Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Nearest neighbor matching: M-out-of-N bootstrapping without bias correction vs. the naive bootstrap

Christopher Walsh [a,*], Carsten Jentsch [b]

[a] Ulm University, Institute of Statistics, Department of Mathematics and Economics, Ulm, Germany
[b] TU Dortmund University, Department of Statistics, Dortmund, Germany

ARTICLE INFO

Article history:
Received 8 November 2022
Revised 18 April 2023
Accepted 20 April 2023
Available online xxx

JEL classification:
C104
C21
ATET
Matching Estimator
M-out-of-N Bootstrap

ABSTRACT

It is well known that the limiting variance of nearest neighbor matching estimators cannot be consistently estimated by a naive Efron-type bootstrap as the conditional variance of the bootstrap estimator does not generally converge to the correct limit in expectation. In essence this is caused by the fact that the bootstrap sample contains ties with positive probability even when the sample size becomes large. This negative result was originally derived in a simple setting by Abadie and Imbens (ECONOMETRICA, pp. 235–267, 76(6), 2008). A proof of concept for a direct M-out-of-N bootstrap on the data is provided in this setting. It is proven that in this setting the conditional variance of a direct M-out-of-N-type bootstrap estimator *without* bias-correction does converge to the correct limit in expectation. The key to the proof lies in the fact that asymptotically with probability one there are no ties in the bootstrap sample. The potential of the direct M-out-of-N-type bootstrap is investigated in simulations.

1. Introduction

Matching estimators are intuitively simple procedures to estimate average treatment effects within the potential outcomes framework. Asymptotic theory for matching estimators was derived in Abadie and Imbens (2006, 2011, 2012). The asymptotic variance is typically rather complicated and depends on numerous nuisance parameters. This, along with possible finite sample improvements, motivates the desire to apply resampling based procedures to estimate the (asymptotic) distribution in the matching context. However, in a highly influential paper Abadie and Imbens (2008) showed that the standard errors obtained from a naive Efron-type bootstrap will in general be invalid. They demonstrated this for nearest neighbor matching for the average treatment effect on the treated (ATET). Specifically, using a simple data generating process (DGP), they obtained closed form expressions for: (i) the (finite and) limiting variance of the matching estimator and (ii) the limit of the expectation of the conditional variance of an Efron-type bootstrap estimator, that is constructed by recomputing the matching estimator on a bootstrap sample obtained by separately sampling with replacement from the control group and the treatment group. As the two expressions do not coincide the Efron-type bootstrap variance estimator is not valid. Of course, this is a hugely significant negative result as it implies that in more general settings such an Efron-type resampling procedure will quite possibly also fail to reproduce the limit distribution of the matching estimator.

* Corresponding author.
E-mail addresses: christopher.walsh@uni-ulm.de (C. Walsh), jentsch@statistik.tu-dortmund.de (C. Jentsch).

https://doi.org/10.1016/j.ecosta.2023.04.005
2452-3062/

This negative result on the validity of Efron-type resampling procedures in the matching context is accompanied by a conjecture in Abadie and Imbens (2008) stating that the limiting distribution should be correctly estimable by either the wild bootstrap of Härdle and Mammen (1993) or the M-out-of-N bootstrap as considered e.g. in Bickel et al. (1997). Indeed, Otsu and Rai (2017) proposed and proved the validity of a weighted bootstrap procedure that can be interpreted as a wild bootstrap. Rather than resampling the data directly, their procedure resamples individual contributions of the bias-corrected matching estimator. These entities are approximations to the martingale differences of the martingale representation for the bias-corrected matching estimator in Abadie and Imbens (2012). The same idea based on resampling the individual contributions rather than the data could be used to construct a valid *indirect* M-out-of-N bootstrap procedure for the *bias-corrected* matching estimator. Subsampling the individual contributions of the *bias-corrected* matching estimator should also be valid given certain conditions on the size of the subsamples along the general results of Politis and Romano (1994), which in turn can then be extended to an M-out-of-N bootstrap when this is asymptotically equivalent to subsampling as for instance shown Politis et al. (1999). Key to the above is that the individual contributions of the *bias-corrected* matching estimator are resampled. Interestingly, it turns out that this bias correction step before resampling is required *even* when the original matching estimator is exactly unbiased for all finite sample sizes. As shown in Walsh et al. (2021) it is not immediately clear how to actually bias-correct in practice. Moreover, simple adhoc procedures may severely distort the resampling based standard errors. Due to this drawback and the fact that it is not clear whether the original conjecture was actually referring to such a bias-correct-first, resample-later approach, we will investigate whether it is at all possible to use a *direct* M-out-of-N bootstrap without a preliminary bias-correction. To this end, we will show that a direct M-out-of-N bootstrap clears the "first hurdle", by proving that in the setting considered in Abadie and Imbens (2008), the conditional variance of our M-out-of-N-type bootstrap estimator that directly resamples the original data is indeed an asymptotically unbiased estimator of the limit variance if the resample size satisfies $M = o(N^{1/2})$. Given the technicality of the proof in Abadie and Imbens (2008), it is to be expected that our proof is also rather technical. However, surprisingly, our proof for the M-out-of-N-type bootstrap does differ quite substantially from the arguments employed in Abadie and Imbens (2008) for the Efron-type bootstrap. The key result in our proof is that, whereas even asymptotically the Efron-type bootstrap samples contain ties, thus leading to the failure to correctly replicate the matching procedure, the direct M-out-of-N-type bootstrap samples will not contain any ties asymptotically. Although it is not immediately clear whether our results can be easily extended to more general settings, it nonetheless opens the possibility that direct application of an M-out-of-N-type bootstrap may be possible, thus *avoiding* the need to bias-correct first. For the time being we content ourselves with showing that the direct M-out-of-N-type bootstrap variance estimator is asymptotically unbiased in the setting used by Abadie and Imbens (2008) to assert the failure of the naive Efron-type bootstrap.

The setting used by Abadie and Imbens (2008) along with the resulting asymptotic properties for the nearest neighbor matching estimator of the ATET and the Efron-type bootstrap variance estimator are given in Section 2. Our proposed direct M-out-of-N-type bootstrap procedure is presented in Section 3 along with the main theoretical result showing that our direct procedure that does not require prior bias-correction can be used to asymptotically unbiasedly estimate the limiting variance of the nearest neighbor matching estimator for the ATET in the setting of Abadie and Imbens (2008). The outline of the proof via a sequence of lemmas is given in the appendix at the end of the manuscript. The detailed theoretical arguments are relegated to the supplementary material available in the online version of this article. Section 4 provides simulations to illustrate the behavior of our estimator and the dependence of its performance on the interplay between the balancedness of the design and the resampling size. Section 5 concludes.

2. Setup

In the basic binary treatment effects setup, $Y_i(0)$ and $Y_i(1)$ are the potential outcomes under control and after treatment, respectively, for unit $i = 1, \dots, N$. For each unit, we observe $Z_i = (Y_i, W_i, X_i')'$, where W_i is the treatment indicator ($W_i = 1$, if the unit is treated, and $W_i = 0$ otherwise), $Y_i = W_i Y_i(1) + (1 - W_i)Y_i(0)$ is the observed outcome and X_i is a vector of (continuously distributed) covariates. Let $(Y(1), Y(0), W, X')'$ be the population random variables from which the data are drawn. We will be interested in estimating the average treatment effect on the treated (ATET)

$$\tau^t = \mathbb{E}[Y(1) - Y(0) \mid W = 1].$$

Given data $\mathbf{Z} = \{(Y_i, W_i, X_i')\}_{i=1}^N$, the simple nearest neighbor matching estimator for the ATET is given by

$$\widehat{\tau}^t = \frac{1}{N_1} \sum_{i=1}^{N} W_i \{Y_i - \widehat{Y}_i(0)\},$$

where $N_1 = \sum_{i=1}^N W_i$ is the number of treated units and $\widehat{Y}_i(0) = Y_{j(i)}$ is the imputed value for the non-observed $Y_i(0)$ with $j(i)$ being the index of the unit that is matched to unit i. The index $j(i)$ is the index of the control unit that is closest to treated unit i, where closeness is measured by the distance of the regressor (vector) to X_i, that is

$$j(i) = \underset{j \in \{1,\dots,N\}: W_j = 0}{\arg\min} ||X_j - X_i||.$$

To allow for the possibility that there is more than one closest unit, one can write $\widehat{Y}_i(0) = \frac{1}{\#\mathcal{J}(i)} \sum_{j\in\mathcal{J}(i)} Y_j$, with $\#\mathcal{J}(i)$ the number of closest units to treated unit i and $\mathcal{J}(i) = \arg\min_{j\in\{1,\dots,N\}:W_j=0} ||X_j - X_i||$, the set of indices of those closest control units. With $K_j = \sum_{i=1}^N W_i \frac{I\{j\in\mathcal{J}(i)\}}{\#\mathcal{J}(i)}$ defined as the (average) number of times (control) unit j is a match to a treated unit, we can rewrite the matching estimator as

$$\widehat{\tau}^t = \frac{1}{N_1}\sum_{i=1}^N (W_i - (1-W_i)K_i)Y_i. \tag{1}$$

The naive Efron-type bootstrap that was shown to fail in Abadie and Imbens (2008) resamples the data and calculates the estimator in (1) on the bootstrap data as follows.

Efron-type bootstrap

- **Step 1:** Split the sample $\mathbf{Z} = \{(Y_i, W_i, X_i')\}_{i=1}^N$ into the treatment group (N_1 units with $W_i = 1$) and the control group (N_0 units with $W_i = 0$). Sample with replacement N_1 units from the treated group and N_0 units from the control group. Combine the two sample groups to get the bootstrap sample $\{(Y_i^*, W_i^*, X_i^{*\prime})\}_{i=1}^N$.
- **Step 2:** Calculate the matching estimator on the bootstrap sample

$$\widehat{\tau}^{t,*} = \frac{1}{N_1}\sum_{i=1}^N (W_i^* - (1-W_i^*)K_i^*)Y_i^*, \tag{2}$$

where $K_j^* = \sum_{i=1}^N W_i^* \frac{I\{j\in\mathcal{J}^*(i)\}}{\#\mathcal{J}^*(i)}$ with $\mathcal{J}^*(i) = \underset{j\in\{1,\dots,N\}:W_j^*=0}{\arg\min} ||X_j^* - X_i^*||$.

In a typical bootstrap sample, some control units will be sampled multiple times. Hence, in contrast to when matching in the original sample, there will be ties (with probability one) when matching in the bootstrap sample.

Abadie and Imbens (2008) asserted the failure of the naive Efron-type bootstrap for DGPs given by Assumption 1.

Assumption 1. Let $\mathbf{Z} = \{(Y_i, W_i, X_i')\}_{i=1}^N$ be a sample of $N = N_1 + N_0$ independent draws from $Y, X \mid W = w$ for $w = 0, 1$ such that:

(i) The N_1 treated and N_0 control units occur in a fixed ratio $\alpha_N = N_1/N_0$ with $\alpha_N \to \alpha > 0$ as $N \to \infty$.
(ii) The regressor satisfies $X \sim \mathcal{U}[0,1]$.
(iii) The propensity score $p(x) = \Pr(W = 1|X = x)$ is constant.
(iv) $Y = WY(1) + (1-W)Y(0)$ and the potential outcomes satisfy:
  (a) $Y(1)$ is degenerate with $\Pr(Y(1) = \tau^t) = 1$ for some fixed τ^t.
  (b) $Y(0) \mid X = x \sim \mathcal{N}(0,1)$ for all $x \in [0,1]$.

The conditions in Assumption 1, correspond to Assumptions 3.1 - 3.4 in Abadie and Imbens (2008) with the exception that in (i) we have added that the ratio of treated to control units converges to some $\alpha > 0$, which is actually needed to study the asymptotics. In the setting of Assumption 1, Abadie and Imbens (2008) obtained simple expressions summarized in Lemma 1 for: (i) the (finite and) limiting variance of the scaled and centered matching estimator $\sqrt{N_1}(\widehat{\tau}^t - \tau^t)$ and (ii) the limit of the expectation of the naive Efron-type bootstrap variance estimator $\mathbb{V}\text{ar}^*[\sqrt{N_1}\widehat{\tau}^{t,*} \mid \mathbf{Z}]$, where $\mathbb{V}\text{ar}^*[\,\cdot \mid \mathbf{Z}]$ denotes the variance over the resampling mechanism conditional on the data.

Lemma 1 (Asymptotic results for $\widehat{\tau}^t$ and $\widehat{\tau}^{t,*}$)**.** *Given Assumption 1, it holds that*

(i) $\mathbb{V}\text{ar}\Big[\sqrt{N_1}(\widehat{\tau}^t - \tau^t)\Big] = 1 + \frac{3}{2}\frac{(N_1-1)(N_0+8/3)}{(N_0+1)(N_0+2)} \to 1 + \frac{3}{2}\alpha.$
(ii) $\mathbb{E}\Big[\mathbb{V}\text{ar}^*[\sqrt{N_1}\widehat{\tau}^{t,*} \mid \mathbf{Z}]\Big] \to 1 + \frac{3}{2}\alpha\frac{5e^{-1}-2e^{-2}}{3(1-e^{-1})} + 2e^{-1}.$

Simple calculations show that if $\alpha = \bar{\alpha} := \frac{4(1-e^{-1})e^{-1}}{3-8e^{-1}+2e^{-2}} \approx 2.84$, then $\mathbb{V}\text{ar}^*[\sqrt{N_1}\widehat{\tau}^{t,*} \mid \mathbf{Z}]$ is an asymptotically unbiased estimator of the limit variance, whereas it will be too large (small) on average in large samples if $\alpha < \bar{\alpha}$ ($\alpha > \bar{\alpha}$), thus establishing the failure of the naive Efron-type bootstrap.

3. A direct M-out-of-N-type bootstrap

Our direct M-out-of-N-type bootstrap will resample $M_1 < N_1$ treated units and $M_0 < N_0$ control units. From the multitude of possible choices for the resample sizes M_0 and M_1, we propose to use a choice that attempts to keep the balancedness between the treated and control groups as close as possible to the one in the original data. Specifically, we would like to resample M_1 treated and M_0 controls from the original data with $M_1/M_0 = N_1/N_0 = \alpha_N$. This requires choosing $M_0 = 1/(1+\alpha_N)M$ and $M_1 = \alpha_N/(1+\alpha_N)M$ with $M = M_0 + M_1$ the total number of resampled units. As the expression for M_1 and M_0 will typically be non-integer, these need to be replaced by integers in our proposed direct M-out-of-N-type bootstrap matching estimator.

Direct M-out-of-N-type bootstrap

- **Step 1:** Fix a $\gamma \in (0,1)$ and set $M = \lfloor N^{\gamma} + \frac{1}{2} \rfloor$ whilst ensuring that $M_1 = \lfloor \alpha_N/(1+\alpha_N)M + \frac{1}{2} \rfloor$ and $M_0 = M - M_1$ are non-zero with $\lfloor x \rfloor$ denoting the integer part of x.
- **Step 2:** Split the sample $\mathbf{Z} = \{(Y_i, W_i, X_i')\}_{i=1}^{N}$ into the treatment and the control groups. Sample with replacement M_0 units from the control group and M_1 units from the treated group. Combine the two sampled groups to get the bootstrap sample $\{(Y_i^*, W_i^*, X_i^{*\prime})\}_{i=1}^{M}$ with $M = M_0 + M_1$.
- **Step 3:** Calculate the matching estimator on the bootstrap sample

$$\widehat{\tau}_M^{t,*} = \frac{1}{M_1} \sum_{i=1}^{M} (W_i^* - (1 - W_i^*)K_i^*)Y_i^*, \tag{3}$$

with $K_j^* = \sum_{i=1}^{M} W_i^* \frac{I\{j \in \mathcal{J}^*(i)\}}{\#\mathcal{J}^*(i)}$ and $\mathcal{J}^*(i) = \underset{j \in \{1,\dots,N\}: W_j^* = 0}{\arg\min} ||X_j^* - X_i^*||$.

In Step 1, $\lfloor x + \frac{1}{2} \rfloor$ rounds x to its closest integer. The parameter γ in Step 1 needs to be chosen to ensure that we have at least one observation in each treatment arm of the bootstrap sample. M_1 and M_0 are set such that the ratio of treated to controls closely matches that ratio in the original sample and $\frac{M_1}{M_0} \to \alpha$ as $N \to \infty$. The parameter γ captures the degree of undersampling as $M \approx N^{\gamma}$. Note, that if we set $\gamma = 1$, then we would get $M = N$ and the procedure would yield the naive Efron-type bootstrap. In Step 3, the matching takes into account that ties may be present in the bootstrap sample. In contrast to the naive Efron-type bootstrap, it will be seen that asymptotically there are no ties. This allows us to establish our main result, that the variance estimator $\mathbb{V}\text{ar}^*[\sqrt{M_1}\widehat{\tau}_M^{t,*} \mid \mathbf{Z}]$ is asymptotically unbiased.

Theorem 1. *Given Assumption 1. If $M_0 = o(\sqrt{N_0})$ and $\frac{M_1}{M_0} \to \alpha$, then, as $N_0 \to \infty$,*

$$\mathbb{E}\Big[\mathbb{V}\text{ar}^*[\sqrt{M_1}(\widehat{\tau}_M^{t,*} - \widehat{\tau}^t) \mid \mathbf{Z}]\Big] \to 1 + \frac{3}{2}\alpha.$$

Thus, the conditional variance of the direct M-out-of-N-type bootstrap estimator is an asymptotically unbiased estimator of the limiting variance of $\sqrt{N_1}(\widehat{\tau}^t - \tau^t)$ given in Lemma 1(i). The requirement $M_0 = o(\sqrt{N_0})$ is used to ensure that asymptotically the bootstrap sample contains no ties. More specifically it is used to establish (S.5) in Section S.1.3 of the supplementary material which shows that the lower bound on the probability that the bootstrap contains no ties converges to 1. The requirement cannot be weakened in the proof: If $M_0 = O(\sqrt{N_0})$, then via (not reported) simulations one can easily see that there is a strictly positive probability in the limit that the bootstrap will contain ties asymptotically. Especially for large α, this will result in underestimation of the limiting variance as is also evidenced by the simulation results reported in Figure 1 below for the M-out-of-N-type bootstrap with $\gamma = 0.5$.

Sketch of Proof of Theorem 1. The variance estimator based on the direct M-out-of-N-type bootstrap estimator is given by $\mathbb{V}\text{ar}^*[\sqrt{M_1}(\widehat{\tau}_M^{t,*} - \widehat{\tau}^t) \mid \mathbf{Z}]$. As we can write

$$\begin{aligned}\sqrt{M_1}(\widehat{\tau}_M^{t,*} - \widehat{\tau}^t) &= \sqrt{M_1}(\widehat{\tau}_M^{t,*} - \tau^t) - \sqrt{M_1}(\tau^t - \widehat{\tau}) \\ &= \sqrt{M_1}(\widehat{\tau}_M^{t,*} - \tau^t) + \sqrt{\frac{M_1}{N_1}}\sqrt{N_1}(\widehat{\tau}^t - \tau^t).\end{aligned} \tag{4}$$

The last term is asymptotically negligible as $M_1/N_1 \to 0$ by construction and due to the asymptotic normality of the matching estimator given in Lemma 3.1 of Abadie and Imbens (2008). Thus, we can concentrate on the variance of the leading term, that is, $\mathbb{V}\text{ar}^*[\sqrt{M_1}(\widehat{\tau}_M^{t,*} - \tau^t) \mid \mathbf{Z}]$. From the assumptions on the DGP in Assumption 1, the definition of the direct M-out-of-N-type bootstrap estimator in (3) and the clever and unorthodox notation used by Abadie and Imbens (2008), it is possible to write our bootstrap estimator in terms of the original data as

$$\begin{aligned}\widehat{\tau}_M^{t,*} &= \frac{1}{M_1} \sum_{i=1}^{N} (W_i R_{b_M,i} - (1 - W_i)K_{b_M,i})Y_i \\ &= \tau^t - \frac{1}{M_1} \sum_{i=1}^{N} (1 - W_i)K_{b_M,i}Y_i(0),\end{aligned}$$

where $K_{b_M,i} = \sum_{j=1}^{N} W_j B_{M,i}(X_j) R_{b_M,j}$ is the number of (bootstrap) treated units the (original) control unit i is matched to when the (original) control unit i is in the bootstrap, and is zero otherwise. $R_{b_M,j}$ counts the number of times the (original) treated unit (with regressor value X_j) is contained in the bootstrap sample. $B_{M,i}(X_j)$ is an indicator that shows whether the (original) control unit i is in the bootstrap sample and is matched to a treated unit in the bootstrap with covariate value X_j. Straightforward calculations then show that the expectation of the conditional variance of the leading term of the direct M-out-of-N-type bootstrap estimator in (4) can be written as

$$\mathbb{E}\Big[\mathbb{V}\text{ar}^*[\sqrt{M_1}(\widehat{\tau}_M^{t,*} - \tau^t) \mid \mathbf{Z}]\Big] = \frac{N_0}{M_1}\mathbb{E}\Big[\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{Z}]\Big], \tag{5}$$

where $\mathbb{E}^*[\,\cdot \mid \mathbf{Z}]$ denotes the expectation over the resampling mechanism conditional on the original data. The proof then proceeds by establishing that $\frac{N_0}{M_1}\mathbb{E}\Big[\mathbb{E}^*[K^2_{b_{M,i}} \mid \mathbf{Z}]\Big] \to 1 + \frac{3}{2}\alpha$. Further details are given in Appendix A. □

As the arguments to show the inconsistency of the naive Efon-type bootstrap in Abadie and Imbens (2008) are quite technical, it is to be expected that the proof of Theorem 1 is also rather technical. Surprisingly, however, the proof requires arguments that differ substantially from the derivation of the limit of the expectation of the naive Efron-type bootstrap variance estimator given in Abadie and Imbens (2008). The first difference concerns the fact that the expression in (5) is not of the form "target + error", whereas for the naive Efron-type bootstrap estimator Abadie and Imbens (2008) show that

$$\mathbb{E}\Big[\mathbb{V}\text{ar}^*[\sqrt{N_1}(\widehat{\tau}^{t,*} - \widehat{\tau})^2 \mid \mathbf{Z}]\Big] = N_1\mathbb{V}\text{ar}[\widehat{\tau}^t] - 2\frac{N_0}{N_1}\mathbb{E}[K_i\mathbb{E}^*[K_{b_{N,i}} \mid \mathbf{Z}]] + \frac{N_0}{N_1}\mathbb{E}[\mathbb{E}^*[K^2_{b_{N,i}} \mid \mathbf{Z}]]. \tag{6}$$

Moreover, the arguments used to derive the limit of $\frac{N_0}{M_1}\mathbb{E}\Big[\mathbb{E}^*[K^2_{b_{M,i}} \mid \mathbf{Z}]\Big]$ in (5) are substantially different to those used by Abadie and Imbens (2008) to analyse the similar-looking term $\frac{N_0}{N_1}\mathbb{E}[\mathbb{E}^*\Big[K^2_{b_{N,i}} \mid \mathbf{Z}\Big]$ in (6). In particular, in their Lemma A.6 they establish the limit of $\mathbb{E}[\mathbb{E}^*\Big[K^2_{b_{N,i}} \mid \mathbf{Z}\Big]$, which is of no use for $\mathbb{E}\Big[\mathbb{E}^*[K^2_{b_{M,i}} \mid \mathbf{Z}]\Big]$ in our case as one would merely establish that this term goes to zero. Instead, we make use of a degeneracy result for the distribution of the number of distinct bootstrap units in our low intensity M-out-of-N sampling scheme. This result on the so-called occupancy distribution ensures that asymptotically our bootstrap samples contain only distinct units with probability one, which in contrast to the Efron-type bootstrap, allows the matching mechanism to be replicated correctly asymptotically.

Remark 1. Suppose Assumption 1 holds with part (iv)(a) replaced by $Y(1) \mid X = x \sim \mathcal{N}(\tau^t, \sigma^2)$ for all $x \in [0,1]$, some fixed τ^t, and some $\sigma^2 \in (0,\infty)$. Then, we have

$$\mathbb{V}\text{ar}\Big[\sqrt{N_1}(\widehat{\tau}^t - \tau^t)\Big] \to \sigma^2 + 1 + \frac{3}{2}\alpha$$

and

$$\mathbb{E}\Big[\mathbb{V}\text{ar}^*[\sqrt{M_1}(\widehat{\tau}^{t,*}_M - \widehat{\tau}^t) \mid \mathbf{Z}]\Big] \to \sigma^2 + 1 + \frac{3}{2}\alpha.$$

Hence, the bootstrap variance $\mathbb{V}\text{ar}^*[\sqrt{M_1}(\widehat{\tau}^{t,*} - \widehat{\tau})^2 \mid \mathbf{Z}]$ is then also asymptotically unbiased for $\mathbb{V}\text{ar}\Big[\sqrt{N_1}(\widehat{\tau}^t - \tau^t)\Big]$.

Proof. The proof is given in the Appendix. □

4. Simulations

We simulate data satisfying Assumption 1 to illustrate the performance of the M-out-of-N-type variance estimator as a point estimator for the limit variance. We also look at the coverage probabilities of the confidence intervals based on the M-out-of-N-type bootstrap. For every fixed ratio of treated to controls $\alpha \in \{0.2, 1, \bar{\alpha}, 5\}$, we simulate $S = 1\,000$ data sets with treatment effect $\tau^t = 1$ and sample sizes $N \in \{100, 250, 500, 1\,000, 2\,000\}$. In each simulation run and for each sample size, we calculate: (i) the nearest neighbor matching estimator $\widehat{\tau}^t$ (given in (1)); (ii) the naive Efron-type bootstrap estimator $\widehat{\tau}^{t,*}$ (given in (2)); and (iii) the direct M-out-of-N-type bootstrap estimator $\widehat{\tau}^{t,*}_M$ (given in (3)) for $\gamma \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$. Recall that the Efron-type bootstrap corresponds to the direct M-out-of-N-type bootstrap with $M = N$, which in turn corresponds to setting $\gamma = 1$. Thus, we will also use the notation $\widehat{\tau}^{t,*}_M$ to denote the Efron-type bootstrap. All the bootstrap estimators were recomputed using $B = 1\,000$ bootstrap resamples. If we wish to make the dependence on α, N and the simulation run indexed by $s = 1, \ldots, S$ explicit, we will write $\widehat{\tau}^t_s(\alpha; N)$ and $\widehat{\tau}^{t,*}_{s,M}(\alpha; N)$, respectively. R code to implement our M-out-of-N bootstrap estimator along with code to run the simulation study can be obtained from https://github.com/ChriWalsh/MooN-matching.

4.1. Performance of bootstrap-based variance estimators

Let $\widehat{\tau}^{t,*,b}_{s,M}(\alpha; N)$ be the M-out-of-N-type bootstrap based estimator of the ATET in simulation run $s = 1, \ldots, S$ calculated from the bootstrap sample indexed by $b = 1, \ldots, B$. Then, for each simulation run, $s = 1, \ldots, S$ we construct bootstrap-based variance estimators for the limiting variance of the ATET for a ratio of treated to controls of α using

$$\widehat{v}^*_{s,M}(\alpha; N) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\tau}^{t,*,b}_{s,M}(\alpha; N) - \frac{1}{B}\sum_{i=1}^{B}\widehat{\tau}^{t,*,b}_{s,M}(\alpha; N)\right)^2. \tag{7}$$

The results for the different bootstrap procedures are given in Figure 1. Each row corresponds to a particular resampling procedure. The first row labelled Efron gives the results for the naive Efron-type bootstrap, which corresponds to our direct M-out-of-N-type procedure with $\gamma = 1$. The following rows are labelled MooN(γ) and give the results for our direct M-out-of-N-type bootstrap for a specific choice of γ. Each column contains the results for a specific α, the ratio of treated to

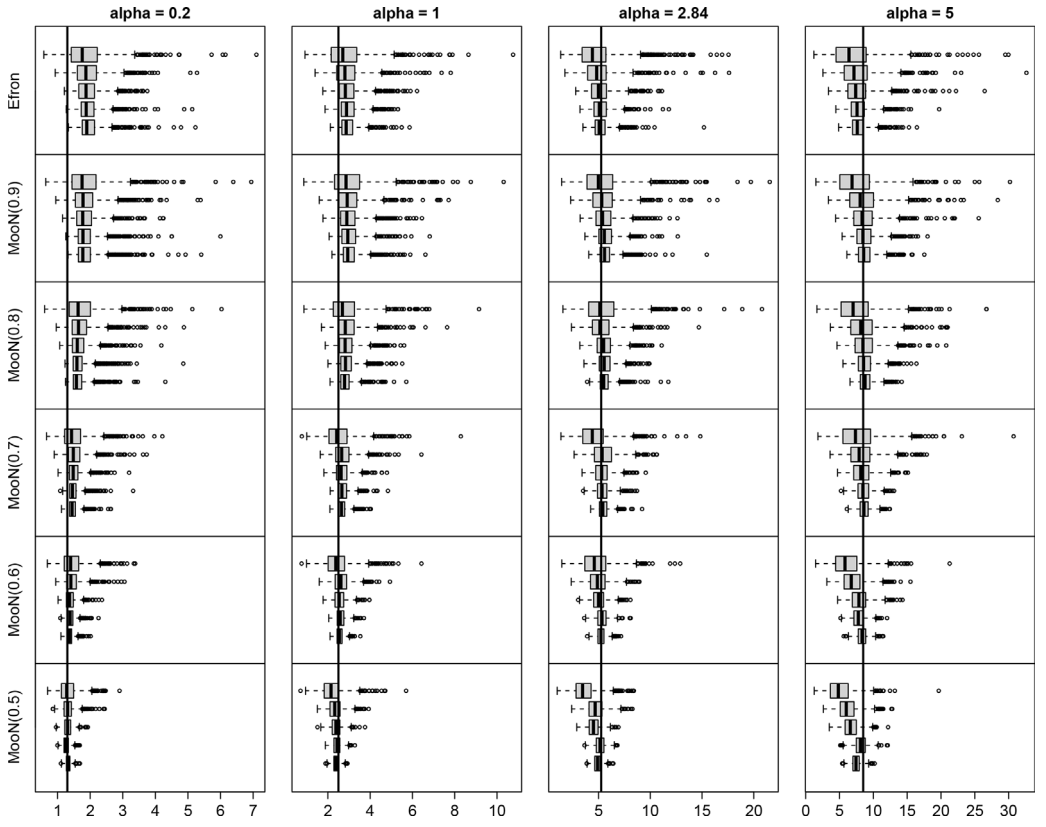

Fig. 1. Boxplots of resampling based variance estimators. The columns correspond to simulations with $\alpha \in \{0.2, 1, \bar{\alpha}, 5\}$. The rows indicate which procedure was used. Each cell contains five boxplots of the distribution of the corresponding variance estimators over the simulations for the sample sizes $N \in \{100, 250, 500, 1\,000, 2\,000\}$ from top to bottom.

controls, with the vertical line indicating $1 + \frac{3}{2}\alpha$, the value of the limit variance. For every procedure and α the results are presented as a block of boxplots of the variance estimates over the simulations. The boxplots within each cell correspond to the varying sample sizes considered, with the one for the smallest sample size $N = 100$ at the top and those for the subsequent larger sample sizes below. As we increase the sample size, the variability of all the resampling based variance estimators decreases. Furthermore, we see that decreasing γ and thus the degree of undersampling also leads to a reduction in variability of the variance estimators.

From the first row, we see that the naive Efron-type bootstrap variance is asymptotically unbiased for $\alpha = \bar{\alpha} \approx 2.84$, but is on average too large for $\alpha < \bar{\alpha}$ and too small for $\alpha > \bar{\alpha}$. From the results for our direct M-out-of-N-type bootstrap variance estimators we see that in the $\alpha = \bar{\alpha}$ case all the procedures work well for the largest sample size and only for $\gamma = 0.5$ do we see that the limit variance is underestimated on average for the smaller sample sizes. When $\alpha > \bar{\alpha}$, so that we have "fewer" controls than treated, our variance estimator works well for γ large enough implying that we need to make sure that our bootstrap sample size M is large enough, but still less than N. In contrast, when $\alpha < \bar{\alpha}$, we see that our variance estimator works well for γ not too large. In our particular simulation setup, the choice of $\gamma = 0.6$ and $\gamma = 0.7$ worked well across all settings. Naturally, this merely says that for our specific setting these choices worked well. Finding a data dependent choice of γ and showing that this works well in practice remains an open question. One possibility may be to adapt the data-driven rule for the i.i.d. setting proposed in Bickel and Sakov (2008), although this is by no means straightforward.

4.2. Performance of bootstrap-based confidence intervals

In typical situations variance estimators are not of interest per se. Rather they are used to construct confidence intervals for the parameter of interest. Given the known asymptotic normality for the matching estimator as derived in Abadie and Imbens (2006) replacing the asymptotic variance of the matching estimators by the asymptotically unbiased M-out-of-N-type bootstrap variance estimators should yield confidence intervals with the correct coverage. To investigate this, Table 1 collects coverage probabilities for the asymptotically motivated Gaussian based 95% confidence intervals for τ^t given by

$$CI_{M,0.95}(\tau^t; \alpha; N) = \left[\widehat{\tau}^t(\alpha; N) - 1.96 \cdot \frac{\sqrt{\widehat{v}^*_M(\alpha; N)}}{\sqrt{N_1}}, \widehat{\tau}^t(\alpha; N) + 1.96 \cdot \frac{\sqrt{\widehat{v}^*_M(\alpha; N)}}{\sqrt{N_1}} \right], \tag{8}$$

where $\widehat{v}^*_M(\alpha; N)$ denotes the M-out-of-N-type bootstrap-based variance estimator in (7).

Table 1
Coverage probabilities of 95% confidence intervals for τ^t given in (8) using resampling based variance estimates over the simulations. The panels correspond to simulations with $\alpha \in \{0.2, 1, \bar{\alpha}, 5\}$. Within each panel each row corresponds to a sample size $N \in \{100, 250, 500, 1\,000, 2\,000\}$. Each column contains the coverage probabilities using resampling based variance estimators via the stated procedure. The bold number corresponds to the procedure that was closest to the asymptotic 95% coverage within each row.

		M-out-of-N (MooN)					Efron
	γ	0.5	0.6	0.7	0.8	0.9	1
α	N						
0.2	100	**0.982**	0.991	0.990	0.996	0.997	0.995
	250	**0.976**	0.984	0.991	0.996	0.997	0.995
	500	**0.965**	0.971	0.981	0.991	0.996	0.992
	1000	**0.957**	0.973	0.987	0.993	0.994	0.994
	2000	**0.947**	0.956	0.973	0.990	0.995	0.996
1	100	**0.953**	0.966	0.970	0.981	0.987	0.977
	250	**0.954**	0.969	0.976	0.982	0.981	0.974
	500	**0.953**	0.966	0.971	0.983	0.985	0.981
	1000	**0.958**	0.968	0.975	0.986	0.985	0.979
	2000	**0.946**	0.955	0.965	0.974	0.978	0.970
$\bar{\alpha}$	100	0.900	**0.948**	0.940	0.960	0.961	0.946
	250	0.948	0.954	0.971	0.963	0.964	**0.951**
	500	0.933	**0.950**	0.958	0.967	0.960	0.953
	1000	**0.951**	0.959	0.958	0.967	0.969	0.954
	2000	**0.945**	0.959	0.964	0.967	0.967	0.960
5	100	0.872	0.913	0.939	**0.940**	0.936	0.918
	250	0.891	0.915	0.936	**0.943**	0.942	0.924
	500	0.907	0.936	0.946	**0.951**	0.956	0.932
	1000	0.941	0.935	0.944	0.945	**0.948**	0.940
	2000	0.942	**0.957**	0.963	0.967	0.969	0.949

Each panel of Table 1 corresponds to a particular α. Each column corresponds to a particular resampling procedure. Each row contains the coverage probabilities for the stated sample size with the bold value the one closest to 95%. The results are in accordance with those in Figure 1: For $\alpha < \bar{\alpha}$, the variances tend to be overestimated for large γ, which directly results in confidence intervals that overcover. Analogously for $\alpha > \bar{\alpha}$, the variances tend to be underestimated for small γ and confidence intervals that undercover.

5. Conclusion

We have proposed a direct M-out-of-N-type bootstrap to estimate the variance of the nearest neighbor matching estimator for the ATET. Our direct M-out-of-N-type bootstrap resamples directly from the data *without* the need to bias-correct first. We are able to show that the direct M-out-of-N-type bootstrap variance estimator is asymptotically unbiased in the setting used to show the failure of the naive Efron-type bootstrap, thus, providing a proof of concept for our direct M-out-of-N-type bootstrap procedure. Extensions to more general settings are desirable and are left for future work as they are not immediate and it is not clear how our results can be carried over. The key to our proof is the fact that contrary to the naive Efron-type bootstrap, our direct M-out-of-N-type bootstrap resamples do not contain any ties asymptotically, which circumvents the problem of not replicating the matching process in the bootstrap. Finally, in a small simulation study we illustrate the potential of our M-out-of-N-type variance estimator and the associated 95% confidence intervals for the ATET. The simulations raise the additional important open question as to how the choice of the bootstrap sample size is to be performed in practice for the M-out-of-N-type bootstrap. To summarize, we have shown that the proposed M-out-of-N-type bootstrap does not fail for the example in Abadie and Imbens (2008), thus mereting its further investigation.

Acknowledgements

We are grateful for the comments and suggestions of two anonymous referees and an associate editor, which helped to improve the paper. We also thank Shaikh Tanvir Hossain for numerous helpful discussions in the early stages of the project. Christopher Walsh is funded by the German Research Foundation (DFG) through project 501082519 and gratefully acknowledges the financial support from the collaborative research center (SFB 823) “Statistical Modelling of Nonlinear dynamic Processes” of the German Science Foundation (DFG) during his time at TU Dortmund University. Financial support of the MERCUR project “Digitale Daten in der sozial- und wirtschaftswissenschaftlichen Forschung” is gratefully acknowledged by Carsten Jentsch. Finally, we are grateful for the computing time provided on the Linux HPC cluster at TU Dortmund University (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the German Research Foundation (DFG) as project 271512359 and for computing time under project p0020105 granted by the NHR4CES Resource Allocation Board and provided on the supercomputer CLAIX at RWTH Aachen University as part of the NHR4CES infrastructure.

Appendix A

A1. Proof of Theorem 1

The proof of Theorem 1 is immediate given the following three lemmas.

Lemma 2. *Given Assumption 1,*

$$\begin{aligned}\mathbb{E}[\mathbb{V}\mathrm{ar}^*[\sqrt{M_1}(\widehat{\tau}_M^{t,*}-\tau^t)^2 \mid \mathbf{Z}] \mid \mathbf{X},\mathbf{W}] &= \frac{N_0}{M_1}\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{X},\mathbf{W}] \\ &= \frac{N_0}{M_1}\mathbb{E}^*[(\sum_{j=1}^{N} W_j B_{M,i}(X_j)R_{b_M,j})^2 \mid \mathbf{X},\mathbf{W}] \\ &= \frac{N_0}{M_1}\sum_{j=1}^{N} W_j \mathbb{E}^*[B_{M,i}(X_j)R_{b_M,j}^2 \mid \mathbf{X},\mathbf{W}] \\ &+ \frac{N_0}{M_1}\sum_{j=1}^{N}\sum_{l\neq j} W_j W_l \mathbb{E}^*[B_{M,i}(X_j)B_{M,i}(X_l)R_{b_M,j}R_{b_M,l} \mid \mathbf{X},\mathbf{W}].\end{aligned}$$

Proof. The proof relies on steps similar to those in Abadie and Imbens (2008). Details are given in Section S.1.1 of the supplementary material. □

Lemma 3. *Given the assumptions in Theorem 1,*

$$\frac{N_0}{M_1}\sum_{j=1}^{N} W_j \mathbb{E}^*[B_{M,i}(X_j)R_{b_M,j}^2 \mid \mathbf{X},\mathbf{W}] = \left(\frac{N_1-1+M_1}{N_1}\right) \to 1.$$

Proof. The proof again relies on steps similar to those in Abadie and Imbens (2008). Details are given in Section S.1.2 of the supplementary material. □

Lemma 4. *Given the assumptions in Theorem 1,*

$$\frac{N_0}{M_1}\sum_{j=1}^{N}\sum_{l\neq j} W_j W_l \mathbb{E}^*[B_{M,i}(X_j)B_{M,i}(X_l)R_{b_M,j}R_{b_M,l} \mid \mathbf{X},\mathbf{W}] \to \frac{3}{2}\alpha.$$

Proof. The proof relies on properties specific to the M-out-of-N resampling, which result in bootstrap samples that asymptotically contain no ties with probability one. Details are given in Section S.1.3 of the supplementary material. □

A2. Proof of Remark 1

Following the arguments in Abadie and Imbens (2008) with $Y(1)$ as in the remark, immediately yields

$$\mathbb{V}\mathrm{ar}\left[\sqrt{N_1}(\widehat{\tau}^t-\tau^t)\right] \to \sigma^2 + 1 + \frac{3}{2}\alpha.$$

In contrast to the degenerate $Y(1)$ case, for the M-out-of-N bootstrap estimator, we now get

$$\widehat{\tau}_M^{t,*}-\tau^t = \frac{1}{M_1}\sum_{i=1}^{N} W_i R_{b_M,i}(Y_i(1)-\tau^t) - \frac{1}{M_1}\sum_{i=1}^{N}(1-W_i)K_{b_M,i}Y_i(0),$$

where we have used $\sum_{i=1}^{N} W_i R_{b_M,i} = M_1$. Using the auxiliary distributional results in Lemma 5(i) of Section S.2 of the supplementary material, and steps similar to those at the beginning of the proof of Lemma 2, one gets

$$\mathbb{E}[\mathbb{V}\mathrm{ar}^*[\sqrt{M_1}(\widehat{\tau}_M^{t,*}-\tau^t)^2 \mid \mathbf{Z}] \mid \mathbf{X},\mathbf{W}] = \left(1-\frac{1}{N_1}\right)\sigma^2 + \frac{N_0}{M_1}\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{X},\mathbf{W}].$$

The statement of Remark 1, then follows with Lemmas 2–4.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ecosta.2023.04.005

References

Abadie, A., Imbens, G.W., 2006. Large sample properties of matching estimators for average treatment effects. Econometrica 74 (1), 235–267.
Abadie, A., Imbens, G.W., 2008. On the failure of the bootstrap for matching estimators. Econometrica 76 (6), 1537–1557.
Abadie, A., Imbens, G.W., 2011. Bias-corrected matching estimators for average treatment effects. Journal of Business & Economic Statistics 29 (1), 1–11.
Abadie, A., Imbens, G.W., 2012. A martingale representation for matching estimators. Journal of the American Statistical Association 107 (498), 833–843.
Bickel, P.J., Götze, F., van Zwet, W.R., 1997. Resampling fewer than n observations: gains, losses and remedies for losses. Statistica Sinica 7 (1), 1–31.
Bickel, P.J., Sakov, A., 2008. On the choice of m in the m out of n bootstrap and confidence bounds for extrema. Statistica Sinica 18 (3), 967–985.
Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21 (4), 1926–1947.
Otsu, T., Rai, Y., 2017. Bootstrap inference of matching estimators for average treatment effects. Journal of the American Statistical Association 112 (520), 1720–1732.
Politis, D.N., Romano, J.P., 1994. Large sample confidence regions based on subsamples under minimal assumptions. Annals of Statistics 22 (4), 2031–2050.
Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer-Verlag, New York.
Walsh, C., Jentsch, C., Hossain, S.T., 2021. Weighted bootstrap consistency for matching estimators: The role of bias-correction. SFB 823 Discussion Paper. TU Dortmund University, Dortmund.