

IZA DP No. 5361

Bootstrap Inference for K -nearest Neighbour Matching Estimators

Xavier de Luna
Per Johansson
Sara Sjöstedt-de Luna

November 2010

Bootstrap Inference for K -Nearest Neighbour Matching Estimators

Xavier de Luna

Umeå University

Per Johansson

*Uppsala University,
IFAU and IZA*

Sara Sjöstedt-de Luna

Umeå University

Discussion Paper No. 5361
November 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Bootstrap Inference for K -Nearest Neighbour Matching Estimators^{*}

Abadie and Imbens (2008, *Econometrica*) showed that classical bootstrap schemes fail to provide correct inference for K -nearest neighbour (KNN) matching estimators of average causal effects. This is an interesting result showing that bootstrap should not be applied without theoretical justification. In this paper, we present two resampling schemes, which we show provide valid inference for KNN matching estimators. We resample “estimated individual causal effects” (EICE), i.e. the difference in outcome between matched pairs, instead of the original data. Moreover, by taking differences in EICEs ordered with respect to the matching covariate, we obtain a bootstrap scheme valid also with heterogeneous causal effects where mild assumptions on the heterogeneity are imposed. We provide proofs of the validity of the proposed resampling based inferences. A simulation study illustrates finite sample properties.

JEL Classification: C14, C21

Keywords: block bootstrap, subsampling, average causal/treatment effect

Corresponding author:

Per Johansson
Department of Economics
Uppsala University
Box 513
SE-75120 Uppsala
Sweden
E-mail: per.johansson@ifau.uu.se

^{*} We are grateful to Yuri Belyaev and the participants of the 3rd joint IFAU/IZA conference on labor market policy evaluation, 1-2 October 2010, for useful comments. The first author acknowledges the financial support of the Swedish Research Council through the Swedish Initiative for Research on Microdata in the Social and Medical Sciences (SIMSAM), the Ageing and Living Condition Program and grant 70246501. The second author acknowledges financial support of the Swedish Council for Working Life and Social Research (grant 20042005).

1 Introduction

K -nearest neighbour (KNN) matching estimators (Abadie and Imbens, 2006) are popular for estimating, non-parametrically, the average causal effect of a binary variable on an outcome in observational studies, where confounders are observed and controlled for. Abadie and Imbens (2008) showed that classical bootstrap schemes fail to provide correct inference for KNN matching estimators. In particular, bootstrap variance estimators were shown to be biased. The resampling schemes considered by Abadie and Imbens were typical in the sense that bootstrap copies were obtained by sampling with replacement from the data (assumed to be a random sample). Their finding is interesting because it highlights the fact that bootstrap inference does not always work and hence must be taken with caution in situations lacking theoretical justification.

In particular, more complex resampling schemes may be needed in cases not falling within the usual range of applicability of the bootstrap. This is the case for KNN matching estimators which, as Abadie and Imbens (2008) noted, are unsmooth functions of the data. In this paper, we present two resampling schemes, which we show provide correct inference. We resample “estimated individual causal effects” (EICE), i.e. the differences in outcomes between matched pairs, whose average forms the matching estimator of interest. This has two major advantages. Because matching is performed only once on the original data, the bootstrap scheme is extremely fast to perform (in contrast to bootstrapping the original data, which implies that matching has to be performed for each bootstrap replicate). However, most importantly, bootstrapping EICEs as described formally below yields valid inference. Note that the KNN matching estimator is a smooth function of the EICEs. Still the latter cannot be naively resampled because they are dependent. This is tackled by ordering the EICEs with respect to the matching covariate (or a score summarizing several covariates) and by using a circular block bootstrapping scheme, which is used for stationary and non-stationary time series (e.g., Künsch, 1989, Carlstein, 1986, Lahiri, 1992 and Sjöstedt, 2000). In contrast to typical time series situations the bootstrapped EICEs have a known dependence structure and we therefore propose using this knowledge to find an appropriate blocking scheme.

The above sketched resampling scheme does typically not work if the EICEs have mean (conditional on the covariate and the assignment to the causal agent) which varies, i.e. we have heterogeneous causal effects. We overcome this difficulty by taking differences in EICEs ordered with respect to the matching covariate, and again using a block bootstrap strategy on these differences. This second bootstrap scheme yields correct inference under rather general forms of heterogeneity in the causal effects.

Two inferential procedures are considered for constructing confidence intervals, either using a subsampling variance estimate together with the asymptotic normality of the estimator, or using bootstrap estimated quantiles of the distribution of the estimator. We provide proofs of the validity of the different resampling based inferences proposed, relying on previous results obtained on block-bootstrapping for non-stationary sequences (Sjöstedt, 2000). The resampling inference studied herein constitutes a new and not straightforward application area of such results which have previously been used in time series and spatial data contexts (Ekström and Sjöstedt-de Luna, 2004).

In the next section KNN matching estimators are introduced in the context of the potential outcome framework. Section 3 summarizes our theoretical justifications of the bootstrap schemes. A simulation study illustrating finite sample properties is presented in Section 4. Abadie and Imbens (2006) matching based variance estimators are used as benchmarks. All proofs are delayed to the Appendix.

2 Matching estimators for average causal effects

Consider the situation where we observe the variables Y , Z , and X for a random sample of individuals, where Z is binary (causal agent: treatment, intervention, etc.), Y is an outcome on which the causal effect of Z is to be evaluated, and X is a vector of covariates not affected by Z .

Assume that the sample consists in n individuals with $Z = 1$ (group of interest, often called treated) and N individuals with $Z = 0$ (reference group), indexed such that $Z_i = 1$ for individuals $i = 1, \dots, n$ and $Z_i = 0$ for $i = n + 1, \dots, n + N$.

The effect evaluation we consider here consists in estimating the following average

causal effect (in the literature often called average treatment effect on the treated)

$$\tau = E(Y_i(1) - Y_i(0) | Z_i = 1),$$

where $Y_i(1)$ and $Y_i(0)$ are the so called potential outcomes, i.e. outcomes arising when individuals are assigned to $Z_i = 1$ and $Z_i = 0$ respectively; see Neyman (1923), Rubin (1974), Imbens (2004). Note that only one of the two potential outcomes $Y_i(1)$ and $Y_i(0)$ is observed for individual i .

We assume that if a given individual i is assigned a given value for Z_i , this does not affect the values taken by the potential outcomes for this individual or any other individual in the study (stable individual value assumption, Rubin, 1991). Moreover, the following assumptions are assumed to hold in the sequel, thereby granting, for instance, that τ is identified (e.g., Rosenbaum and Rubin, 1983, Abadie and Imbens, 2006).

(A.1) : *Conditional on the assignment to the causal agent $Z_i = j$, $(Y_i(j), X_i)$ are independently drawn from the distribution law $\mathcal{L}\{(Y_i(j), X_i) \mid Z_i = j\}$, for $j = 0, 1$. Let also $n^s / (n^s + N) \rightarrow \alpha$ as $n \rightarrow \infty$, $0 < \alpha < 1$, for some $s \geq 1$.*

(A.2) : *For all x in \mathcal{X} , where \mathcal{X} is the support of the distribution of the random variable X_i :*

- i) Z_i and $Y_i(0)$ are independently distributed given $X_i = x$,*
- ii) $\Pr(Z_i = 1 | X_i = x) < 1$.*

Assumption (A.2-i) is violated if there are unobserved confounders, that is variables that affect both Z_i and $Y_i(0)$ which are not included in X_i . By assumption (A.2-ii), we ensure that all those in the group of interest could as well have been in the reference group for a given X_i .

Another commonly targeted average causal effect is $E(Y_i(1) - Y_i(0))$. The latter is equal to τ , for instance, when $Y_i(1) - Y_i(0) = \tau$ (constant individual causal effect) for all individuals in the population. However, in general the latter does not hold and stronger assumptions are needed to identify $E(Y_i(1) - Y_i(0))$; see, e.g., Imbens (2004). Moreover,

in many applications the group of interest has far fewer individuals than the reference group and it is therefore most realistic to focus on τ rather than on $E(Y_i(1) - Y_i(0))$.

We now define the KNN matching estimator as

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^n (Y_i(1) - \hat{Y}_i(0)), \quad (1)$$

where $\hat{Y}_i(0)$ is a K -nearest neighbour estimator of the unobserved outcome $Y_i(0)$. Thus, for $K = 1$ we have

$$\hat{Y}_i(0) = Y_j(0) \text{ such that } j = \underset{j=n+1, \dots, n+N}{\operatorname{argmin}} |X_j - X_i|,$$

where $|\cdot|$ is a vector norm. Generally, for $K > 0$ and for $i = 1, \dots, n$, denote by $j_K(i)$ the index $j \in \{n+1, \dots, n+N\}$ that makes $\sum_{l=n+1}^{n+N} \mathbb{1}\{|X_l - X_i| \leq |X_j - X_i|\} = K$, where $\mathbb{1}\{A\}$ is the indicator function which is equal to one when A is true and zero otherwise. The set of indices for the K -nearest matches for individual i is then $\mathcal{J}_K(i) = \{j_1(i), j_2(i), \dots, j_K(i)\}$. Then, a K -nearest neighbour estimator of the unobserved outcome $Y_i(0)$ is $\hat{Y}_i(0) = \frac{1}{K} \sum_{j \in \mathcal{J}_K(i)} Y_j(0)$, i.e. the average of the K observed reference individuals which are closest to individual i in terms of X .

Abadie and Imbens (2006) derived the asymptotic properties of (1), and under given regularity conditions the KNN matching estimator is asymptotically normal. They consider the marginal variance $\operatorname{Var}(\hat{\tau})$ as well as the conditional variance $\operatorname{Var}(\hat{\tau}|\mathbf{X}, \mathbf{Z})$, where \mathbf{X} and \mathbf{Z} are vectors containing the observed values X_i and Z_i , $i = 1, \dots, n+N$, respectively, and introduce consistent estimators for these two variances (Abadie and Imbens, 2006, Theorems 6 and 7), which we shall use as benchmarks in the Monte Carlo study below.

3 Resampling estimated individual causal effects

We now introduced bootstrapping and subsampling schemes that can be used to perform inference on τ .

Denote by

$$D_{in} = Y_i(1) - \hat{Y}_i(0)$$

the estimated individual causal effects (EICE). Hence, the KNN matching estimator (1) can be written as $\hat{\tau} = 1/n \sum_{i=1}^n D_{in}$. Note that the EICEs depend on X_i through the matching process.

(H.1) X_i is a scalar- and continuous-valued random variable with compact and convex support \mathcal{X} and density function $f(x)$ such that $0 < f(x) < \infty$ for $x \in \mathcal{X}$.

In the multivariate covariate case, the covariate vector is typically replaced by a one-dimensional continuous summarizing score (e.g., Rosenbaum and Rubin, 1983, Hansen, 2008, Waernbaum, 2010) to avoid the curse of dimensionality, thereby falling back into our context.

From now on we consider the EICEs to be ordered according to their corresponding X_i values:

$$D_{1n}, D_{2n}, \dots, D_{nn}, \quad \text{where}$$

$$X_{1n} \leq X_{2n} \leq \dots \leq X_{nn},$$

with X_{in} , $i = 1, \dots, n$, the sequence of ordered (ascendant) X'_i 's. The EICEs are locally dependent, because two EICEs may be computed using one or several identical individuals from the reference group. This dependence implies that we cannot bootstrap the EICEs as if they were independently distributed and we henceforth consider block resampling schemes.

3.1 Block bootstrap

We now describe a (circular) block bootstrap scheme and give conditions under which it is theoretically justified for estimating the variance of $\hat{\tau}$ and for constructing confidence intervals; see Politis and Romano (1992) and Sjöstedt (2000).

Construct consecutive blocks of data of size $b < n$ such that $\mathcal{B}_j = \{D_{jn}, D_{j+1,n}, \dots, D_{j+b-1,n}\}$, $j = 1, \dots, n$, where $D_{n+j,n} = D_{jn}$; see Figure 3. Furthermore, let $D_{\cdot jn} = \sum_{i=j}^{j+b-1} D_{in}$. A resampling copy (a pseudo sample) of $\{D_{\cdot jn}\}_{j=1}^n$ is denoted $\{D_{\cdot jn}^*\}_{j=1}^n$ and is constructed by drawing n items¹ with replacement from $\{D_{\cdot jn}\}_{j=1}^n$. Define $\bar{D}_{\cdot n}^* = \frac{1}{nb} \sum_{j=1}^n D_{\cdot jn}^*$

¹We could instead draw n^* items such that $n^*b \approx n$. This is not necessary but probably slightly better.

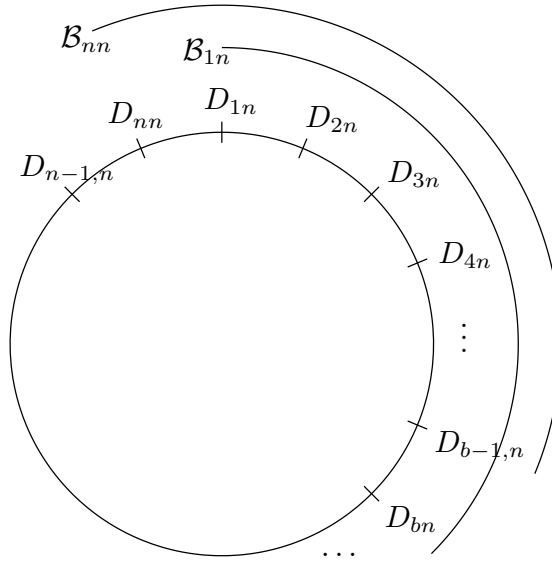


Figure 1: Circular blocking scheme.

and $\bar{D}_{\cdot n} = \frac{1}{nb} \sum_{j=1}^n D_{jn}$. Because of the circular blocking scheme we have that $\bar{D}_{\cdot n} = \hat{\tau}$. Under some additional assumptions given below we show that the distribution $\mathcal{L}\{\sqrt{bn}(\bar{D}_{\cdot n}^* - \bar{D}_{\cdot n}) | \text{data}\}$ asymptotically mimics the centered distribution of $\sqrt{n}\hat{\tau}$ conditional on \mathbf{X} and \mathbf{Z} .

As noted above, the sequence $\{D_{1n}, \dots, D_{nn}\}$ is locally dependent due to the fact that the same individuals from the reference group can be used in different estimates $\hat{Y}_i(0)$. Thus, there is a dependence between D_{in} and all its neighbours using the same reference individuals, forming thereby a cluster of dependent EICEs. By letting m_n be equal to the maximum size of the n clusters, we have that $\{D_{1n}, \dots, D_{nn}\}$ are m_n -dependent, i.e., D_{in} and D_{jn} are independent when $|i - j| > m_n$, given \mathbf{X} and \mathbf{Z} .

We will use the following assumptions.

(H.2) For all n , we have that $\sup_{n=1,2,\dots} m_n < m < \infty$ a.s.

(H.3) There is a $\delta > 0$ such that $\sup_{j=1,\dots,n} E[|D_{jn}|^{2+\delta} | \mathbf{X}, \mathbf{Z}] < \infty$ a.s.

Let $\tau_{in} = E(Y_i(1) - Y_i(0) | \mathbf{X}, \mathbf{Z}) = \tau(X_{in})$, the expected individual causal effect

estimated by D_{in} , and define $\tau_{.jn} = \sum_{i=j}^{j+b-1} \tau_{in}$, $\bar{\tau}_{.jn} = \frac{1}{b} \tau_{.jn}$, and $\bar{\tau}_{.n} = \frac{1}{n} \sum_{i=1}^n \tau_{in}$.

We need further the following assumptions.

$$(H.4) \quad i) \frac{b}{n} \sum_{j=1}^n (\bar{\tau}_{.jn} - \bar{\tau}_{.n})^2 = o(1) \text{ a.s.},$$

$$ii) b = b(n) \rightarrow \infty \text{ as } n \rightarrow \infty, \text{ and } b(n) = o(n^{1-r}), r > 0.$$

$$(H.5) \quad E(Y_i(0) \mid X_i = x_i, Z_i = 0) \text{ is Lipschitz on } \mathcal{X}.$$

Assumption (H.4-i) allows for vanishing variation (asymptotically) and could thus be called ‘‘asymptotically homogeneous causal effect assumption.’’ Thus, a different re-sampling scheme is introduced in the next section to cover heterogeneous causal effects. Finally, (H.4-ii) tells us how the block size must increase with sample size n in order to achieve consistency, while (H.5) allows us to have control on the matching bias $E(Y_i(0) - \hat{Y}_i(0) \mid \mathbf{X}, \mathbf{Z})$.

Below we use the concept of weakly approaching sequences in probability ($wa(P)$, introduced by Belyaev and Sjöstedt-de Luna, 2000), which is a generalization of the well known concept of weak convergence, but without the need to have a limiting distribution; see the Appendix for definitions.

Theorem 1 *Assume (A.1-2) and (H.1-5). Then, as $n \rightarrow \infty$*

$$\mathcal{L}\{\sqrt{bn}(\bar{D}_{.n}^* - \bar{D}_{.n}) \mid \mathbf{X}, \mathbf{Z}, \mathbf{Y}\} \overset{wa(P)}{\leftrightarrow} \mathcal{L}\{\sqrt{n}(\hat{\tau} - \bar{\tau}_{.n}) \mid \mathbf{X}, \mathbf{Z}\},$$

where \mathbf{Y} is the vector containing the observed outcomes $Y_i(1)$, $i = 1, \dots, n$, and $Y_i(0)$, $i = n + 1, \dots, n + N$.

The latter result tells us how we can mimic the distribution law $\mathcal{L}\{\sqrt{n}(\hat{\tau} - \bar{\tau}_{.n}) \mid \mathbf{X}, \mathbf{Z}\}$ using bootstrap (see below). Note that the target distribution is conditional on \mathbf{X} and \mathbf{Z} and centered on the parameter $\bar{\tau}_{.n}$, sometimes called sample average causal effect in the literature (e.g. Imbens, 2004, Imbens and Wooldridge, 2009). In cases where we have homogeneous expected individual causal effects, i.e. where $\tau_i = \tau$ for all i , we have $\bar{\tau}_{.n} = \tau$.

Theorem 1 justifies the following bootstrap confidence intervals. Draw B resampling copies as described above yielding $\{\sqrt{bn}(\bar{D}_{.n}^{*g} - \bar{D}_{.n})\}_{g=1}^B$. These B draws yield an empirical distribution whose quantiles q_α^* are used to construct a $(1 - \alpha)$ confidence interval for τ , e.g. as $(\hat{\tau} - q_{1-\alpha/2}^*/\sqrt{n}, \hat{\tau} + q_{\alpha/2}^*/\sqrt{n})$. The B draws could also be used to obtain a variance estimator of $\hat{\tau}$. However, such a variance estimator can readily be obtained without drawing resampling copies, utilizing a subsampling estimator. We need the following assumption.

(H.6) For all $x \in \mathcal{X}$ and $Z \in \{0, 1\}$, $\sigma^2(x, z) = \text{Var}(Y_i(1)Z_i + Y_i(0)(1 - Z_i) \mid X_i = x, Z_i = z) < \infty$.

Note that (H.6) holds, for instance, when (H.1) holds and $\sigma^2(x, z)$ is Lipschitz on \mathcal{X} for $z \in \{0, 1\}$.

Theorem 2 Under assumptions (A.1-2) and (H.1-6) we have that

$$\frac{b}{n} \sum_{j=1}^n (\bar{D}_{.jn} - \bar{D}_{.n})^2 - \text{Var}(\sqrt{n}\hat{\tau} \mid \mathbf{X}, \mathbf{Z}) \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty, \quad (2)$$

where $\bar{D}_{.jn} = D_{.jn}/b$.

This variance estimator may be used together with the asymptotic normality of $\hat{\tau}$ (Abadie and Imbens, 2006) to construct confidence intervals for τ .

3.2 Block difference bootstrap

We want to allow for heterogeneity in the individual expected causal effects and thus want to relax assumption (H.4-i), allowing instead for smoothly varying $\tau(X_{in})$. For such situations we need to resample block-differences in order to achieve asymptotically correct inference. Let $D'_{.jn} = D_{.jn} - D_{.j+2b,n}$, $j = 1, \dots, n$, denote block differences separated by distance $2b$. A resampling copy $\{D'_{.jn}\}_{j=1}^n$ of $\{D'_{.jn}\}_{j=1}^n$ is constructed by randomly drawing n items with replacement from $\{D'_{.jn}\}_{j=1}^n$. Let $\bar{D}'_{.n} = \frac{1}{2bn} \sum_{j=1}^n D'_{.jn}$.

Further, we use below the following assumptions.

(H.7) *i)* $E(Y_i(1) \mid X_i = x_i, Z_i = 1)$ is Lipschitz on \mathcal{X} ,

ii) $b(n) \rightarrow \infty$ as $n \rightarrow \infty$ and $b(n) = o(n^{2/3})$.

Assumptions (H.5) and (H.7-i) imply that $\tau(X_{in})$ is Lipschitz on \mathcal{X} . This may be called a “smoothly varying causal effect assumption” and replaces below the asymptotically homogeneous causal effect assumption (H.4-i).

Theorem 3 *Assume (A.1-2), (H.1-3), (H.5) and (H.7). Then, as $n \rightarrow \infty$*

$$\mathcal{L}\{\sqrt{2bn}\bar{D}'_n \mid \mathbf{X}, \mathbf{Z}, \mathbf{Y}\} \stackrel{wa(P)}{\leftrightarrow} \mathcal{L}\{\sqrt{n}(\hat{\tau} - \bar{\tau}_n) \mid \mathbf{X}, \mathbf{Z}\}.$$

The resampling distribution $\mathcal{L}\{\sqrt{2bn}\bar{D}'_n \mid \mathbf{X}, \mathbf{Z}, \mathbf{Y}\}$ can be estimated by generating B bootstrap copies $\{\sqrt{2bn}\bar{D}'_n^{*g}\}_{g=1}^B$ and using the resulting empirical distribution. The latter is used to construct a confidence interval for τ . Here again a subsampling variance estimator is available without the need to bootstrap.

Theorem 4 *Under assumptions (A.1-2), (H.1-3), and (H.5-7), we have*

$$\frac{1}{2bn} \sum_{j=1}^n D'_{jn}{}^2 - \text{Var}(\sqrt{n}\hat{\tau} \mid \mathbf{X}, \mathbf{Z}) \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty. \quad (3)$$

Note that the marginal variance is obtained by adding $1/n \sum_{i=1}^n (Y_i - \hat{Y}_i(0) - \hat{\tau})^2$ (i.e., an estimate of the variance of τ_{in}) to $\frac{1}{2bn} \sum_{j=1}^n D'_{jn}{}^2$; see Abadie and Imbens (2006, Sec. 4.2).

4 Monte Carlo study

To illustrate the finite sample properties of the methods introduced in this paper we simulate data from a range of different data generating mechanisms (DGM) and present results on $K = 1$ nearest neighbour matching estimators. For each individual i , values for the variables are simulated using a combination of the mechanisms described below, where the covariate is generated as $X_i \sim U(0, 1)$.

Table 1: Specification of the simulated data generating mechanisms.

	mechanism	n/N	$\tau(x)^a$	τ
DGM1.a	(T.1,Y0.1,Y1.1)	1	c	2
DGM1.b	(T.2,Y0.1,Y1.1)	0.1	c	2
DGM2.a	(T.1,Y0.1,Y1.2)	1	c	2
DGM2.b	(T.2,Y0.1,Y1.2)	0.1	c	2
DGM3.a	(T.1,Y0.1,Y1.3)	1	nc	1.8 ^b
DGM3.b	(T.2,Y0.1,Y1.3)	0.1	nc	1.8 ^b

^ac: constant; nc: non-constant.

^bApproximate values obtained via simulation.

Treatment assignment Z given X

$$(T.1) \Pr(Z_i = 1|X_i = x_i) = (1 + \exp(0.5 - 2x_i))^{-1},$$

$$(T.2) \Pr(Z_i = 1|X_i = x_i) = 0.25((1 + \exp(0.5 - 2x_i)))^{-1}.$$

Outcome without treatment

$$(Y0.1) Y_i(0)|X_i = x_i \sim N(-1 + 2x_i, 1).$$

Outcome under treatment

$$(Y1.1) Y_i(1) = Y_i(0) + 2,$$

$$(Y1.2) Y_i(1)|X_i = x_i \sim N(1 + 2x_i, 1),$$

$$(Y1.3) Y_i(1)|X_i = x_i \sim N(4x_i, 1),$$

The DGMs used in our study are described in Table 1. Constant and different forms of heterogeneity in the treatment effects are considered. Sample sizes considered are $n = 500$ and 2000 . For (T.1), data is simulated such that $n = N$ and for (T.2) such that $N = 10n$.

Due to the dependence in the EICEs and in order to achieve consistency, block size b must increase as n increases (assumptions (H.4-ii) and (H7-ii)). The choice of b is, in our particular case, simplified by the fact that we know the dependence structure for a given sample. In particular, m_n in assumption (H.2) is the maximum cluster size of dependent EICEs. This information can be used to decide upon a block size b . Here we investigate the choice $b = cm_n$, where c is a tuning parameter. In the simulations, we vary c within $\{1/4, 1/2, 3/4, 1, 3/2, 2, 5/2, 3, 4, 5, 7\}$.

The results of the Monte Carlo experiments based on 10'000 replicates (with fixed \mathbf{X} and \mathbf{Z}) are displayed in Tables 2-4. AI-C and AI-M stands for the conditional and marginal variance estimators, respectively, introduced by Abadie and Imbens (2006, Theorem 6 and 7), while BB and BDB stands for block bootstrap and block difference bootstrap respectively. To save space, we display only the results for $c = 3/2$, which yielded best empirical coverages over a wide range of situations. The complete results may be obtained from the authors.

Both AI-M and the BB scheme fail for DGM3 which was expected. The former is a marginal variance estimate (conditional and marginal variance differ only for DGM3) while our Monte Carlo study is validating conditional inference (the replicates are conditioned on \mathbf{X} and \mathbf{Z} fixed). Moreover, BB is valid under assumption (H.4-i), which is violated under DGM3 since the average causal effect is a function of the covariate. Abadie and Imbens (2006, Theorem 7) conditional variance estimator performs remarkably well in all situations considered, both in terms of variance (of $\sqrt{n}\hat{\tau}$) estimation and empirical coverage of 90% and 95% confidence interval for τ . Finally, bootstrap is generally outperformed by AI-C, although the difference in results decreases with increasing sample sizes.

For homogeneous causal effects (DGM1-2) the results are not sensitive to value of $c \geq 3/2$. This is not the case for DGM3 (results not shown). Although, $c = 3/2$ works well for all considered situations, one may want to use data-driven choices of block size; see, e.g., Hall, Horowitz, and Jing (1995), Sherman (1998) and Ekström and Sjöstedt-de Luna (2004).

5 Conclusion

The main contribution of this paper is to present resampling schemes yielding valid inference for K -nearest neighbour estimators of the average causal effect on the treated. We distinguish two cases, “asymptotically homogeneous causal effects” and “smoothly varying causal effects,” for which we need to introduce different resampling schemes. We show that inference can be carried out by using a bootstrap estimator of the distribution

Table 2: Results with design DGM1; see Table 1: Mean estimated variances are found in column $\sqrt{ns^2}$; estimators introduced in Abadie and Imbens (2006, Theorem 6 and 7) are denoted AI-C (conditional variance) and AI-M (marginal variance); BB(c) and BDB(c) denote resampling schemes using blocks of size cm , with m being the maximum cluster size of dependent EICEs; “True” yields the variance estimated over the 10’000 replicates. Coverages (cov) are obtained using the variance estimates and the normal approximation (column $\pm 1.64s$ and $\pm 1.96s$), as well as quantiles estimated with bootstrap (column “quant”).

	variance		cov 90%		cov 95%	
	$\sqrt{ns^2}$	s.e.	$\pm 1.64s$	quant	$\pm 1.96s$	quant
<hr/>						
$n = N = 500$						
“True”	4.25					
AI-C	4.25	0.04	0.898		0.946	
AI-M	4.26	0.04	0.896		0.946	
BB(3/2)	3.82	0.04	0.872	0.874	0.926	0.928
BDB(3/2)	3.96	0.04	0.876	0.879	0.927	0.929
<hr/>						
$n = N = 2000$						
“True”	4.03					
AI-C	4.04	0.04	0.899		0.952	
AI-M	4.04	0.04	0.900		0.951	
BB(3/2)	3.76	0.04	0.886	0.886	0.939	0.940
BDB(3/2)	3.79	0.04	0.884	0.885	0.938	0.941
<hr/>						
$n = 500$						
$N = 5000$						
“True”	2.25					
AI-C	2.21	0.02	0.898		0.948	
AI-M	2.22	0.02	0.899		0.948	
BB(3/2)	2.15	0.02	0.891	0.892	0.941	0.944
BDB(3/2)	2.17	0.02	0.890	0.894	0.943	0.944
<hr/>						
$n = 2000$						
$N = 20000$						
“True”	2.25					
AI-C	2.23	0.02	0.896		0.947	
AI-M	2.23	0.02	0.897		0.947	
BB(3/2)	2.19	0.02	0.888	0.893	0.942	0.944
BDB(3/2)	2.20	0.02	0.892	0.893	0.942	0.943

Table 3: Results with design DGM2 from Table 1: See also caption from Table 2.

	variance		cov 90%		cov 95%	
	$\sqrt{n_1}s^2$	s.e.	$\pm 1.64s$	quant	$\pm 1.96s$	quant
<hr/>						
$n = N = 500$						
“True”	4.31					
AI-C	4.25	0.04	0.895		0.946	,
AI-M	4.26	0.04	0.895		0.946	
BB(3/2)	3.81	0.04	0.869	0.870	0.925	0.928
BDB(3/2)	3.96	0.04	0.872	0.875	0.926	0.930
$n = N = 2000$						
“True”	4.02					
AI-C	4.04	0.04	0.899		0.951	
AI-M	4.04	0.04	0.899		0.952	
BB(3/2)	3.75	0.04	0.881	0.883	0.939	0.940
BDB(3/2)	3.79	0.04	0.881	0.884	0.938	0.941
$n = 500$ $N = 5000$						
“True”	2.26					
AI-C	2.21	0.02	0.901		0.948	
AI-M	2.22	0.02	0.901		0.948	
BB(3/2)	2.14	0.02	0.892	0.895	0.941	0.942
BDB(3/2)	2.17	0.02	0.895	0.896	0.942	0.943
$n = 2000$ $N = 20000$						
“True”	2.21					
AI-C	2.23	0.02	0.897		0.953	
AI-M	2.23	0.02	0.898		0.951	
BB(3/2)	2.19	0.02	0.894	0.894	0.947	0.950
BDB(3/2)	2.20	0.02	0.892	0.894	0.948	0.950

Table 4: Results with design DGM3 from Table 1: See also caption from Table 2.

	variance		cov 90%		cov 95%	
	$\sqrt{n_1}s^2$	s.e.	$\pm 1.64s$	quant	$\pm 1.96s$	quant
<hr/>						
$n = N = 500$						
“True”	4.31					
AI-C	4.26	0.04	0.883		0.936	
AI-M	4.57	0.05	0.895		0.947	
BB(3/2)	8.85	0.09	0.970	0.971	0.986	0.988
BDB(3/2)	5.74	0.06	0.918	0.921	0.959	0.961
$n = N = 2000$						
“True”	4.02					
AI-C	4.04	0.04	0.889		0.944	
AI-M	4.35	0.04	0.904		0.953	
BB(3/2)	9.23	0.09	0.980	0.980	0.993	0.994
BDB(3/2)	4.31	0.04	0.897	0.897	0.945	0.950
$n = 500$ $N = 5000$						
“True”	2.26					
AI-C	2.22	0.02	0.871		0.931	
AI-M	2.53	0.03	0.896		0.949	
BB(3/2)	3.70	0.04	0.947	0.951	0.977	0.979
BDB(3/2)	2.33	0.02	0.875	0.877	0.933	0.937
$n = 2000$ $N = 20000$						
“True”	2.21					
AI-C	2.23	0.02	0.879		0.935	
AI-M	2.54	0.03	0.901		0.952	
BB(3/2)	4.64	0.05	0.974	0.9876	0.990	0.992
BDB(3/2)	2.30	0.02	0.880	0.885	0.938	0.939

of the matching estimator. Alternatively, we also show how to obtain a subsampling estimator for the variance of the matching estimator. While bootstrap provides conditional inference (given \mathbf{X} and \mathbf{Z}), the subsampling variance (together with the normal approximation) can be used to perform conditional or (with a correction) marginal inference.

Finally, we conjecture that a cluster resampling scheme similar to the one introduced in Belyaev (2005) for a nearest neighbour classifier could also be used in our context to obtain valid inference, because the EICEs can be organized into clusters which are independent of each others. Such a resampling scheme is, however, difficult to study theoretically since the number of clusters and their size are sample dependent. We plan to pursue this line of research elsewhere.

A Appendix: Proofs

We first define the concept of weakly approaching sequences (introduced by Belyaev and Sjöstedt-de Luna, 2000), which is a generalization of the well known concept of weak convergence, but without the need to have a limiting distribution. Let $C_b(\mathbb{R})$ denote all continuous real-valued bounded functions on \mathbb{R} . For two sequences of random variables $\{X_n \in \mathbb{R}\}, \{Y_n \in \mathbb{R}\}$ we say that $\mathcal{L}(X_n)$ *weakly approaches* $\mathcal{L}(Y_n)$ if, for each function $h \in C_b(\mathbb{R})$, $E[h(X_n)] - E[h(Y_n)] \rightarrow 0$ as $n \rightarrow \infty$. We denote this type of convergence by $\mathcal{L}(X_n) \xrightarrow{wa} \mathcal{L}(Y_n)$. A similar definition exists for random distribution laws: Consider the two sequences $\{X_n, \mathbb{Z}_n\}$ and $\{Y_n\}$, where the random elements \mathbb{Z}_n belong to some space \mathcal{Z}_n , and X_n, \mathbb{Z}_n are defined on the same probability space. Then the sequence of regular conditional distribution laws $\{\mathcal{L}(X_n|\mathbb{Z}_n)\}$, given \mathbb{Z}_n , *weakly approaches* $\{\mathcal{L}(Y_n)\}$ *in probability* along $\{\mathbb{Z}_n\}$ if $E[h(X_n)|\mathbb{Z}_n] - E[h(Y_n)] \xrightarrow{P} 0$ as $n \rightarrow \infty$. This type of convergence is denoted by $\mathcal{L}(X_n|\mathbb{Z}_n) \xrightarrow{wa(P)} \mathcal{L}(Y_n)$. For more general definitions and a collection of properties, see Belyaev and Sjöstedt-de Luna (2000) and Sjöstedt-de Luna (2005). The proofs of Theorems 1-4 rely to a large extent on results by Sjöstedt (2000) for m-dependent sequences. We assume $s = 1$ in (A.1), which is a worst case scenario.

Proof of Theorem 1. Let $T_{in} = E[D_{in}|\mathbf{X}, \mathbf{Z}]$, $\delta_{in} = T_{in} - \tau_{in}$, and furthermore let

$\bar{T}_{\cdot jn} = \sum_{i=j}^{j+b-1} T_{in}/b$, $\bar{T}_{\cdot n} = \sum_{i=1}^n T_{in}/n$, $\bar{\delta}_{\cdot jn} = \sum_{i=j}^{j+b-1} \delta_{in}/b$, and $\bar{\delta}_{\cdot n} = \sum_{i=1}^n \delta_{in}/n$. Then

$$\begin{aligned} \frac{b}{n} \sum_{j=1}^n (\bar{T}_{\cdot jn} - \bar{T}_{\cdot n})^2 &= \frac{b}{n} \sum_{j=1}^n (\bar{\delta}_{\cdot jn} - \bar{\delta}_{\cdot n})^2 + \frac{b}{n} \sum_{j=1}^n (\bar{\tau}_{\cdot jn} - \bar{\tau}_{\cdot n})^2 \\ &\quad + \frac{2b}{n} \sum_{j=1}^n (\bar{\delta}_{\cdot jn} - \bar{\delta}_{\cdot n})(\bar{\tau}_{\cdot jn} - \bar{\tau}_{\cdot n}). \end{aligned}$$

Note that

$$\frac{1}{n} \sum_{j=1}^n (\bar{\delta}_{\cdot jn} - \bar{\delta}_{\cdot n})^2 = \frac{1}{n} \sum_{j=1}^n \bar{\delta}_{\cdot jn}^2 - \bar{\delta}_{\cdot n}^2.$$

Let $\mu_0(X) = E[Y(0)|X, Z = 0]$. Then

$$\delta_{in} = \frac{1}{K} \sum_{k=1}^K (\mu_0(X_{in}) - \mu_0(X_{j_k(i)})).$$

It follows from Jensen's inequality that

$$\left| \sum_{j=1}^b c_j \right|^a \leq b^{a-1} \sum_{j=1}^b |c_j|^a. \quad (4)$$

Therefore, by (4), (H.1) and (H.5) and from Lemma 2 in Abadie and Imbens (2006) we have, for some positive constant $c < \infty$, that

$$E[\delta_{in}^2] \leq \frac{1}{K} \sum_{k=1}^K E[|\mu_0(X_{in}) - \mu_0(X_{j_k(i)})|^2] \leq \frac{c}{K} \sum_{k=1}^K E[|X_{in} - X_{j_k(i)}|^2] = O(N^{-2}). \quad (5)$$

Note that (A.1) implies that $O(N^{-2}) = O(n^{-2})$. Hence it follows, due to independence, (4) and (5) that

$$E[\bar{\delta}_{\cdot jn}^2] \leq O(n^{-2}) \text{ and } E[\bar{\delta}_{\cdot n}^2] \leq O(n^{-2}). \quad (6)$$

Combining the above results and using Chebyshev's inequality and (H.4-ii) we thus have that

$$P\left(\frac{b}{n} \sum_{j=1}^n (\bar{\delta}_{\cdot jn} - \bar{\delta}_{\cdot n})^2 > \varepsilon\right) \leq E\left[\frac{b}{n} \sum_{j=1}^n (\bar{\delta}_{\cdot jn} - \bar{\delta}_{\cdot n})^2\right] / \varepsilon \leq O(bn^{-2}) = O\left(\frac{1}{n^{1+r}}\right).$$

Now

$$\sum_{n=1}^{\infty} P\left(\frac{b}{n} \sum_{j=1}^n (\bar{\delta}_{\cdot jn} - \bar{\delta}_{\cdot n})^2 > \varepsilon\right) = \sum_{n=1}^{\infty} O\left(\frac{1}{n^{1+r}}\right) < \infty,$$

which implies that

$$\frac{b}{n} \sum_{j=1}^n (\bar{\delta}_{.jn} - \bar{\delta}_{.n})^2 \rightarrow 0 \quad \text{a.s.}, \quad (7)$$

see, e.g., Shiryaev (1984, p. 252-253). From Cauchy-Schwartz inequality together with (H.4) and (7) we have that

$$\frac{b}{n} \sum_{j=1}^n (\bar{\delta}_{.jn} - \bar{\delta}_{.n})(\bar{\tau}_{.jn} - \bar{\tau}_{.n}) \leq \sqrt{\frac{b}{n} \sum_{j=1}^n (\bar{\delta}_{.jn} - \bar{\delta}_{.n})^2} \sqrt{\frac{b}{n} \sum_{j=1}^n (\bar{\tau}_{.jn} - \bar{\tau}_{.n})^2} \rightarrow 0 \quad \text{a.s.},$$

and thus

$$\frac{b}{n} \sum_{j=1}^n (\bar{T}_{.jn} - \bar{T}_{.n})^2 \rightarrow 0 \quad \text{a.s.} \quad (8)$$

By (H.2), (H.3) and (8) we have from Theorem 2 in Sjöstedt (2000) that, for every $\varepsilon > 0$ and for all $h(\cdot) \in C_b(\mathbb{R})$,

$$P \left(\left| E[h(\sqrt{bn}(\bar{D}_{.n}^* - \bar{D}_{.n})) | \mathbf{Y}, \mathbf{X}, \mathbf{Z}] - E[h(\sqrt{n}(\bar{D}_{.n} - \bar{T}_{.n})) | \mathbf{X}, \mathbf{Z}] \right| > \varepsilon \mid \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \right) = o(1) \quad \text{a.s.}$$

By dominated convergence it follows that

$$\lim_{n \rightarrow \infty} P \left(\left| E[h(\sqrt{bn}(\bar{D}_{.n}^* - \bar{D}_{.n})) | \mathbf{Y}, \mathbf{X}, \mathbf{Z}] - E[h(\sqrt{n}(\bar{D}_{.n} - \bar{T}_{.n})) | \mathbf{X}, \mathbf{Z}] \right| > \varepsilon \right) = 0,$$

for every $\varepsilon > 0$ and for all $h(\cdot) \in C_b(\mathbb{R})$. Hence,

$$\mathcal{L}(\sqrt{bn}(\bar{D}_{.n}^* - \bar{D}_{.n})) | \mathbf{Y}, \mathbf{X}, \mathbf{Z} \xrightarrow{wa(p)} \mathcal{L}(\sqrt{n}(\bar{D}_{.n} - \bar{T}_{.n})) | \mathbf{X}, \mathbf{Z} \quad \text{as } n \rightarrow \infty.$$

We have that

$$\mathcal{L}(\sqrt{n}(\bar{D}_{.n} - \bar{T}_{.n})) | \mathbf{X}, \mathbf{Z} = \mathcal{L}(\sqrt{n}(\bar{D}_{.n} - \bar{\tau}_{.n})) | \mathbf{X}, \mathbf{Z} + \mathcal{L}(\sqrt{n}\bar{\delta}_{.n}) | \mathbf{X}, \mathbf{Z}.$$

By (6) and Chebyshev's inequality $\sqrt{n}\bar{\delta}_{.n} \xrightarrow{P} 0$ as $n \rightarrow \infty$, and thus the result follows. ■

Proof of Theorem 2. Let $U_{in} = D_{in} - T_{in}$. Then

$$\begin{aligned} \frac{b}{n} \sum_{j=1}^n (\bar{D}_{.jn} - \bar{D}_{.n})^2 &= \frac{b}{n} \sum_{j=1}^n (\bar{U}_{.jn} - \bar{U}_{.n})^2 + \frac{b}{n} \sum_{j=1}^n (\bar{T}_{.jn} - \bar{T}_{.n})^2 \\ &\quad + \frac{2b}{n} \sum_{j=1}^n (\bar{U}_{.jn} - \bar{U}_{.n})(\bar{T}_{.jn} - \bar{T}_{.n}), \end{aligned}$$

where $\bar{U}_{.jn} = \sum_{i=j}^{j+b-1} U_{in}/b$, and $\bar{U}_{.n} = \sum_{j=1}^n U_{jn}/n$. From (H.2)-(H.4) and by the same arguments as in the proof of Theorem 2 in Sjöstedt (2000) we have that for all $\varepsilon > 0$

$$P \left(\left| \frac{b}{n} \sum_{j=1}^n (\bar{U}_{.jn} - \bar{U}_{.n})^2 - \text{Var}[\sqrt{n}\hat{\tau}|\mathbf{X}, \mathbf{Z}] \right| > \varepsilon \mid \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \right) = o(1) \text{ a.s.}$$

Dominated convergence then implies that

$$\frac{b}{n} \sum_{j=1}^n (\bar{U}_{.jn} - \bar{U}_{.n})^2 - \text{Var}[\sqrt{n}\hat{\tau}|\mathbf{X}, \mathbf{Z}] \xrightarrow{p} 0 \text{ as } n \rightarrow \infty. \quad (9)$$

From equation (13) in Abadie and Imbens (2006) we have that

$$\text{Var}[\sqrt{n}\hat{\tau}|\mathbf{X}, \mathbf{Z}] = \frac{1}{n} \sum_{i=1}^n \sigma^2(X_{in}, Z_{in}) + \frac{1}{n} \sum_{i=n+1}^N \frac{Q_K^2(i)}{K^2} \sigma^2(X_{in}, Z_{in}), \quad (10)$$

where $Q_K^2(i)$ denotes the number of times individual i (in the reference group) is used as a match given that K matches per individual (in the group of interest) are used. The first term on the right hand side of (10) is $O(1)$ by (H.6). From Lemma 3 in Abadie and Imbens (2006) we have that $(N/n)E[Q_K^2(i)|Z_{in} = 0]$ is bounded, which thus makes the last term in (10) of magnitude $O_P(1)$, and therefore

$$\text{Var}[\sqrt{n}\hat{\tau}|\mathbf{X}, \mathbf{Z}] = O_p(1). \quad (11)$$

Hence,

$$\frac{b}{n} \sum_{j=1}^n (\bar{U}_{.jn} - \bar{U}_{.n})^2 = O_p(1). \quad (12)$$

By Cauchy-Schwartz inequality, (12) and (8) we have that

$$\frac{b}{n} \sum_{j=1}^n (\bar{U}_{.jn} - \bar{U}_{.n})(\bar{T}_{.jn} - \bar{T}_{.n}) \leq \sqrt{\frac{b}{n} \sum_{j=1}^n (\bar{U}_{.jn} - \bar{U}_{.n})^2} \sqrt{\frac{b}{n} \sum_{j=1}^n (\bar{T}_{.jn} - \bar{T}_{.n})^2} = o_p(1). \quad (13)$$

Hence, combining (9), (H.4) and (13) yields the desired result. ■

Proof of Theorem 3. Let $R_{in} = D_{in} - \delta_{in}$ such that $D_{in} = R_{in} + \delta_{in}$, and note that $E[R_{in}|\mathbf{X}, \mathbf{Z}] = \tau_{in}$. It then follows that

$$\mathcal{L}(\sqrt{2bn}D'_{.n}|\mathbf{X}, \mathbf{Z}, \mathbf{Y}) = \mathcal{L}(\sqrt{2bn}R'_{.n}|\mathbf{X}, \mathbf{Z}, \mathbf{Y}) + \mathcal{L}(\sqrt{2bn}\delta'_{.n}|\mathbf{X}, \mathbf{Z}, \mathbf{Y}),$$

where $R'_{\cdot n}$ and $\delta'_{\cdot n}$ are constructed as $D'_{\cdot n}$, while replacing D_{in} by R_{in} and δ_{in} , respectively. Assumptions (H.2-3), (H.5) and (H.7-ii) imply that (7) holds and thus by Remark 3 in Sjöstedt (2000) we have that

$$\mathcal{L}(\sqrt{2bn}\delta'_{\cdot n}|\mathbf{X}, \mathbf{Z}, \mathbf{Y}) \xrightarrow{wa(p)} \mathcal{L}(\sqrt{n}(\bar{\delta}_{\cdot n} - E[\bar{\delta}_{\cdot n}|\mathbf{X}, \mathbf{Z}]|\mathbf{X}, \mathbf{Z}) = 0.$$

That

$$\mathcal{L}(\sqrt{2bn}R'_{\cdot n}|\mathbf{X}, \mathbf{Z}, \mathbf{Y}) \xrightarrow{wa(p)} \mathcal{L}(\sqrt{n}(\hat{\tau} - \bar{\tau}_{\cdot n})|\mathbf{X}, \mathbf{Z}) \text{ as } n \rightarrow \infty,$$

follows by similar arguments as in the proof of Theorem 1, using Theorem 1 in Sjöstedt (2000), and noting that by (H.5) and (H.7) τ_{in} is Lipschitz. Hence,

$$\mathcal{L}(\sqrt{2bn}D'_{\cdot n}|\mathbf{X}, \mathbf{Z}, \mathbf{Y}) \xrightarrow{wa(p)} \mathcal{L}(\sqrt{n}(\hat{\tau} - \bar{\tau}_{\cdot n})|\mathbf{X}, \mathbf{Z}) \text{ as } n \rightarrow \infty.$$

■

Proof of Theorem 4. We have that

$$\frac{1}{2bn} \sum_{j=1}^n (D'_{\cdot jn})^2 = \frac{1}{2bn} \sum_{j=1}^n (U'_{\cdot jn} + T'_{\cdot jn})^2 = \frac{1}{2bn} \sum_{j=1}^n (U'_{\cdot jn})^2 + \frac{1}{2bn} \sum_{j=1}^n (T'_{\cdot jn})^2 + \frac{1}{bn} \sum_{j=1}^n U'_{\cdot jn} T'_{\cdot jn},$$

where $U'_{\cdot jn} = b(\bar{U}_{\cdot jn} - \bar{U}_{\cdot j+2b,n})$ and $T'_{\cdot jn} = b(\bar{T}_{\cdot jn} - \bar{T}_{\cdot j+2b,n})$. Since $T_{in} = \tau_{in} + \delta_{in}$, by repeated use of (4) we have that

$$\begin{aligned} \frac{1}{bn} \sum_{j=1}^n (T'_{\cdot jn})^2 &\leq \frac{2b}{n} \sum_{j=1}^n (\bar{\tau}_{\cdot jn} - \bar{\tau}_{\cdot i+2b,n})^2 + \frac{2b}{n} \sum_{j=1}^n (\bar{\delta}_{\cdot jn} - \bar{\delta}_{\cdot i+2b,n})^2 \\ &\leq \frac{2}{n} \sum_{j=1}^n \sum_{i=j}^{j+b-1} (\tau_{in} - \tau_{i+2b,n})^2 + \frac{8b}{n} \sum_{j=1}^n \bar{\delta}_{\cdot jn}^2. \end{aligned}$$

From (6) we have that $E[b \sum_{j=1}^n \bar{\delta}_{\cdot jn}^2/n] = O(b/n^2)$ which tends to zero as $n \rightarrow \infty$, and hence, $b \sum_{j=1}^n \bar{\delta}_{\cdot jn}^2/n \xrightarrow{p} 0$ as $n \rightarrow \infty$. Furthermore, (H.5) and (H.7) implies that τ_{in} is Lipschitz, and thus, for some positive constant $c_L < \infty$,

$$\frac{1}{n} \sum_{j=1}^n \sum_{i=j}^{j+b-1} E[(\tau_{in} - \tau_{i+2b,n})^2] \leq \frac{c_L^2}{n} \sum_{j=1}^n \sum_{i=j}^{j+b-1} E[(X_{in} - X_{i+2b,n})^2]$$

$$\leq \frac{2bc_L^2}{n} \sum_{j=1}^n \sum_{i=j}^{j+b-1} \sum_{k=0}^{2b-1} E[(X_{i+k,n} - X_{i+k+1,n})^2] = O(b^3/n^2),$$

by (4) and Lemma 2 of Abadie and Imbens (2006). But $O(b^3/n^2) \rightarrow 0$ as $n \rightarrow \infty$, by (H.7-ii). Hence,

$$\frac{1}{2bn} \sum_{j=1}^n (T'_{.jn})^2 \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty. \quad (14)$$

(H.2-3), (H.5) and (H.7) together with Lemma 3 in Sjöstedt (2000) ensures that, for any $\varepsilon > 0$

$$P \left(\left| \frac{1}{2bn} \sum_{j=1}^n (U'_{.jn})^2 - \text{Var}[\sqrt{n}\hat{\tau}|\mathbf{X}, \mathbf{Z}] \right| > \varepsilon \mid \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z} \right) = o(1) \quad \text{a.s.}$$

By dominated convergence we thus have that

$$\frac{1}{2bn} \sum_{j=1}^n (U'_{.jn})^2 - \text{Var}[\sqrt{n}\hat{\tau}|\mathbf{X}, \mathbf{Z}] \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty. \quad (15)$$

From similar arguments as for (11) and (12) it follows that $\sum_{j=1}^n (U'_{.jn})^2 / (2bn) = O_p(1)$.

Now, by the Cauchy-Schwartz inequality, (14) and (15)

$$\frac{1}{bn} \sum_{j=1}^n U'_{.jn} T'_{.jn} \leq \sqrt{\frac{1}{2bn} \sum_{j=1}^n (U'_{.jn})^2} \sqrt{\frac{1}{2bn} \sum_{j=1}^n (T'_{.jn})^2} = o_p(1).$$

Hence, the desired result follows. ■

References

- Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, **74**, 235–267.
- Abadie, A. and Imbens, G. W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica*, **76**, 1537–1557.
- Belyaev, Y. K. (2005). Application of clustered resampling methods in assessing accuracy of cross-validated estimators of cross-classification probabilities of nearest-neighbor classifiers. *Research Report*, 2005:1, Center of Biostochastics, Swedish University of Agricultural Sciences.

- Belyaev, Y. K. and Sjöstedt-de Luna, S. (2000). Weakly approaching sequences of random distributions. *Journal of Applied Probability*, **37**, 807–822.
- Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics*, **14**, 1171–1194.
- Ekström, M. and Sjöstedt-de Luna, S. (2004). Subsampling methods to estimate the variance of sample means based on nonstationary spatial data with varying expected values. *Journal of the American Statistical Association*, **99**, 82–95.
- Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, **82**, 561–574.
- Hansen, B. B. (2008). The prognostic analogue of the propensity score. *Biometrika*, **95**, 481–488.
- Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, **86**, 4–29.
- Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, **47**, 5–86.
- Künsch, H. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, **17**, 1217–1241.
- Lahiri, S. N. (1992). Edgeworth correction by ‘moving blocks’ bootstrap for stationary and nonstationary data. In LePage R. and Billard L., eds., *Exploring the Limits of Bootstrap*. New York: John Wiley and Sons.
- Neyman, J. (1923). Sur les applications de la théorie des probabilités aux expériences agricoles: Essai des principes. *Roczniki Nauk Rolniczych*, **X**, 1–51. In Polish, English translation by D. Dabrowska and T. Speed in *Statistical Science*, **5**, 465–472, 1990.
- Politis, N. and Romano, J. (1992). A circular block-resampling procedure for stationary data. In LePage, R. and Billard, L. (Eds.), *Exploring the Limits of Bootstrap*, New York: John Wiley and Sons., pp. 263-270.

- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, **66**, 688–701.
- Rubin, D. B. (1991). Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism. *Biometrics*, **47**, 1213–1234.
- Sherman, M. (1998). Data-based choice of batch size for simulation output analysis. *Simulation*, **71**, 38–47.
- Shiryayev, A. N. (1984). *Probability*. Berlin: Springer-Verlag.
- Sjöstedt, S. (2000). Resampling m-dependent random variables with applications to forecasting. *Scandinavian Journal of Statistics*, **27**, 543–561.
- Sjöstedt-de Luna, S. (2005). Some properties of weakly approaching sequences of distributions. *Statistics and Probability Letters*, **75**, 119–126.
- Waernbaum, I. (2010). Propensity score model specification for estimation of average treatment effects. *Journal of Statistical Planning and Inference*, **140**, 1948–1956.