Statistica Sinica Preprint No: SS-2024-0306	
Title	Rematching Estimators For Average Treatment Effects
Manuscript ID	SS-2024-0306
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0306
Complete List of Authors	Lam Lam Hui and Kin Wai Chan
Corresponding Authors	Kin Wai Chan
E-mails	kinwaichan@cuhk.edu.hk

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

REMATCHING ESTIMATORS FOR AVERAGE TREATMENT EFFECTS

Lam Lam Hui and Kin Wai Chan

Department of Statistics and Data Science, The Chinese University of Hong Kong

Abstract: Matching estimators are widely applied in practice for their great intuitive appeal. However, simple matching estimators with a fixed number of matches (M_0) are generally inefficient. In this article, we propose matching estimators with a variable number of matches to gain efficiency via rematching. Rather than increasing M_0 to gain precision, which introduces an increase in bias, the key is to rematch the treated units from the opposite direction to utilize unmatched control units. Our rematching estimators are applicable to both the average treatment effect and its counterpart for the treated population. The proposed rematching estimators are proven asymptotically valid and uniformly more efficient than matching estimators with the same M_0. Simulation results confirm that the proposed rematching estimators substantially improve the simple matching estimators in finite samples. As an empirical illustration, we apply the estimators proposed in this article to the National Supported Work data.

Key words and phrases: Average treatment effects, causal inference, potential outcome.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

1. Introduction

Matching estimators are widely used to estimate the average effects of a program, medical treatment, or policy intervention by empirical researchers because of their great intuitive appeal. In contrast to other studies on propensity score matching estimators (Rosenbaum and Rubin, 1983), the term 'matching estimator' in this article is reserved for estimators that match each unit to a number of units in the opposite treatment group by multivariate distance matching. Under the setting of matching with replacement, Abadie and Imbens (2006, 2008, 2011, 2012) conducted various research on simple matching estimators with a fixed number of matches. While simple matching estimators are commonly used in practice, research has shown that some undesirable properties exist for simple matching estimators (Abadie and Imbens, 2006). In contrast to some regression adjustment estimators (e.g., Hahn (1998); Heckman et al. (1998); Imbens et al. (2005); Chen et al. (2008)) and weighting estimators (e.g., Horvitz and Thompson (1952); Robins and Rotnitzky (1995); Hirano et al. (2003); Abadie (2005)), matching estimators may not be fully efficient (Abadie and Imbens, 2006). According to a recent work by Lin et al. (2023), when matching is performed on a vector of covariates, allowing $M_0 \to \infty$ with the sample size leads to the semiparametric efficiency of bias-corrected match-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

ing estimators if the outcome model is appropriately specified. However, when the outcome model is misspecified, increasing the number of matched units through matching in the original direction may introduce matches of lower quality leading to a bias increase, which may render the causal interpretation of the estimates invalid. It motivates us to develop a new matching procedure that better utilizes observations and gains efficiency without increasing the bias too much.

The major contribution of this article is that we can achieve variance reduction by incorporating an additional rematching step with a small value of M_0, as opposed to using a much larger M_0 in a standard matching estimator. The use of a smaller M_0 ensures high-quality matching units, leading to a smaller bias. Based on a landmark paper by Abadie and Imbens (2006), which provided the first large sample analysis of simple matching estimators, we show the asymptotic validity of our rematching-based estimators when matched on a scalar covariate. By incorporating a bias correction method by Abadie and Imbens (2011), our method can be asymptotically valid even for matching based on a vector of covariates. As the simulation results show, our bias-corrected rematching estimators lead to a substantial reduction in variance compared with bias-corrected simple matching estimators when the bias is negligible after correction and have a lower

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

bias when the outcome model is misspecified. In summary, our proposed rematching estimators with a small M_0 are better alternatives to the matching estimators with a large M_0.

2. Background and setup

Matching estimators are often used to estimate treatment effects in observational studies where experimental data are unavailable. And the now dominant approach to analysing causal effects in observational studies was formulated by Rubin (1973a,b, 1974, 1977, 1978). For N observed units, indexed by $i = 1, \ldots, N$, the ith outcome variable is

$$Y_i = \begin{cases} Y_i(0) & \text{if } D_i = 0; \\ Y_i(1) & \text{if } D_i = 1, \end{cases}$$

where D_i indicates the treatment received ($D_i = 1$ if treated and $D_i = 0$ otherwise), and $Y_i(d)$ is the potential outcome under $D_i = d$. The sizes of the control and the treated groups are N_0 and N_1, respectively, with $N_0 + N_1 = N$. A vector of covariates X_i is also observed for the ith unit. Based on $\{(Y_i, D_i, X_i)\}_{i=1}^N$, we conduct the inference on the average treatment effect

$$\tau = E\{Y_i(1) - Y_i(0)\},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

and its counterpart for the treated population defined as

$$\tau^t = E\{Y_i(1) - Y_i(0) \mid D_i = 1\}.$$

In this article, we consider continuous covariates only and ignore the possibility of ties. Let $\mathbb{1}\{\cdot\}$ be the indicator function and $\|\cdot\|$ be the Euclidean norm. By simple matching estimators, we estimate the potential outcomes by

$$\hat{Y}_i(0) = \begin{cases} Y_i & \text{if } D_i = 0; \\ M_0^{-1} \sum_{j \in \mathcal{J}_{M_0}(i)} Y_j & \text{if } D_i = 1, \end{cases}$$

$$\hat{Y}_i(1) = \begin{cases} M_0^{-1} \sum_{j \in \mathcal{J}_{M_0}(i)} Y_j & \text{if } D_i = 0; \\ Y_i & \text{if } D_i = 1, \end{cases}$$

where

$$\mathcal{J}_{M_0}(i) = \{j_1(i), \ldots, j_{M_0}(i)\} \tag{2.1}$$

is the set of indices corresponding to the closest M_0 matches for the unit i in the opposite treatment group, i.e., $j_m(i)$ is the index $j \in \{1, \ldots, N\}$ that solves $D_j = 1 - D_i$ and

$$\sum_{\ell : D_\ell = 1 - D_i} \mathbb{1}\{\|X_\ell - X_i\| \leq \|X_j - X_i\|\} = m \quad (m = 1, \ldots, M_0).$$

Let $K_{M_0}(i)$ denotes the number of times unit i is matched, i.e.,

$$K_{M_0}(i) = \sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\}. \tag{2.2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

The average treatment effects are estimated by

$$\hat{\tau}_0 = \frac{1}{N}\sum_{i=1}^{N} \hat{Y}_i(1) - \hat{Y}_i(0) = \frac{1}{N}\sum_{i=1}^{N}(2D_i - 1)\left\{1 + \frac{K_{M_0}(i)}{M_0}\right\} Y_i, \tag{2.3}$$

$$\hat{\tau}_0^t = \frac{1}{N_1}\sum_{\substack{1\le i\le N \\ D_i=1}} Y_i - \hat{Y}_i(0) = \frac{1}{N_1}\sum_{i=1}^{N}\left\{D_i - (1-D_i)\frac{K_{M_0}(i)}{M_0}\right\} Y_i.$$

To remove the bias of matching, Abadie and Imbens (2011) combined matching with a bias correction proposed in Rubin (1973b) and Quade (1982), which produces bias-corrected matching estimators $\hat{\tau}_{0,\mathrm{bc}}$ and $\hat{\tau}^t_{0,\mathrm{bc}}$.

Despite being commonly used by practitioners, the matching estimators $\hat{\tau}_0$ and $\hat{\tau}_0^t$ are not free of problems. It is found that when there is a larger reservoir of potential controls than treated and M_0 is not large enough, the matching estimators are inefficient since some control units will likely be discarded, leading to an efficiency loss. However, increasing M_0 for precision may repeatedly include far away units as low-quality matches, accompanied by a significant bias increase if the bias is not well-corrected; see Figure 1. It thus motivates us to revisit the matching procedure and propose a new generation of matching estimators to gain efficiency without increasing the bias too much through better utilization of unmatched control units.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

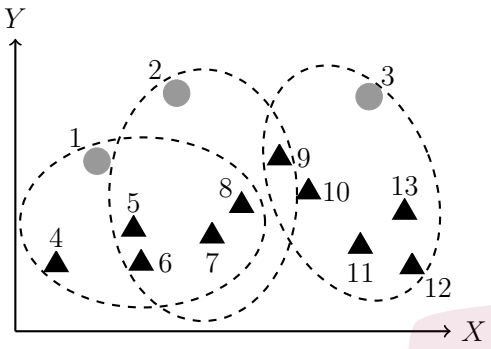

Figure 1: An example of the simple matching procedure that finds matches from the controls (triangles) for each treated unit (circles) with the horizontal axis and the vertical axis representing the scalar covariate X and the outcome Y, respectively. The markers represent the values of (X_i, Y_i), $i = 1, \ldots, 13$. We set $M_0 = 5$ so that each control is matched at least once. When τ^t is the estimand of interest, the matching results are represented by dotted loops.

3. Rematching estimators

3.1 A matching-and-rematching procedure

We propose rematching estimators to avoid efficiency loss due to points discarding when there is a large reservoir of potential controls. The efficiency gain is obtained without increasing the bias too much by rematching the treated units from an opposite direction. The discussion in this section is

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

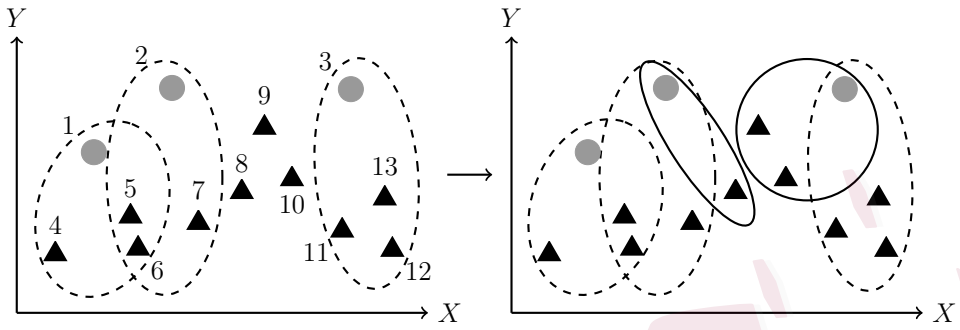

Figure 2: An example of the matching-and-rematching procedure that finds matches from the controls (triangles) for each treated unit (circles) with the horizontal axis and the vertical axis representing the scalar covariate X and the outcome Y, respectively. The markers represent the values of (X_i, Y_i), $i = 1, \ldots, 13$. When $M_0 = 3$ and τ^t is the estimand, the results of the first and the second matching are represented by dotted and solid loops, respectively.

based on τ^t. Because often, matching estimators have been used when (1) the interest is in τ^t, and (2) a large reservoir of potential controls is available (Imbens and Wooldridge, 2009).

Figure 2 is a visualization of the whole matching process that produces our proposed estimators. After the first matching, which is the simple matching, a dotted loop will cover both a control unit i and a treated unit ℓ if the unit i is matched to unit ℓ in the matching, e.g., the 1st treated unit

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

is matched to the 4th, 5th, and 6th controls if $M_0 = 3$. However, no dotted loops will cover unit i if it remains unmatched after the first matching, e.g., the 8th, 9th, and 10th units. When there is a large reservoir of potential controls, some controls will likely be discarded without being used, leading to an efficiency loss. To gain efficiency, we propose to rematch the treated units after the first matching by matching those unmatched controls to them, which means we do the rematching from an opposite direction. Instead of performing M_0-nearest neighbour search for each fixed treated unit as in the first matching process, we find only the nearest neighbour for each fixed unmatched control. Because though matching for unmatched points can provide efficiency gain, those unmatched points are generally not matches of good quality and should not be overused. The final matching results after rematching are in the right panel of Figure 2. As an example of the rematching results represented by solid loops, the 9th and 10th units are rematched to the 3rd unit which is the closest to the 9th and the 10th units.

The rematching step provides a channel to extract additional information from the observations that are originally unused. Intuitively, it has two appealing features. First, the rematching step utilizes more observations that may lead to a reduction in variance. Second, the rematching step

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

tends to match higher-quality observations in the opposite direction than lower-quality matches in the original matching direction. Consequently, by construction, the resulting estimator has the potential to strike a good balance between variance and bias.

Regarding the choice of M_0, "little is known about the optimal number of matches, or about data-dependent ways of choosing it" (Imbens and Wooldridge, 2009). It is suggested that people fix M_0 at one when using rematching estimators on finite samples as empirical researchers typically do when using matching estimators. In the asymptotic context, increasing M_0 with the sample size, which shares a similar idea to that of Lin et al. (2023) is advisable for rematching estimators. However, we do not suggest a diverging M_0 considering the large bias when the outcome model is not well-specified. More attention is attached to the particular choice of $M_0 = 1$ in this article as our simulated data and real data are of small sample sizes.

3.2 Main proposal

Let $\emptyset$ be an empty set. Every unmatched unit is matched to its nearest neighbour with the index

$$\mathcal{J}_{\text{re}}(i) = \begin{cases} \mathcal{J}_1(i) & \text{if } K_{M_0}(i) = 0; \\ \emptyset & \text{if } K_{M_0}(i) \neq 0. \end{cases} \tag{3.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Let $M_{\text{re}}(\ell)$ be the number of times unit ℓ is matched to unmatched units, i.e.,

$$M_{\text{re}}(\ell) = \sum_{i=1}^{N} \mathbb{1}\{\ell \in \mathcal{J}_{\text{re}}(i)\}. \tag{3.2}$$

The rematching process increases the number of matched samples in the control group by matching each treated unit ℓ to a set of controls of a variable size

$$M(\ell) = M_0 + M_{\text{re}}(\ell),$$

where M_0 is the fixed number of nearest neighbours found in the first matching and $M_{\text{re}}(\ell)$ is the variable number of unmatched units matched to unit ℓ in the rematching. Then, an equal matching weight of $1/M(\ell)$ will be given to all control units that are matched to unit ℓ, which deviates the matching weight of those $M(\ell)$ controls from $1/M_0$ or 0 to a non-zero variable value; see Remark 1. We now define the weighted version of the total number of times the unit i is matched to treated units as

$$K(i) = \left[\sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\} \frac{M_0}{M_{\text{re}}(\ell) + M_0}\right] + \mathbb{1}\{K_{M_0}(i) = 0\} \frac{M_0}{M_{\text{re}}(\mathcal{J}_{\text{re}}(i)) + M_0} \tag{3.3}$$

so that $K(i)/M_0$ is the total matching weight assigned to unit i. If there is $\ell \in [1, N]$ such that $\mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\} = 1$, or equivalently, $K_{M_0}(i) \neq 0$, then unit i is matched in the first matching, and after rematching, $K(i)/M_0 =$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

$\sum_{\ell=1}^{N} \mathbb{1}\{i \in \mathcal{J}_{M_0}(\ell)\}/\{M_{\text{re}}(\ell) + M_0\}$. Otherwise, the originally unmatched unit i will have a matching weight of $1/\{M_{\text{re}}(\mathcal{J}_{\text{re}}(i))+M_0\}$ as it is rematched to unit $\mathcal{J}_{\text{re}}(i)$. Define the set of indices of unmatched units that unit i is rematched to as

$$\mathcal{L}_{\text{re}}(i) = \{\ell \in \{1, \ldots, N\} : i \in \mathcal{J}_{\text{re}}(\ell)\}.$$

The set of indices of units that unit i is matched to is then given by

$$\mathcal{J}(i) = \mathcal{J}_{M_0}(i) \cup \mathcal{L}_{\text{re}}(i).$$

The following proposition gives an equivalent form of $K(i)$, which is concise and accessible.

Proposition 1. *The term $K(i)$ $(i = 1, \ldots, N)$ in (3.3) can be equivalently computed as*

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

Under the proposed matching-and-rematching scheme, the potential outcomes are estimated as

$$\tilde{Y}_i(0) = \begin{cases} Y_i & \text{if } D_i = 0; \\ M(i)^{-1} \sum_{j \in \mathcal{J}(i)} Y_j & \text{if } D_i = 1, \end{cases}$$

$$\tilde{Y}_i(1) = \begin{cases} M(i)^{-1} \sum_{j \in \mathcal{J}(i)} Y_j & \text{if } D_i = 0; \\ Y_i & \text{if } D_i = 1. \end{cases}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Remark 1. Besides giving an equal weight of $1/M(\ell)$ to controls matched to unit ℓ, there are some possible methods to treat the M_0 matches and the $M_{\text{re}}(\ell)$ matches differently, which prevents the rematching process from incorporating large biases produced by matches of poor quality. For instance, we may use a kernel function on the $M_{\text{re}}(\ell)$ matches (or all matches) to give larger weights to controls with smaller distances. Since it is beyond the scope of this article, we leave this technical modification for future study.

Our proposed rematching estimators for average treatment effects τ and τ^t are given by

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N} \tilde{Y}_i(1) - \tilde{Y}_i(0), \quad \hat{\tau}^t = \frac{1}{N_1} \sum_{\substack{1\le i\le N \\ D_i=1}} Y_i - \tilde{Y}_i(0),$$

respectively. Both of which average within-match differences in the potential outcome between the treated and the control units. Proposition 2 below provides equivalent forms of $\hat{\tau}$ and $\hat{\tau}^t$.

Proposition 2. *The rematching estimators can be equivalently written as*

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N} (2D_i - 1)\left\{1 + \frac{K(i)}{M_0}\right\} Y_i, \tag{3.4}$$

$$\hat{\tau}^t = \frac{1}{N_1}\sum_{i=1}^{N} \left\{D_i - (1-D_i)\frac{K(i)}{M_0}\right\} Y_i. \tag{3.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Denote $\mu(d,x) = E(Y \mid D = d, X = x)$. When the bias correction is needed, we follow the study of Abadie and Imbens (2011) and estimate the bias as follows:

$$\hat{B} = \frac{1}{N}\sum_{i=1}^{N}\frac{(2D_i-1)}{M(i)}\sum_{j\in\mathcal{J}_M(i)}\{\hat{\mu}(1-D_i,X_i)-\hat{\mu}(1-D_i,X_j)\},$$
$$\hat{B}^t = \frac{1}{N_1}\sum_{i=1}^{N}\frac{D_i}{M(i)}\sum_{j\in\mathcal{J}_M(i)}\{\hat{\mu}(0,X_i)-\hat{\mu}(0,X_j)\},$$

where $\hat{\mu}(d,x)$ $(d = 0,1)$ is a non-parametric series regression estimator estimating $\mu(d,x)$. Then the bias-corrected rematching estimators are

$$\hat{\tau}_{\text{bc}} = \hat{\tau} - \hat{B}, \quad \hat{\tau}^t_{\text{bc}} = \hat{\tau}^t - \hat{B}^t.$$

Algorithm 1 summarizes the procedure that produces our proposed rematching estimators. An extra bias correction step may be needed depending on the data.

4. Asymptotic properties

4.1 Decomposition and bias analysis

Let (Y_i, X_i, D_i), $i = 1,\ldots,N$, be independent and identical copies of (Y, X, D). Denote $\sigma^2(d,x) = \text{var}(Y \mid D = d, X = x)$. Also denote the residual as $\epsilon_i = Y_i - \mu(D_i, X_i)$ for each i. We first decompose $\hat{\tau} - \tau$ as

$$\hat{\tau} - \tau = \{\bar{\tau}(X) - \tau\} + R + B, \tag{4.1}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Algorithm 1: Matching-and-Rematching Algorithm

Data: $Z = \{(Y_i, D_i, X_i)\}_{i=1}^N$ and M_0.

Result: $\hat{\tau}^t$ and $\hat{\tau}$.

1 Step 1 (Matching step): For $i \in \{1, \ldots, N\}$, perform M_0-nearest neighbour search to find $\mathcal{J}_{M_0}(i)$ defined in (2.1).

2 Step 2 (Weighting step): For $i \in \{1, \ldots, N\}$, compute the number of times unit i is matched, i.e., $K_{M_0}(i)$ defined in (2.2).

3 Step 3 (Rematching step): For $i \in \{1, \ldots, N\}$, find the index of the nearest neighbour for potentially unmatched unit i as in (3.1).

4 Step 4 (Re-weighting step): For $\ell \in \{1, \ldots, N\}$, obtain the number of times unit ℓ is matched to unmatched units as in (3.2). For $i \in \{1, \ldots, N\}$, obtain $K(i)$ defined in (3.3).

5 Step 5: Estimate τ and τ^t by rematching estimators defined in (3.4) and (3.5).

where

$$\bar{\tau}(X) = \frac{1}{N}\sum_{i=1}^{N}\{\mu(1, X_i) - \mu(0, X_i)\},$$

$$R = \frac{1}{N}\sum_{i=1}^{N}(2D_i - 1)\left\{1 + \frac{K(i)}{M_0}\right\}\epsilon_i,$$

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

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Here, $\bar{\tau}(X)$ is the average conditional treatment effect, R is a weighted average of the residuals, and B is the conditional bias relative to $\tau(X)$. The decomposition (4.1) is also employed by Abadie and Imbens (2006). Similarly, we can decompose $\hat{\tau}^t - \tau^t$ as

$$\hat{\tau}^t - \tau^t = \left\{\bar{\tau}^t(X) - \tau^t\right\} + R^t + B^t,$$

where

$$\begin{aligned}
\bar{\tau}^t(X) &= \frac{1}{N_1}\sum_{i=1}^{N} D_i\{\mu(1,X_i) - \mu(0,X_i)\}, \\
R^t &= \frac{1}{N_1}\sum_{i=1}^{N}\left\{D_i - (1-D_i)\frac{K(i)}{M_0}\right\}\epsilon_i, \\
B^t &= \frac{1}{N_1}\sum_{i=1}^{N} D_i\left[\frac{1}{M(i)}\sum_{j\in\mathcal{J}(i)}\{\mu(0,X_i) - \mu\left(0,X_j\right)\}\right].
\end{aligned}$$

All asymptotic results are stated as $N \to \infty$ unless otherwise stated. Regularity conditions for our asymptotic theory are included in the supplement; see Assumptions S1–S6.

Let $\mathcal{X} \subset \mathbb{R}^k$ be the support of X and $\mathcal{X}_0 \subset \mathbb{R}^k$ be the support of X given $D = 0$. We establish bounds on the stochastic order of the conditional bias terms as follows.

Theorem 1 (Conditional Bias)**.** *Suppose Assumption S1 holds.*

1. *If Assumption S2 holds and $x \mapsto \mu(d,x)$ $(d = 0,1)$ is Lipschitz continuous on $\mathcal{X}$, then $B = O_p(N^{-1/k})$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

2. *If Assumption S3 holds and* $x \mapsto \mu(0, x)$ *is Lipschitz continuous on* $\mathcal{X}_0$, *then* $B^t = O_p(N_0^{-1/k})$.

The theorem above shows that the bias of our proposed rematching estimators is of the same order as that of simple matching estimators, which makes the causal interpretation of our proposals as credible as that of simple matching estimators. Moreover, Theorem 1 has a great impact on the asymptotic normality of rematching estimators.

From part 1 of the theorem, $B = O_p(N^{-1})$ when the dimension of the continuous covariate is $k = 1$. As we shall see in Section 4.3, $N^{1/2}\{\bar{\tau}(X) - \tau\} = N^{1/2}R = O_p(1)$ is asymptotically normal under regularity conditions. Consequently, the asymptotic normality of $N^{1/2}(\hat{\tau} - \tau)$ is achieved, hence, the rematching estimator is $N^{1/2}$-consistent. When there are more than one continuously distributed covariate, i.e., $k > 1$, bias correction is needed in order to attain the $N^{1/2}$-consistency of the rematching estimator.

Part 2 of this theorem is particularly useful since matching estimators nearly always estimate the average effect for the treated when a large reservoir of controls is available. Generally, the bias is ignorable when there is only one continuous covariate or the number of controls is sufficiently large. To be precise, if the two group sizes go to infinity at different rates such that $N_1 = O_p(N_0^s)$, then $B^t = O_p\{N_1^{1/(sk)}\}$. Therefore, if $s < 2/k$, then

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

$B^t = o_p(N_1^{-1/2})$, which will be dominated in the large sample distribution of $\hat{\tau} - \tau$ by $\bar{\tau}^t(X) - \tau$ and R^t, which are of size $O_p(N_1^{-1/2})$.

4.2 Variance improvement

We first investigate the conditional variance of $\hat{\tau}$. Conditional on $X_{1:N} = (X_1, \ldots, X_N)^{\mathrm{T}}$ and $D_{1:N} = (D_1, \ldots, D_N)^{\mathrm{T}}$, the weighted version of the number of times a unit is used as a match, $K(i)$, is deterministic. Hence, according to (3.4), the conditional variance of $\hat{\tau}$ is

$$\mathrm{var}(\hat{\tau} \mid D_{1:N}, X_{1:N}) = \frac{1}{N^2} \sum_{i=1}^{N} \left\{ 1 + \frac{K(i)}{M_0} \right\}^2 \sigma^2(D_i, X_i).$$

Similarly, according to (3.5), its counterpart for the average effect of the treatment on the treated is

$$\mathrm{var}(\hat{\tau}^t \mid D_{1:N}, X_{1:N}) = \frac{1}{N_1^2} \sum_{i=1}^{N} \left\{ D_i - (1 - D_i) \frac{K(i)}{M_0} \right\}^2 \sigma^2(D_i, X_i).$$

Let $V^R = N\mathrm{var}(\hat{\tau} \mid D_{1:N}, X_{1:N})$ and $V^{R,t} = N_1\mathrm{var}(\hat{\tau}^t \mid D_{1:N}, X_{1:N})$. The lemma below shows the finiteness of the expectation of these conditional variances.

Lemma 1 (Conditional variances)**.** *Suppose Assumption S1 holds.*

1. *Suppose Assumption S2 holds. For $i \in \{1, \ldots, N\}$, $K(i) = O_p(1)$, and $E\{K_M^q(i)\}$ is bounded uniformly in N for any $q > 0$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

2. *Suppose Assumption S3 holds. For* $i \in \{1, \ldots, N\}$, $(N_0/N_1)E\{K_M^q(i) \mid D_i = 0\}$ *is bounded uniformly in* N *for any* $q > 0$.

3. *Suppose* $x \mapsto \sigma^2(d, x)$ *is Lipschitz continuous in* $\mathcal{X}$, *for* $d = 0, 1$. *If Assumption S2 holds, then* $E(V^R) = O(1)$. *If Assumption S3 holds, then* $E(V^{R,t}) = O(1)$.

According to Theorem 1, $B = O_p(1/N)$ and $B^t = O_p(1/N_0)$ when $k = 1$. And $B^t = o_p(N_1^{-1/2})$ when the number of controls is sufficiently large compared to the number of treated. In these cases, the bias terms B and B^t are dominated in the large sample distributions of $\hat{\tau} - \tau$ and $\hat{\tau}^t - \tau^t$, respectively, and given the value of N_1, the unconditional variances of $\hat{\tau}$ and $\hat{\tau}^t$ are

$$\text{var}(\hat{\tau}) \sim \frac{E(V^R) + V^{\tau(X)}}{N}, \quad \text{var}(\hat{\tau}^t) \sim \frac{E(V^{R,t}) + V^{\tau(X),t}}{N_1},$$

respectively, where $a_N \sim b_N$ means $a_N/b_N \to 1$ as $N \to \infty$; and

$$V^{\tau(X)} = E\left\{\mu(1, X) - \mu(0, X) - \tau\right\}^2,$$
$$V^{\tau(X),t} = E\left[\left\{\mu(1, X) - \mu(0, X) - \tau^t\right\}^2 \mid D = 1\right].$$

According to Abadie and Imbens (2006), investigating the asymptotic variance of matching estimators when $k \geq 2$ is challenging. We leave the general case for future study. The theorem below shows the asymptotic efficiency gain for the special case with a scalar covariate, i.e., $k = 1$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Theorem 2 (Asymptotic efficiency gain)**.** *Suppose $k = 1$. If Assumptions S1, S2 and S4 hold, and $f_0(x)$ and $f_1(x)$ are continuous on $\mathcal{X}$, then*

$$\begin{aligned} N\text{var}(\hat{\tau}) &\leq N\text{var}(\hat{\tau}_0) \\ &= \left(1 + \frac{1}{2M}\right) E\left[\frac{\sigma_1^2(X_i)}{e(X_i)} + \frac{\sigma_0^2(X_i)}{1 - e(X_i)}\right] + V^{\tau(X)} \\ &\quad - \frac{1}{2M} E[e(X_i)\sigma_1^2(X_i) + (1 - e(X_i))\sigma_0^2(X_i)] + o(1), \end{aligned}$$

where $e(x) = \text{pr}(D = 1|X = x)$ is the propensity score.

The strict inequality holds when $K_{M_0}(i) = 0$ for some i, i.e., when rematching can be done. The exact form of $N\text{var}(\hat{\tau}_0)$ in the theorem above is given by Abadie and Imbens (2006). The asymptotic unconditional variance of $\hat{\tau}$ is not available due to the complicated form of $K(i)$; see Remark 2.

Remark 2. Unlike $K_{M_0}(i)$, which has a simple form and follows a binomial distribution, our $K(i)$ has a more complicated form, making it difficult to derive the unconditional variance of the estimator. Specifically, we need to consider the distribution and the first two moments of $X/(Y + M_0)$, where X and Y are two dependent binomial random variables. It is difficult to find moments of the ratio of two binomial distributions, not to mention rewriting the moments in terms of propensity scores like what Abadie and Imbens (2006) and Hahn (1998) did. Thus, it is hard to derive explicitly the exact efficiency loss relative to the semi-parametric efficiency bound in

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Hahn (1998) or the exact efficiency gain relative to the standard matching estimator in Abadie and Imbens (2006).

To derive a representation of the unconditional variance, we need the following notation. Let $\mathrm{Bin}(n,p)$ denote the binomial distribution with $n \in \mathbb{N}$ trails and success probability $p \in [0,1]$. Let $\mathrm{Bern}(p)$ denote the Bernoulli distribution with success probability p. Now, we let $p = \mathrm{pr}(D = 1)$ denoting the treated ratio and let $f_d(x)$ denote the density function of X under $D = d$. The theorem below gives the form of unconditional variances.

Theorem 3 (Unconditional variances)**.** *Suppose that $k = 1$, $f_0(x) = f_1(x)$, and $M_0 N_1 < N_0$. Further suppose that $x \mapsto \sigma^2(d,x) =: \sigma_d^2$ is a constant as a function of x for each $d \in \{0,1\}$. Define the random variables K, I, and H as follows given the values of N_0 and N_1:*

$$K \sim \mathrm{Bin}(M_0 N_0, 1/N_1), \quad I \sim \mathrm{Bern}(M_0 N_1 / N_0),$$
$$[H \mid I] \sim \mathrm{Bin}(N_0 - M_0 N_1 + I - 1, 1/N_1).$$

1. *If Assumptions S2 and S4 hold, then*

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

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

2. If Assumptions S3 and S4 hold, then

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

Note that K and $[K_{M_0}(i) \mid D_i = 1]$ have identical distributions, and so do I and $[K_{M_0}(i) \mid D_i = 0]$. Also, $[H \mid I]$, $M_{\text{re}}(\ell)$, and $M_{\text{re}}(\mathcal{J}_{re}(i)) - 1$ follow the same distribution. To study the variance improvement of our proposed estimators over simple matching estimators under this assumption, we need the condition that $M_0 N_1 < N_0$ to simulate scenarios with unmatched controls. When there are no unmatched controls after the simple matching, the rematching estimator $\hat{\tau}^t$ and the simple matching estimator $\hat{\tau}_0^t$ produce the same estimate. According to the proposition below, our proposal has a uniform variance improvement over the simple matching estimator.

Proposition 3. *Under the conditions in Theorem 3, the difference between* $N_1\text{var}(\hat{\tau}_0^t)$ *and* $N_1\text{var}(\hat{\tau}^t)$ *satisfies that*

$$N_1\left\{\text{var}(\hat{\tau}_0^t) - \text{var}(\hat{\tau}^t)\right\}$$
$$\to \sigma_0^2(1-p)E\left\{\left(\frac{I}{M_0}\right)^2 - \left(\frac{I}{H+M_0} + \frac{1-I}{H+1+M_0}\right)^2\right\} > 0.$$

Based on Theorem 3, we compare rematching estimators with simple matching estimators by simulation experiments. Since matching estimators

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

are used nearly always when the quantity of interest is the average effect on the treated and there are more controls than treated, we only include the results for the estimators of τ^t and consider a design with the treated ratio $p = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$, and $M_0 = 1, 2, 3, 4, 5$. For each case, we set the sample size as $N = 100$ and the conditional error variance as $\sigma^2(d, x) = 1$. Assume we have prior knowledge that the average treatment effect conditional on the covariates is a constant, i.e., $\mu(1, x) - \mu(0, x) = \tau^t$ for all x, and hence, $V^{\tau(x),t} = 0$. We use the percentage decrease in variance, i.e., $\{\text{var}(\hat{\tau}_0^t) - \text{var}(\hat{\tau}^t)\}/\text{var}(\hat{\tau}_0^t)$, as a measure of the ability of our rematching process to gain efficiency.

Figure 3 visualizes the reduction in variance across different treated ratios and different numbers of fixed matches. Since Theorem 3 assumes there are unmatched units after the simple matching, i.e., $M_0 N_1 < N_0$, the maximum possible M_0 depends on the treated ratio $p = N_1/(N_0 + N_1)$. In any case, the proposed estimator $\hat{\tau}^t$ is always more precise than the traditional $\hat{\tau}_0^t$. Fixing M_0 at a particular value, the smaller the treated ratio, the greater the percentage reduction in variance. Given a specific p, the smaller the number of fixed matches, the greater the variance reduction percentage. The results with $M_0 = 1$ are worth noticing as people typically fix M_0 at one when using matching estimators. In particular, when $M_0 = 1$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

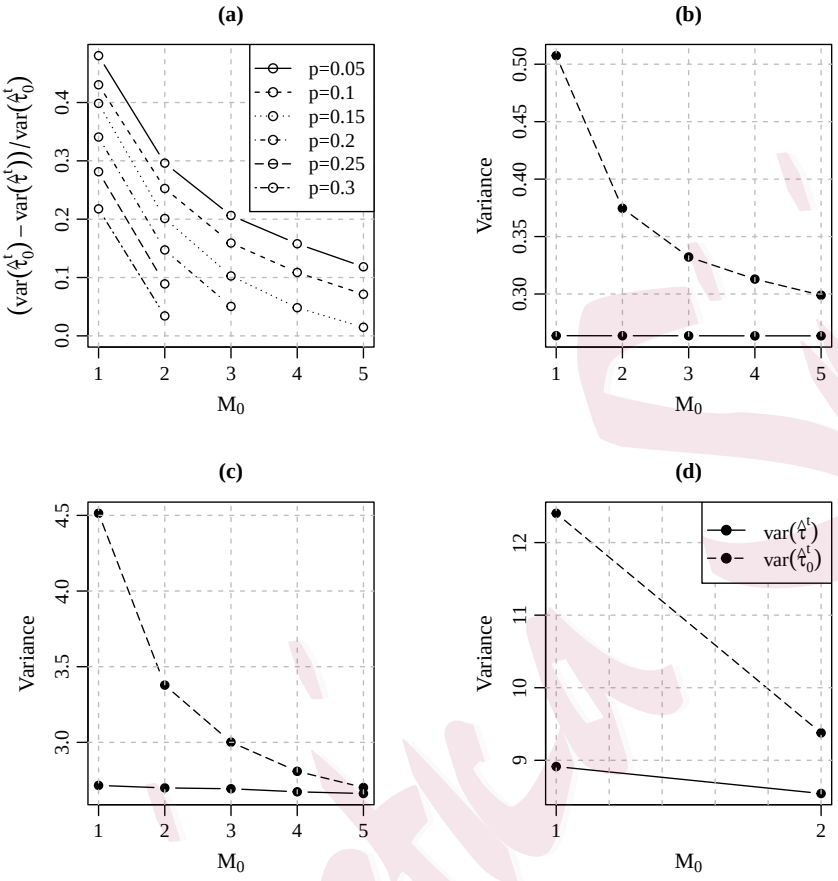

Figure 3: Plots of simulation results of variances of $\hat{\tau}_0^t$ and $\hat{\tau}^t$ across different M_0 in various settings: (a) Plot of percentage variance reduction with $p = 0.05, 0.1, 0.15, 0.2, 0.25, 0.3$; (b)–(d) Variance plots of $\hat{\tau}_0^t$ (dashed line) and $\hat{\tau}^t$ (solid line) fixing p at $0.05, 0.15, 0.25$, respectively. The variances are approximated by simulation based on 2^{14} replications.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

and $p = 0.05$, the variance is reduced by around 50%; see also Remark 3. Although previous results are based on the condition that $f_0(x) = f_1(x)$, it is interesting to discuss the impact of the density ratio on the variance reduction; see Remark 4.

Remark 3. Since $M_0 < N_0/N_1$ and $f_0(x) = f_1(x)$, the variance of the rematching estimator $\hat{\tau}^t$ is roughly a constant across M_0. To reach roughly the same variance as the rematching estimator, we need to increase M_0 to the largest value that satisfies $M_0 < N_0/N_1$ as specified in Theorem 3, i.e., $M_0 < (1-p)/p$. However, even if M_0 is increased to the largest possible value, the variance of the matching estimator is still larger than the variance of the rematching estimator.

Remark 4. Intuitively, if the density ratio is large across different X, i.e., the covariate distribution is quite different between the treated and the control groups, our estimator with rematching will utilize more information than the matching estimator when a small to moderate M_0 is chosen. Because when $f_0(x)/f_1(x)$ is large, the number of neighbours within a particular distance is quite different among units. Then, our estimator with a small M_0 is more efficient than the matching estimator with a large but not large enough M_0 since a really large M_0 is needed to include more information in this case. But increasing M_0 infinitely is risky because additional

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

bias may be introduced. Moreover, the higher the density ratio, the more pronounced the bias caused by the increase in M_0 for the sake of efficiency.

4.3 Consistency and asymptotic normality

Here, we show that the proposed matching estimators are consistent for τ and τ^t. Without the bias term or after the bias correction, they are $N^{1/2}$-consistent and asymptotically normal; see Theorems 4, 5, and 6.

Let $\mathrm{N}(\mu, \sigma^2)$ denote the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}^+$.

Theorem 4 (Consistency)**.** *Suppose Assumption S1 holds.*

1. *If Assumptions S2 and S4.1 hold, then* $\hat{\tau} - \tau \to 0$ *in probability.*

2. *If Assumptions S3 and S4.1 hold, then* $\hat{\tau}^t - \tau^t \to 0$ *in probability.*

Theorem 5 (Asymptotic Normality)**.** *Suppose Assumptions S1 and S4 hold.*

1. *If Assumption S2 holds, then* $\{V^R + V^{\tau(X)}\}^{-1/2} N^{1/2}(\hat{\tau} - B - \tau) \to \mathrm{N}(0,1)$ *in distribution.*

2. *If Assumption S3 holds, then* $\{V^{R,t} + V^{\tau(X),t}\}^{-1/2} N_1^{1/2}(\hat{\tau}^t - B^t - \tau^t) \to \mathrm{N}(0,1)$ *in distribution.*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Theorem 5, which adopts a similar form as Theorem 4 in Abadie and Imbens (2006), states the asymptotic normality of $\hat{\tau}$ and $\hat{\tau}^t$ after subtracting B and B^t. Though a similar form is adopted, the formulas of some terms including V^R, $V^{R,t}$, B, and B^t adopt different forms from those in Abadie and Imbens (2006) since our matching-and-rematching procedure can produce matching results that differ greatly from simple matching results.

Theorem 6 (Consistency and Asymptotic Normality for $\hat{\tau}_{\text{bc}}$ and $\hat{\tau}^t_{\text{bc}}$)**.** *Suppose that Assumptions S1 and S4–S6 hold.*

1. *Suppose Assumption S2 hold, then* $N^{1/2}(B - \hat{B}) \to 0$ *in probability, and* $\{V^R + V^{\tau(X)}\}^{-1/2} N^{1/2}(\hat{\tau}_{\text{bc}} - \tau) \to \text{N}(0,1)$ *in distribution.*

2. *Suppose Assumption S3 hold, then* $N_1^{1/2}(B^t - \hat{B}^t) \to 0$ *in probability, and* $\{V^{R,t} + V^{\tau(X),t}\}^{-1/2} N_1^{1/2}(\hat{\tau}^t_{\text{bc}} - \tau^t) \to \text{N}(0,1)$ *in distribution.*

The theorem above implies that we can estimate the bias of rematching estimators at a similar speed to that of estimating the bias of simple matching estimators, which is faster than $N^{1/2}$ for the estimated average treatment effect and $N_1^{1/2}$ for the estimated average treatment effect on the treated. Also, the normalized variance remains the same after the bias correction.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

4.4 Variance estimation

By the matching strategy mentioned in Abadie and Imbens (2006), we estimate $\sigma^2(D_i, X_i)$ as

$$\hat{\sigma}^2(D_i, X_i) = \frac{J}{J+1}\left(Y_i - \frac{1}{J}\sum_{j=1}^{J} Y_{l_j(i)}\right)^2,$$

where J is the number of matches used, $l_j(i)$ is the index of the jth closest match of the same treatment group for unit i, $\ell_J(i) = \{l_1(i), \ldots, l_J(i)\}$, and

$$\ell_J(i) = \left\{ j \in \{1, ..., N\} : D_j = D_i \text{ and } \sum_{l \neq j : D_l = D_i} \mathbb{1}\{\|X_l - X_i\| \leq \|X_j - X_i\|\} \leq J \right\}.$$

Then, we propose to estimate $V = V^R + V^{\tau(X)}$ and $V^t = V^{R,t} + V^{\tau(X),t}$ by

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

respectively. The consistency is guaranteed as follows, which allows us to perform statistical inference on τ and τ^t. For example, confidence interval for τ and τ^t can be computed.

Theorem 7. *If Assumptions S1, S2 and S4 hold, then* $|\hat{V} - V| = o_p(1)$. *If Assumptions S1, S3 and S4 hold, then* $|\hat{V}^t - V^t| = o_p(1)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

5. Simulation experiments

We evaluate the performance of rematching estimators by Monte Carlo simulation with ten design cases: Cases 1(a)–(e) concern a scalar covariate, whereas Cases 2(a)–(e) concern multiple covariates. Due to space constraints, only four cases are shown in Figure 4 with the complete results deferred to the supplement. We follow previous studies and focus on the average treatment effect on the treated τ^t. Based on Frölich (2004), Busso et al. (2014) and Otsu and Rai (2017), we consider the following data-generating process: For $i = 1, \ldots, N$ and $d = 0, 1$,

$$Y_i(d) = \tau d + m(X_i) + \epsilon_i, \quad D_i = \mathbb{1}\{P(X_i) \geq \upsilon_i\}\, \mathbb{1}(\xi_i > c),$$

where $\epsilon_i \sim \mathrm{N}(0, 0.5^2)$, $\upsilon_i \sim \mathrm{Beta}(1,1)$, and $\xi_i \sim \mathrm{Beta}(1,1)$ are mutually independent; $x \mapsto P(x)$ is a function for specifying the propensity score; $x \mapsto m(x)$ is a mean function; and c is a constant. Here $\mathrm{Beta}(\alpha, \beta)$ denotes the beta distribution with parameters $\alpha, \beta > 0$. For Cases 1(a)–(e), let $X_i \sim \mathrm{Beta}(1.2, 1.2)$. For Cases 2(a)–(e), let X_i be a 6-dimensional covariate containing both continuous and discrete components. The detailed simulation designs, including the settings of $P(\cdot)$, $m(\cdot)$, c, and the distribution of X_i, are available in the supplement.

We first consider cases with a scalar X. The bias term is negligible since

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

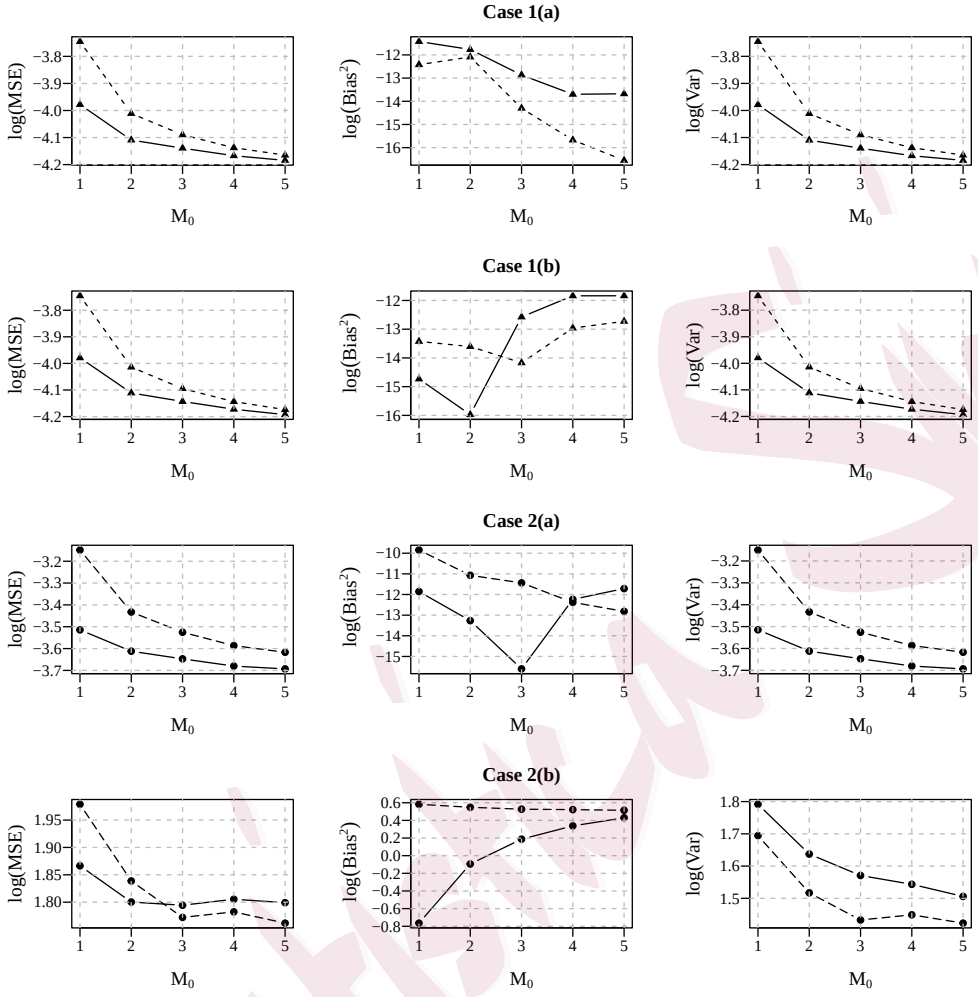

Figure 4: A graph showing the performance comparison between $\hat{\tau}_0^t$ (triangle with a dashed line) and $\hat{\tau}^t$ (triangle with a solid line) when X is a scalar and between their bias-corrected versions, $\hat{\tau}_{0,\mathrm{bc}}^t$ (circle with a dashed line) and $\hat{\tau}_{\mathrm{bc}}^t$ (circle with a solid line), when X is multi-dimensional.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

$B^t = O_p(1/N_0)$ is of sufficiently low order. From Figure 4, we find that our proposed estimator $\hat{\tau}^t$ performs better than the simple matching estimator $\hat{\tau}_0^t$ across different choices of M_0 in all scalar covariate cases, i.e., Cases 1(a) and 1(b), as the rematching process allows more efficiency gain while maintaining a negligible bias when X is a scalar. Also, the mean squared error is comprised mainly of the variance, which makes the leftmost and the rightmost graphs almost identical in each case. The gain in precision by increasing M_0 diminishes rapidly after 4, which is consistent with the findings of Rosenbaum (2020). The declining trend in the mean squared error (MSE) and variance, along with the convergence of both methods as M_0 increases, may give the impression that the rematching process is redundant as higher M_0 already ensures a more precise estimation by $\hat{\tau}_0^t$. However, as we shall see later, the bias can increase to a non-negligible magnitude with M_0 when there are multiple covariates. This is undesirable given the focus on decreasing the bias in the literature of causal inference (Imbens and Wooldridge, 2009).

For all multivariate cases, we compare the performance of the rematching estimator, the simple matching estimator, and their bias-corrected versions. We also include in the comparison the genetic matching estimator (Diamond and Sekhon, 2013), which uses a state-of-art iterative algorithm

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

to maximize a criterion related to covariate balance. Since the mean squared errors of $\hat{\tau}^t_{\text{bc}}$ and $\hat{\tau}^t_{0,\text{bc}}$ are much lower compared with others, we move the complete simulation result to the supplement and show only the results for $\hat{\tau}^t_{\text{bc}}$ and $\hat{\tau}^t_{0,\text{bc}}$ in Figure 4. The performance of the bias-corrected estimators depends on the regression-based bias estimation. When the deviation from the true outcome model of the specified model is small, Case 2(a) for instance, it is favourable to increase M_0 for efficiency, which is consistent with the findings in Lin et al. (2023), meaning that we can increase the number of fixed matches for $\hat{\tau}^t_{0,\text{bc}}$ with the sample size for better estimation. However, using our proposed estimator gives an ideal result even at a very small number of fixed matches (e.g., $M_0 = 1$). Also, we may not always be able to specify the model correctly in practice. In Case 2(b), where the bias is not well-corrected, an interesting finding is that our proposal $\hat{\tau}_{\text{bc}}$ is able to achieve a lower bias across M_0. Considering that bias is a good measure of the covariate balance and a smaller bias means a more valid causal interpretation, our proposal with a small M_0 is generally better than the original proposal.

In all cases, our matching-and-rematching procedure gains efficiency while maintaining low bias, which sits well with the focus on the credibility of the causal inference in the literature (Imbens and Wooldridge, 2009).

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Practically, people simply fix $M_0 = 1$ when using matching estimators. For this particular choice of M_0, our matching-and-rematching procedure gives better estimation results since it allows more efficiency gain when the bias is negligible or the outcome model is well-specified and less precision loss when the outcome model is misspecified.

6. Empirical application

We apply rematching estimators to analyze the National Supported Work data, an evaluation of a job training program analyzed by LaLonde (1986), Heckman and Hotz (1989), Dehejia and Wahba (1999), Imbens (2003), and Smith and Todd (2005). The dataset is available on Rajeev Dehejia's website (`http://users.nber.org/~rdehejia/nswdata2.html`). It contains experimental and non-experimental samples; see the supplement for details. As is common in previous studies, we focus on the average effect of the program on earnings for the treated.

Table 1 summarizes the results for the estimates and standard errors. Although the true treatment effect is unknown, the difference-in-mean estimate for the experimental data, which is 1.79 with a standard error of 0.67, can be regarded as a benchmark. The 95% confidence interval obtained by a normal approximation to the limiting distribution is $[0.48, 3.11]$. For the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Table 1: Experimental and nonexperimental estimates for the NSW data.

	$M_0 = 1$		$M_0 = 4$		$M_0 = 16$		$M_0 = 64$		$M_0 = N_0$	
	Est.	(SE)	Est.	(SE)	Est.	(SE)	Est.	(SE)	Est.	(SE)
Experimental										
$\hat{\tau}_0^t$	1.82	(0.85)	2.08	(0.72)	1.97	(0.68)	2.23	(0.69)	1.79	(0.67)
$\hat{\tau}_{0,\mathrm{bc}}^t$	1.87	(0.84)	1.90	(0.73)	1.74	(0.68)	1.63	(0.69)	1.77	(0.66)
$\hat{\tau}^t$	1.94	(0.81)	2.04	(0.72)	1.97	(0.68)	2.23	(0.69)	1.79	(0.67)
$\hat{\tau}_{\mathrm{bc}}^t$	1.99	(0.80)	1.88	(0.72)	1.75	(0.68)	1.63	(0.69)	1.77	(0.66)
Non-experimental										
$\hat{\tau}_0^t$	1.14	(1.25)	1.23	(1.08)	0.66	(1.02)	−0.32	(0.86)	−3.65	(0.75)
$\hat{\tau}_{0,\mathrm{bc}}^t$	1.82	(1.26)	2.06	(1.08)	1.96	(1.03)	2.32	(0.88)	1.83	(0.77)
$\hat{\tau}^t$	1.05	(1.22)	1.13	(1.08)	0.61	(1.02)	−0.32	(0.86)	−3.65	(0.75)
$\hat{\tau}_{\mathrm{bc}}^t$	2.02	(1.22)	2.11	(1.08)	1.96	(1.04)	2.29	(0.88)	1.83	(0.77)

non-experimental sample, the estimated treatment effects by our proposed bias-adjusted estimator and by the bias-adjusted simple matching estimator are all inside the experimental 95% confidence interval. It is found that the standard errors of our proposal $\hat{\tau}_{\mathrm{bc}}^t$ are smaller than those of the bias-adjusted matching estimator by Abadie and Imbens (2011) when $M_0 = 1$, which is consistent with our goal of increasing efficiency by rematching. However, because each pair of the treated and the control groups are close in size, which results in a small number of unmatched controls for large M_0, the improvement in standard error is insignificant for $M_0 > 1$ according to

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Table 1. Moreover, setting $M_0 = 1$ results in the use of the highest quality matches, reducing bias in both matching and rematching estimators. Therefore, we consider $M_0 = 1$ for our analysis below.

In particular, we are interested in testing whether the treatment effect is positive, i.e., testing $H_0 : \tau^t = 0$ against $H_1 : \tau^t > 0$. The p-value can be computed by the normal approximation to the asymptotic distribution. Based on the non-experimental sample, we find that when $M_0 = 1$, the p-values based on the classical matching estimator $\hat{\tau}^t_{0,\text{bc}}$ and our proposed rematching estimator $\hat{\tau}^t_{\text{bc}}$ are $1 - \Phi(1.82/1.26) = 7.43\%$ and $1 - \Phi(2.02/1.22) = 4.88\%$, respectively, where $\Phi(\cdot)$ is the distribution function of $\text{N}(0, 1)$. It means that our proposed test successfully identifies the treatment effect at 5% significance level while $\hat{\tau}^t_{0,\text{bc}}$ fails in doing so.

In a nutshell, our proposed estimator $\hat{\tau}^t_{\text{bc}}$ is generally more efficient, resulting in a more powerful test and maintaining statistical validity.

7. Conclusion

In this article, we propose new matching estimators of treatment effects and derive their large sample properties. In contrast to the simple matching, our matching-and-rematching procedure gains efficiency without increasing the bias too much by rematching for the treated units from an opposite direc-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

tion, which matches each treated unit with a variable number of unmatched controls and increases the matched sample size. Our method is applicable to both the average treatment effect and its counterpart for the treated population. Simulation results indicate that our method works well in finite samples, suggesting it may be a useful estimator in practice. Finally, an application to the National Supported Work data reveals an interesting test result.

Supplementary Material

The supplement contains technical assumptions, proofs of main results, additional simulation results, a detailed description of the real-data application, and a summary of existing work; see Sections S1–S5, respectively.

Acknowledgments

The authors would like to thank the anonymous referees, an Associate Editor, and the Editor for their constructive comments that improved the scope and theoretical development of the paper. We are particularly grateful to the referees for providing valuable suggestions to improve the argument in the proof of Theorem 1.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Funding

Chan was partially supported by the Early Career Scheme 24306919 and the General Research Fund 14307922, provided by the Research Grants Council of HKSAR.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *Review of Economic Studies 72*(1), 1–19.

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 (2008). On the failure of the bootstrap for matching estimators. *Econometrica 76*(6), 1537–1557.

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.

Busso, M., J. DiNardo, and J. McCrary (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics 96*(5), 885–897.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Chen, X., H. Hong, and A. Tarozzi (2008). Semiparametric efficiency in GMM models with auxiliary data. *Annals of Statistics 36*(2), 808–843.

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.

Frölich, M. (2004). Finite-sample properties of propensity-score matching and weighting estimators. *Review of Economics and Statistics 86*(1), 77–90.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*, 315–331.

Heckman, J. J. and V. J. Hotz (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association 84*(408), 862–874.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies 65*(2), 261–294.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*(260), 663–685.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review 93*(2), 126–132.

Imbens, G. W., W. K. Newey, and G. Ridder (2005). Mean-square-error calculations for average treatment effects. manuscript.

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

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review 76*, 604–620.

Lin, Z., P. Ding, and F. Han (2023). Estimation based on nearest neighbor matching: from density ratio to average treatment effect. *Econometrica 91*(6), 2187–2217.

Otsu, T. and Y. Rai (2017). Bootstrap inference of matching estimators for average treatment effects. *Journal of the American Statistical Association 112*(520), 1720–1732.

Quade, D. (1982). Nonparametric analysis of covariance by matching. *Biometrics 38*, 597–611.

Robins, J. M. and A. Rotnitzky (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association 90*(429), 122–129.

Rosenbaum, P. R. (2020). Modern algorithms for matching in observational studies. *Annual*

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0306

Review of Statistics and Its Application 7, 143–176.

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.

Rubin, D. B. (1973a). Matching to remove bias in observational studies. *Biometrics 29*, 159–183.

Rubin, D. B. (1973b). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics 29*, 185–203.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*(5), 688–701.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics 2*(1), 1–26.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics 6*(1), 34–58.

Smith, J. A. and P. E. Todd (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics 125*(1-2), 305–353.

Lam Lam Hui and Kin Wai Chan

Department of Statistics and Data Science, The Chinese University of Hong Kong, Shatin, Hong Kong

E-mail: `lamlamhui@link.cuhk.edu.hk` and `kinwaichan@cuhk.edu.hk`