SFB 823

Discussion Paper

Weighted bootstrap consistency for matching estimators: The role of bias-correction

Christopher Walsh, Carsten Jentsch, Shaikh Tanvir Hossain

Nr. 8/2021

SFB 823

Weighted bootstrap consistency for matching estimators: the role of bias-correction *

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

Abstract

We show that the purpose of consistent bias-correction for matching estimators of treatment effects is two-fold. Firstly, it is known to improve point estimation to get rid of asymptotically non-negligible bias terms. Secondly, we show that it is also inevitable to ensure the validity of weighted (or wild) bootstrap procedures for statistical inference. In fact, we provide a simple setting, where although the nearest neighbor matching estimator of the average treatment effect is exactly unbiased even in finite samples, valid weighted bootstrap inference requires bias-correction. As a direct and practically important consequence, an inadequate bias-correction will not only lead to biased point estimates, it will also distort inference leading e.g. to invalid confidence intervals. In simulations, we show that the choice of the bias-correction estimator that practitioners still have to make, can severely affect the weighted

*We are grateful for the computing time provided on the Linux HPC cluster at Technical University Dortmund (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.

bootstrap's performance when estimating the asymptotic variance in finite samples. In particular, simple rules such as estimating the bias based on linear regressions in the treatment arms can lead to very poor weighted bootstrap based variance estimates.

Key words: ATE, matching estimator, bootstrap consistency, weighted bootstrap, wild bootstrap, bias-correction **JEL codes:** C14, C21

1 Introduction

Matching estimators are intuitively simple procedures to estimate average treatment effects within the potential outcomes framework. The asymptotic properties of these estimators were established in Abadie and Imbens (2006, 2011, 2012). The expression for the variance of the asymptotic distribution is 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 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 showed this by considering a very simple data generating process (DGP), for which they were able to derive simple closed form expressions of: (i) the limiting variance of the nearest neighbor matching estimator of the average treatment effect of the treated (ATET) and (ii) the limit of the expectation of the conditional variance of a naive Efron-type bootstrap estimator. As the two expressions differ, their DGP constitutes a counterexample showing that the naive Efron-type bootstrap variance estimator is not valid.

In addition to this negative result concerning the validity of resampling procedures in the matching context, they provide two possible solutions in the form of a conjecture stating that using either the wild bootstrap of Härdle and Mammen (1993) or the M-out-of-N bootstrap (Bickel et al. (1997)) can cure this invalidity and provide a remedy to correctly estimate the limiting distribution of matching estimators. Walsh et al. (2021) proved that an M-out-of-N bootstrap procedure can indeed be used to unbiasedly estimate the limit variance in the counterexample DGP setting of Abadie and Imbens (2008). As for the other possible solution, Otsu and Rai (2017) proposed a weighted bootstrap procedure that can be interpreted as a wild bootstrap and proved its bootstrap consistency. Thus, as they indeed write, their paper formally confirms the conjecture that the wild bootstrap can be used to correctly estimate the limiting distribution of matching estimators.

In this paper, we shed light on the mechanism of the weighted bootstrap proposed by Otsu and Rai (2017) leading to the validity of their procedure. In order to bring out the intuition for the validity of their procedure we follow in the vein of Abadie and Imbens (2008) by looking at a class of DGPs for which it is possible to derive simple expressions for the asymptotics of the nearest neighbor matching estimator of the average treatment effect (ATE) as well as for the weighted bootstrap variance estimator. However, in contrast to the setting considered (only for ATET) in Abadie and Imbens (2008), the number of treated and controls cannot be fixed for the ATE. Hence, we need to extend the setup considered in Abadie and Imbens (2008). The derivation of the simple expressions for the asymptotic distribution of the nearest neighbor matching estimator of the ATE is substantially more complicated and relies on arguments on asymptotic expansions of inverse moments of binomial random variables.

The key to the validity of the weighted bootstrap is seen to be that it resamples the individual contributions of the bias-corrected matching estimator rather than resampling, for instance, the original data or the individual contributions of the classical matching estimator. Thus, the procedure requires to do bias-correction before resampling. Surprisingly, this is also the case for settings where it is known that the classical matching estimator is (asymptotically) unbiased. The reason for this somewhat surprising result is that the individual contributions of the bias-corrected matching estimator are approximately uncorrelated, when the bias is estimated sufficiently precisely. In fact, if one were to correct with the actual bias, then the individual contributions are uncorrelated. In contrast, the individual contributions of the classical matching estimator are not uncorrelated even if the classical matching estimator is (asymptotically) unbiased. Hence, by doing bias-correction first, the resamples obtained by the weighted bootstrap are based on draws from a collection of approximately uncorrelated random variables. Therefore, the choice of the resampling weights of the weighted bootstrap is less important. Specifically, it is not necessary to explicitly use the wild bootstrap weights and a simple Efron bootstrap of the individual contributions is also sufficient. Finally, as resampling the individual contributions from the classical matching estimator is not valid even when the estimator is unbiased, it is seen that treating the observed matches as an additional characteristic and resampling them along with the original data will not yield a valid bootstrap estimator.

The validity of the weighted bootstrap is an asymptotic result. In particular, it hinges on the fact that asymptotically the actual bias is estimated consistently. Theoretically, this is achieved in Abadie and Imbens (2011) and Otsu and Rai (2017) by constructing a bias estimator based on flexible, nonparametric series regression estimators. In finite samples, one of course needs to choose the truncation parameter in these series estimators, but, in practice, it is often advocated that using a linear regression with all the regressors or with their squares and crossproducts will be sufficient. These choices correspond to a series estimator using polynomial basis functions up to the first or second order only. In our simulations, we will demonstrate that such adhoc choices may lead to distorted inference results. We do so by varying the first two conditional moments in the distribution of the data. Using different bias estimators, we are able to demonstrate the importance of doing accurate bias-correction for the weighted bootstrap to work properly. Finally, the simulations will also highlight the benefit of the properly performed weighted bootstrap vis-à-vis the plug-in variance estimator proposed in Abadie and Imbens (2006).

The remainder of the paper is structured as follows. Section 2 provides the basic treatment effect setup along with the classical matching estimator and its bias-corrected version. The weighted bootstrap procedure of Otsu and Rai (2017) and some related resampling procedures are presented in Section 3. Some DGPs along with the corresponding distributional results of the nearest neighbor matching estimator for the ATE are given in Section 4 allowing us to explicitly determine the asymptotic variance in our simulation study. Results pertaining to the behavior of the weighted bootstrap are collected in Section 5 including the main result showing that the key to validity is an appropriate bias-correction. The results of the simulation study to illustrate how the performance of the weighted bootstrap depends on the appropriateness of the estimated bias are given in Section 6. Finally, Section 7 concludes. The detailed proofs are collected in the appendix at the end of the manuscript.

2 Setup

We consider the basic treatment effects setup. For each unit $i = 1, \ldots, N$ let $Y_i(0)$ and $Y_i(1)$ be the unobserved 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. The observed data is drawn from (Y, W, X). Here, we are interested in estimating the the average treatment effect (ATE) given by

$$\tau = \mathbb{E}[Y(1) - Y(0)],$$

which is also the parameter of interest in Otsu and Rai (2017). Let $\mathbf{Z} := \{(Y_i, W_i, X_i)\}_{i=1}^N$ be an i.i.d. random sample from the population (Y, W, X). The classical matching estimator for the ATE using M fixed matches with replacement is given by

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^{N} \{\hat{Y}_i(1) - \hat{Y}_i(0)\}, \tag{2.1}$$

where

$$\hat{Y}_i(1) = W_i Y_i + (1 - W_i) \sum_{j \in \mathcal{J}_M(i)} \frac{Y_j}{M}$$

and

$$\hat{Y}_i(0) = W_i \sum_{j \in \mathcal{J}_M(i)} \frac{Y_j}{M} + (1 - W_i) Y_i$$

are imputation based estimates of the potential outcomes with $\mathcal{J}_M(i)$ the index set of the first M matches for unit i

$$\mathcal{J}_M(i) = \Big\{j \in \{1, \ldots, N\} : W_j = 1 - W_i, \sum_{l: W_l = W_j} \mathbb{1}\{||X_l - X_i|| \le ||X_j - X_i||\} \le M\Big\}.$$

Finally, with $K_M(i) = \sum_{l=1}^N \mathbb{1}(j \in \mathcal{J}_M(l))$ denoting the number of times unit i was a match, the classical matching estimator can be written as

$$\hat{\tau} = \frac{1}{N} \sum_{i=1}^N (2W_i - 1) \left(1 + \frac{K_M(i)}{M}\right) Y_i. \tag{2.2}$$

In order to derive the asymptotic properties of the matching estimator $\hat{\tau}$, it is typically assumed that one has an i.i.d. sample. In addition, one typically assumes that the regressors are continuously distributed, that the so-called common support condition and certain moment conditions hold. Denote the first two conditional moments of the outcome given the treatment status and the covariate value by $\mu(w, x) = \mathbb{E}[Y \mid W = w, X = x]$ and $\sigma^2(w, x) = \mathbb{V}\text{ar}[Y \mid W = w, X = x]$, then a typical set of conditions is given in Assumption 1.

Assumption 1. *The data* $\mathbf{Z} = \{(Y_i, W_i, X)\}_{i=1}^N$ *consists of* N *i.i.d. draws from the distribution of* (Y, W, X)*, where* $Y = WY(1) + (1 - W)Y(0)$ *with:*

(i) X *is continuously distributed on a compact and convex set* $\mathbb{X} \subset \mathbb{R}^k$*. The density of* X *is bounded and bounded away from zero on* $\mathbb{X}$*.*

(ii) W *is independent of* $(Y(0), Y(1))$ *conditional on* $X = x$ *for almost every* $x \in \mathbb{X}$*. There exists* $\eta > 0$ *such that* $\Pr(W = 1 \mid X = x) \in (\eta, 1 - \eta)$ *for almost every* $x \in \mathbb{X}$*.*

(iii) For each $w \in \{0, 1\}$*,* $\mu(w, \cdot)$ *and* $\sigma^2(w, \cdot)$ *are Lipschitz continuous on* $\mathbb{X}$*;* $\sigma^2(w, \cdot)$ *is bounded away from zero on* $\mathbb{X}$ *and* $\mathbb{E}[Y^4 | W = w, X = \cdot]$ *is uniformly bounded on* $\mathbb{X}$*.*

These conditions correspond to the set of assumptions given in Abadie and Imbens (2006, 2011) and Otsu and Rai (2017). Identification of the ATE is guaranteed by the common support condition in Assumption 1(ii). Asymptotically, one could weaken Assumption 1(i) and allow for discretely distributed covariates taking finitely many values with all the results being established for subsamples based on the values of the discrete covariates. In finite samples, allowing for discrete covariates in such a fashion may lead to difficulties as the common support condition may fail in the sample. In this case one could treat the discrete covariates as if they were continuously distributed. The smoothness conditions in Assumption 1(iii) are used to establish the consistency and asymptotic normality of the matching estimators. Under the conditions in Assumption 1, Abadie and Imbens (2006) derived consistency and asymptotic normality of the classical matching estimator $\hat{\tau}$. In particular, they established that

$$\frac{\sqrt{N}(\hat{\tau} - B_N - \tau)}{\sigma_N} \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\sigma_N^2 = \frac{1}{N}\sum_{i=1}^{N}\left(1 + \frac{K_M(i)}{M}\right)^2 \sigma^2(W_i, X_i) + \mathbb{E}\left[\left(\mu(1,X) - \mu(0,X) - \tau\right)^2\right]$$

and the bias term is given by

$$\begin{aligned} B_N &= \frac{1}{N}\sum_{i=1}^{N}\frac{2W_i - 1}{M}\sum_{j\in\mathcal{J}_M(i)}\big(\mu(X_i, 1-W_i) - \mu(X_j, 1-W_i)\big) \\ &= \frac{1}{N}\sum_{i=1}^{N}(2W_i - 1)\Big(\mu(X_i, 1-W_i) - \frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}\mu(X_j, 1-W_i)\Big). \end{aligned}$$

Moreover, they showed that unless one uses a single regressor ($k = 1$), the bias term B_N of the classical matching estimator $\hat{\tau}$ dominates the asymptotic distribution and the classical matching estimator will not be $\sqrt{N}$-consistent. In particular, under Assumption 1, their Theorem 1 holds, which states that $B_N = O_p(N^{-1/k})$. The bias term depends on the conditional means in both treatment arms. Thus, if we have estimators for these conditional means, denoted by $\hat{\mu}(x, w)$ for $w \in \{0, 1\}$, then we can estimate the bias by

$$\widehat{B}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{2W_i - 1}{M}\sum_{j\in\mathcal{J}_M(i)}\big(\hat{\mu}(X_i, 1-W_i) - \hat{\mu}(X_j, 1-W_i)\big) \tag{2.3}$$

and we can define the bias-corrected matching estimator by $\tilde{\tau} := \hat{\tau} - \widehat{B}_N$. It is immediately clear that if the bias estimator can be shown to satisfy $\sqrt{N}(\hat{B}_N - B_N) =$

$o_P(1)$, then it follows that

$$\frac{\sqrt{N}(\tilde{\tau}-\tau)}{\sigma_N} \xrightarrow{d} \mathcal{N}(0,1).$$

Abadie and Imbens (2011) use a flexible, nonparametric series regression estimator to estimate the conditional means and provide sufficient conditions to ensure $\sqrt{N}(\hat{B}_N - B_N) = o_P(1)$. Plugging-in (2.2) and (2.3) and re-arranging, the bias-corrected matching estimator can be written as

$$\begin{aligned}\tilde{\tau} &= \frac{1}{N}\sum_{i=1}^{N}(2W_i-1)\left(1+\frac{K_M(i)}{M}\right)\big(Y_i-\hat{\mu}(X_i,W_i)\big)\\ &\quad + (2W_i-1)\big(\hat{\mu}(X_i,W_i)-\hat{\mu}(X_i,1-W_i)\big)\\ &=: \frac{1}{N}\sum_{i=1}^{N}\tilde{\tau}_i, \end{aligned} \tag{2.4}$$

where we will call

$$\begin{aligned}\tilde{\tau}_i &= (2W_i-1)\left(1+\frac{K_M(i)}{M}\right)\big(Y_i-\hat{\mu}(X_i,W_i)\big)\\ &\quad + (2W_i-1)\big(\hat{\mu}(X_i,W_i)-\hat{\mu}(X_i,1-W_i)\big)\end{aligned} \tag{2.5}$$

the *individual contribution to the bias-corrected matching estimator.* Similarly, we can write

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N}(2W_i-1)\left(1+\frac{K_M(i)}{M}\right)Y_i =: \frac{1}{N}\sum_{i=1}^{N}\hat{\tau}_i, \tag{2.6}$$

and will call

$$\hat{\tau}_i = (2W_i-1)\left(1+\frac{K_M(i)}{M}\right)Y_i \tag{2.7}$$

the individual contribution to the classical matching estimator.

3 The weighted bootstrap estimator

The weighted bootstrap estimator proposed by Otsu and Rai (2017) is defined as a randomly weighted average of the difference between the individual contributions of the bias-corrected matching estimator and the bias-corrected estimator. Recalling the decomposition of the bias-corrected matching estimator in (2.4) in terms of the individual contributions given in (2.5) the weighted bootstrap estimator is defined as

$$\sqrt{N}\tilde{T}^* = \sum_{i=1}^{N} \eta_i^* (\tilde{\tau}_i - \tilde{\tau}), \tag{3.1}$$

for a given choice of randomly drawn resampling weights $\eta_i^*, i = 1, \ldots, N$.

In order to derive the validity of the weighted bootstrap, Otsu and Rai (2017) consider the setting of Abadie and Imbens (2011). In particular, the conditions in Assumption 1 are assumed to hold, the bias is estimated using nonparametric series estimators $\hat{\mu}(x, w)$, for $w \in \{0, 1\}$ and the conditions ensuring that $\sqrt{N}(\hat{B}_N - B_N) = o_p(1)$ are satisfied. In this setup, Otsu and Rai (2017) showed that, whenever the resampling weights satisfy the conditions given in Assumption 2, the weighted bootstrap is valid in the sense that $d(\sqrt{N}\tilde{T}^*, \sqrt{N}(\tilde{\tau} - \tau)) \xrightarrow{\text{P}} 0$ with d the Kolmogorov distance.

Assumption 2. *The resampling weights $\eta_i^*, i = 1, \ldots, N$ satisfy:*

(i) $(\eta_1^*, \ldots, \eta_N^*)$ *is exchangeable and independent of the data* $\mathbf{Z} = \{(Y_i, W_i, X_i)\}_{i=1}^N$.

(ii) $\sum_{i=1}^N (\eta_i^* - \overline{\eta^*})^2 \xrightarrow{\text{P}^*} 1$, *where* $\overline{\eta^*} = \frac{1}{N}\sum_{i=1}^N \eta_i$

(iii) $\max_{i=1,\ldots,N} |\eta_i^* - \overline{\eta^*}| \xrightarrow{\text{P}^*} 0$.

(iv) $\mathbb{E}^*\left[(\eta_i^*)^2\right] = O\left(N^{-1}\right)$ *for all* $i = 1, \ldots, N$.

Choosing wild bootstrap-type weights $\eta_i^* = \epsilon_i^*/\sqrt{N}$, where $\{\epsilon_i^*\}_{i=1}^N$ are i.i.d. random variables with a zero mean and unit variance is admissible in Assumption 2. By calling the resulting procedure based on this choice of weights *the wild bootstrap*, allows Otsu and Rai (2017) to conclude that the conjecture of Abadie and Imbens (2008) has been formally confirmed.

However, the conditions in Assumption 2 also allow for the choice of $\eta_i^* = M_i^*/\sqrt{N}$ with $(M_1^*, \ldots, M_N^*)$ a multinomially distributed random vector based on N trials and N equally likely cells. This choice of resampling weights is nothing else but a rescaling of a simple Efron bootstrap applied to $(\tilde{\tau}_{i,c}, i = 1, \ldots, N)$, where $\tilde{\tau}_{i,c} := \tilde{\tau}_i - \tilde{\tau}$ as

$$\sum_{i=1}^{N} \frac{M_i^*}{\sqrt{N}} (\tilde{\tau}_i - \tilde{\tau}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} M_i^* \tilde{\tau}_{i,c} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{\tau}_{i,c}^*$$

where $(\tilde{\tau}^*_{i,c}, i = 1, \ldots, N)$ denotes an Efron bootstrap sample obtained by independently drawing with replacement from $(\tilde{\tau}_{i,c}, i = 1, \ldots, N)$. In fact, it can even be written in terms of a simple Efron bootstrap of the individual contributions of the bias-corrected matching estimator as $\sum_{i=1}^N M_i^* = N$ implies that

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} M_i^* (\tilde{\tau}_i - \tilde{\tau}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} M_i^* \tilde{\tau}_i - \sqrt{N}\tilde{\tau} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{\tau}_i^* - \sqrt{N}\tilde{\tau}$$

where $(\tilde{\tau}^*_i, i = 1, \ldots, N)$ is an Efron bootstrap sample obtained by independently drawing with replacement from the (uncentered) individual contributions $(\tilde{\tau}_i, i = 1, \ldots, N)$.

As the choice of the weights is not restricted to the wild bootstrap-type weights and an Efron bootstrap of the $(\tilde{\tau}_{i,c}, i = 1, \ldots, N)$ is also valid, this already indicates that it is not the "wildness" that makes the weighted bootstrap procedure work. In fact, as we will see Section 5 the key for the validity is that the bias-corrected individual contributions are resampled as opposed to resampling the individual contributions of the classical matching estimator or to resampling the original data.

In the last part of this section we will show in what way the valid bootstrap estimator using Efron-type weights can be interpreted as an Efron bootstrap based on an "augmented" data set. The naive Efron-type bootstrap of the data $\{(Y_i, W_i, X_i)\}_{i=1}^N$ considered in Abadie and Imbens (2008) fails because the bootstrap counterpart of $K_M(i)$ fails to correctly reproduce the matching distribution $K_M(i)$. As a possible solution, it may be conceivable to use an Efron bootstrap that treates the matches $K_M(i)$ as a characteristic of the original data. However, if one uses an Efron-type bootstrap on the "augmented" data $\{(Y_i, W_i, X_i, K_M(i))\}_{i=1}^N$ and then plugs in the bootstrap variables into the formula of the classical matching estimator, we get exactly the same as using an Efron bootstrap on the individual contributions of the classical matching estimator

$$\sqrt{N}\hat{T}^* = \sum_{i=1}^{N} \frac{M_i^*}{\sqrt{N}} (\hat{\tau}_i - \hat{\tau}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{\tau}^*_{i,c} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{\tau}_i^* - \sqrt{N}\hat{\tau}, \tag{3.2}$$

where $(\hat{\tau}^*_{i,c} = \hat{\tau}^* - \hat{\tau}, i = 1, \ldots, N)$ with $(\hat{\tau}^*_i, i = 1, \ldots, N)$ the Efron bootstrap sample of the individual contributions of the classical matching estimator. Thus, when the weighted bootstrap on the individual contributions of the classical matching estimator is invalid, it is also invalid to use an Efron on the "augmented" data $\{(Y_i, W_i, X_i, K_M(i))\}_{i=1}^N$ along with the formula for the classical matching estimator – even in settings when the classical matching estimator is unbiased in finite samples. Finally, from

$$\sqrt{N}\tilde{T}^* = \sum_{i=1}^{N} \frac{M_i^*}{\sqrt{N}}(\tilde{\tau}_i - \tilde{\tau}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \tilde{\tau}_{i,c}^* = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \tilde{\tau}_i^* - \sqrt{N}\tilde{\tau}.$$

we see that the valid weighted bootstrap with Efron-type weights can be interpreted as an Efron bootstrap on the "augmented" data $\{(Y_i, W_i, X_i, K_M(i))\}_{i=1}^N$ that uses the formula of the bias-corrected estimator. Notice, that the bias-correction is not re-calculated in the bootstrap sample, so that this is another "characteristic" of the original data that is kept. Therefore, the weighted bootstrap with multinomially distributed weights, can be thought of as an Efron boostrap on the "augmented" data $\{(Y_i, W_i, X_i, K_M(i), \hat{\mu}(X_i, W_i), \hat{\mu}(X_i, 1 - W_i)\}_{i=1}^N$.

4 A simple DGP allowing for closed-form asymptotic expressions for the ATE

In order to show that the bias-correction is necessary for the validity of the weighted bootstrap, we will follow in the vein of Abadie and Imbens (2008). In particular, we will consider a DGP for which the classical nearest neighbor matching estimator is unbiased even in finite samples and for which we can derive closed from expressions for (i) the limit variance of the nearest neighbor matching estimator and (ii) the limit of the expectation of the conditional variance of the weighted bootstrap when applied to the individual contributions of the classical matching estimator. As these two expressions turn out to be different, this proves that a weighted bootstrap applied to the individual contributions of the classical matching estimator (without bias correction) is *not* valid, although the estimator is actually unbiased and, in particular, for the purpose of point estimation there is no bias to correct for. Note, that the simple expressions cannot be derived using the setup considered by Abadie and Imbens (2008), which was used to get corresponding results when estimating the ATET. We have to modify the DGP to allow for i.i.d. draws from

(Y, W, X). In particular, this entails that we will no longer have a fixed ratio of treated to control units in the sample, which in turn makes the derivation of the expressions substantially more difficult.

Assumption 3. *Let* $\mathbf{Z} = \{(Y_i, W_i, X_i)\}_{i=1}^N =: (\mathbf{Y}, \mathbf{W}, \mathbf{X})$ *be a sample of* N *independent draws from* (Y, W, X)*, where:*

(i) The regressor satisfies $X \sim \mathcal{U}[0, 1]$.

(ii) $W \sim \mathcal{B}ern(p)$ *with* $p = \alpha/(1+\alpha)$ *for some finite positive* α.

(iii) $Y = WY(1) + (1-W)Y(0)$ *with the potential outcomes satisfying:*

(a) $Y(1)$ *is degenerate with* $\Pr(Y(1) = c) = 1$ *for some fixed* c.

(b) $Y(0) \mid X = x \sim \mathcal{N}(0, 1)$ *for all* $x \in [0, 1]$.

As we are only considering one continuous covariate in Assumption 3, we know that $\sqrt{N} B_N = o_P(1)$, so that it will not contribute to the asymptotic distribution of the classical matching estimator. In fact, as $\mu(x, 0) = 0$ and $\mu(x, 1) = c$ for all x the bias of the classical matching estimator is exactly zero, because

$$B_N = \frac{1}{N}\sum_{i=1}^N (2W_i - 1)\frac{1}{M}\sum_{j \in \mathcal{J}(i)} \big(\mu(X_i, 1 - W_i) - \mu(X_j, 1 - W_i)\big) = 0.$$

Assumption 3(iii) implies that the ATE τ equals c. It thus follows that

$$\frac{\sqrt{N}(\hat{\tau} - \tau)}{\sigma_N} \xrightarrow{d} \mathcal{N}(0, 1),$$

where, because $\mu(x, 1) - \mu(x, 0) - \tau = 0$, $\sigma^2(x, 1) = 0$ and $\sigma^2(x, 0) = 1$ for all x, we have

$$\begin{aligned}
\sigma_N^2 &= \frac{1}{N}\sum_{i=1}^N \left(1 + \frac{K_M(i)}{M}\right)^2 \sigma^2(X_i, W_i) + \mathbb{E}\left[\big(\mu(X, 1) - \mu(X, 0) - \tau\big)^2\right] \\
&= \frac{1}{N}\sum_{i=1}^N (1 - W_i)\left(1 + \frac{K_M(i)}{M}\right)^2.
\end{aligned}$$

In the following, we will consider the nearest neighbor matching estimator based on a single match, that is with $M = 1$. To lighten notation, we will write $K_i := K_1(i)$. In this case, it is possible to establish the distributional results for the matching estimator of the ATE under Assumption 3.

Proposition 1 (Distributional results for nearest neighbor matching estimator $\hat{\tau}$)**.** *Given Assumption 3 the nearest neighbor matching estimator for the ATE,*

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N}(2W_i - 1)(1 + K_i) Y_i$$

satisfies

$$\sqrt{N}(\hat{\tau} - \tau) \xrightarrow{d} \mathcal{N}\left(0, 1 + \frac{\alpha}{1+\alpha}\left(2 + \frac{3}{2}\alpha\right)\right).$$

The proposition can be seen as a companion result to the simple expressions derived by Abadie and Imbens (2008) in the ATET case. In order to derive the results in the ATE case given in Proposition 1, we have to modify their proposed DGP, which then requires the use of substantially more complicated arguments based on some non-trivial results on asymptotic approximations for reciprocal moments of binomial random variables. From the proposition we can see that the classical nearest neighbor matching estimator is asymptotically normal with a limit variance that depends solely on the parameter α, which governs the expected ratio of treated to control units in the sample. Under the additional assumption that the bias estimator $\widehat{B}_N$ satisfies $\sqrt{N}(\widehat{B}_N - B_N) = \sqrt{N}\widehat{B}_N = o_P(1)$, the bias-corrected nearest neighbor matching estimator with $(M = 1)$ will have the same asymptotic limit, that is

$$\sqrt{N}(\tilde{\tau} - \tau) \xrightarrow{d} \mathcal{N}\left(0, 1 + \frac{\alpha}{1+\alpha}\left(2 + \frac{3}{2}\alpha\right)\right).$$

Proof of Proposition 1. Given the DGP in Assumption 3, we have already seen that the ATE τ is given by c. Some simple calculations show that for the classical nearest neighbor matching estimator under the DGP of Assumption 3 one gets

$$\begin{aligned}\hat{\tau} &= \frac{1}{N}\sum_{i=1}^{N} W_i (1 + K_i)\tau - (1 - W_i)(1 + K_i) Y_i(0) \\ &= \tau - (1 - W_i)(1 + K_i) Y_i(0),\end{aligned}$$

where the last line follows from $\sum_{i=1}^{N} W_i + \sum_{i=1}^{N} K_i W_i = N_1 + N_0 = N$. Thus $\hat{\tau} - \tau = -\frac{1}{N}\sum_{i=1}^{N}(1 - W_i)(1 + K_i) Y_i(0)$ and and it follows that

$$\sqrt{N}(\hat{\tau} - \tau) \mid \mathbf{W}, \mathbf{X} \sim \mathcal{N}\left(0, \frac{1}{N}\sum_{i=1}^{N}(1 - W_i)(1 + K_i)^2\right).$$

The total law of variance leaves us with

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\sqrt{N}(\hat{\tau}-\tau)] &= \mathbb{E}\left[\mathbb{V}\mathrm{ar}[\sqrt{N}(\hat{\tau}-\tau) \mid \mathbf{X},\mathbf{W}]\right] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}[\sqrt{N}(\hat{\tau}-\tau) \mid \mathbf{X},\mathbf{W}]\right] \\ &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}(1-W_i)\,(1+K_i)^2\right].\end{aligned}$$

Thus, we are left to show that the last expression converges to the limit variance given in the proposition. As $\{W_i\}_{i=1}^N$ are i.i.d. and the $\{K_i\}_{i=1}^N$ are exchangeable, we get

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\sqrt{N}(\hat{\tau}-\tau)] &= \mathbb{E}[1-W_i] + 2\mathbb{E}[(1-W_i)K_i] + \mathbb{E}[(1-W_i)K_i^2] \\ &= \frac{1}{1+\alpha}\left(1 + 2\mathbb{E}[K_i \mid W_i = 0] + \mathbb{E}[K_i^2 \mid W_i = 0]\right).\end{aligned}$$

When deriving the limit expressions of the terms $\mathbb{E}[K_i \mid W_i = 0]$ and $\mathbb{E}[K_i^2 \mid W_i = 0]$ we cannot directly appeal to the results in Abadie and Imbens (2008) as there, N_0 and N_1, the number of control units and treated units are assumed to be fixed fractions of the sample size. Instead, we have to use a conditioning argument and results on asymptotic expansions of reciprocal moments of binomial random variables to get

$$\mathbb{E}[K_i \mid W_i = 0] \to \alpha \quad \text{and} \quad \mathbb{E}[K_i^2 \mid W_i = 0] \to \alpha + \frac{3}{2}\alpha^2 \tag{4.1}$$

as $N \to \infty$ from which the proposition follows. The details of the lengthy technical arguments leading to (4.1) are given in Appendix A. □

The setting in Assumption 3 will serve to show that the weighted bootstrap applied to the individual contributions of the classical matching estimator (without bias correction) is not valid even if the classical matching estimator is unbiased in finite samples.

In Section 6, we will use simulations to investigate the performance of the weighted bootstrap estimator when the bias is poorly estimated. In order to do so, we will consider DGPs that satisy Assumption 4, where (iii)(b) in Assumption 3 has been replaced by $Y(0)|X = x \sim \mathcal{N}(\mu(x,0), \sigma^2(x,0))$ for specific choices of $\mu(x,0)$ and $\sigma(x,0)$. Note, that if $\mu(x,0) = 0$ and $\sigma^2(x,0) = 1$, then the conditions are the same as those in Assumption 3.

Assumption 4. *Let* $\mathbf{Z} = \{(Y_i, W_i, X_i)\}_{i=1}^N =: (\mathbf{Y}, \mathbf{W}, \mathbf{X})$ *be a sample of* N *independent draws from* (Y, W, X)*, where:*

(i) The regressor satisfies $X \sim \mathcal{U}[0,1]$.

(ii) $W \sim \mathcal{B}ern(p)$ *with* $p = \alpha/(1+\alpha)$ *for some finite positive* α.

(iii) $Y = WY(1) + (1-W)Y(0)$ *with the potential outcomes satisfying:*

(a) $Y(1)$ *is degenerate with* $\Pr(Y(1) = c) = 1$ *for some fixed* c.

(b) $Y(0)|X = x \sim \mathcal{N}(\mu(x,0), \sigma^2(x,0))$ *for specific choices of* $\mu(x,0)$ *and* $\sigma(x,0)$.

Given any DGP satisfying Assumption 4, we now get $\tau = c - \mathbb{E}[\mu(X,0)]$. As we are still only considering one regressor, i.e. $B_N = O_p(N^{-1})$, so that the bias will still not contribute to the asymptotic distribution of the classical nearest neighbor matching estimator. Moreover, we get

$$\frac{\sqrt{N}(\hat{\tau} - \tau)}{\sigma_N} \xrightarrow{d} \mathcal{N}(0,1)$$

where (using $\sigma^2(x,1) = 0$ and $\mu(x,1) = c$), we have

$$\begin{aligned}\sigma_N^2 &= \frac{1}{N}\sum_{i=1}^N (1+K_i)^2\,\sigma^2(X_i, W_i) + \mathbb{E}\left[\left(\mu(X,1) - \mu(X,0) - \tau\right)^2\right] \\ &= \frac{1}{N}\sum_{i=1}^N (1-W_i)\,(1+K_i)^2\,\sigma^2(X_i,0) + \mathbb{E}\left[\left(\mu(X,0) - \mathbb{E}[\mu(X,0)]\right)^2\right].\end{aligned}$$

Although the expression for σ_N^2 is more complicated than under the more restrictive Assumption 3, it is still possible to calculate the limit of its expectation as in the proof of Proposition 1 provided $\mu(x,0)$ and $\sigma^2(x,0)$ are polynomials in x. In particular, as the K_i depend on the regressors in the treated group and these are drawn independently of the control observations, we have $\mathbb{E}[(1-W_i)(1+K_i)^2\sigma^2(0,X_i)] = \mathbb{E}[(1-W_i)(1+K_i)^2]\mathbb{E}[\sigma^2(0,X_i)]$. Thus, in addition to the steps of the proof of Proposition 1, one has to calculate certain moments of a uniformly distributed random variable.

5 Bias-correction affects bootstrap validity

In this section we prove that the weighted bootstrap without bias-correction is in general invalid. We will do so by considering the setting of Assumption 3, where the classical matching estimator is already unbiased even in finite samples. In this setting we will derive the limit of the expectation of the conditional variance of the weighted bootstrap based on resampling the individual contributions of the classical matching estimator (without bias correction) and see that it does not converge to the correct limit.

Denote by $\mathbb{E}^*[\,\cdot\,]$ and $\mathbb{V}\mathrm{ar}^*[\,\cdot\cdot]$ the expectation and the variance, respectively, over the resampling mechanism conditional on the orginal data. Now, suppose that the DGP is given as in Assumption 3 and that the bias estimator based on series estimators $\hat{\mu}(x, w)$ for $w \in \{0, 1\}$ satisfies $\sqrt{N}(\widehat{B}_N - B_N) = \sqrt{N}\widehat{B}_N = o_P(1)$. From the validity of the weighted bootstrap in this setting, we know that $\mathbb{E}\left[\mathbb{V}\mathrm{ar}^*\left[\sqrt{N}\tilde{T}^*\right]\right] \to 1 + \frac{\alpha}{1+\alpha}\left(2 + \frac{3}{2}\alpha\right)$ for any choice of resampling weights satisfying Assumption 2.

Next, we will see that although the classical matching estimator is unbiased in the setting of Assumption 3, a weighted bootstrap procedure based on resampling the individual contributions of the classical nearest neighbor matching estimator will not be valid. Specifically, given the decomposition of the classical nearest neighbor matching estimator for the ATE in (2.6) in terms of the individual contributions, we will show that the conditional variance of the weighted bootstrap estimator based on resampling the corresponding contributions given in (2.7) will not converge to the correct limit. In particular, we will consider the weighted bootstrap procedure

$$\sqrt{N}\hat{T}^* := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i^* (\hat{\tau}_i - \hat{\tau}) \tag{5.1}$$

with $\{\epsilon_i^*\}_{i=1}^N$ a sequence of i.i.d. random variables that are independent of the data with $\mathbb{E}^*[\epsilon_i^*] = 0$ and $\mathbb{V}\mathrm{ar}^*[\epsilon_i^*] = 1$. Hence, this weighted bootstrap corresponds to a weighted bootstrap using wild bootstrap weights applied to the individual contributions of the classical matching estimator.

Theorem 1 (Inconsistency of $\sqrt{N}\hat{T}^*$)**.** *Under the DGP of Assumption 3, the weighted bootstrap on the individual contributions of the classical nearest neighbor matching* *estimator* $\sqrt{N}\hat{T}^*$ *in* (5.1) *satisfies* $\mathbb{E}^*\left[\sqrt{N}\hat{T}^*\right] = 0$ *and as* $N \to \infty$*,*

$$\mathbb{E}\left[\mathbb{V}\mathrm{ar}^*\left[\sqrt{N}\hat{T}^*\right]\right] \to 1 + \frac{\alpha}{1+\alpha}\left(2 + \frac{3}{2}\alpha\right) + \tau^2\frac{1}{1+\alpha}\left(2 + \frac{3}{2\alpha}\right).$$

We immediatelty see that $\sqrt{N}\hat{T}^*$ is inconsistent as the expectation of the conditional variance does not converge to the asymptotic variance of the matching estimator as given in Proposition 1. This result is quite surprising at first as the classical matching estimator is actually unbiased under the DGP of Assumption 3. Hence, in order to use the weighted bootstrap for valid inference, it is still necessary to perform bias correction first and then resample the individual contributions of the bias-corrected matching estimator although there is no bias to correct for.
As the limit in Theorem 1 is of the form $\lim_{N\to\infty} \mathbb{V}\mathrm{ar}[\sqrt{N}(\hat{\tau} - \tau)] + \tau^2 \frac{1}{1+\alpha}(2 + \frac{3}{2\alpha})$, large ATE τ and small values of α will result in a larger bias when estimating the limit variance.

Proof of Theorem 1. As the resampling weights $\{\epsilon_i^*\}_{i=1}^N$ are i.i.d. and independent of the data with $\mathbb{E}^*[\epsilon_i^*] = 0$ and $\mathbb{V}\mathrm{ar}^*[\epsilon_i^*] = 1$, we get

$$\mathbb{E}^*[\sqrt{N}\hat{T}^*] = \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}^*[\epsilon_i^*](\hat{\tau}_i - \hat{\tau}) = 0.$$

With $\mathbb{C}\mathrm{ov}^*[\,\cdot,\cdot]$ the covariance over the resampling mechanism conditional on the data the independence of the resampling weights implies that

$$\begin{aligned}\mathbb{V}\mathrm{ar}^*[\sqrt{N}\hat{T}^*] &= \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \mathbb{C}\mathrm{ov}^*[\epsilon_i^*, \epsilon_j^*](\hat{\tau}^t - \hat{\tau})(\hat{\tau}_j - \hat{\tau}) \\ &= \frac{1}{N}\sum_{i=1}^N (\hat{\tau}_i - \hat{\tau})^2 = \frac{1}{N}\sum_{i=1}^N \hat{\tau}_i^2 - (\hat{\tau})^2.\end{aligned}$$

Under the DGP of Assumption 3, we have $\hat{\tau} = \tau - (1 - W_i)(1 + K_i) Y_i(0)$, which when plugged into the second term and upon taking expectations yields

$$\begin{aligned}\mathbb{E}[(\hat{\tau})^2] &= \mathbb{E}[(\tau - \frac{1}{N}\sum_{i=1}^{N}(1-W_i)(1+K_i)Y_i(0))^2] \\ &= \tau^2 - 2\tau\mathbb{E}[\frac{1}{N}\sum_{i=1}^{N}(1-W_i)(1+K_i)Y_i(0)] \\ &+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(1-W_i)(1-W_j)(1+K_i)(1+K_j)Y_i(0)Y_j(0) \\ &= \tau^2 + \frac{1}{N}\mathbb{E}[(1-W_i)(1+K_i)^2] \to \tau^2,\end{aligned}$$

where for the third equality we have used that the $Y_i(0)$ are zero mean i.i.d. and independent of the X_i and the limit in the last line is due to $\mathbb{V}\mathrm{ar}[\sqrt{N}(\hat{\tau}-\tau)] = \mathbb{E}[(1-W_i)(1+K_i)^2] = O(1)$.

Using the DGP of Assumption 3, we get

$$\begin{aligned}\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\hat{\tau}_i^2\right] &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[(W_i(1+K_i)Y_i - (1-W_i)(1+K_i)Y_i)^2] \\ &= \tau^2\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[W_i(1+K_i)^2] + \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[(1-W_i)(1+K_i)^2] \\ &= \tau^2\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[W_i(1+K_i)^2] + \mathbb{V}\mathrm{ar}[\sqrt{N}(\hat{\tau}-\tau)].\end{aligned}$$

The limit of $\mathbb{V}\mathrm{ar}[\sqrt{N}(\hat{\tau}-\tau)]$, which is of course the target, was given in Proposition 1. As for the limit of $\tau^2\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[W_i(1+K_i)^2]$, we get upon multiplying out and due to the exchangeability of the $\{K_i\}_{i=1}^{N}$ and as the $\{W_i\}_{i=1}^{N}$ are i.i.d. that

$$\tau^2\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[W_i(1+K_i)^2] = \tau^2\left(\frac{\alpha}{1+\alpha} + 2\frac{\alpha}{1+\alpha}\mathbb{E}[K_i \mid W_i = 1]\right) + \frac{\alpha}{1+\alpha}\mathbb{E}[K_i^2 \mid W_i = 1].$$

Mirroring the derivation for the marginal moments of $K_i \mid W_i = 0$ given in Abadie and Imbens (2008), we get for a fixed $n_1 \in \{1, \dots, N\}$ and $n_0 = N - n_1$,

$$\mathbb{E}[K_i \mid W_i = 1, \sum_{j=1}^{N} W_j = n_1] = \frac{n_0}{n_1}$$

and

$$\mathbb{E}[K_i^2 \mid W_i = 1, \sum_{j=1}^{N} W_j = n_1] = \frac{n_0}{n_1} + \frac{3}{2}\frac{n_0(n_0-1)(n_1+8/3)}{n_1(n_1+1)(n_1+2)}.$$

Then using the same conditioning argument and the results on asymptotic expansions of reciprocal moments of binomial random variables as in Appendix A, we get

$$\mathbb{E}[K_i \mid W_i = 1] \to \frac{1}{\alpha}$$

and

$$\mathbb{E}[K_i^2 \mid W_i = 1] \to \frac{1}{\alpha} + \frac{3}{2}\left(1 - 2\frac{1+\alpha}{\alpha} + \left(\frac{1+\alpha}{\alpha}\right)^2\right).$$

Putting everything together, as $\mathbb{V}\mathrm{ar}[\sqrt{N}(\hat{\tau} - c)] \to 1 + \frac{\alpha}{1+\alpha}(2 + \frac{3}{2}\alpha)$, we get

$$\begin{aligned}
&\mathbb{E}[\mathbb{V}\mathrm{ar}^*[\sqrt{N}\tau_{WB}^t]] \\
&\to \tau^2 \frac{\alpha}{1+\alpha}\left(1 + \frac{2}{\alpha} + \frac{1}{\alpha} + \frac{3}{2}\left(1 - 2\frac{1+\alpha}{\alpha} + \left(\frac{1+\alpha}{\alpha}\right)^2\right)\right) + 1 + \frac{\alpha}{1+\alpha}\left(2 + \frac{3}{2}\alpha\right) - \tau^2 \\
&= 1 + \frac{\alpha}{1+\alpha}\left(2 + \frac{3}{2}\alpha\right) + \tau^2\left(\frac{1}{1+\alpha}\left(2 + \frac{3}{2}\cdot\frac{1}{\alpha}\right)\right).
\end{aligned}$$

□

The result in Theorem 1 shows that the DGP in Assumption 3 serves as a counterexample for the validity of the weigthed bootstrap, when applied to the individual contributions of the classical matching estimator (without bias correction) instead of to the individual contributions of the bias-corrected estimator. This result is even more surprising given the fact that in the setting of Assumption 3, the classical matching estimator is unbiased. In order to gain some additional insight into this somewhat surprising result let us compare the individual contributions of the classical nearest neighbor matching estimator $\hat{\tau}_i = (2W_i - 1)(1 + K_i)Y_i$ with those of an *oracle* bias-corrected estimator, defined by $\check{\tau} = \frac{1}{N}\sum_{i=1}^N \check{\tau}_i$ with

$$\check{\tau}_i = (2W_i - 1)(1 + K_i)(Y_i - \mu(X_i, W_i)) + (2W_i - 1)\left(\mu(X_i, W_i) - \mu(X_i, 1 - W_i)\right), \tag{5.2}$$

where in contrast to the bias-corrected estimator $\tilde{\tau}$ the conditional means $\mu(x, w)$ for $w \in \{0, 1\}$ are treated as known. In the following, we will see that although $\sqrt{N}(\hat{\tau} - \tau) = \sqrt{N}(\check{\tau} - \tau) + o_p(1)$ in the setting of Assumption 3, the stochastic properties of the individual contributions $\hat{\tau}_i - \hat{\tau}$ and $\check{\tau}_i - \check{\tau}$ differ quite substantially.

Proposition 2 (First and second moment structure of $\check{\tau}_i$)**.** *Given Assumption 3, the individual contributions of the oracle bias-corrected estimator given in (5.2) satisfy*

(i) $\mathbb{E}[\check{\tau}_i] = \tau$

(ii) $\mathbb{C}\mathrm{ov}[\check{\tau}_i \check{\tau}_j] = 0$ *for* $i \neq j$*.*

(iii) $\mathbb{V}\mathrm{ar}[\check{\tau}_i] = \mathbb{V}\mathrm{ar}[\sqrt{N}(\check{\tau} - \tau)]$

Proof. The straightforward though lengthy proof is given in Appendix B. □

From Proposition 2 it follows that

$$\mathbb{V}\mathrm{ar}[\sqrt{N}(\check{\tau} - \tau)] = \mathbb{V}\mathrm{ar}[\sqrt{N}\check{\tau}] = \mathbb{V}\mathrm{ar}[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\check{\tau}_i] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{V}\mathrm{ar}[\check{\tau}_i]$$

Thus, Proposition 2 states that the individual contributions of the oracle bias-corrected matching estimator $\{\check{\tau}_i\}_{i=1}^N$ are uncorrelated, centered at the ATE, $\mathbb{E}[\check{\tau}_i] = \tau$, and their variance corresponds to the target variance $\mathbb{V}\mathrm{ar}[\check{\tau}_i] = \mathbb{V}\mathrm{ar}[\sqrt{N}(\check{\tau} - \tau)]$. Thus, it is intuitively clear that any random weighted average of the $\{\check{\tau}_i - \check{\tau}\}_{i=1}^N$ will be valid. Heuristically, when the bias-corrected individual contribution $\tilde{\tau}_i$ approximates $\check{\tau}_i$ well enough, asymptotically the $\{\tilde{\tau}_i\}_{i=1}^N$ will satisfy the properties in Proposition 2 from which it follows that the weighted bootstrap is valid. Notice, that the result in Proposition 2 is not so surprising as Abadie and Imbens (2011) showed that the oracle bias-corrected matching estimator has a martingale representation. In particular, they showed that $\sqrt{N}(\hat{\tau} - B_N - \tau) = \sum_{i=1}^{2N} \xi_{N,i}$ with

$$\xi_{N,i} = \begin{cases} \frac{1}{\sqrt{N}}(2W_i - 1)\left(\mu(X_i, W_i) - \mu(X_i, 1 - W_i) - \tau\right) & , 1 \leq i \leq N \\ \frac{1}{\sqrt{N}}(2W_{i-N} - 1)\left(1 + \frac{K_M(i-N)}{M}\right)\left(Y_{i-N} - \mu(X_{i-N}, W_{i-N})\right) & , N+1 \leq i \leq 2N. \end{cases}$$

Moreover, given the specific filtration in Abadie and Imbens (2011), the $\{\xi_{N,i}\}$ are martingale differences, and thus uncorrelated. As $\check{\tau}_i = \sqrt{N}\left(\xi_{N,i} + \tau + \xi_{N,i+N}\right)$, it follows that the individual contributions are also uncorrelated. In contrast to this, the individual contributions of the classical matching estimator $\hat{\tau}_i$ do not have these nice properties. In particular, they are not uncorrelated with the correlation depending on the correlation structure of the matches in a complicated way as seen in (B.1) of Appendix B.

6 Simulations

In practice, in order to calculate the weighted bootstrap estimator $\sqrt{N}\widetilde{T}^*$ in (3.1), one needs to choose a bias estimator $\widehat{B}_N$. The asymptotic validity of the weighted bootstrap is derived using a bias estimator based on nonparametric series estimators for $\mu(x,w)$ for $w \in \{0,1\}$. In particular, it is assumed that the truncation parameter in the series estimation grows at a specific rate to ensure that $\sqrt{N}(\widehat{B}_N - B_N) = o_P(1)$. In finite samples one of course still needs to choose at which point to truncate. In this section we will conduct a simulation study to illustrate how sensitive the weighted bootstrap performance can be to the selected bias-correction. In particular, we will see how popular simple choices such as using linear regressions, which corresponds to only taking polynomial bases up to order one into account, may lead to a very poor performance of the weighted bootstrap variance estimator. All the simulations are based on a DGP satisfying the conditions in Assumption 4 with $c = 1$ and specific choices of $\mu(x,0)$ and $\sigma^2(x,0)$.

All the simulations in this section were based on $S = 10\,000$ simulation runs. We varied the number of observations according to $N \in \{100, 250, 500, 1\,000, 2\,000\}$. Furthermore, we varied the expected balancedness of the design by considering $\alpha \in \{10, 5, 2, 1, 0.5, 0.2, 0.1\}$. For each simulated data set, we compute the classical nearest neighbor matching estimator

$$\hat{\tau} = \sum_{i=1}^{N} (2W_i - 1)(1 + K_i)Y_i$$

and two different bias-corrected nearest neighbor matching estimators

$$\tilde{\tau} = \sum_{i=1}^{N} (2W_i - 1)\,(1 + K_i)\,\big(Y_i - \hat{\mu}(X_i, W_i)\big) + (2W_i - 1)\big(\hat{\mu}(X_i, W_i) - \hat{\mu}(X_i, 1 - W_i)\big)$$

that we will denote by $\tilde{\tau}_{\mathrm{c}}$ and $\tilde{\tau}_{\mathrm{lin}}$. The two bias-corrected matching estimators differ in the model used to estimate $\mu(x,w)$ to obtain the predictions $\hat{\mu}(x,w)$ and thus in the choice of the bias estimator used. The estimator $\tilde{\tau}_{\mathrm{lin}}$ uses the popular simple bias estimator based on linear least squares estimation of $\mu(x,w)$. The second estimator $\tilde{\tau}_{\mathrm{c}}$ is even simpler using a bias estimator based on estimating $\mu(x,w)$ by the sample average of the response in treatment arm w. In addition to using the two bias-corrected nearest neighbor matching estimators we will also consider using the *infeasible* oracle bias-corrected estimator

$$\check{\tau} = \sum_{i=1}^{N} (2W_i - 1)\,(1 + K_i)\,\big(Y_i - \mu(X_i, W_i)\big) + (2W_i - 1)\big(\mu(X_i, W_i) - \mu(X_i, 1 - W_i)\big).$$

All the above matching estimators can be written as

$$\omega = \frac{1}{N} \sum_{i=1}^{N} \omega_i \tag{6.1}$$

with ω_i the individual contribution of the respective estimator.

We will consider two versions of the weighted bootstrap. For the first, we will use the Efron type weighted bootstrap based on using $\eta_i^* = M_i^*/\sqrt{N}$, where $(M_1^*, \ldots, M_N^*)$ is multinomially distributed with N trials and N equally likely cells. Applied to the generic matching estimator ω in (6.1), it is given by

$$\sqrt{N}\omega^{*,\mathrm{EF}} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} M_i^* (\omega_i - \omega).$$

Secondly, we will use the wild bootstrap type weights $\eta_i^* = \epsilon_i^*/\sqrt{N}$ where the $\{\epsilon_i^*\}_{i=1}$ follow the Mammen two-point distribution. Applied to the generic matching estimator ω in (6.1), the weighted bootstrap estimator is given by

$$\sqrt{N}\omega^{*,\mathrm{WB}} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i^* (\omega_i - \omega).$$

The focus of the simulation study is to determine how the different choices for $\hat{\mu}(x, w)$ that are used in the bias-correction will affect the estimated variance of the weighted bootstrap given by $\mathbb{V}\mathrm{ar}^*[\omega^{*,k}]$ for $k \in \{\mathrm{WB}, \mathrm{EF}\}$. In each simulation run, these variance estimates will be calculate based on $B = 1\,000$ bootstrap resamples. For each $k \in \{\mathrm{WB}, \mathrm{EF}\}$ the estimate is given by

$$\widehat{\mathbb{V}\mathrm{ar}}^*[\omega^{*,k}] = N \frac{1}{B-1} \sum_{b=1}^{B} \left(\omega_b^{*,k} - \frac{1}{B} \sum_{b=1}^{B} \omega_b^{*,k} \right)^2,$$

where $\sqrt{N}\omega_b^{*,k}$ is the weighted bootstrap estimate from bootstrap replication $b = 1, \ldots, B$.

As an alternative to the above weighted bootstrap based variance estimators we will also consider the plug-in variance estimator proposed in Theorem 7 of Abadie and Imbens (2006) that is not based on resampling, which we denote by $\widehat{\mathbb{V}\mathrm{ar}}^{(\mathrm{AI})}$.

Table 1: Variance estimation results, when $\mu(x, 0) = 0$ and $\sigma^2(x, 0) = 1$ based on simulations with $N = 2\,000$ for various α.

α	10	5	2	1	0.5	0.2	0.1
Asy. Variance	16.45	8.92	4.33	2.75	1.92	1.38	1.20
$\widehat{\mathbb{V}\mathrm{ar}}^{(\mathrm{AI})}$	16.48	8.92	4.33	2.75	1.92	1.38	1.19
$\widehat{\mathbb{V}\mathrm{ar}}^*[\check{\tau}^{*,\mathrm{EF}}]$	16.45	8.93	4.33	2.75	1.92	1.38	1.20
$\widehat{\mathbb{V}\mathrm{ar}}^*[\hat{\tau}^{*,\mathrm{EF}}]$	16.64	9.31	5.25	4.50	5.24	9.32	16.69
$\widehat{\mathbb{V}\mathrm{ar}}^*[\tilde{\tau}^{*,\mathrm{EF}}_{\mathrm{lin}}]$	16.27	8.88	4.32	2.74	1.92	1.38	1.19
$\widehat{\mathbb{V}\mathrm{ar}}^*[\tilde{\tau}^{*,\mathrm{EF}}_{\mathrm{c}}]$	16.36	8.90	4.33	2.74	1.92	1.38	1.20
$\widehat{\mathbb{V}\mathrm{ar}}^*[\check{\tau}^{*,\mathrm{WB}}]$	16.45	8.93	4.33	2.75	1.91	1.38	1.19
$\widehat{\mathbb{V}\mathrm{ar}}^*[\hat{\tau}^{*,\mathrm{WB}}]$	16.64	9.31	5.25	4.50	5.25	9.31	16.70
$\widehat{\mathbb{V}\mathrm{ar}}^*[\tilde{\tau}^{*,\mathrm{WB}}_{\mathrm{lin}}]$	16.28	8.88	4.32	2.74	1.91	1.38	1.19
$\widehat{\mathbb{V}\mathrm{ar}}^*[\tilde{\tau}^{*,\mathrm{WB}}_{\mathrm{c}}]$	16.36	8.91	4.33	2.74	1.91	1.38	1.19

All the above variance estimators will be compared in terms of their ability to unbiasedly estimate the asymptotic variance of the matching estimator. This will be done by comparing their average over the simulations with the asymptotic variance.

6.1 Baseline setting: $\mu(x, 0) = 0$ and $\sigma^2(x, 0) = 1$

The first setting corresponds to Assumption 3 used to show that bias-correction is necessary even when the classical matching estimator is unbiased. The results for the various variance estimators are given in Table 1 for the largest considered sample size of $N = 2\,000$. The first row gives the target value of the asymptotic variances of the nearest neighbor matching estimator for different values of α as derived in Proposition 1. The second row shows that the mean over the simulations of the variance estimator of Abadie and Imbens (2006) is very close to the target variance. The next four rows provide results for the various bootstrap based variance estimators using the Efron type weights. We immediately see that the results for the infeasible oracle bias-corrected estimator $\check{\tau}$ are very good. We also see, as shown in the theory, that, the estimator based on the individual contributions of the classical matching estimator $\hat{\tau}$ performs very poorly especially for small values of α. As for the feasible bias-corrected estimators $\tilde{\tau}_{\text{lin}}$ and $\tilde{\tau}_{\text{c}}$, both estimate the bias correctly and it is seen that the variance estimators based on these perform very well. Notice that $\tilde{\tau}_{\text{lin}}$ uses an overparametrized model to estimate the conditional mean $\mu(x, w)$ for $w \in \{0, 1\}$ as includes an additional term that is linear in the regressor although this is not necessary. The additional estimation uncertainty induced by this leads to the slightly poorer performance of $\widehat{\mathbb{V}\text{ar}}^*[\tilde{\tau}_{\text{lin}}^{*,\text{EF}}]$ compared to $\widehat{\mathbb{V}\text{ar}}^*[\tilde{\tau}_{\text{c}}^{*,\text{EF}}]$, which is most clearly seen for large values of α. Finally, the last four rows contain the results for the resampling based estimators using the wild bootstrap type weights. Comparing these with the results for the Efron type weights, we see that they are virtually identical. Thus, choosing wild bootstrap weights does not improve the performance of the variance estimator over Efron-type weights. In fact, for all the remaining considered simulation designs this was also the case. Hence, we will only report the results obtained from using the Efron-type weigths.

6.2 Linear case: $\mu(x, 0) = -1 - 2x$ and $\sigma^2(x, 0) = 1$

In this setting, the bias-corrected estimator $\tilde{\tau}_{\text{c}}$ will not estimate the bias correctly. Following the arguments outlined at the end of Section 4 it is possible to show that $\tau = \mathbb{E}[\tau(X)] = \mathbb{E}[1 - (-1 - 2X)] = 3$ and $\mathbb{E}[(\tau(X) - \tau)^2] = 1/3$, which yields $\mathbb{V}\text{ar}\left[\sqrt{N}(\hat{\tau} - \tau)\right] \to 4/3 + \frac{\alpha}{1+\alpha}\left[2 + \frac{3}{2}\alpha\right]$. This is used to calculate the values of the asympotic variance given in the first row of Table 2, which contains the simulation results for the largest considered sample size of $N = 2\,000$. In contrast to the baseline setting, the variance estimator of Abadie and Imbens (2006) is seen to be positively biased for all α. Again the infeasible oracle bias-corrected estimator is unbiased. The weighted bootstrap variance estimator based on the classical matching estimator performs very poorly over all α. The performance of the feasible bias-

Table 2: Variance estimation results, when $\mu(x,0) = -1-2x$ and $\sigma^2(x,0) = 1$ based on simulations with $N = 2\,000$ for various α.

α	10	5	2	1	0.5	0.2	0.1
Asy. Variance	16.79	9.25	4.67	3.08	2.25	1.72	1.53
$\widehat{\mathbb{V}\text{ar}}^{(\text{AI})}$	23.78	15.64	9.77	6.91	4.81	2.99	2.23
$\widehat{\mathbb{V}\text{ar}}^*[\check{\tau}^{*,\text{EF}}]$	16.78	9.26	4.66	3.08	2.25	1.72	1.53
$\widehat{\mathbb{V}\text{ar}}^*[\hat{\tau}^{*,\text{EF}}]$	80.14	40.07	16.02	8.42	5.56	7.31	13.87
$\widehat{\mathbb{V}\text{ar}}^*[\tilde{\tau}_{\text{lin}}^{*,\text{EF}}]$	16.61	9.21	4.65	3.08	2.25	1.72	1.53
$\widehat{\mathbb{V}\text{ar}}^*[\tilde{\tau}_{\text{c}}^{*,\text{EF}}]$	21.98	11.92	5.77	3.66	2.56	1.84	1.59

corrected variance estimator based on $\tilde{\tau}_{\text{lin}}$ that uses the correct model to estimate the bias is nearly as good as the infeasible estimator. However, the estimator based on $\tilde{\tau}_{\text{c}}$, that uses the incorrect model to estimate components of the bias performs worse. It still offers an improvement over the estimator based on the classical matching estimator and on the variance estimator of Abadie and Imbens (2006), but it is clearly biased for all α.

6.3 Quadratic case: $\mu(x,0) = -3 + 8(x-0.5)^2$ and $\sigma^2(x,0) = 1$

In this setting, both the feasible bias-corrected estimators will not estimate the bias correctly. For the present setting one gets $\tau = \mathbb{E}[\tau(X)] = \mathbb{E}[1-(3+8(X-0.5)^2)] = 10/3$ and $\mathbb{E}[(\tau(X)-\tau)^2] = 16/45$, which yields $\mathbb{V}\text{ar}\left[\sqrt{N}(\hat{\tau}-\tau)\right] \to 61/45 + \frac{\alpha}{1+\alpha}\left[2+\frac{3}{2}\alpha\right]$. This is used to calculate the values of the asympotic variance given in the first row of Table 3, which contains the simulation results for the largest considered sample size of $N = 2\,000$. The only difference to the previous setting is that now the variance estimators based on both feasible bias-corrected estimators are biased due to the fact that they do not estimate the conditional means in the treatement arms correctly and thus do not estimate the bias correctly. Again,

Table 3: Variance estimation results, when $\mu(x,0) = -3+8(x-0.5)^2$ and $\sigma^2(x,0) = 1$ based on simulations with $N = 2\,000$ for various α.

α	10	5	2	1	0.5	0.2	0.1
Asy. Variance	16.81	9.27	4.69	3.11	2.27	1.74	1.55
$\widehat{\mathbb{V}\mathrm{ar}}^{(\mathrm{AI})}$	25.71	17.40	11.19	7.98	5.53	3.36	2.44
$\widehat{\mathbb{V}\mathrm{ar}}^*[\check{\tau}^{*,\mathrm{EF}}]$	16.80	9.29	4.68	3.10	2.27	1.74	1.55
$\widehat{\mathbb{V}\mathrm{ar}}^*[\hat{\tau}^{*,\mathrm{EF}}]$	102.05	50.97	20.25	10.34	6.26	7.23	13.51
$\widehat{\mathbb{V}\mathrm{ar}}^*[\tilde{\tau}_{\mathrm{lin}}^{*,\mathrm{EF}}]$	22.41	12.13	5.87	3.72	2.60	1.87	1.62
$\widehat{\mathbb{V}\mathrm{ar}}^*[\tilde{\tau}_{\mathrm{c}}^{*,\mathrm{EF}}]$	22.36	12.12	5.87	3.72	2.60	1.87	1.62

similar to the baseline case the inclusion of a linear term merely leads to additional estimation uncertainty as the linear component does not help in estimating the bias in this particular setting.

6.4 Allowing for nonconstant $\sigma^2(x,0)$

Finally, we considered settings with $\sigma^2(x,0) = 4(x-0.5)^2$. As mentioned at the end of Section 4 this choice will alter the limit variance of the nearest neighbor matching estimator in comparison to the cases with $\sigma^2(x,0) = 1$. We looked at all three conditional mean settings with this additional choice for the conditional variance. Qualitatively, the results are not changed. In particular, both weighted bootstrap versions still yield essentially identical results in terms of variance estimation bias.

7 Conclusion

In this article, we have seen that consistent bias-correction not only affects point estimation of matching estimators for treatment effects, but also impacts the validity of weighted bootstrap based inference procedures. In particular, by means of a simple DGP we have shown that the weighted bootstrap applied to the classical (non-bias-corrected) nearest neighbor matching estimator for the ATE is not valid as the corresponding variance estimator does not converge to the correct asymptotic limit variance – even though the classical nearest neighbor matching estimator is unbiased in the considered setting. Thus, although the classical matching estimator in this setup is unbiased even in finite samples, we need to apply the weighted bootstrap procedure to its bias-corrected version in order to do valid inference. We have seen that the main reason that the bias-correction must be used when doing weighted bootstrap inference is that unlike the individual contributions of the classical matching estimator, the individual contributions of the (oracle) bias-corrected estimator are uncorrelated. As the feasible individual contributions of the bias-corrected estimator consistently estimate the contributions based on the oracle bias-correction, whenever the bias-correction estimator is consistent, it follows that the individual contributions of the feasible bias-correction estimator will be asymptotically uncorrelated. Being based purely on asymptotic considerations, this says nothing about their behavior in finite samples. In simulations, we have seen that inadequate bias-correction can seriously distort the weighted bootstrap based variance estimates. Moreover, simple rules to choose a model for the bias-correction may perform poorly. This naturally leads to the question of how to choose a good bias-correction procedure in practice when interest lies in inference, a topic that will be addressed in future work.

A Proof of asymptotics for matching estimator in ATE counterexample

This appendix contains the remaining the arguments to prove Proposition 1, in particular, the details of the arguments leading to the two results in (4.1). These will be adressed separately in two lemmas.

Lemma 1. *Given a DGP satisfying Assumption 3, it holds that*

$$\mathbb{E}[K_i \mid W_i = 0] \to \alpha.$$

Proof. From the results in (Abadie and Imbens (2008), Lemma A.2 and Lemma A.3) we know that

$$\mathbb{E}[K_i \mid W_i = 0, \sum_{j=1}^{N} W_j = n_1] = \frac{n_1}{n_0} = \frac{n_1}{N - n_1}. \tag{A.1}$$

Thus, we get

$$\begin{aligned}
&\mathbb{E}[K_i \mid W_i = 0] \\
&= \sum_{n_1=0}^{N-1} \mathbb{E}[K_i \mid W_i = 0, \sum_{j=1}^{N} W_j = n_1] \mathrm{P}\left(\sum_{j=1}^{N} W_j = n_1 \mid W_i = 1\right) \\
&= \sum_{n_1=0}^{N-1} \frac{n_1}{N - n_1} \mathrm{P}\left(\sum_{j=1}^{N} W_j = n_1 \mid W_i = 1\right) \\
&= \sum_{n_0=1}^{N} \frac{N - n_0}{n_0} \mathrm{P}\left(\sum_{j=1}^{N} (1 - W_j) = n_0 \mid W_i = 1\right) \\
&= N \sum_{n_0=0}^{N-1} \frac{1}{n_0 + 1} \mathrm{P}\left(\sum_{j=1, j\neq i}^{N} (1 - W_j) = n_0\right) - 1.
\end{aligned}$$

Noticing that $\sum_{j=1,j\neq i}^{N}(1 - W_j)$ is a binomial random variable with $N - 1$ trials and success probability of each trial $p = \frac{1}{1+\alpha}$, we can appeal to the closed form formula for the first shifted reciprocal moment of a binomial given in Chao and Strawderman (1972) to get

$$\sum_{n_0=0}^{N-1} \frac{1}{n_0 + 1} \mathrm{P}\left(\sum_{j=1, j\neq i}^{N} (1 - W_j) = n_0\right) = \frac{1 - \left(\frac{\alpha}{1+\alpha}\right)^2}{N \frac{1}{1+\alpha}}$$

which yields

$$\mathbb{E}[K_i \mid W_i = 0] = N \frac{1 - \left(\frac{\alpha}{1+\alpha}\right)^2}{N \frac{1}{1+\alpha}} - 1 = \alpha - (1 + \alpha) \left(\frac{\alpha}{1 + \alpha}\right)^N \to \alpha.$$

□

Lemma 1 proves the first statement in (4.1). The remaining statement in (4.1) is covered by the next lemma, thus completing the proof of Proposition 1.

Lemma 2. *Given a DGP satisfying Assumption 3, it holds that*

$$\mathbb{E}[K_i^2 \mid W_i = 0] \to \alpha + \frac{3}{2}\alpha^2.$$

Proof. From the results in (Abadie and Imbens (2008), Lemma A.2 and Lemma A.3) we know that

$$\mathbb{E}[K_i^2 \mid W_i = 0, \sum_{j=1}^{N} W_j = n_1] = \frac{n_1}{n_0} + \frac{3}{2}\frac{n_1(n_1-1)(n_0+8/3)}{n_0(n_0+1)(n_0+2)}.$$

Thus, we get

$$\begin{aligned}
&\mathbb{E}[K_i^2 \mid W_i = 0] \\
&= \sum_{n_1=0}^{N-1} \mathbb{E}[K_i^2 \mid W_i = 0, \sum_{j=1}^{N} W_j = n_1] \mathrm{P}\left(\sum_{j=1}^{N} W_j = n_1 \mid W_i = 1\right) \\
&= \sum_{n_1=0}^{N-1} \frac{n_1}{n_0} \mathrm{P}\left(\sum_{j=1}^{N} W_j = n_1 \mid W_i = 1\right) \\
&+ \frac{3}{2}\sum_{n_1=0}^{N-1} \frac{n_1(n_1-1)(n_0+8/3)}{n_0(n_0+1)(n_0+2)} \mathrm{P}\left(\sum_{j=1}^{N} W_j = n_1 \mid W_i = 1\right) \\
&= \mathbb{E}[K_i \mid W_i = 0] \\
&+ \frac{3}{2}\sum_{n_1=0}^{N-1} \frac{n_1(n_1-1)(n_0+8/3)}{n_0(n_0+1)(n_0+2)} \mathrm{P}\left(\sum_{j=1}^{N} W_j = n_1 \mid W_i = 1\right),
\end{aligned}$$

where the last equality follows from (A.1). From Lemma 1 we know that the first summand converges to α. So we are left to show that

$$\sum_{n_1=0}^{N-1} \frac{n_1(n_1-1)(n_0+8/3)}{n_0(n_0+1)(n_0+2)} \mathrm{P}\left(\sum_{j=1}^{N} W_j = n_1 \mid W_i = 1\right) \to \alpha^2 \tag{A.2}$$

Using similar arguments as above we arrive at

$$\begin{aligned}
&\sum_{n_0=0}^{N-1} \frac{(N-(n_0+1))(N-(n_0+2))(n_0+11/3)}{(n_0+1)(n_0+2)(n_0+3)} \mathrm{P}\left(\sum_{j=1, j\neq i}^{N} (1-W_j) = n_0\right) \\
&=: \sum_{n_0=0}^{N-1} g_N(n_0) \mathrm{P}\left(\sum_{j=1, j\neq i}^{N} (1-W_j) = n_0\right).
\end{aligned}$$

Using $n_0+11/3 = n_0+3+2/3$, multiplying out and splitting up the sums in $g_N(n_0)$, we get

$$g_N(n_0) = g_N^{(1)}(n_0) + g_N^{(2)}(n_0) + g_N^{(3)}(n_0)\frac{N^2}{(n_0+1)(n_0+2)}$$

with

$$g_N^{(1)}(n_0) = \frac{N^2}{(n_0+1)(n_0+2)}$$
$$g_N^{(2)}(n_0) = -\frac{2}{3}\frac{N}{(n_0+1)(n_0+3)} - \frac{2}{3}\frac{N}{(n_0+2)(n_0+3)}$$
$$g_N^{(3)}(n_0) = 1 - \frac{N}{n_0+1} - \frac{N}{n_0+2} + \frac{2/3}{n_0+3}$$

Recall that $\sum_{j=1,j\neq i}^{N}(1-W_j)$ is a binomial random variable with $N-1$ trials and sucess probability of each trial $p = \frac{1}{1+\alpha}$, thus we can directly use the results from Chao and Strawderman (1972) to get

$$\sum_{n_0=0}^{N-1} g_N^{(3)}(n_0)\mathrm{P}\left(\sum_{j=1,j\neq i}^{N}(1-W_j) = n_0\right) \to 1 - 2(1-\alpha) = -2\alpha - 1.$$

In order to deal with the corresponding term $g_N^{(1)}(n_0)$ we need the second reciprocal moments. In contrast to the first reciprocal moment that are no nice analytic expressions. However, we can appeal to the asymptotic expansions derived in Wuyungaowa and Wang (2008). In particular, their Corollary 1 implies that

$$N^2\sum_{n_0=0}^{N-1}\frac{1}{(n_0+1)^2}\mathrm{P}\left(\sum_{j=1,j\neq i}^{N}(1-W_j) = n_0\right)$$
$$= N^2\frac{1}{(N\frac{1}{1+\alpha})^2}\left(1 + O\left(\frac{1}{N}\right)\right) \to (1+\alpha)^2.$$

As $\frac{N^2}{(n_0+2)^2} \leq g_N^{(1)}(n_0) \leq \frac{N^2}{(n_0+1)^2}$, we have just derived an upper bound for term with $g_N^{(1)}(n_0)$. Notice, that

$$\frac{1}{(n_0+2)^2} = \frac{1}{(n_0+1)^2} + \frac{-2n_0-3}{(n_0+1)^2(n_0+2)^2}$$

and

$$0 \geq \frac{-2n_0-3}{(n_0+1)^2(n_0+2)^2} \geq \frac{-2(n_0+2)}{(n_0+1)^2(n_0+2)^2} \geq \frac{-2}{(n_0+1)^3}.$$

By Theorem 1 in Wuyungaowa and Wang (2008) we get

$$N^2\sum_{n_0=0}^{N-1}\frac{1}{(n_0+1)^3}\mathrm{P}\left(\sum_{j=1,j\neq i}^{N}(1-W_j) = n_0\right) = \frac{N^2}{(N\frac{\alpha}{1+\alpha})^3}\left(1 + O\left(\frac{1}{N}\right)\right) \to 0,$$

which implies for the lower bound of the term involving $g_N^{(1)}(n_0)$ that

$$N^2 \sum_{n_0=0}^{N-1} \frac{1}{(n_0+2)^2} \mathrm{P}\left(\sum_{j=1,j\neq i}^{N} (1-W_j) = n_0\right)$$
$$= N^2 \sum_{n_0=0}^{N-1} \frac{1}{(n_0+1)^2} \mathrm{P}\left(\sum_{j=1,j\neq i}^{N} (1-W_j) = n_0\right) + O\left(\frac{1}{N}\right) \qquad \to (1+\alpha)^2$$

and $N^2 \sum_{n_0=0}^{N-1} g_N^{(1)}(n_0) \mathrm{P}\left(\sum_{j=1,j\neq i}^{N}(1-W_j) = n_0\right) \to (1+\alpha)^2$. Similar arguments to the ones for $g_N^{(1)}(n_0)$ yield

$$N^2 \sum_{n_0=0}^{N-1} g_N^{(2)}(n_0) \mathrm{P}\left(\sum_{j=1,j\neq i}^{N} (1-W_j) = n_0\right) \to 0$$

and putting everything together we get

$$N^2 \sum_{n_0=0}^{N-1} g_N(n_0) \mathrm{P}\left(\sum_{j=1,j\neq i}^{N} (1-W_j) = n_0\right) \to (1+\alpha)^2 - 2\alpha - 1 = \alpha^2$$

establishing (A.2). □

B First and second moments of individual contributions

In this appendix we derive the first two moments of the individual contributions of the infeasible oracle bias-corrected nearest neighbor matching estimator, which then proves Proposition 2. The results are collected in two lemmas.

Lemma 3 (First moments of the individual contributions)**.** *For the individual contributions of the infeasible bias-corrected estimator we have*

$$\mathbb{E}[\check{\tau}_i] = \tau \text{ and } \mathbb{E}[\check{\tau}_i \mid W_i = w] = \mathbb{E}[\tau(X_i) \mid W_i = w] \text{ for } w \in \{0,1\}.$$

Moreover, for any DGP with a homogenous treatment effect, that is with $\tau(x) = \tau$ for all x, we get $\mathbb{E}[\check{\tau}_i \mid W_i = w] = \tau$. (In our simple DGP with $\mu(x,1) = \tau$ for all x and $\mu(x,0) = 0$ for all x, the treatment effect is homogenous.)

Proof. For the components of the (infesible) bias-corrected estimator we have

$$\begin{aligned}
\mathbb{E}[\check{\tau}_i \mid W_i = 1] &= \mathbb{E}\big[\mathbb{E}[(1+K_i)(Y_i - \mu(X_i,1)) + \mu(X_i,1) - \mu(X_i,0) \mid \mathbf{X}, W_i = 1] \mid W_i = 1\big] \\
&= \mathbb{E}\big[(1+K_i)\underbrace{\mathbb{E}[Y_i - \mu(X_i,1) \mid \mathbf{X}, W_i = 1]}_{=0} \mid W_i = 1\big] + \mathbb{E}\big[\mu(X_i,1) - \mu(X_i,0) \mid W_i = 1\big] \\
&= \mathbb{E}\big[\mu(X_i,1) - \mu(X_i,0) \mid W_i = 1\big] \\
&= \mathbb{E}\big[\mathbb{E}[Y_i(1) \mid X_i] - \mathbb{E}[Y_i(0) \mid X_i] \mid W_i = 1\big] \\
&= \mathbb{E}[\tau(X_i) \mid W_i = 1]
\end{aligned}$$

where $\tau(x) := \mathbb{E}[Y(1) - Y(0) \mid X = x]$. Similarly, we get

$$\mathbb{E}[\check{\tau}_i \mid W_i = 0] = \mathbb{E}[\tau(X_i) \mid W_i = 0].$$

such that

$$\begin{aligned}
\mathbb{E}[\check{\tau}_i] &= \Pr(W_i = 1)\mathbb{E}[\tau(X_i) \mid W_i = 1] + \Pr(W_i = 0)\mathbb{E}[\tau(X_i) \mid W_i = 0]) \\
&= \mathbb{E}[\tau(X_i)] = \tau.
\end{aligned}$$

□

The next lemma provides the second moment structure of the individual contributions of the (infeasible) bias-corrected estimator.

Lemma 4 (Second moment properties of individual contributions to (infeasible) bias corrected matching estimator)**.** *For the individual contributions of the bias-corrected estimator we have*

(a) For $i \neq j$, $\mathbb{C}\mathrm{ov}[\check{\tau}_i\check{\tau}_j \mid W_i = w, W_j = u] = 0$ for all $w, u \in \{0,1\}$.

Moreover, as for $i \neq j$, it holds that $\mathbb{E}\big[\mathbb{E}[\check{\tau}_i \mid W_i, W_j]\mathbb{E}[\check{\tau}_j \mid W_i, W_j]\big] = \mathbb{E}\big[\mathbb{E}[\check{\tau}_i \mid W_i]\big]\mathbb{E}\big[\mathbb{E}[\check{\tau}_j \mid W_j]\big]$ we get

$$\mathbb{C}\mathrm{ov}[\check{\tau}_i\check{\tau}_j] = 0 \textit{ for } i \neq j.$$

(b) For $i = j$, we get

$$\mathbb{V}\mathrm{ar}[\check{\tau}_i \mid W_i = w] = \mathbb{E}[(1+K_i)^2(Y_i - \mu(X_i,w))^2 \mid W_i = w] + \mathbb{V}\mathrm{ar}[\tau(X_i) \mid W_i = w].$$

so that as $\mathbb{E}[\check{\tau}_i \mid W_i = w] = \mathbb{E}[\tau(X_i) \mid W_i = w]$ *we get*

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\check{\tau}_i] &= \mathbb{E}[(1+K_i)^2(Y_i - \mu(X_i, W_i))^2] + \mathbb{E}\big[\mathbb{V}\mathrm{ar}[\tau(X_i) \mid W_i\big] + \mathbb{V}\mathrm{ar}\big[\mathbb{E}[\tau(X_i) \mid W_i]\big] \\ &= \mathbb{E}[(1+K_i)^2(Y_i - \mu(X_i, W_i))^2] + \mathbb{V}\mathrm{ar}[\tau(X_i)].\end{aligned}$$

(c) *Finally, notice that due to the result in (a) we get that for the (infeasible) bias corrected matching estimator*

$$\mathbb{V}\mathrm{ar}[\sqrt{N}(\check{\tau} - \tau)] = \mathbb{V}\mathrm{ar}[\sqrt{N}\check{\tau}] = \mathbb{V}\mathrm{ar}[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\check{\tau}_i] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{V}\mathrm{ar}[\check{\tau}_i]$$

Due to the random sampling assumption (and as the K_i are exchangeable) in (b), we get that $\mathbb{V}\mathrm{ar}[\check{\tau}_i] = \mathbb{V}\mathrm{ar}[\check{\tau}_j]$ for all $i, j = 1, \ldots, n$, which with (c) implies that

$$\mathbb{V}\mathrm{ar}[\check{\tau}_i] = \mathbb{V}\mathrm{ar}[\sqrt{N}(\check{\tau} - \tau)].$$

As an aside for our simple DGP, as $Y_i(1) = \tau$, and $Y_i(0) \mid X_i \sim \mathcal{N}(0,1)$ the expression in (b) simplifies to

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\check{\tau}_i] &= \mathrm{P}(W_i = 0)\mathbb{E}[(1+K_i)^2(Y_i - \mu(X_i, 0))^2 \mid W_i = 0] + \mathbb{V}\mathrm{ar}[\tau] \\ &= \mathrm{P}(W_i = 0)\mathbb{E}[(1+K_i)^2 \mid W_i = 0] \\ &= \frac{1}{1+\alpha}(1 + 2\mathbb{E}[K_i \mid W_i = 0] + \mathbb{E}[K_i^2 \mid W_i = 0]) \\ &= \mathbb{V}\mathrm{ar}[\sqrt{N}(\check{\tau} - \tau)] \\ &\to 1 + \frac{\alpha}{1+\alpha}\big(2 + \frac{3}{2}\alpha\big).\end{aligned}$$

From these two lemmas, we see that the $\check{\tau}_i$ are uncorrelated with $\mathbb{E}[\check{\tau}_i] = \tau$ and $\mathbb{V}\mathrm{ar}[\check{\tau}_i] = \mathbb{V}\mathrm{ar}[\sqrt{N}(\check{\tau} - \tau)]$ thus proving Proposition 2.

Proof of Lemma 4. For the contributions to the (infeasible) bias-corrected estimator, we get

$$\begin{aligned}&\mathbb{E}[\check{\tau}_i\check{\tau}_j \mid W_i = 1, W_j = 1, i \neq j] \\ &\quad = \mathbb{E}[(1+K_i)(1+K_j)(Y_i - \mu(X_i, 1))(Y_j - \mu(X_j, 1)) \mid W_i = 1, W_j = 1, i \neq j] \\ &\quad + 2\mathbb{E}[(1+K_i)(Y_i - \mu(X_i, 1))\big(\mu(X_j, 1) - \mu(X_j, 0)\big) \mid W_i = 1, W_j = 1, i \neq j] \\ &\quad + \mathbb{E}[\big(\mu(X_i, 1) - \mu(X_i, 0)\big)\big(\mu(X_j, 1) - \mu(X_j, 0)\big) \mid W_i = 1, W_j = 1, i \neq j]] \\ &\quad = \big(\mathbb{E}[\big(\mu(X_i, 1) - \mu(X_i, 0)\big) \mid W_i = 1]\big)^2 = \big(\mathbb{E}[\tau(X) \mid W = 1]\big)^2.\end{aligned}$$

Thus $\mathbb{C}\text{ov}[\check{\tau}_i\check{\tau}_j \mid W_i = 1, W_j = 1, i \neq j] = 0$.

Similarly, if both units are controls we get

$$\begin{aligned}
&\mathbb{E}[\check{\tau}_i\check{\tau}_j \mid W_i = 0, W_j = 0, i \neq j] \\
&\qquad = \mathbb{E}[(-1)(1+K_i)(-1)(1+K_j)(Y_i - \mu(X_i,0))(Y_j - \mu(X_j,0)) \mid W_i = 0, W_j = 0, i \neq j] \\
&\qquad + 2\mathbb{E}[-(1+K_i)(Y_i - \mu(X_i,1))\big(\mu(X_j,1) - \mu(X_j,0)\big) \mid W_i = 0, W_j = 0, i \neq j] \\
&\qquad + \mathbb{E}[\big(\mu(X_i,1) - \mu(X_i,0)\big)\big(\mu(X_j,1) - \mu(X_j,0)\big) \mid W_i = 0, W_j = 0, i \neq j]] \\
&\qquad = \big(\mathbb{E}[\big(\mu(X_i,1) - \mu(X_i,0)\big) \mid W_i = 0]\big)^2 = \big(\mathbb{E}[\tau(X) \mid W = 0]\big)^2
\end{aligned}$$

such that $\mathbb{C}\text{ov}[\check{\tau}_i\check{\tau}_j \mid W_i = 0, W_j = 0, i \neq j] = 0$.

If the units are of different type, then

$$\begin{aligned}
&\mathbb{E}[\check{\tau}_i\check{\tau}_j \mid W_i = 1, W_j = 0] \\
&\qquad = \mathbb{E}[(1+K_i)(-1)(1+K_j)(Y_i - \mu(X_i,1))(Y_j - \mu(X_j,0)) \mid W_i = 1, W_j = 0] \\
&\qquad + \mathbb{E}[(1+K_i)(Y_i - \mu(X_i,1))\big(\mu(X_j,1) - \mu(X_j,0)\big) \mid W_i = 1, W_j = 0] \\
&\qquad + \mathbb{E}[-(1+K_j)(Y_j - \mu(X_j,0))\big(\mu(X_i,1) - \mu(X_i,0)\big) \mid W_i = 1, W_j = 0] \\
&\qquad + \mathbb{E}[\big(\mu(X_i,1) - \mu(X_i,0)\big)\big(\mu(X_j,1) - \mu(X_j,0)\big) \mid W_i = 1, W_j = 0, i \neq j]] \\
&\qquad = \mathbb{E}[\big(\mu(X_i,1) - \mu(X_i,0) \mid W_i = 1]\mathbb{E}[\big(\mu(X_j,1) - \mu(X_j,0) \mid W_j = 0] \\
&\qquad = \mathbb{E}[\tau(X) \mid W = 1]\mathbb{E}[\tau(X) \mid W = 0]
\end{aligned}$$

and $\mathbb{C}\text{ov}[\check{\tau}_i\check{\tau}_j \mid W_i = 1, W_j = 0] = 0$.

Finally, for the case that $i = j$, we get

$$\mathbb{E}[\check{\tau}_i^2 \mid W_i = w] = \mathbb{E}[(1+K_i)^2(Y_i - \mu(X_i,1))^2 \mid W_i = w] + \mathbb{E}[\big(\mu(X_i,1) - \mu(X_i,0)\big)^2 \mid W_i = w].$$

such that

$$\begin{aligned}
\mathbb{V}\text{ar}[\check{\tau}_i \mid W_i = w] &= \mathbb{E}[(1+K_i)^2(Y_i - \mu(X_i,w))^2 \mid W_i = w] \\
&+ \mathbb{E}[\tau^2(X_i) \mid W_i = w] - \big(\mathbb{E}[\tau(X_i) \mid W_i = w]\big) \\
&= \mathbb{E}[(1+K_i)^2(Y_i - \mu(X_i,w))^2 \mid W_i = w] + \mathbb{V}\text{ar}[\tau(X_i) \mid W_i = w].
\end{aligned}$$

□

We finish this appendix by providing some derivations for the individual contributions of the simple matching estimator. In particular, we will see that even in our simple setting they will be correlated with the correlation structure depending on the correlation structure of the matches in a complicated way.

For the first moment of $\hat{\tau}_i$ it holds that

$$\begin{aligned}\mathbb{E}[\hat{\tau}_i \mid W_i = 1] &= \mathbb{E}\big[\mathbb{E}[(1 + K_i)Y_i \mid \mathbf{X}, W_i = 1] \mid W_i = 1\big] \\ &= \mathbb{E}\big[(1 + K_i)\mu(X_i, 1) \mid W_i = 1\big]\end{aligned}$$

and

$$\mathbb{E}[\hat{\tau}_i \mid W_i = 0] = -\mathbb{E}\big[(1 + K_i)\mu(X_i, 0) \mid W_i = 0\big]$$

such that

$$\begin{aligned}\mathbb{E}[\hat{\tau}_i] &= \mathrm{P}(W_i = 1)\mathbb{E}[\hat{\tau}_i \mid W_i = 1] + \mathrm{P}(W_i = 0)\mathbb{E}[\hat{\tau}_i \mid W_i = 0] \\ &= \mathrm{P}(W_i = 1)\mathbb{E}\big[(1 + K_i)\mu(X_i, 1) \mid W_i = 1\big] - \mathrm{P}(W_i = 0)\mathbb{E}\big[(1 + K_i)\mu(X_i, 0) \mid W_i = 0\big].\end{aligned}$$

For our simple DGP, we had $\mathrm{P}(W_i = 1) = \frac{\alpha}{1+\alpha}$, $\mathbb{E}[K_i \mid W_i = 1] \to \frac{1}{\alpha}$, $\mu(x_i, 0) = 0$ and $\mu(X_i, 1) = \tau$. In this case, we get using results from Appendix A that

$$\begin{aligned}\mathbb{E}[\hat{\tau}_i] &= \mathrm{P}(W_i = 1)\mathbb{E}\big[(1 + K_i)\mu(X_i, 1) \mid W_i = 1\big] - \mathrm{P}(W_i = 0)\mathbb{E}\big[(1 + K_i)\mu(X_i, 0) \mid W_i = 0\big] \\ &= \frac{\alpha}{1+\alpha}\tau\mathbb{E}\big[(1 + K_i) \mid W_i = 1\big] - 0 \to \frac{\alpha}{1+\alpha}\tau(1 + \frac{1}{\alpha}) = \tau.\end{aligned}$$

In contrast to the individual contributions of the (infeasible) bias-corrected estimator the expectation of the $\hat{\tau}_i$ converges to τ as opposed to being equal to τ.

As for the second moment structure of the $\hat{\tau}_i$, we get

$$\begin{aligned}&\mathbb{E}[\hat{\tau}_i\hat{\tau}_j \mid W_i = 1, W_j = 1, i \neq j] \\ &\quad= \mathbb{E}[(1 + K_i)(1 + K_j)Y_iY_j \mid W_i = 1, W_j = 1, i \neq j] \\ &\quad= \mathbb{E}[(1 + K_i)(1 + K_j)\mathbb{E}[Y_iY_j \mid \mathbf{X}, W_i = 1, W_j = 1, i \neq j \mid W_i = 1, W_j = 1, i \neq j]] \\ &\quad= \mathbb{E}[(1 + K_i)(1 + K_j)\mu(X_i, 1)\mu(X_j, 1) \mid W_i = 1, W_j = 1, i \neq j]] \\ &\quad= \big(\mathbb{E}[\mu(X_i, 1) \mid W_i = 1]\big)^2 + 2\mathbb{E}[K_i\mu(X_i, 1) \mid W_i = 1]\mathbb{E}[\mu(X_i, 1) \mid W_i = 1] \\ &\quad+ \mathbb{E}[K_iK_j\mu(X_i, 1)\mu(X_j, 1) \mid W_i = 1, W_j = 1, i \neq j]\end{aligned}$$

The last equality makes use of the random sampling assumption and the fact that the K_i are exchangeable. Thus

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}[\hat{\tau}_i, \hat{\tau}_j \mid W_i = 1, W_j = 1, i \neq j] \\
&\qquad = \mathbb{E}[\hat{\tau}_i \hat{\tau}_j \mid W_i = 1, W_j = 1, i \neq j] - \left(\mathbb{E}[\hat{\tau}_i \mid W_i = 1]\right)^2 \\
&\qquad = \mathbb{E}[(1 + K_i)(1 + K_j)\mu(X_i, 1)\mu(X_j, 1) \mid W_i = 1, W_j = 1, i \neq j]] \\
&\qquad\quad - \left(\mathbb{E}[(1 + K_i)\mu(X_i, 1) \mid W_i = 1]\right)^2 \\
&\qquad = \mathbb{E}[K_i K_j \mu(X_i, 1)\mu(X_j, 1) \mid W_i = 1, W_j = 1, i \neq j] - \left(\mathbb{E}[K_i \mu(X_i, 1) \mid W_i = 1]\right)^2 \\
&\qquad = \mathbb{C}\mathrm{ov}[K_i \mu(X_i, 1), K_j \mu(X_j, 1) \mid W_i = 1, W_j = 1, i \neq j] \\
&\qquad \neq 0.
\end{aligned}
$$

By the same argument, we get

$$\mathbb{C}\mathrm{ov}[\hat{\tau}_i, \hat{\tau}_j \mid W_i = 0, W_j = 0, i \neq j] = \mathbb{C}\mathrm{ov}(K_i \mu(X_i, 0), K_j \mu(X_j, 0) \mid W_i = 0, W_j = 0, i \neq j] \neq 0.$$

and

$$\mathbb{C}\mathrm{ov}[\hat{\tau}_i, \hat{\tau}_j \mid W_i = 1, W_j = 0] = -\mathbb{C}\mathrm{ov}(K_i \mu(X_i, 1), K_j \mu(X_j, 0) \mid W_i = 1, W_j = 0] \neq 0.$$

Notice, that for $i \neq j$

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[\hat{\tau}_i \mid W_i, W_j]\mathbb{E}[\hat{\tau}_j \mid W_i, W_j]] &= \mathbb{E}[\mathbb{E}[\hat{\tau}_i \mid W_i]\mathbb{E}[\hat{\tau}_j \mid W_i]] \\
&= \mathbb{E}[\mathbb{E}[\hat{\tau}_i \mid W_i]]\mathbb{E}[\mathbb{E}[\hat{\tau}_j \mid W_i]],
\end{aligned}
$$

which implies that $\mathbb{C}\mathrm{ov}\left[\mathbb{E}[\hat{\tau}_i \mid W_i, W_j], \mathbb{E}[\hat{\tau}_j \mid W_i, W_j]\right] = 0$. Thus, for $i \neq j$

$$\mathbb{C}\mathrm{ov}[\hat{\tau}_i, \hat{\tau}_j] = \mathbb{E}\left[\mathbb{C}\mathrm{ov}[\hat{\tau}_i, \hat{\tau}_j] \mid W_i, W_j\right].$$

In contrast to the individual contributions of the (infeasible) bias corrected matching estimator these are not zero.

To get a better understanding of these expressions let us assume that $\mu(X_i, w) = \mu(w)$ for all x, which implies homogenous treatment effects. (This is assumed in our simple DGP). Then, we get

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}[\hat{\tau}_i, \hat{\tau}_j \mid W_i = 1, W_j = 1, i \neq j] = \mu(1)^2 \mathbb{C}\mathrm{ov}(K_i, K_j \mid W_i = 1, W_j = 1, i \neq j) \\
&\mathbb{C}\mathrm{ov}[\hat{\tau}_i, \hat{\tau}_j \mid W_i = 0, W_j = 0, i \neq j] = \mu(0)^2 \mathbb{C}\mathrm{ov}(K_i, K_j \mid W_i = 0, W_j = 0, i \neq j) \\
&\mathbb{C}\mathrm{ov}[\hat{\tau}_i, \hat{\tau}_j \mid W_i = 1, W_j = 0, i \neq j] = -\mu(1)\mu(0)\mathbb{C}\mathrm{ov}(K_i, K_j \mid W_i = 1, W_j = 0, i \neq j)
\end{aligned}
\tag{B.1}
$$

and can see that the correlation depens on the correlation of the matching in a complicated way.

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.

Chao, M. T. and W. E. Strawderman (1972): "Negative Moments of Positive Random Variables." *Journal of the American Statistical Association*, 67, 429–431.

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

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

Walsh, C., C. Jentsch, and S. T. Hossain (2021): "Nearest neighbor matching: Does the M-out-of-N bootstrap work when the naive bootstrap fails?" SFB 823 Discussion Paper Nr. 05/2021, Sonderforschungsbereich (SFB) 823.

Wuyungaowa and T. Wang (2008): "Asymptotic expansions for inverse moments of binomial and negative binomial," *Statistics and Probability Letters*, 78, 3018–3022.