SFB 823

Nearest neighbor matching: Does the M-out-of-N bootstrap work when the naive bootstrap fails?

Christopher Walsh, Carsten Jentsch, Shaikh Tanvir Hossain

Discussion Paper

Nr. 5/2021

SFB 823

Nearest neighbor matching: Does the M-out-of-N bootstrap work when the naive bootstrap fails?*

Christopher Walsh[†,‡] Carsten Jentsch[†,§] Shaikh Tanvir Hossain[†,§]

Abstract

In a seminal paper Abadie and Imbens (2008) showed that the limiting variance of the classical nearest neighbor matching estimator cannot be consistently estimated by a naive Efron-type bootstrap. Specifically, they show that the conditional variance of the Efron-type boostrap estimator does not converge to the correct limit in expectation. In essence this is due to drawing with replacement such that original observations appear more than once in the bootstrap sample with positive probability even when the sample size becomes large. In the same paper, it is conjectured that the limiting variance should be consistently estimable by an M-out-of-N bootstrap. Here, we prove that the conditional variance of an M-out-of-N-type bootstrap estimator does indeed converge to the correct limit in expectation in the setting considered in Abadie and Imbens (2008). The key to the proof lies in the fact that asymptotically the M-out-of-N-type bootstrap sample does not contain any observations more than once with probability one. The finite sample performance of the M-out-of-N-type bootstrap is investigated in a simulation study of the DGP considered by Abadie and Imbens (2008).

Key words: ATET, Matching Estimator, M-out-of-N Bootstrap. **JEL codes:** C14, C21.

*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.

[†]TU Dortmund University. Department of Statistics.

[‡]Financial support from the collaborative research center (SFB 823) "Statistical Modelling of Nonlinear dynamic Processes" of German Science Foundation (DFG) is gratefully acknowledged. Corresponding Author. Email: walsh@statistik.tu-dortmund.de

[§]Financial support of the MERCUR project "Digitale Daten in der sozial- und wirtschaftswissenschaftlichen Forschung" is gratefully acknowledged.

1 Introduction

Matching estimators are intuitively simple procedures to estimate average treatment effects within the potential outcomes framework. Asymptotic properties of these estimators were established in Abadie and Imbens (2006, 2011, 2012). The expression for the variance of the asymptotic distribution is typically seen to depend on numerous nuisance parameters. Besides possible finite sample improvements, the need to estimate the nuisance parameters motivates the desire to apply resampling based procedures to estimate the asymptotic variance in the mataching 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 showed this by considering a very simple data generating process (DGP) for which it is clear that the nearest neighbor matching estimator for the average treatment effect of the treated (ATET) will be $\sqrt{N_1}$-consistent and asymptotically normal, where N_1 is the number of treated units. In their setting, the matching estimator is unbiased, thus avoiding the necessity to look at bias-corrected versions of the estimator. Given their postulated DGP they obtain closed form expressions for the limiting variance of the nearest neighbor matching estimator for the ATET as well as for the limit of the expectation of the conditional variance of an Efron-type bootstrap estimator, which is based on separately sampling with replacement from the control group and from the treatment group. As the two limit expressions do not coincide, their simple setting is a counterexample showing that the considered 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.

In addition to this negative result on the validity of Efron-type resampling procedures, Abadie and Imbens (2008) conjecture that using either the wild bootstrap of Härdle and Mammen (1993) or the M-out-of-N bootstrap (Bickel et al. (1997)) should correctly estimate the limiting distribution of nearest neighbor matching estimators. Recently, Otsu and Rai (2017) proposed and proved the validity of a weighted bootstrap procedure that can be interpreted as a wild bootstrap. In contrast to their procedure, which does bias correction before resampling, the present paper investigates an M-out-of-N-type bootstrap estimator that directly resamples the original data. We will prove that in the setting considered in Abadie and Imbens (2008) the conditional variance of this M-out-of-N-type bootstrap estimator is an asymptotically unbiased estimator of the limit variance if the resample size satisfies $M = o(N^{1/2})$. The proof is very technical. The key result is that, whereas the Efron-type bootstrap samples even asymptotically contain ties, thus leading to the failure to correctly replicate the matching procedure, the M-out-of-N-type bootstrap samples will not contain any ties asymptotically.

The finite sample performance of our procedure is investigated in a simulation study based on the DGP used in Abadie and Imbens (2008) for various sample sizes and degrees of balancedness between the size of the treated group and the size of the control group. The simulations show, as would be expected, that the performance of the M-out-of-N-type bootstrap conditional variance estimator depends on the choice of the resample size. Moreover, it is seen that the (ex-post) best choice depends not only on the sample size but on the interplay between the sample size and the degree of balancedness.

The remainder of the paper is structured as follows. Section 2 provides the basic treatment effect setup considered in 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. Our proposed M-out-of-N-type bootstrap procedure is presented in Section 3 along with the main theoretical result showing that this procedure 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 simulation study to illustrate the behavior of the M-out-of-N-type bootstrap variance estimator and the dependence of its performance on the interplay between the balancedness of the design and the resampling size are collected in Section 4. Section 5 concludes. The details of the proof are collected in the appendix at the end of the manuscript.

2 Setup in Abadie and Imbens (2008)

We consider the basic binary treatment effects setup. For each unit $i = 1, \ldots, N$ let $Y_i(0)$ and $Y_i(1)$ be the potential outcomes under control and after treatment, respectively. 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

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

where $N_1 = \sum_{i=1}^N W_i$ is the number of treated units and $\hat{Y}_i(0) = Y_{j(i)}$ 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

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

In order to allow for the possibility that there is not a unique closest unit, as indeed will be the case in the bootstrap samples, one can write $\hat{Y}_i(0) = \frac{1}{\#\mathcal{J}(i)} \sum_{j \in \mathcal{J}(i)} Y_j$, with $\#\mathcal{J}(i)$ the number of closest units to treated unit i and $\mathcal{J}(i) = \arg\min_{j \in \{1,\ldots,N\}: W_j = 0} ||X_j - X_i||$, the set of indices of those closest control units. With $K_j = \sum_{i=1}^N W_i \frac{I\{j \in \mathcal{J}(i)\}}{\#\mathcal{J}(i)}$ the number of times (control) unit j is a match to a treated unit, we can rewrite the matching estimator as

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

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

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 for the bootstrap sample

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

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

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 when matching in the bootstrap sample. Thus, in the bootstrap world there can be multiple bootstrap control units that get matched to the bootstrap treated unit i and the set of indices $\mathcal{J}^*(i)$ will not necessarily be a singleton – as was the case when matching in the original data.

Abadie and Imbens (2008) analysed the asymptotic behavior of the nearest neighbor matching estimator for the ATET, $\hat{\tau}^t$, and its naive Efron-type bootstrap estimator, $\hat{\tau}^{t,*}$, under the simple DGP given in Assumption 1 that allowed to explicitly calculate the limiting variance of the matching estimator $\sqrt{N_1}(\hat{\tau}^t - \tau^t)$ and the probability limit of $\mathbb{V}\mathrm{ar}^*[\sqrt{N_1}\hat{\tau}^{t,*} \mid \mathbf{Z}]$, where $\mathbb{V}\mathrm{ar}^*[\,\cdot \mid \mathbf{Z}]$ denotes the variance over the resampling mechanism conditional on the data.

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* $\alpha = N_1/N_0$*.*

(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* $x \in [0, 1]$*.*

In particular, Abadie and Imbens (2008) showed that in the setting of Assumption 1, the Efron-type bootstrap estimator is invalid as $\mathbb{V}\mathrm{ar}^*[\sqrt{N_1}\hat{\tau}^{t,*} \mid \mathbf{Z}]$ is an asymptotically *biased* estimator of the limiting variance except for the case that $\alpha = \frac{4(1-e^{-1})e^{-1}}{3-8e^{-1}+2e^{-2}}$. This is easily seen from Lemma 1. Part (i) states the limit distribution of $\hat{\tau}^t$ including the simple formula for the asymptotic variance; and part (ii) provides the limit for the expectation of the conditional variance the Efron-type bootstrap estimator.

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

(i) $\sqrt{N_1}(\hat{\tau}^t - \tau^t) \xrightarrow{d} \mathcal{N}(0, 1 + \frac{3}{2}\alpha)$.

(ii) $\mathbb{E}\big[\mathbb{V}\text{ar}^*[\sqrt{N_1}\hat{\tau}^{t,*} \mid \mathbf{Z}]\big] \longrightarrow 1 + \frac{3}{2}\alpha\frac{5e^{-1}-2e^{-2}}{3(1-e^{-1})} + 2e^{-1}$

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

3 The M-out-of-N-type boostrap

Next, we propose our M-out-of-N-type bootstrap procedure. The main idea is that rather than resampling N_1 treated and N_0 controls as in the Efron-type bootstrap we will resample $M_1 < N_1$ treated units and $M_0 < N_0$ control units. There are a lot of different feasible 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 $M_1/M_0 = N_1/N_0 = \alpha$. Denoting by $M = M_0 + M_1$ the total number of resampled units requiring that the balancedness between the treated and control groups is as in the original data implies that $M_0 = 1/(1+\alpha)M$ and $M_1 = \alpha/(1+\alpha)M$. As these last two equations will generally lead to non-integer M_1 and M_0, we will have to take interger parts in the algorithm. Our proposed M-out-of-N-type bootstrap matching estimator for a given $\alpha = N_1/N_0$ is construced as follows:

Step 1: Fix a $\gamma \in (0,1)$ large enough to ensure that $\lfloor 1/(1+\alpha)N^\gamma \rfloor$ and $\lfloor \alpha/(1+\alpha)N^\gamma \rfloor$ are non-zero, where $\lfloor x \rfloor$ is the largest integer less than or equal to x.

Step 2: Split the sample $\mathbf{Z} = \{(Y_i, W_i, X_i')\}_{i=1}^N$ into the treatment and the control group. Sample with replacement $M_0 = \lfloor 1/(1+\alpha)N^\gamma \rfloor$ units from the control group and $M_1 = \lfloor \alpha/(1+\alpha)N^\gamma \rfloor$ units from the treated group. Combine the two sample 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 for the bootstrap sample

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

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

Step 1 of the procedure ensures that we have at least one observation in each treatment arm of the bootstrap sample. The choosen γ captures the degree of undersampling as $M \approx \lfloor N^\gamma \rfloor$. 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, we must again allow for ties when doing the matching in the bootstrap world. However, in contrast to the naive Efron-type bootstrap, it will be seen that there are no ties asymptotically, which ensures that the variance estimator $\mathbb{V}\text{ar}^*[\sqrt{M_1}\hat{\tau}_M^{t,*} \mid \mathbf{Z}]$ is asymptotically unbiased in the setting considered by Abadie and Imbens (2008). The formal result is given in Theorem 1.

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

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

showing that the conditional variance of the M-out-of-N-type bootstrap estimator is an asymptotically unbiased estimator of the limiting variance of $\sqrt{N_1}(\hat{\tau}^t - \tau^t)$ given in Lemma 1(i).

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

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

and the last term is asymptotically negligible we can concentrate on the leading term $\mathbb{V}\text{ar}^*[\sqrt{M_1}(\hat{\tau}_M^{t,*} - \tau^t) \mid \mathbf{Z}]$. From the assumptions on the DGP in Assumption 1, the definition of the M-out-of-N-type bootstrap estimator in (3) and the 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

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

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 if 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 boostrap 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 M-out-of-N-type bootstrap estimator in (4) can be written as

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

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 $\frac{N_0}{M_1}\mathbb{E}\big[\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{Z}]\big] \to 1 + \frac{3}{2}\alpha$. The details are given in Appendix A. $\square$

If we replaced M by N (and M_1 by N_1) in (3), the definition of our bootstrap estimator then we would get the naive Efron-type bootstrap estimator, that is $\hat{\tau}_N^{t,*} = \hat{\tau}^{t,*}$. Similarly, replacing the subscript M, in $K_{b_M,i}$,$B_{M,i}$ and $R_{b_M,i}$ by N when considering these entities for the naive Efron-type bootstrap estimator it is possible to show using results from Abadie and Imbens (2008) that the expectation of the conditional variance of the Efron-type bootstrap estimator can be written as

$$\mathbb{E}\left[\mathbb{V}\mathrm{ar}^*[\sqrt{N_1}(\hat{\tau}^{t,*} - \hat{\tau})^2 \mid \mathbf{Z}]\right] = N_1\mathbb{V}\mathrm{ar}[\hat{\tau}^t] - 2\frac{1}{\alpha}\mathbb{E}[K_i\mathbb{E}^*[K_{b_N,i} \mid \mathbf{Z}]] + \frac{1}{\alpha}\mathbb{E}[\mathbb{E}^*[K_{b_N,i}^2 \mid \mathbf{Z}]].$$

In contrast to this equation, the expression in (5) is not of the form "target + error". Secondly, (5) does not contain any terms of the form $\mathbb{E}\big[K_i\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{Z}]\big]$, which are quite complicated to handle. Finally, the arguments used to establish the limit of $\frac{N_0}{M_1}\mathbb{E}\big[\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{Z}]\big]$ are substantially different to those used by Abadie and Imbens (2008) to analyse $\frac{N_0}{N_1}\mathbb{E}[\mathbb{E}^*\big[K_{b_N,i}^2 \mid \mathbf{Z}]\big]$. In particular, their arguments cannot be directly applied here. Instead, we make use of a degeneracy result for the occupation distribution in the low intensity sampling scheme. This degeneracy result ensures that asymptotically the M-out-of-N-type bootstrap samples contain only distinct units with probability one. This in turn allows for the matching mechanism to be replicated correctly asymptotically. The difficulty to replicate the matching mechanism correctly is due to the fact that even though the covariate is continuously distributed in the population, the distribution of the covariate in the bootstrap sample is discrete. Thus, when matching in the bootstrap sample, there will be ties. In the Efron-type bootstrap in Abadie and Imbens (2008) the reason for the failure is that, with positive probability, control units are used more than once even as the sample size increases. Thus the matching mechanism is not replicated correctly by the Efron-type bootstrap. In contrast, as just discussed the M-out-of-N-type bootstrap samples will asymptotically not contain any ties.

4 Simulations

In order to investigate the finite sample behavior of the M-out-of-N-type bootstrap estimator $\hat{\tau}_M^{t,*}$ (given in (3)) for different choices of resample size M under the DGP in Assumption 1 we conduct an extensive simulation study to see how well the limit variance of the nearest neighbor matching estimator is approximated by the conditional variance of the resampling estimators when varying the sample size ($N = N_0 + N_1$) and the degree of balancedness between the treated and control groups measured by $\alpha = N_1/N_0$. The simulations will re-affirm that the naive Efron-type bootstrap estimator is invalid except when $N_1/N_0 = \bar{\alpha} \approx 2.84$. We will also see that although as shown in Theorem 1 the conditional variance of the M-out-of-N-type bootstrap estimator is an asymptotically unbiased estimator of the limit variance of the nearest neighbor matching estimator for the ATET, its behaviour is very much dependent on the interplay between the balancedness of the number of treated and control units and the degree of undersampling as measured by γ, where $M \approx \lfloor N^\gamma \rfloor$ denotes the total number of units in the bootstrap sample.

The simulations were based on $S = 10\,000$ simulation runs. We considered sample sizes of $N \in \{100, 250, 500, 1\,000, 2\,000\}$. The balancedness of the design was varied by considering $\alpha \in \{10, 5, \bar{\alpha}, 2, 1, 0.5, 0.2, 0.1\}$. In each simulation run we simulated data sets according to the DGP given in Assumption 1 for all the considered values of N and α with τ^t fixed at a value of 1 throughout. For each simulated data set, we calculated the nearest neighbor matching estimator $\hat{\tau}^t$ (given in (1)). We also computed the naive Efron-type estimator $\hat{\tau}^{t,*}$ (according to (2)) and the M-out-of-N-type bootstrap estimator $\hat{\tau}_M^{t,*}$ (as in (3)) for various choices of $\gamma \in \{0.4, 0.5, 0.6, 0.7\}$. The bootstrap estimators were recomputed using $B = 1\,000$ bootstrap resamples.

Table 1: Exact variance of $\sqrt{N_1}(\hat{\tau}^t - \tau^t)$ as given in (6) for various N and α. The last row is $1 + \frac{3}{2}\alpha$, which is the value of the limiting variance.

N\\α	10	5	$\bar{\alpha}$	2	1	0.5	0.2	0.1
100	13.81	8.07	4.95	3.84	2.46	1.71	1.27	1.13
250	15.50	8.33	5.14	3.93	2.48	1.74	1.29	1.14
500	15.66	8.38	5.20	3.98	2.49	1.74	1.29	1.14
1000	15.91	8.46	5.24	3.98	2.50	1.75	1.30	1.15
2000	15.95	8.47	5.25	3.99	2.50	1.75	1.30	1.15
INF	16.00	8.50	5.26	4.00	2.50	1.75	1.30	1.15

4.1 The target

Interest lies in seeing how well the resampling procedures estimate the (limiting) variance of the nearest neighbor matching estimator for the ATET. From Lemma 1(i), we see that the limiting variance of the matching estimator is given by $1 + \frac{3}{2}\alpha$. In fact, we could also investigate how well the finite sample variance of the matching estimator is approximated, because from Lemma 3.1(i) of Abadie and Imbens (2008) it follows that under the DGP of Assumption 1

$$\mathbb{V}\text{ar}\left[\sqrt{N_1}(\hat{\tau}^t - \tau^t)\right] = 1 + \frac{3}{2}\frac{(N_1 - 1)(N_0 + 8/3)}{(N_0 + 1)(N_0 + 2)}. \tag{6}$$

Table 1 provides the exact variance given in (6) for the choices of N and α considered in the simulations along with the value of the limiting variance for different α. Notice that the asymptotics bite earlier when α is small. This is to be expected as the randomness in the estimator under the DGP of Assumption 1 is entirely due to the imputation of the counterfactual outcome under no treatment, which will be better if there are relatively more controls, which corresponds to α being small. Note, that in settings where the asymptotic distribution is less accurate any resampling procedure will tend to perform worse. Before comparing the condtional variance estimators based on the resampling procedures and the corresponding finite sample and asymptotic target, it is insightful to see how well these targets are approximated by our simulations. Empirical moments of the estimators calculated over the simulations should be close to the population moments of the estimator. Thus, if we index each simulation run by $s = 1, \ldots, S$ and denote by $\hat{\tau}^t_s(N, \alpha)$ the nearest neighbor matching estimator based on the simulated data set with index s using a sample size of N and a degree of balancedness α, then

$$\mathbb{V}\mathrm{ar}_S[\sqrt{N_1}\hat{\tau}^t(N,\alpha)] := \frac{1}{S-1}\sum_{s=1}^{S}\left(\sqrt{N_1}\hat{\tau}_s^t(N,\alpha) - \frac{1}{S}\sum_{s=1}^{S}\sqrt{N_1}\hat{\tau}_s^t(N,\alpha)\right)^2 \tag{7}$$

should be close to the variance of $\sqrt{N_1}(\hat{\tau}(N,\alpha)^t - \tau^t)$, whose exact value is given in Table 1. The value of the empirical variance in (7) for the different values of N and α is given in Table 2. Comparing these with the values for the population variance given in Table 1, we see that the approximation using the empirical moments is in general very good, but especially for the smaller sample sizes there is still a small error due to the simulations, which we should bare in mind when determining if the variance estimators get close to the target.

Table 2: Emprical variance $\sqrt{N_1}(\hat{\tau}(N,\alpha)^t - \tau^t)$ over the $S = 10\,000$ simulations as in (7) for various N and α.

N\\α	10	5	$\bar{\alpha}$	2	1	0.5	0.2	0.1
100	13.83	8.05	5.02	3.86	2.47	1.68	1.29	1.15
250	15.72	8.36	5.16	3.85	2.38	1.75	1.26	1.12
500	15.87	8.27	5.13	3.95	2.42	1.70	1.27	1.13
1000	15.69	8.47	5.13	3.91	2.46	1.75	1.28	1.15
2000	15.99	8.29	5.15	4.01	2.49	1.76	1.32	1.13

4.2 Performance of bootstrap based variance estimators

Having established the appropriate target and seeing how well it is approximated in the simulations, we now investigate how well the target is estimated by the bootstrap based procedures. We will focus on determining how close the expectation of the conditional variance of the bootstrap estimators are to the target. Unlike the variance of the nearest neighbor matching estimator we only have asymptotic results (Theorem 1 and Lemma 1(ii)) and no exact expressions in finite samples. Hence, for a fixed sample size we have to approximate the expectation of the conditional variance of the resampling based procedure by calculating its empirical mean over the simulations. Let $\hat{v}(\alpha; N)$ be a generic estimator of the limiting variance of the nearest neighbor matching estimator based on N

observations and degree of balancedness α. Then the empirical mean of the variance estimator over the simulations is given by

$$\mathbb{E}_S[\hat{v}(\alpha; N)] := \frac{1}{S}\sum_{s=1}^{S} \hat{v}_s(\alpha; N), \tag{8}$$

where $\hat{v}_s(\alpha; N)$ is the corresponding variance estimator based on the simulated data set indexed by s. If the estimator $\hat{v}(\alpha; N)$ is asymptotically unbiased then as N tends to infinity it should hold that (up to simulation error) $\mathbb{E}_S[\hat{v}(\alpha; N)]$ converges to the limit variance. Naturally, for any given N we could compare $\mathbb{E}_S[\hat{v}(\alpha; N)]$ in (8) with the corresponding limit variance $1 + \frac{3}{2}\alpha$ to determine the sum of the asymptotic bias and the bias resulting from using an asymptotic approximation. Moreover, by comparing $\mathbb{E}_S[\hat{v}(\alpha; N)]$ for any given N with the corresponding exact variance of the matching estimator in Table 1 we could determine the finite sample bias of the variance estimator. However, as our focus lies in seeing whether the variance estimators are asymptotically unbiased we will report the results for the largest considered sample size $N = 2\,000$ only. Before showing the results, we need to specify how we constructed the variance estimators based on the bootstrap procedures. Let $\hat{\tau}^{t,*}_{s,b}(\alpha; N)$ be a generic bootstrap based estimator of the nearest neighbor matching estimator based on the simulated data set index by s with sample size N and degree of balancedness α, which is calculated using the bootstrap data set indexed by $b = 1, \ldots, B$. For this generic estimator, we define the estimator of the limiting variance by

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

Applying (8) to $(\hat{v}^*_s(\alpha; N), s = 1, \ldots, S)$ we get an estimate of the expectation of the conditional variance of the (generic) bootstrap procedure.

The results of the simulation study for $N = 2\,000$ are collected in Table 3. The first two rows state the exact and the limit variance of the matching estimator for different values of balancedness as previously given in the last two rows of Table 1. The next two lines contain the simulation based approximated expectation of the conditional variance of the naive Efron-type bootstrap estimator and its theoretical limit as given in Lemma 1(ii). We can clearly see that for $\alpha = \bar{\alpha}$, the variance estimator is asymptotically unbiased, whereas it is too large (small) on average if $\alpha < \bar{\alpha}$ ($\alpha > \bar{\alpha}$). The final block gives the simulation based approximated expectation of the conditional variance of our M-out-of-N-type bootstrap estimator for the different choices of $\gamma \in \{0.4, 0.5, 0.6, 0.7\}$. Recall that

Table 3: Results on $\mathbb{E}_S[\hat{v}(\alpha; N)]$ in (8) for various estimators given the largest considered sample size, $N = 2000$.

Est.\\α	10	5	$\bar{\alpha}$	2	1	0.5	0.2	0.1
Target	15.95	8.47	5.25	3.99	2.50	1.75	1.30	1.15
Asy.Target	16.00	8.50	5.26	4.00	2.50	1.75	1.30	1.15
Efron	14.00	7.88	5.23	4.21	2.97	2.35	1.98	1.85
Efron limit	14.14	7.94	5.26	4.22	2.98	2.36	1.98	1.86
MooN(0.4)	10.92	6.01	4.03	3.43	2.31	1.54	1.19	1.01
MooN(0.5)	14.02	7.27	4.79	3.78	2.45	1.68	1.28	1.15
MooN(0.6)	14.65	8.25	5.18	3.98	2.52	1.79	1.35	1.20
MooN(0.7)	15.53	8.62	5.33	4.18	2.68	1.93	1.48	1.31

γ governed the degree of undersampling by determining $M \approx \lfloor N^\gamma \rfloor$ the number of resampled units in the procedure. From Theorem 1, we know that the M-out-of-N-type bootstrap variance estimator is asymptotically unbiased provided the appropriate growth conditions for the resample size M are met. Not surprisingly, the results from the simulation study show that the performance of the M-out-of-N-type bootstrap variance estimator very much depends on the actual M chosen. In particular, when the resample size is too small the performance is very poor as seen for the results using $\gamma = 0.4$. Discarding this choice, we see that the performance of the remaining variance estimators in terms of bias very much depends on the interplay between α and γ. In particular, a larger choice of γ (and thus of M) seems to be better whenever α is larger. The interpretation for this finding is quite simple. Recall that the randomness in the estimator under the DGP of Assumption 1 is entirely due to the imputation of the counterfactual outcome under no treatment, which will be better if there are relatively more controls, which corresponds to α being small. Thus for small α, when there are many controls relative to treated in the original data, it is better to resample only a relatively small number of observations by taking a smaller γ as this minimizes the chance of having ties in the matching, whilst still ensuring that the bootstrap sample will have a sufficient number controls to do the imputation of the counterfactual well enough. In contrast when α is large we have only a few control units relative to treated in the original data. In this case we need to resample relatively

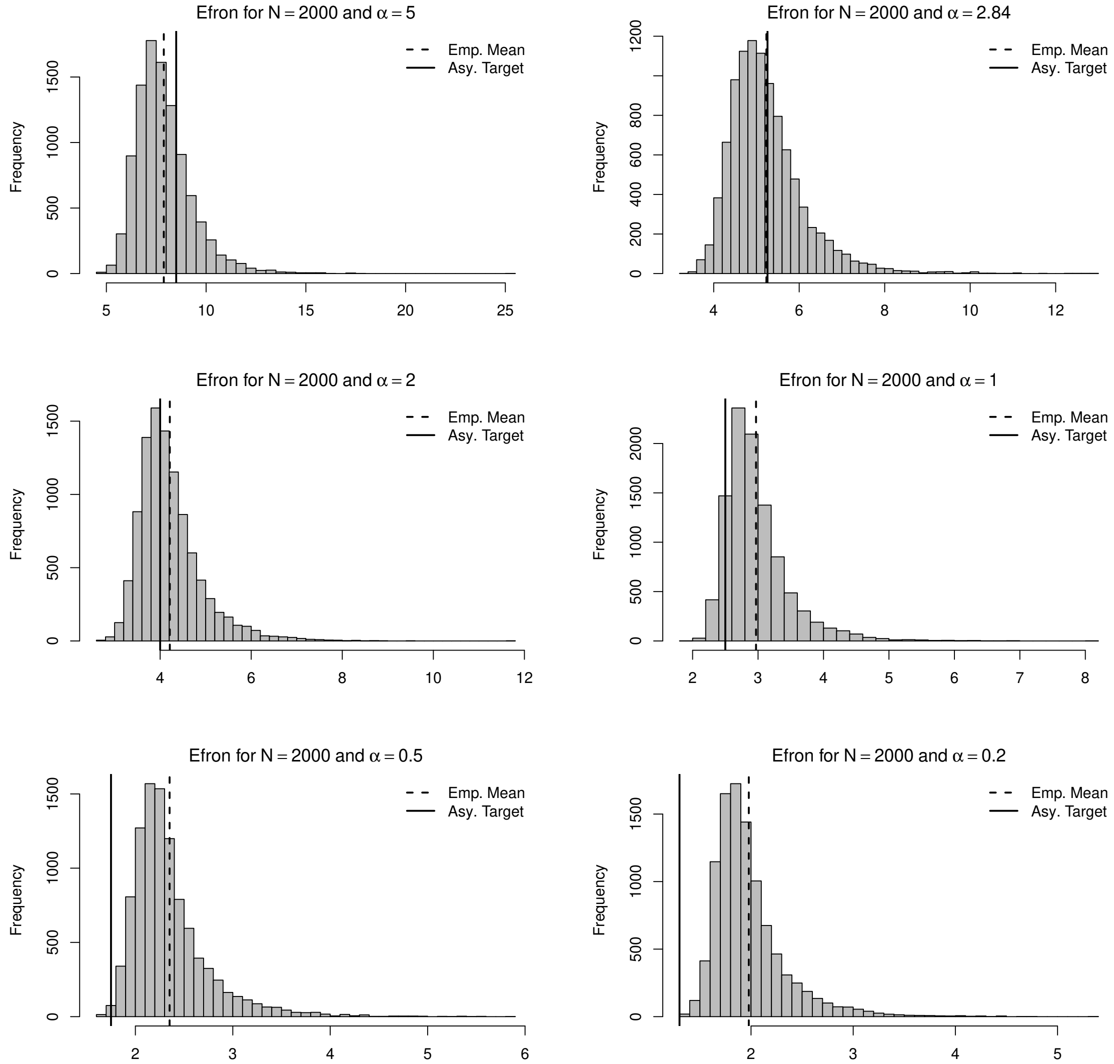

Figure 1: Histograms of naive Efron based bootstrap variance estimates with $\gamma = 0.6$ for $N = 2000$ and various α.

many observations by choosing a larger γ to make sure that there are sufficient controls to do the imputation of the counterfactual well enough. Finally, if we ignore the extreme balancedness settings $\alpha \in \{0.1, 10\}$, then for our simulation setting a choice of $\gamma = 0.6$ provides satisfactory results.

Finally, restricting attention to the non-extreme balancedness settings with $\alpha \in \{5, \bar{\alpha}, 2, 1, 0.5, 0.2\}$, we compare the distribution over the $S = 10\,000$ simulations of the variance estimates based on the naive Efron-type bootstrap and our M-out-of-N-type bootstrap with $\gamma = 0.6$, that was seen to produce satistfactory results. The distributional results for the Efron-type bootstrap are given in Figure 1 for the largest sample size $N = 2\,000$. Each panel shows the distribution of the variance estimates in terms of a histogram for a particular value of balancedness between the treated and control. The solid vertical line corresponds to the asymptotic target of $1 + \frac{3}{2}\alpha$. The dashed vertical line corresponds to the simulation based expectation of the variance estimator. Again, we can clearly see that for $\alpha = \bar{\alpha}$, the variance estimator is asymptotically unbiased, whereas it is too large (small) on average if $\alpha < \bar{\alpha}$ ($\alpha > \bar{\alpha}$). Furthermore, the distribution of the variance estimates is quite clearly positively skewed.

In Figure 2, we have plotted the corresponding results on the distribution of the variance estimates based on our M-out-of-N-type bootstrap with $\gamma = 0.6$. To ease the comparison, we have plotted the histograms on the same scale as those for the Efron-type estimator. As previously seen, the simulation based expectation of the variance estimator is much closer to the asymptotic target. Moreover, we can see that the distribution of the variance estimates is considerably less dispersed and quite substantially less skewed than for the Efron-type estimates.

5 Conclusion

We have proposed an M-out-of-N-type bootstrap on the data to estimate the variance of the nearest neighbor matching estimator for the ATET. We have also shown that in the counterexample DGP used by Abadie and Imbens (2008) to demonstrate that the naive Efron-type based variance estimator is asymptotically biased the variance estimator based on our M-out-of-N-type bootstrap estimator is in fact asymptotically unbiased. In simulations based on the DGP considered by Abadie and Imbens (2008) it was seen that the actual performance of the M-out-of-N-type bootstrap variance estimator heavily depends on the choice of the bootstrap sample size. In particular, we saw that bootstrap sample sizes need to be chosen larger when there are relatively few control units in the original data. Thus, although our result provides evidence for the conjecture made in Abadie and Imbens (2008) in the setting they considered, it also highlights how delicate the actual performance is with respect to the chosen bootstrap sample size. Sadly, all we can say concerning the choice of M is that for the present setting unless the degree of balancedness is extreme, one should get satisfactory results by choosing the bootstrap sample size according to $\gamma = 0.6$ in our procedure. Naturally, the next steps are to find a good data driven choice for the bootstrap sample size M as well as

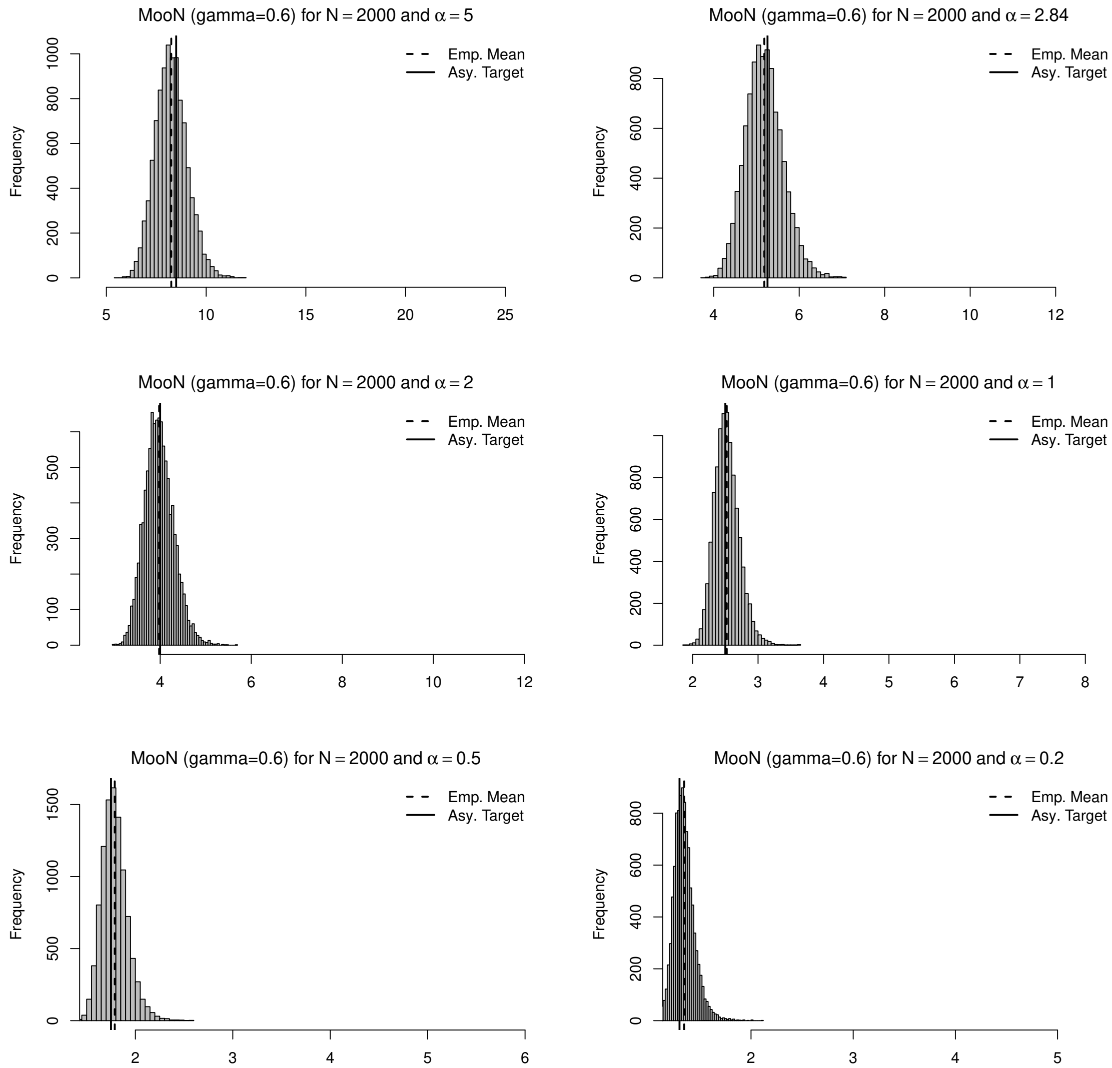

Figure 2: Histograms of M-out-of-N-type bootstrap variance estimates with $\gamma = 0.6$ for $N = 2000$ and various α.

establishing the theoretical results in more general settings. Nonetheless, the current paper provides in essence a proof of concept that using the M-out-of-N-type bootstrap for matching estimators of the ATET works. Moreover, the proof of the theoretical result also highlights that the key is that the bootstrap samples do not contain any ties asymptotically, which avoids the problem of not replicating the matching process in the bootstrap sample as was the case for the naive Efron-type bootstrap. Finally, although not of immediate practical use, the intuition for requiring a larger bootstrap sample when there are relatively few controls in the original data should nonetheless carry over to more general settings.

A Proof of Theorem 1

Using the clever and unorthodox notation introduced by Abadie and Imbens (2008) it was seen in the main text that

$$\hat{\tau}_M^{t,*} = \tau^t - \frac{1}{M_1}\sum_{i=1}^{N}(1-W_i)K_{b_M,i}Y_i(0) \tag{9}$$

with $K_{b_M,i} = \sum_{j=1}^{N} W_j B_{M,i}(X_j)R_{b_M,j}$ and its constituents defined in the main text. Moreover, as argued in the main text, we can concentrate on $\mathbb{V}\text{ar}^*[\sqrt{M_1}(\hat{\tau}_M^{t,*} - \tau^t) \mid \mathbf{Z}]$. Rearranging (9), we obtain

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

and from this with $\mathbf{Z} = (\mathbf{Y}, \mathbf{X}_0, \mathbf{X}_1, \mathbf{W})$, where $\mathbf{X}_0$ collect the covariate values for the controls in the sample and $\mathbf{X}_1$ for the treated, we get upon taking expectations over the resampling that

$$\mathbb{E}^*[\tau^t - \hat{\tau}_M^{t,*} \mid \mathbf{Z}] = \frac{1}{M_1}\sum_{i=1}^{N}(1-W_i)\mathbb{E}^*[K_{b_M,i} \mid \mathbf{Z}]Y_i(0).$$

Integrating out $\mathbf{Y}$ and using the fact that $\mathbb{E}^*[K_{b_M,i} \mid \mathbf{Z}] = \mathbb{E}^*[K_{b_M,i} \mid \mathbf{X}_0, \mathbf{X}_1, \mathbf{W}]$, we get

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}^*[\tau^t - \hat{\tau}_M^{t,*} \mid \mathbf{Z}] \mid \mathbf{X}_0, \mathbf{X}_1, \mathbf{W}] \\
&= \mathbb{E}[\frac{1}{M_1}\sum_{i=1}^{N}(1-W_i)\mathbb{E}^*[K_{b_M,i} \mid \mathbf{Z}]Y_i(0) \mid \mathbf{X}_0, \mathbf{X}_1, \mathbf{W}] \\
&= \frac{1}{M_1}\sum_{i=1}^{N}(1-W_i)\mathbb{E}^*[K_{b_M,i} \mid \mathbf{Z}]\mathbb{E}[Y_i(0) \mid \mathbf{X}_0, \mathbf{X}_1, \mathbf{W}] \\
&= 0.
\end{aligned}$$

Thus, $\mathbb{E}\big[\mathbb{E}^*[\sqrt{M_1}(\hat{\tau}_M^{t,*} - \tau^t) \mid \mathbf{Z}]\big] = 0$.

Similarly, we get

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}^*[(\hat{\tau}_M^{t,*} - \tau^t)^2 \mid \mathbf{Z}] \mid \mathbf{X}_0, \mathbf{X}_1, \mathbf{W}] \\
&= \mathbb{E}[\mathbb{E}^*[\frac{1}{M_1^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(1-W_i)(1-W_j)K_{b_M,i}K_{b_M,j}Y_i(0)Y_j(0) \mid \mathbf{Z}] \mid \mathbf{X}_0, \mathbf{X}_1, \mathbf{W}] \\
&= \frac{1}{M_1^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(1-W_i)(1-W_j)\mathbb{E}^*[K_{b_M,i}K_{b_M,j} \mid \mathbf{Z}]\mathbb{E}[Y_i(0)Y_j(0)) \mid \mathbf{X}_0, \mathbf{X}_1, \mathbf{W}] \\
&= \frac{1}{M_1^2}\sum_{i=1}^{N}(1-W_i)\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{Z}] \\
&= \frac{N_0}{M_1^2}\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{X}_0, \mathbf{X}_1, \mathbf{W}]
\end{aligned}$$

$$= \frac{N_0}{M_1^2}\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{Z}^+],$$

where in the last line we have defined $\mathbf{Z}^+ := (\mathbf{X}_0, \mathbf{X}_1, \mathbf{W})$ to lighten notation. From the above we get

$$\mathbb{E}[\mathbb{V}\text{ar}^*[\sqrt{M_1}(\hat{\tau}_M^{t,*} - \tau^t)^2 \mid \mathbf{Z}] \mid \mathbf{X}_0, \mathbf{X}_1, \mathbf{W}] = \frac{N_0}{M_1}\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{Z}^+].$$

A.1 Derivation of limit of $\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{Z}^+]$

Plugging in the expression for $K_{b_M,i}$ we get

$$\begin{aligned}
\mathbb{E}^*[K_{b_M,i}^2 \mid \mathbf{Z}^+] &= \mathbb{E}^*[(\sum_{j=1}^{N} W_j B_{M,i}(X_j) R_{b_M,j})^2 \mid \mathbf{Z}^+] \\
&= \sum_{j=1}^{N} W_j \mathbb{E}^*[B_{M,i}(X_j) R_{b_M,j}^2 \mid \mathbf{Z}^+] \\
&+ \sum_{j=1}^{N}\sum_{l\neq j} W_j W_l \mathbb{E}^*[B_{M,i}(X_j) B_{M,i}(X_l) R_{b_M,j} R_{b_M,l} \mid \mathbf{Z}^+] \\
&=: (A1)_M + (A2)_M.
\end{aligned}$$

A.1.1 Derivation of limit of $(A1)_M$

For the summand in the term $(A1)_M$, we get, with $D_{b_M,i}$ the indicator of whether the (control) unit i is in the bootstrap sample or not, that

$$\begin{aligned}
\mathbb{E}^*[B_{M,i}(X_j) R_{b_M,j}^2 \mid \mathbf{Z}^+] &= \mathbb{E}^*[B_{M,i}(X_j) R_{b_M,j}^2 \mid \mathbf{Z}^+, D_{b_M,i} = 1] \mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+) \\
&= \mathbb{E}^*[R_{b_M,j}^2 \mid \mathbf{Z}^+, B_{M,i}(X_j) = 1, D_{b_M,i} = 1] \\
&\quad \times \mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1) \mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+),
\end{aligned}$$

where P^* denotes the probability function over the resampling mechanism.

Due to the (conditional) independence of $(R_{b_M,j}, R_{b_M,l})$ and $(D_{b_M,i}, B_{M,i}(X_j), B_{M,i}(X_l))$, we get from Lemma 2 (i) in Appendix B that

$$\begin{aligned}
&\mathbb{E}^*[R_{b_M,j}^2 \mid \mathbf{Z}^+, B_{M,i}(X_j) = 1, D_{b_M,i} = 1] = \mathbb{E}^*[R_{b_M,j}^2 \mid \mathbf{Z}] \\
&= \mathbb{V}\text{ar}^*[R_{b_M,j} \mid \mathbf{Z}] + (\mathbb{E}^*[R_{b_M,j} \mid \mathbf{Z}])^2 \\
&= M_1\left(\frac{1}{N_1}\left(1 - \frac{1}{N_1}\right)\right) + \left(\frac{M_1}{N_1}\right)^2 = \frac{M_1}{N_1}\left(\frac{N_1 - 1 + M_1}{N_1}\right)
\end{aligned}$$

$$\rightarrow 0.$$

Notice that this expression only depends on the number of treated observations N_1 and the resample size M_1, but not on $\mathbf{Z}^+$. Similarly,

$$\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+) = 1 - (1 - 1/N_0)^{M_0} =: \mathrm{P}^*(D_{b_M,i} = 1)$$

only depends on the number of control units N_0 and the resample size M_0, but not on the data $\mathbf{Z}^+$. To finish analysing $(A1)_M$, we need to deal with $\mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)$. Dealing with $\mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)$ directly is complicated. However, in analogy to the Efron case considered in Abadie and Imbens (2008), it is much simpler to deal with $\mathbb{E}\big[\mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1, N_{b_M,0}) \mid \mathbf{W}, D_{b_M,i} = 1, N_{b_M,0}\big]$, where $N_{b_M,0}$ denotes the distinct number of control units in the boostrap sample. By Bayes' Theorem and as $\mathrm{P}^*(N_{b_M,0} = n \mid \mathbf{Z}^+, D_{b_M,i} = 1) := \mathrm{P}^*(N_{b_M,0} = n \mid D_{b_M,i} = 1)$ does not depend on the data, we get

$$\begin{aligned}
&\mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)\\
&= \sum_{n=1}^{M_0} \mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1, N_{b_M,0} = n)\mathrm{P}^*(N_{b_M,0} = n \mid \mathbf{Z}^+, D_{b_M,i} = 1)\\
&= \sum_{n=1}^{M_0} \mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1, N_{b_M,0} = n)\mathrm{P}^*(N_{b_M,0} = n \mid D_{b_M,i} = 1)\\
&= \sum_{n=1}^{M_0} \mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1, N_{b_M,0} = n)\\
&\qquad\qquad \times \frac{\mathrm{P}^*(D_{b_M,i} = 1 \mid N_{b_M,0} = n)\mathrm{P}^*(N_{b_M,0} = n)}{\mathrm{P}^*(D_{b_M,i} = 1)}\\
&= \frac{1}{\mathrm{P}^*(D_{b_M,i} = 1)} \sum_{n=1}^{M_0} \mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1, N_{b_M,0} = n)\frac{n}{N_0}\mathrm{P}^*(N_{b_M,0} = n),
\end{aligned}$$

where we have used $\mathrm{P}^*(D_{b_M,i} = 1 \mid N_{b_M,0} = n) = \frac{n}{N_0}$ on the last line. Integrating out $\mathbf{X}_0$ and $\mathbf{X}_1$ allows us to use Lemma 2(iv) in Appendix B yielding

$$\begin{aligned}
&\mathbb{E}\big[\mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b,i} = 1) \mid \mathbf{W}, D_{b,i} = 1\big]\\
&\quad = \frac{1}{\mathrm{P}^*(D_{b_M,i} = 1)} \sum_{n=1}^{M_0} \mathbb{E}\big[\mathrm{P}^*(B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1, N_{b_M,0} = n) \mid \mathbf{W}, D_{b_M,i} = 1\big]\\
&\qquad \times \frac{n}{N_0}\mathrm{P}^*(N_{b_M,0} = n)\\
&\quad = \frac{1}{\mathrm{P}^*(D_{b_M,i} = 1)} \sum_{n=1}^{N_0} \frac{1}{n}\frac{n}{N_0}\mathrm{P}^*(N_{b_M,0} = n) = \frac{1}{N_0}\frac{1}{\mathrm{P}^*(D_{b_M,i} = 1)}
\end{aligned}$$

From these results, we get

$$\mathbb{E}[\mathbb{E}^*[B_{M,i}(X_j)R^2_{b_M,j} \mid \mathbf{Z}] \mid \mathbf{W}] = \frac{M_1}{N_1}\left(\frac{N_1 - 1 + M_1}{N_1}\right)\frac{1}{N_0\mathrm{P}^*(D_{b_M,i} = 1)}\mathrm{P}^*(D_{b_M,i} = 1)$$
$$= \frac{1}{N_0}\frac{M_1}{N_1}\left(\frac{N_1 - 1 + M_1}{N_1}\right),$$

from which it follows that

$$(A1)_M = \frac{N_1}{N_0}\frac{M_1}{N_1}\left(\frac{N_1 - 1 + M_1}{N_1}\right) = \frac{M_1}{N_0}\left(\frac{N_1 - 1 + M_1}{N_1}\right) \tag{10}$$

whose contribution to $\mathbb{E}[\mathbb{V}\mathrm{ar}^*[\sqrt{M_1}(\hat{\tau}^{t,*}_M - \tau)^2 \mid \mathbf{Z}] \mid \mathbf{W}]$ is $\frac{N_0}{M_1}\frac{M_1}{N_0}\left(\frac{N_1-1+M_1}{N_1}\right) \to 1$.

Thus, if we can show that the contribution to $\mathbb{E}[\mathbb{V}\mathrm{ar}^*[\sqrt{M_1}(\hat{\tau}^{t,*}_M - \tau)^2 \mid \mathbf{Z}] \mid \mathbf{W}]$ from the $(A2)_M$ term goes to $\frac{3}{2}\alpha$, then we are finished.

A.1.2 Derivation of limit of $(A2)_M$

In order to derive a limit expression for the term related to $(A2)_M$ note that

$$\mathbb{E}^*[B_{M,i}(X_j)B_{M,i}(X_l)R_{b_M,j}R_{b_M,l} \mid \mathbf{Z}^+]$$
$$= \mathbb{E}^*[R_{b_M,j}R_{b_M,l} \mid \mathbf{Z}^+, B_{M,i}(X_j) = 1, B_{M,i}(X_j) = 1, D_{b_M,i} = 1]$$
$$\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+).$$

Due to the (conditional) independence of $(R_{b_M,j}, R_{b_M,l})$ and $(D_{b_M,i}, B_{M,i}(X_j), B_{M,i}(X_l))$, we get

$$\mathbb{E}^*[R_{b_M,j}R_{b_M,l} \mid \mathbf{Z}^+, B_{M,i}(X_j) = 1, B_{M,i}(X_j) = 1, D_{b_M,i} = 1] = \mathbb{E}^*[R_{b_M,j}R_{b_M,l} \mid \mathbf{Z}^+]$$
$$= \mathbb{C}\mathrm{ov}^*[R_{b_M,j}, R_{b_M,l} \mid \mathbf{Z}^+] + \mathbb{E}^*[R_{b_M,j} \mid \mathbf{Z}^+]\mathbb{E}^*[R_{b_M,l} \mid \mathbf{Z}^+].$$

By Lemma 2 (i) of Appendix B, as the $\{R_{b_M,j} : W_j = 1\}$ are (conditionally) exchangeable, with $\sum_{j:W_j=1} R_{b_M,j} = M_1$, we get $\mathbb{E}^*[R_{b_M,j} \mid \mathbf{Z}^+] = \mathbb{E}^*[R_{b_M,l} \mid \mathbf{Z}^+] = \frac{M_1}{N_1}$. Furthermore, we get

$$\mathbb{V}\mathrm{ar}^*[R_{b_M,j}] = \mathbb{C}\mathrm{ov}^*[R_{b_M,j}, R_{b_M,j}] = \mathbb{C}\mathrm{ov}^*[M_1 - \sum_{l\neq j} R_{b_M,l}, R_{b_M,j}]$$
$$= -(N_1 - 1)\mathbb{C}\mathrm{ov}^*[R_{b_M,l}, R_{b_M,j}]$$

which upon rearranging yields $\mathbb{C}\mathrm{ov}^*[R_{b_M,j}, R_{b_M,l} \mid \mathbf{Z}^+] = -\mathbb{V}\mathrm{ar}^*[R_{b_M,j} \mid \mathbf{Z}^+]/(N_1 - 1) = -M_1/N_1^2$ for $j \neq l$. Thus,

$$\mathbb{E}^*[R_{b_M,j}R_{b_M,l} \mid \mathbf{Z}^+, B_{M,i}(X_j) = 1, B_{M,i}(X_l) = 1, D_{b_M,i} = 1] = \frac{M_1}{N_1}\left(\frac{M_1}{N_1} - \frac{1}{N_1}\right)$$

which does not depend on $\mathbf{Z}^+$. Notice, that from Lemma 2 (ii) of Appendix B, $\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+) = 1 - (1 - 1/N_0)^{M_0}$ also does not depend on the data $\mathbf{Z}^+$.

Next, let us analyse $\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)$.

Remark 1. *We cannot follow the proof of Abadie and Imbens (2008). To see why, note that by Lemma 2(vi) and (v) in Appendix B we get*

$$\mathbb{E}^*\left[\frac{N_{b_M,0}}{N_0}\right] \to 0$$

and

$$\frac{1}{N_0}\mathbb{V}\mathrm{ar}^*[N_{b_M,0}] \to 0.$$

From these it is clear that, in contrast to the standard bootstrap case, $\mathbb{E}[N_0/N_{b_M,0}]$ *will not converge. In the standard bootstrap case the limit of* $\mathbb{E}[N_0/N_{b,0}]$ *was derived in Lemma A.5 of Abadie and Imbens (2008) and was used to establish*

$$N_0^2\mathbb{E}[\mathrm{P}^*(B_i(X_j)B_i(X_l) = 1 \mid \mathbf{Z}, D_{b,i} = 1, N_{b,0}) \mid \mathbf{W}, D_{b,i} = 1, N_{b,0}] \xrightarrow{\mathrm{P}^*} \frac{3}{2}\left(\frac{1}{1 - \exp(-1)}\right)^2.$$

In order to deal with $\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_j) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+)$ note that

$$\begin{aligned}
&\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)\\
&= \sum_{n=1}^{M_0} \mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1, N_{b_M,0} = n)\mathrm{P}^*(N_{b_M,0} = n \mid \mathbf{Z}^+, D_{b_M,i} = 1)\\
&= \sum_{n=1}^{M_0} \mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1, N_{b_M,0} = n)\\
&\qquad \frac{\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+, N_{b_M,0} = n)\mathrm{P}^*(N_{b_M,0} = n)}{\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+)}\\
&= \frac{1}{\mathrm{P}^*(D_{b_M,i} = 1)} \sum_{n=1}^{M_0} \mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1, N_{b_M,0} = n)\\
&\qquad \times \frac{n}{N_0}\mathrm{P}^*(N_{b_M,0} = n)
\end{aligned}$$

where we have used $\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+, N_{b_M,0} = n) = n/N_0$ in the last line. Integrating out $\mathbf{X}_0$ and $\mathbf{X}_1$ allows us to use Lemma 2(v) of Appendix B yielding

$$\mathbb{E}\big[\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+) \mid \mathbf{W}, D_{b_M,i} = 1\big]$$

$$= \mathrm{P}^*(D_{b_M,i}=1)\mathbb{E}\big[\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l)=1 \mid \mathbf{Z}^+, D_{b_M,i}=1) \mid \mathbf{W}, D_{b_M,i}=1\big]$$
$$= \sum_{n=1}^{M_0} \frac{n}{N_0}\mathrm{P}^*(N_{b_M,0}=n)$$
$$\times\, \mathbb{E}\big[\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l)=1 \mid \mathbf{Z}^+, D_{b_M,i}=1, N_{b_M,0}=n) \mid \mathbf{W}, D_{b_M,i}=1, N_{b_M,0}=n\big]$$
$$= \sum_{n=1}^{M_0} \frac{n}{N_0}\mathrm{P}^*(N_{b_M,0}=n)\frac{3}{2}\frac{(n+8/3)}{n(n+1)(n+2)} = \frac{1}{N_0}\frac{3}{2}\sum_{n=1}^{M_0}\frac{(n+8/3)}{(n+1)(n+2)}\mathrm{P}^*(N_{b_M,0}=n)$$

Next notice, that by Lemma 2(iii) of Appendix B and the fact that $S(M_0, M_0) = 1$, where $S(m,k) = \left\{{m \atop k}\right\}$ denotes the Stirling number of the second kind, we get that

$$\begin{aligned}
\mathrm{P}^*(N_{b_M,0}=M_0) &= \binom{N_0}{N_0-M_0}\frac{M_0!}{N_0^{M_0}}S(M_0,M_0) \\
&= \frac{N_0!}{(N_0-M_0)!M_0!}\frac{M_0!}{N_0^{M_0}} \\
&= \frac{N_0(N_0-1)\cdots(N_0-M_0+1)}{N_0^{M_0}} \\
&= \frac{(N_0-1)}{N_0}\cdots\frac{(N_0-M_0+1)}{N_0}.
\end{aligned}$$

Note that $\mathrm{P}^*(N_{b_M,0}=M_0)$ is bounded from above by 1 and from below by

$$\left(\frac{N_0-M_0}{N_0}\right)^{M_0} = \left(1-\frac{M_0}{N_0}\right)^{M_0}\left(1-\frac{\frac{M_0^2}{N_0}}{M_0}\right)^{M_0} = \left(1-\frac{c_{M_0}}{M_0}\right)^{M_0},$$

where $c_{M_0} := \frac{M_0^2}{N_0}$. Hence, if $M_0 = o(\sqrt{N_0})$ holds, we have $c_{M_0}\to 0$ if $M_0\to\infty$ such that

$$\left(1-\frac{c_{M_0}}{M_0}\right)^{M_0} \to \exp(-0) = 1,$$

leading to $\mathrm{P}^*(N_{b_M,0}=M_0)\to 1$. Furthermore, it is clear that for $n \neq M_0$, it holds that $\mathrm{P}^*(N_{b_M,0} = n) = O(1/N_0)$. Using these results, for any small $\delta\in(0,1)$, we get

$$\begin{aligned}
&M_0\sum_{n=1}^{M_0}\frac{(n+8/3)}{(n+1)(n+2)}\mathrm{P}^*(N_{b_M,0}=n) \\
&\quad= \frac{M_0(M_0+8/3)}{(M_0+1)(M_0+2)}\mathrm{P}^*(N_{b_M,0}=M_0) + M_0\sum_{n=1}^{M_0-1}\frac{(n+8/3)}{(n+1)(n+2)}\mathrm{P}^*(N_{b_M,0}=n) \\
&\quad= \frac{M_0(M_0+8/3)}{(M_0+1)(M_0+2)}\mathrm{P}^*(N_{b_M,0}=M_0) + O\left(M_0\sum_{n=1}^{M_0-1}\frac{(n+8/3)}{(n+1)(n+2)N_0}\right) \\
&\quad= \frac{M_0(M_0+8/3)}{(M_0+1)(M_0+2)}\mathrm{P}^*(N_{b_M,0}=M_0) + O\left(\frac{M_0}{N_0^{1-\delta}}\sum_{n=1}^{M_0-1}\frac{(n+8/3)}{(n+1)(n+2)N_0^{\delta}}\right) \\
&\quad= \frac{M_0(M_0+8/3)}{(M_0+1)(M_0+2)}\mathrm{P}^*(N_{b_M,0}=M_0) + O\left(\frac{M_0}{N_0^{1-\delta}}\right),
\end{aligned}$$

where in the last line we have used

$$\sum_{n=1}^{M_0-1} \frac{(n+8/3)}{(n+1)(n+2)N_0^{\delta}} \leq \sum_{n=1}^{M_0-1} \frac{(n+8/3)}{(n+1)(n+2)n^{\delta}} \leq \sum_{n=1}^{\infty} O\left(\frac{1}{n^{1+\delta}}\right) < \infty.$$

In particular, we can choose $\delta = \frac{1}{2}$, which with $M_0 = o(\sqrt{N_0})$ implies

$$M_0 \sum_{n=1}^{M_0} \frac{(n+8/3)}{(n+1)(n+2)} \mathrm{P}^*(N_{b_M,0} = n) = \frac{M_0(M_0+8/3)}{(M_0+1)(M_0+2)} \mathrm{P}^*(N_{b_M,0} = M_0) + o(1).$$

Therefore, it follows that

$$\begin{aligned}
&M_0 N_0 \mathbb{E}\big[\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+) \mid \mathbf{W}, D_{b_M,i} = 1\big] \\
&\qquad = M_0 N_0 \frac{1}{N_0} \frac{3}{2} \sum_{n=1}^{M_0} \frac{(n+8/3)}{(n+1)(n+2)} \mathrm{P}^*(N_{b_M,0} = n) \\
&\qquad = \frac{3}{2} M_0 \sum_{n=1}^{M_0} \frac{(n+8/3)}{(n+1)(n+2)} \mathrm{P}^*(N_{b_M,0} = n) \\
&\qquad \to \frac{3}{2}.
\end{aligned}$$

Thus, putting everything together including $M_1 = \alpha M_0$ we get

$$\begin{aligned}
&\mathbb{E}\big[(A2)_M \mid \mathbf{W}\big] \\
&= \sum_{j=1}^{N} \sum_{l \neq j} W_j W_l \mathbb{E}\Big[\mathbb{E}^*\big[B_{M,i}(X_j)B_{M,i}(X_l)R_{b_M,j}R_{b_M,l} \mid \mathbf{Z}^+\big] \mid \mathbf{W}\Big] \\
&= N_1(N_1-1)\mathbb{E}\Big[\mathbb{E}^*\big[B_{M,i}(X_j)B_{M,i}(X_l)R_{b_M,j}R_{b_M,l} \mid \mathbf{Z}^+\big] \mid \mathbf{W}\Big] \\
&= N_1(N_1-1)\mathbb{E}\Big[\mathbb{E}^*[R_{b_M,j}R_{b_M,l} \mid \mathbf{Z}^+, B_{M,i}(X_j) = 1, B_{M,i}(X_j) = 1, D_{b_M,i} = 1] \\
&\qquad \mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1)\mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+)\big] \mid \mathbf{W}\Big] \\
&= N_1(N_1-1)\frac{M_1}{N_1}\left(\frac{M_1}{N_1} - \frac{1}{N_1}\right)\frac{1}{M_0 N_0} M_0 N_0 \\
&\qquad \times \mathbb{E}\Big[\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}^+, D_{b_M,i} = 1) \\
&\qquad \times \mathrm{P}^*(D_{b_M,i} = 1 \mid \mathbf{Z}^+) \mid \mathbf{Z}^+)\big] \mid \mathbf{W}\Big] \\
&= \frac{N_1-1}{N_0}\frac{M_1}{M_0}\left(\frac{M_1}{N_1} - \frac{1}{N_1}\right)\frac{3}{2} M_0 \sum_{n=1}^{M_0} \frac{(n+8/3)}{(n+1)(n+2)} \mathrm{P}^*(N_{b_M,0} = n) \\
&= \left(\alpha - \frac{1}{N_0}\right)\alpha\left(\frac{M_0}{N_0} - \frac{1}{N_1}\right)\frac{3}{2} M_0 \sum_{n=1}^{M_0} \frac{(n+8/3)}{(n+1)(n+2)} \mathrm{P}^*(N_{b_M,0} = n)
\end{aligned}$$

from which the contribution of $(A2)_M$ to $\mathbb{E}[\mathbb{V}\mathrm{ar}^*[\sqrt{M_1}(\hat{\tau}_M^{t,*} - \tau)^2 \mid \mathbf{Z}] \mid \mathbf{W}]$ satisfies

$$\frac{N_0}{M_1}\mathbb{E}\big[(A2)_M \mid \mathbf{W}\big]$$

$$= \frac{N_0}{M_1}\left(\alpha - \frac{1}{N_0}\right)\alpha\left(\frac{M_0}{N_0} - \frac{1}{N_1}\right)\frac{3}{2}M_0\sum_{n=1}^{M_0}\frac{(n+8/3)}{(n+1)(n+2)}\mathrm{P}^*(N_{b_M,0}=n)$$

$$= \left(\alpha - \frac{1}{N_0}\right)\alpha\left(\frac{N_0}{M_1}\frac{M_0}{N_0} - \frac{N_0}{M_1}\frac{1}{N_1}\right)\frac{3}{2}M_0\sum_{n=1}^{M_0}\frac{(n+8/3)}{(n+1)(n+2)}\mathrm{P}^*(N_{b_M,0}=n)$$

$$= \left(\alpha - \frac{1}{N_0}\right)\alpha\left(\frac{M_0}{M_1} - \frac{N_0}{N_1 M_1}\right)\frac{3}{2}M_0\sum_{n=1}^{M_0}\frac{(n+8/3)}{(n+1)(n+2)}\mathrm{P}^*(N_{b_M,0}=n)$$

$$= \left(\alpha - \frac{1}{N_0}\right)\left(1 - \frac{1}{M_1}\right)\frac{3}{2}M_0\sum_{n=1}^{M_0}\frac{(n+8/3)}{(n+1)(n+2)}\mathrm{P}^*(N_{b_M,0}=n)$$

$$\to \frac{3}{2}\alpha.$$

Thus we get the desired $\mathbb{E}[\mathbb{V}\mathrm{ar}^*[\sqrt{M_1}(\hat{\tau}_M^{t,*} - \tau)^2 \mid \mathbf{Z}] \mid \mathbf{W}] \to 1 + \frac{3}{2}\alpha$. □

B Distributional results for the M-out-of-N resampling mechanism

A key ingredient for the proof, that we make use of numerous times is Lemma 2, which is an analog to Lemma A.4 in Abadie and Imbens (2008) for the M-out-of-N resampling scheme.

Lemma 2 (Distributional results for M-out-of-N entities). *For $M_0 \le N_0$, $M_1 \le N_1$ and $w \in \{0,1\}$ we get*

(i) $R_{b_M,i} \mid W_i = w, \mathbf{Z} \sim \mathcal{B}(M_w, 1/N_w)$

(ii) $D_{b_M,i} \mid W_i = w, \mathbf{Z} \sim \mathcal{B}(1, 1-(1-1/N_w)^{M_w})$

(iii) Let $N_{b_M,0}$ the number of distinct control units in the bootstrap sample. The occupancy distribution is given by

$$\mathrm{P}^*(N_{b_M,0}=n) = \binom{N_0}{N_0-n}\frac{n!}{N_0^{M_0}}S(M_0,n)$$

where $S(m,k) = \left\{ {m \atop k} \right\} = \frac{1}{k!}\sum_{i=1}^{k}(-1)^i\binom{k}{i}(k-i)^m$ is the Stirling number of the second kind.

(iv) Let $N_{b_M,0}$ be the number of distinct control units in the bootstrap sample, then for (control) i

$$\mathbb{E}\big[\mathrm{P}^*(B_{M,i}(X_j)) = 1 \mid \mathbf{Z}, D_{b_M,i}=1, N_{b_M,0}) \mid \mathbf{W}, D_{b_M,i}=1, N_{b_M,0}\big] = \frac{1}{N_{b_M,0}}.$$

(v) Let $N_{b_M,0}$ be the number of distinct control units in the bootstrap sample, then for (control) i and distinct treated units $j \neq l$, we have

$$\mathbb{E}\big[\mathrm{P}^*(B_{M,i}(X_j)B_{M,i}(X_l) = 1 \mid \mathbf{Z}, D_{b_M,i} = 1, N_{b_M,0}) \mid \mathbf{W}, D_{b_M,i} = 1, N_{b_M,0}\big]$$
$$= \frac{3N_{b_M,0} + 8}{2N_{b_M,0}(N_{b_M,0}+1)(N_{b_M,0}+2)}$$

(vi) The expectation of the occupancy distribution is given by

$$\mathbb{E}^*[N_{b_M,0}] = N_0 - N_0(1 - \frac{1}{N_0})^{M_0}.$$

(vii) The variance of the occupancy distribution is given by

$$\mathbb{V}\mathrm{ar}^*[N_{b_M,0}] = N_0(N_0-1)(1-\frac{2}{N_0})^{M_0} + N_0(1-\frac{1}{N_0})^{M_0} - N_0^2(1-\frac{1}{N_0})^{2M_0}.$$

Proof. Parts (i), (ii), (iv) and (v) are obvious given the results in lemma A.4 of Abadie and Imbens (2008). Parts (iii), (vi) and (vii) follow immediately from Johnson and Kotz (1977). Note that for part (vi) one has

$$\mathbb{E}^*[N_{b_M,0}] = \sum_{i=1}^{N_0} \mathbb{E}^*[D_{b_M,i}].$$

□

References

Abadie, A. and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267.

——— (2008): "On the Failure of the Boostrap for Matching Estimators," *Econometrica*, 76, 1537–1557.

——— (2011): "Bias-Corrected Matching Estimators for Average Treatment Effects," *Journal of Business & Economic Statistics*, 29, 1–11.

——— (2012): "A Martingale Representation for Matching Estimators," *Journal of the American Statistical Association*, 107, 833 – 843.

Bickel, P. J., F. Götze, and W. R. van Zwet (1997): "Resampling fewer than n observations: gains, losses and remedies for losses." *Statistica Sinica*, 7, 1–31.

Härdle, W. and E. Mammen (1993): "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926–1947.

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

Otsu, T. and Y. Rai (2017): "Bootstrap Inference of Matching Estimators for Average Treatment Effects," *Journal of the American Statistical Association*, 112, 1720–1732.