885 | 0.938 | 0.889 | 0.935 | 0.890 | 0.936 | 0.884 | 0.938 | 0.885 |
| | 1(b) | 0.939 | 0.885 | 0.938 | 0.892 | 0.935 | 0.891 | 0.935 | 0.885 | 0.937 | 0.882 |
| | 1(c) | 0.939 | 0.885 | 0.939 | 0.888 | 0.936 | 0.891 | 0.937 | 0.885 | 0.940 | 0.882 |
| | 1(d) | 0.937 | 0.885 | 0.939 | 0.890 | 0.935 | 0.891 | 0.939 | 0.886 | 0.939 | 0.885 |
| | 1(e) | 0.940 | 0.883 | 0.938 | 0.889 | 0.937 | 0.889 | 0.936 | 0.887 | 0.937 | 0.883 |
| $\hat{C}^t_0$ | 1(a) | 0.940 | 0.888 | 0.942 | 0.891 | 0.938 | 0.884 | 0.939 | 0.886 | 0.937 | 0.883 |
| | 1(b) | 0.941 | 0.886 | 0.941 | 0.891 | 0.938 | 0.885 | 0.937 | 0.884 | 0.937 | 0.886 |
| | 1(c) | 0.941 | 0.888 | 0.941 | 0.890 | 0.938 | 0.888 | 0.939 | 0.884 | 0.939 | 0.882 |
| | 1(d) | 0.939 | 0.887 | 0.943 | 0.891 | 0.938 | 0.885 | 0.938 | 0.886 | 0.939 | 0.885 |
| | 1(e) | 0.941 | 0.887 | 0.941 | 0.891 | 0.939 | 0.884 | 0.938 | 0.883 | 0.937 | 0.883 |

more weights given to data matched with high frequency.

Case 2(e) in Figure S4 shows a decreasing trend in bias possibly because the mean function is the most complex and increasing $M_0$ introduces more bias which is hard to correct since the outcome model is not well-estimated even more data are used. Our rematching estimator performs better because it avoids repeatedly reusing matches that are too far away and hence does not introduce much bias when using more information.

We also evaluate the performance of the asymptotic variance estimator proposed. The implementation of the variance estimation follows a similar manner to the software documented in Abadie et al. (2004). We consider the simulation Cases 1(a)–(e). The $100(1-\alpha)\%$ confidence intervals for the rematching estimator $\hat{\tau}^t$ and the simple matching estimator $\hat{\tau}_0^t$ are computed as

$$\hat{C}^t = \left[\hat{\tau}^t - q_{1-\alpha/2}(\hat{V}^t)^{1/2}, \quad \hat{\tau}^t + q_{\alpha/2}(\hat{V}^t)^{1/2}\right],$$
$$\hat{C}_0^t = \left[\hat{\tau}_0^t - q_{1-\alpha/2}(\hat{V}_0^t)^{1/2}, \quad \hat{\tau}_0^t + q_{\alpha/2}(\hat{V}_0^t)^{1/2}\right],$$

respectively, where $q_{1-\alpha/2}$ and $q_{\alpha/2}$ are the $(1-\alpha/2)$th and $(\alpha/2)$th quantiles of the standard normal distribution, and $\hat{V}_0^t$ is defined in Section S5.3. For each estimator, we report the coverage rates of the 95% and the 90% confidence intervals in Table S3. It is found that the coverages of confidence intervals are very close to the nominal levels across different values of the number of fixed matches and different simulation cases, which means the matching-based variance estimator works well for both estimators.

## S4. National Supported Work data

In particular, we use nine variables: RE78 (earnings in 1978), treatment indicator (1 if treated, 0 if not treated), age, education, Black (1 if black, 0 otherwise), Hispanic (1 if Hispanic, 0 otherwise), married (1 if married, 0 otherwise), RE74 (earnings in 1974), RE75 (earnings in 1975), where RE78 is the outcome variable and the last seven variables are covariates. The dataset we use contains three groups including the experimental treated and control groups from a randomized evaluation of the NSW program, and a subset of the nonexperimental control group from the Panel Study of Income Dynamics (PSID). According to Dehejia and Wahba (1999), there is a lack of overlap in the pretreatment covariates of the PSID data and the experimental treated group. A subset of the nonexperimental control group named PSID-2 is therefore extracted by Lalonde from PSID to bridge the gap between the treated and the nonexperimental control groups.

We compute experimental matching estimates using the experimental treated and control groups and nonexperimental matching estimates using the experimental treated group and PSID-2. For each pair of the treated and control groups, we implement matching estimators based on the Mahalanobis distance and estimate their standard errors. The estimates include matching estimates without bias correction and bias-adjusted matching estimates, first by the simple matching and then by the matching-and-rematching procedure. For the bias correction, we estimate the regression function $\hat{\mu}(0, x)$ by the ordinary least squares of the linear regression on all covariates. The computation details for finding matches, doing regression adjustments and estimating standard errors, follow a similar manner to the software documented in Abadie et al. (2004). For the number of fixed matches $M_0$, we consider the cases $M_0 = 1, 4, 16, 64, N_0$, where $N_0$ is 260 and 253, respectively, for the experimental and the nonexperimental controls. The matching estimates reduce to difference-in-means estimates when $M_0 = N_0$.

## S5. Results from Abadie and Imbens's works

### S5.1 Consistency and asymptotic normality

Abadie and Imbens (2006) showed that $\hat{\tau}_0$ is consistent and asymptotically normal under Assumptions S1, S2, and S4, i.e.,

$$(V_0)^{-1/2} N^{1/2} (\hat{\tau}_0 - B_0 - \tau) \to \mathrm{N}(0, 1)$$

in distribution, where $B_0$ and $V_0$ are asymptotic bias and variance terms, respectively, defined as

$$\begin{aligned}
B_0 &= \frac{1}{N} \sum_{i=1}^{N} (2D_i - 1) \left[ \frac{1}{M_0} \sum_{j \in \mathcal{J}_{M_0}(i)} \{\mu(1 - D_i, X_i) - \mu(1 - D_i, X_j)\} \right], \\
V_0 &= V_0^R + V^{\tau(X)}, \\
V_0^R &= \frac{1}{N} \sum_{i=1}^{N} \left\{ 1 + \frac{K_{M_0}(i)}{M_0} \right\}^2 \sigma^2(D_i, X_i), \\
V^{\tau(X)} &= E\{\mu(1, X) - \mu(0, X) - \tau\}^2.
\end{aligned}$$

Under Assumptions S1, S3, and S4, similar results hold for its counterpart for the treated population as follows:

$$(V_0^t)^{-1/2} N_1^{1/2} (\hat{\tau}_0^t - B_0^t - \tau^t) \to \mathrm{N}(0,1)$$

in distribution, where $B_0^t$ and $V_0^t$ are asymptotic bias and variance terms, respectively, defined as

$$\begin{aligned}
B_0^t &= \frac{1}{N_1} \sum_{i=1}^{N} D_i \left[ \frac{1}{M_0} \sum_{j \in \mathcal{J}_{M_0}(i)} \{\mu(0, X_i) - \mu(0, X_j)\} \right], \\
V_0^t &= V_0^{R,t} + V^{\tau(X),t}, \\
V_0^{R,t} &= \frac{1}{N_1} \sum_{i=1}^{N} \left\{ D_i - (1 - D_i) \frac{K_{M_0}(i)}{M_0} \right\}^2 \sigma^2(D_i, X_i), \\
V^{\tau(X),t} &= E\left[ \{\mu(1, X) - \mu(0, X) - \tau^t\}^2 \mid D = 1 \right].
\end{aligned}$$

### S5.2 Bias correction

Abadie and Imbens (2011) proposed to estimate $B_0$ and $B_0^t$ by

$$\begin{aligned}
\hat{B}_0 &= \frac{1}{N} \sum_{i=1}^{N} \frac{(2D_i - 1)}{M_0} \sum_{j \in \mathcal{J}_{M_0}(i)} \{\hat{\mu}(1 - D_i, X_i) - \hat{\mu}(1 - D_i, X_j)\}, \\
\hat{B}_0^t &= \frac{1}{N_1} \sum_{i=1}^{N} \frac{D_i}{M_0} \sum_{j \in \mathcal{J}_{M_0}(i)} \{\hat{\mu}(0, X_i) - \hat{\mu}(0, X_j)\},
\end{aligned}$$

respectively, where $\hat{\mu}(d, x)$ is a non-parametric series regression estimator that satisfies Assumption S5. By Theorem 2 of Abadie and Imbens (2011),

$$N^{1/2}(B_0 - \hat{B}_0) \to 0$$

in probability, and

$$\{V_0^R + V^{\tau(X)}\}^{-1/2} N^{1/2}(\hat{\tau}_0 - \hat{B}_0 - \tau) \to \mathrm{N}(0,1)$$

in distribution, and by Theorem 2′ of Abadie and Imbens (2011),

$$N_1^{1/2}(B_0^t - \hat{B}_0^t) \to 0$$

in probability, and

$$\{V_0^{R,t} + V^{\tau(X),t}\}^{-1/2} N_1^{1/2}(\hat{\tau}_0^t - \hat{B}_0^t - \tau^t) \to \mathrm{N}(0,1)$$

in distribution, meaning that the bias adjustment does not affect the asymptotic variances.

### S5.3 Variance estimation

As suggested by Abadie and Imbens (2006), the asymptotic variances can be consistently estimated by

$$\begin{aligned}
\hat{V}_0 &= \frac{1}{N}\sum_{i=1}^{N}\{\hat{Y}_i(1) - \hat{Y}_i(0) - \hat{\tau}_0\}^2 \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\left[\left\{\frac{K_{M_0}(i)}{M_0}\right\}^2 + \left(\frac{2M_0 - 1}{M_0}\right)\left\{\frac{K_{M_0}(i)}{M_0}\right\}\right]\hat{\sigma}^2(D_i, X_i), \\
\hat{V}_0^t &= \frac{1}{N_1}\sum_{\substack{1\le i\le N \\ D_i = 1}}\{Y_i - \hat{Y}_i(0) - \hat{\tau}_0^t\}^2 \\
&\quad + \frac{1}{N_1}\sum_{i=1}^{N}(1 - D_i)\left[\frac{K_{M_0}(i)\{K_{M_0}(i) - 1\}}{M_0^2}\right]\hat{\sigma}^2(D_i, X_i),
\end{aligned}$$

where $\hat{\sigma}^2(D_i, X_i)$ is defined in the main article.

### S5.4 Martingale representation

Abadie and Imbens (2012) provide a martingale representation for matching estimators. This martingale representation holds no matter how matching is done. According to Abadie and Imbens (2012), this martingale representation is useful for analysing the asymptotic distribution of the estimator and correcting the standard error of a sample mean when missing data are imputed using the "hot deck". However, since they are beyond the scope of this article, we leave them for future study.

## References

Abadie, A., D. Drukker, J. L. Herr, and G. W. Imbens (2004). Implementing matching estimators for average treatment effects in stata. *The Stata Journal 4*(3), 290–311.

Abadie, A. and G. Imbens (2002). Simple and bias-corrected matching estimators for average treatment effects. manuscript.

Abadie, A. and G. W. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*(1), 235–267.

Abadie, A. and G. W. Imbens (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics 29*(1), 1–11.

Abadie, A. and G. W. Imbens (2012). A martingale representation for matching estimators. *Journal of the American Statistical Association 107*(498), 833–843.

Cohn, H. and Y. Zhao (2014). Sphere packing bounds via spherical codes. *Duke Math. J. 163*, 1965–2002.

Dehejia, R. H. and S. Wahba (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association 94*(448), 1053–1062.

Diamond, A. and J. S. Sekhon (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics 95*(3), 932–945.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Kozakova, I., R. Meester, and S. Nanda (2006). The size of components in continuum nearest-neighbor graphs. *The Annals of Probability 34*, 528–538.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Zong, C. (1998). The kissing numbers of convex bodies—a brief survey. *Bulletin of the London Mathematical Society 30*, 1–10.

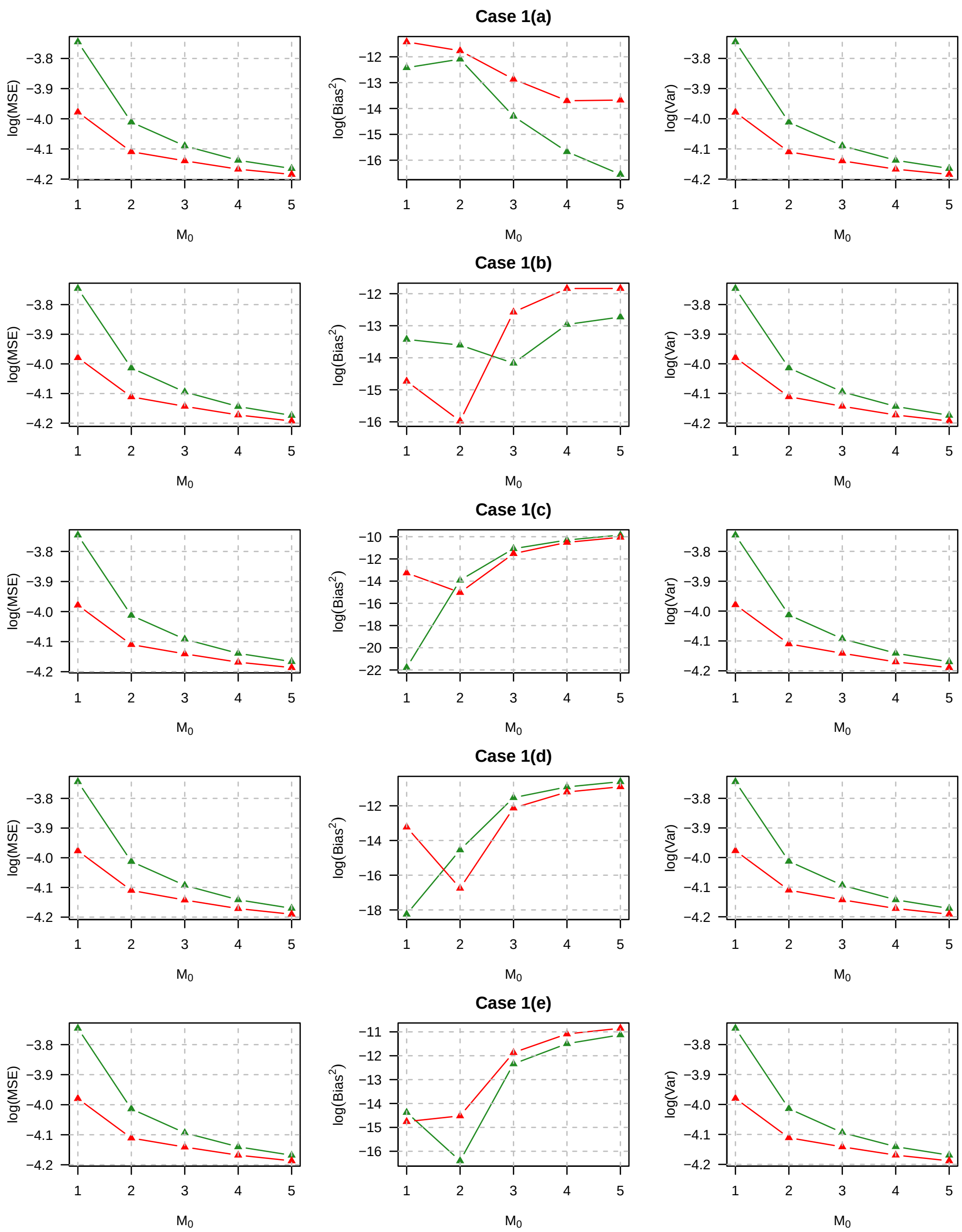


Figure S1: A graph comparing the performance between $\hat{\tau}_0^t$ (green triangle), $\hat{\tau}^t$ (red triangle) when $X$ is a scalar.

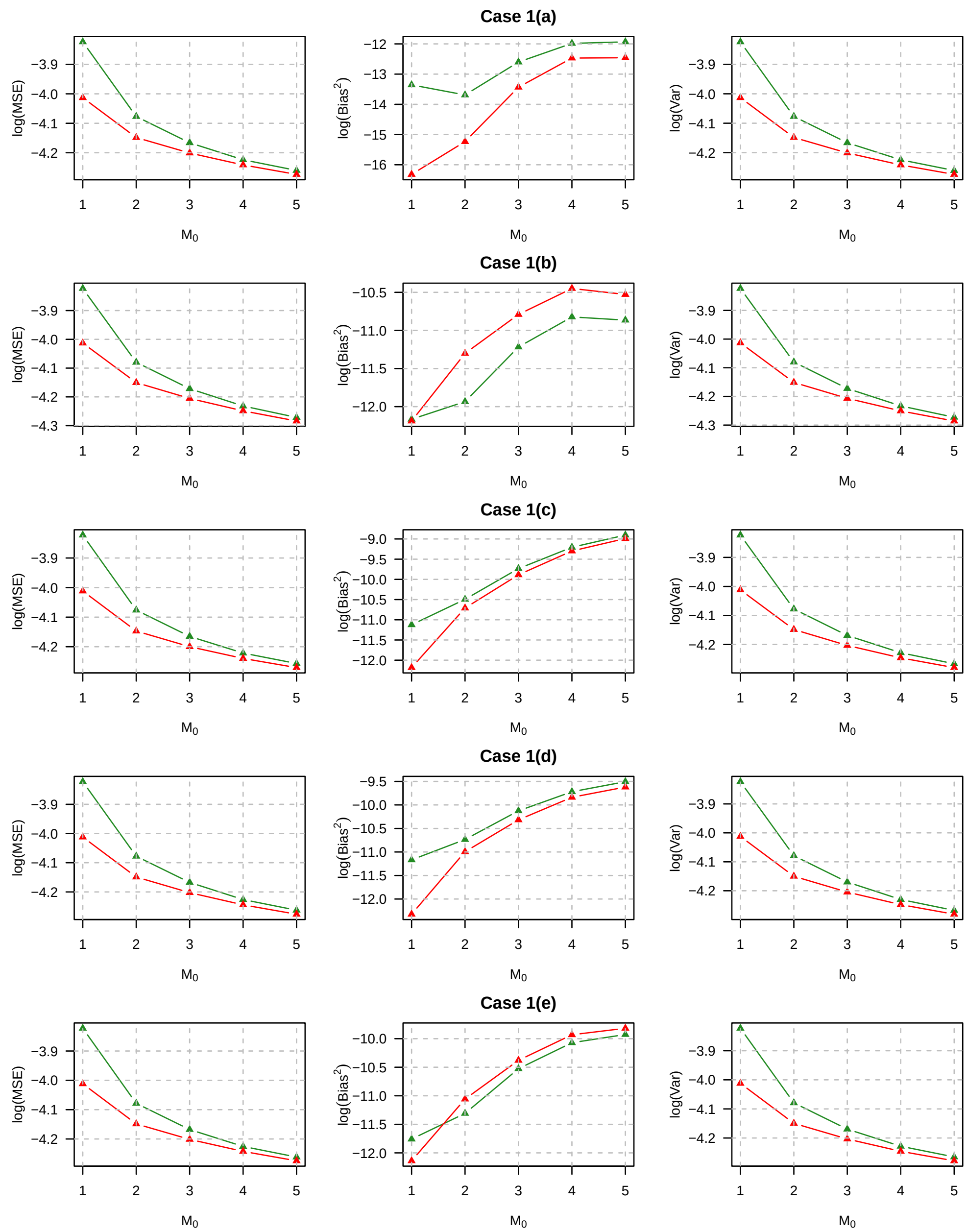


Figure S2: A graph comparing the performance between $\hat{\tau}_0^t$ (green triangle), $\hat{\tau}^t$ (red triangle) when $X$ is a scalar and $c = 0.3$.

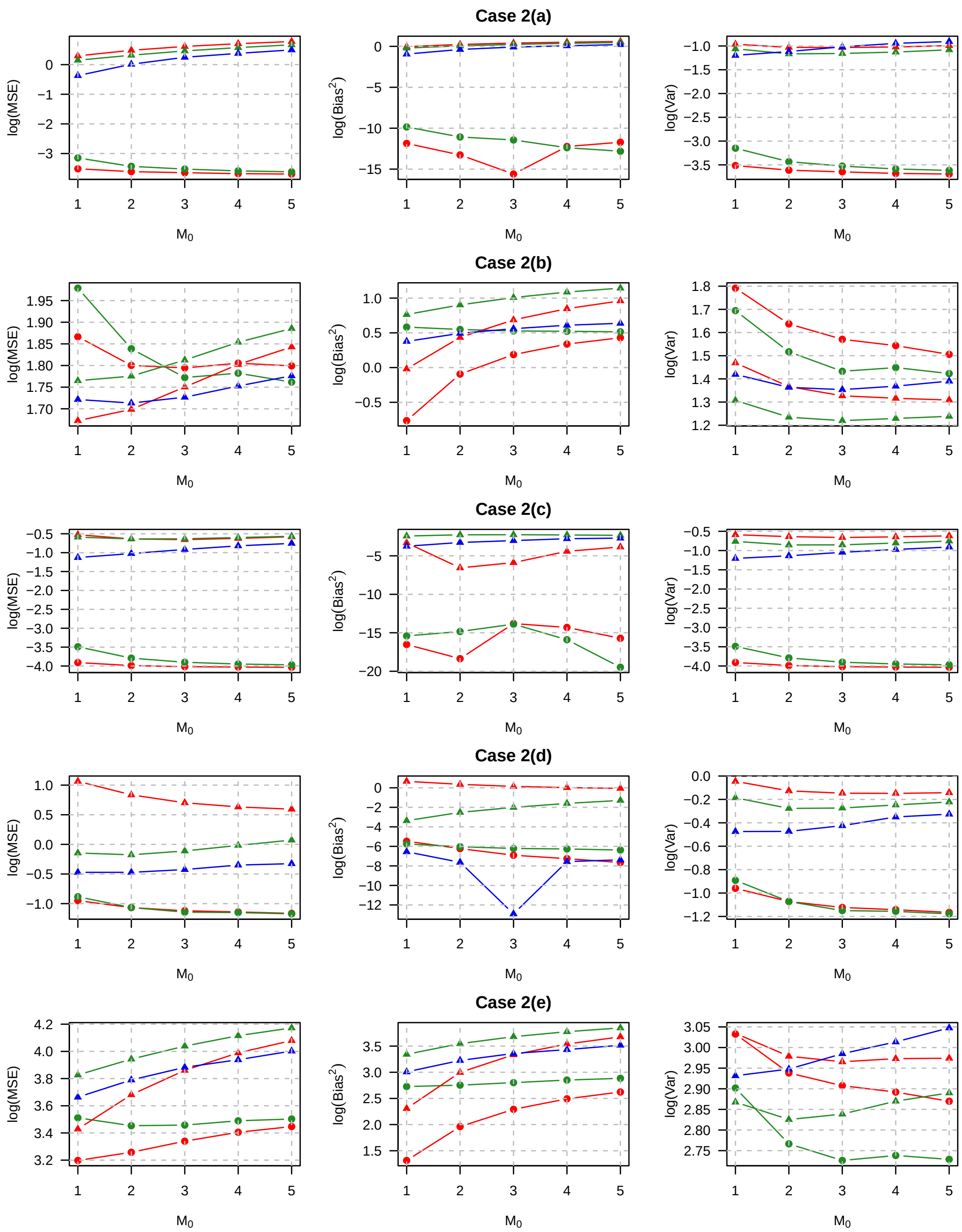


Figure S3: A graph comparing the performance among $\hat{\tau}_0^t$ (green triangle), $\hat{\tau}^t$ (red triangle), $\hat{\tau}_{0,\mathrm{bc}}^t$ (green dot), $\hat{\tau}_{\mathrm{bc}}^t$ (red dot), and the genetic matching estimator $\hat{\tau}_{gen}^t$ (blue triangle) when $X$ is multi-dimensional.

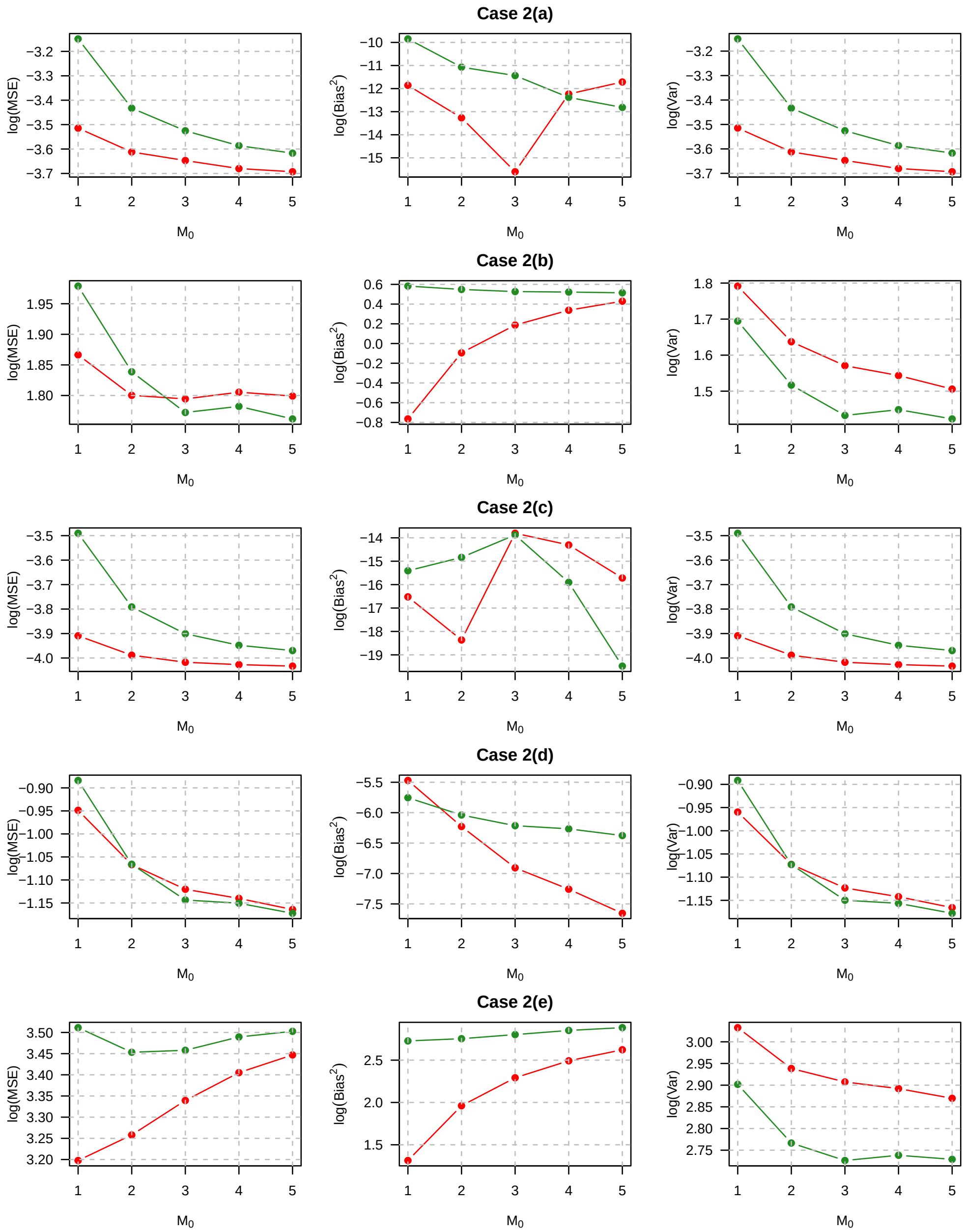


Figure S4: A graph comparing the performance between $\hat{\tau}^t_{0,\mathrm{bc}}$ (green dot) and $\hat{\tau}^t_{\mathrm{bc}}$ (red dot) when $X$ is multi-dimensional.