Bootstrap inference for K-nearest neighbour matching estimators*

Xavier de Luna,[a] Per Johansson[b] and Sara Sjöstedt-de Luna[a]

[a]Umeå University, [b]Uppsala University and [b]IFAU, Uppsala
xavier.deluna@stat.umu.se
per.johansson@ifau.uu.se
sara@math.umu.se

September 1, 2010

Preliminary version.

Abstract:

Abadie and Imbens (2008, Econometrica) showed that classical bootstrap schemes fail to provide correct inference for K-nearest neighbour matching (KNNM) estimators. 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 KNNM 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.

*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.

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

1 Introduction

K-nearest neighbour matching (KNNM) 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 KNNM 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 KNNM 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 KNNM 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 −heterogeneous causal effect. 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 wider generality, for instance, allowing for 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 KNNM 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 units $i = 1, \dots, n$ 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 units are assigned to $z_i = 1$ and $z_i = 0$ respectively; see Neyman (1923), Rubin (1974), Imbens (2004). Here, and in the sequel, $y_i(1), y_i(0), z_i$, and x_i denote both the random variable (modelling the random sampling from a population) and their realizations, depending on the context.

We assume that if a given unit i is assigned a given value for z_i, this does not affect the values taken by the potential outcomes for this unit or any other unit in the study (stable unit 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 $\mathcal{L}\{(y_i(j), x_i) \mid z_i = j\}$, $j = 0, 1$. 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 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 units 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 KNNM estimator as

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}(y_i(1) - \hat{y}_i(0)), \tag{1}$$

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

$$\hat{y}_i(0) = y_j(0) \text{ such that } j = \underset{j=n+1,\ldots,n+N}{\operatorname{argmin}} |x_j - x_i|,$$

where $|\cdots|$ is a vector norm. Generally, for $K > 0$ and for $i = 1, \ldots, n$, denote by $j_K(i)$ the index $j \in \{n+1, \ldots, n+N\}$ 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 $\mathcal{J}_K(i) = \{j_1(i), j_2(i), \ldots, j_K(i)\}$. 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 KNNM estimator is asymptotically normal. They consider the marginal variance $Var(\hat{\tau})$ as well as the conditional variance $Var(\hat{\tau}|X, Z)$, where X and Z are vectors containing the observed values x_i and z_i, $i = 1, \ldots, n+N$, 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

$$e_{in} = y_i(1) - \hat{y}_i(0)$$

the estimated individual causal effects (EICE). Hence, the KNNM estimator (1) can be written as $\hat{\tau} = 1/n \sum_{i=1}^{n} e_{in}$. 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:

$$e_{1n},\ e_{2n},\ \ldots,\ e_{nn}, \quad \text{where}$$

$$x_{1n} \leq x_{2n} \leq \ldots \leq x_{nn},$$

with x_{in}, $i = 1, \ldots, n$, 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 $\mathcal{B}_j = \{e_{jn}, e_{j+1,n}, \ldots, e_{j+b-1,n}\}$, $j = 1, \ldots, n$, where $e_{n+j,n} = e_{jn}$; see Figure 3. Furthermore, let $e_{\cdot jn} = \sum_{i=j}^{j+b-1} e_{in}$. A resampling copy (a pseudo sample) of $\{e_{\cdot jn}\}_{j=1}^{n}$ is denoted $\{e^*_{\cdot jn}\}_{j=1}^{n}$ and is constructed by

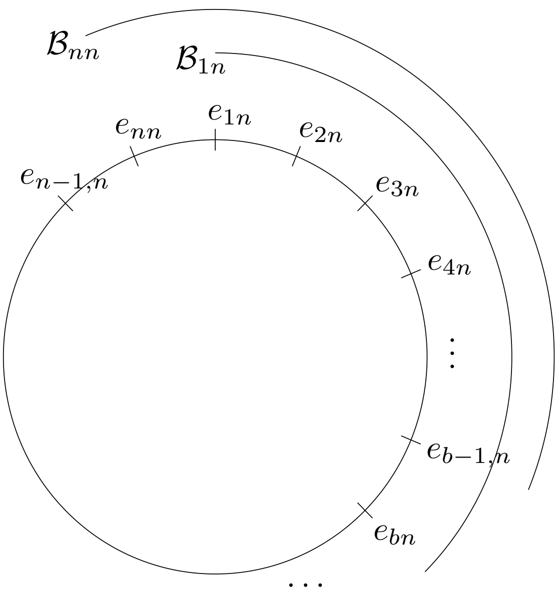

Figure 1: Circular blocking scheme.

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

As noted above, the sequence $\{e_{1n}, \dots, e_{nn}\}$ is locally dependent due to the fact that the same units from the reference group can be used in different estimates $\hat{y}_i(0)$. Thus, there is a dependence between e_{in} and all its neighbours using the same reference units, forming thereby a cluster of dependent EICEs. By letting m_n be equal to the maximum size of the n clusters, we have that $\{e_{1n}, \dots, e_{nn}\}$ are m_n-dependent, i.e., e_{in} and e_{jn} are independent when $|i - j| > m_n$, given X and 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 $\sup_{\substack{n=1,2,\dots \\ j=1,\dots,n}} E\left[|e_{jn}|^{2+\delta} \mid X, Z\right] < \infty$ a.s.

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

Let $\tau_{in} = E(y_i(1) - y_i(0) \mid X, Z) = \tau(x_{in})$, the expected individual causal effect estimated by e_{in}, and define $\tau_{\cdot jn} = \sum_{i=j}^{j+b-1} \tau_{in}$, $\bar{\tau}_{\cdot jn} = \frac{1}{b}\tau_{\cdot jn}$, and $\bar{\tau}_{\cdot n} = \frac{1}{n}\sum_{i=1}^{n} \tau_{in}$.

We need further the following assumptions.

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

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

(H.5) $E(y_i(0) \mid x_i, z_i = 0)$ 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 resampling scheme is introduced in the next section to cover heterogeneous causal effects. Finally, (H4-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 $E(y_i(0) - \hat{y}_i(0) \mid X, Z)$.

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* $n \to \infty$

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

where Y *is the vector containing the observed outcomes* $y_i(1)$, $i = 1, \ldots, n$, *and* $y_i(0)$, $i = n+1, \ldots, n+N$.

The latter result tells us how we can mimic the distribution law $\mathcal{L}\{\sqrt{n}(\hat{\tau} - \bar{\tau}_{\cdot n}) \mid X, Z\}$ using bootstrap (see below). Note that the target distribution is conditional on X and Z and centered on the parameter $\bar{\tau}_{\cdot n}$, 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 $\bar{\tau}_{\cdot n} = \tau$.

Theorem 1 justifies the following bootstrap confidence intervals. Draw B resampling copies as described above yielding $\{\sqrt{bn}(\bar{e}^{*g}_{\cdot n} - \bar{e}_{\cdot n})\}_{g=1}^{B}$. 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 $z \in \{0,1\}$, $\sigma^2(x,z) < \infty$, where $\sigma^2(x,z) = Var(y_i(1)z_i + y_i(0)(1-z_i) \mid x_i = x, z_i = z)$.

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*

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

where $\bar{e}_{\cdot jn} = e_{\cdot jn}/b$.

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 $e'_{\cdot jn} = e_{\cdot jn} - e_{\cdot j+2b,n}$, $j = 1, \ldots, n$, denote block differences separated by distance $2b$. A resampling copy $\{e'^*_{\cdot jn}\}_{j=1}^n$ of $\{e'_{\cdot jn}\}_{j=1}^n$ is constructed by randomly drawing n items with replacement from $\{e'_{\cdot jn}\}_{j=1}^n$. Let $\bar{e}'^*_{\cdot n} = \frac{1}{2bn}\sum_{j=1}^n e'^*_{\cdot jn}$.

Further, we use below the following assumptions.

(H.7) i) $E(y_i(1) \mid x_i, z_i = 1)$ is Lipschitz on $\mathcal{X}$,

ii) $b(n) \to \infty$ as $n \to \infty$ 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* $n \to \infty$

$$\mathcal{L}\{\sqrt{2bn}\bar{e}'^{*}_{\cdot n}|X, Z, Y\} \overset{wa(P)}{\leftrightarrow} \mathcal{L}\{\sqrt{n}(\hat{\tau} - \bar{\tau}_{\cdot n}) \mid X, Z\}.$$

The resampling distribution $\mathcal{L}\{\sqrt{2bn}\bar{e}'_{\cdot n}|X, Z, Y\}$ can be estimated by generating B bootstrap copies $\{\sqrt{2bn}\bar{e}'^{*g}_{\cdot n}\}_{g=1}^{B}$ 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), (H.5) and (H.6-7), we have*

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

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 $\frac{1}{2bn}\sum_{j=1}^{n} e'^{2}_{\cdot jn}$; 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

[a]c: constant; nc: non-constant.

[b]Approximate values obtained via simulation.

Treatment assignment z given x

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

Outcome without treatment

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

Outcome under treatment

(Y1.1) $y_i(1) = y_i(0) + 2$,
(Y1.2) $y_i(1)|x_i \sim N(1 + 2x_i, 1)$,
(Y1.3) $y_i(1)|x_i \sim N(4x_i, 1)$,

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 X and Z) are displayed in Tables 2-4 (Tables are under compilation and will be included soon). 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 BDP 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 X and Z fixed). 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 $c \leq 3/2$. This is not the case for DGM3 where variance estimation and empirical coverages deteriorate when departing from $c = 3/2$.

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. 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.

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 experiences 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 nonrandomized 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.

Shiryaev, A. N. (1984). *Probability.* Berlin: Springer-Verlag.

Sjöstedt, S. (2000). Resampling m-dependent random variables with applications to forecasting. *Scandinavian Journal of Statistics*, **27**, 43–562.

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.

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 $h \in C_b(\mathbb{R})$, $E[h(X_n)] - E[h(Y_n)] \to 0$ as $n \to \infty$. We denote this type of convergence by $\mathcal{L}(X_n) \overset{wa}{\leftrightarrow} \mathcal{L}(Y_n)$. 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 $E[h(X_n)|\mathbb{Z}_n] - E[h(Y_n)] \overset{P}{\to} 0$ as $n \to \infty$. This type of convergence is denoted by $\mathcal{L}(X_n|\mathbb{Z}_n) \overset{wa(P)}{\longleftrightarrow} \mathcal{L}(Y_n)$. 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 $t_{in} = E[e_{in}|X, Z]$, $\delta_{in} = t_{in} - \tau_{in}$, 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

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

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 Jensens inequality that

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

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

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

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

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

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

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

Now

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

which implies that

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

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

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

and thus

$$\frac{b}{n}\sum_{j=1}^{n}(\bar{t}_{\cdot jn}-\bar{t}_{\cdot n})^2 \to 0 \ \ \text{a.s.} \tag{8}$$

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

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

By dominated convergence it follows that

$$\lim_{n\to\infty} P\left(\left|E[h(\sqrt{bn}(\bar{e}^*_{\cdot n}-\bar{e}_{\cdot n}))|Y,X,Z] - E[h(\sqrt{n}(\bar{e}_{\cdot n}-\bar{t}_{\cdot n})|X,Z]\right| > \varepsilon \ \right) = 0,$$

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

$$\mathcal{L}(\sqrt{bn}(\bar{e}^*_{\cdot n}-\bar{e}_{\cdot n}))|Y,X,Z) \overset{wa(p)}{\longleftrightarrow} \mathcal{L}(\sqrt{n}(\bar{e}_{\cdot n}-\bar{t}_{\cdot n})|X,Z) \ \ \text{as } n \to \infty.$$

We have that

$$\mathcal{L}(\sqrt{n}(\bar{e}_{\cdot n}-\bar{t}_{\cdot n})|X,Z) \ = \mathcal{L}(\sqrt{n}(\bar{e}_{\cdot n}-\bar{\tau}_{\cdot n})|X,Z) \ + \mathcal{L}(\sqrt{n}\bar{\delta}_{\cdot n}|X,Z).$$

By (6) and Chebyshevs inequality $\sqrt{n}\bar{\delta}_{\cdot n} \overset{P}{\to} 0$ as $n \to \infty$, and thus the result follows. ■

Proof of Theorem 2. Let $u_{in} = e_{in} - t_{in}$. Then

$$\frac{b}{n}\sum_{j=1}^{n}(\bar{e}_{\cdot jn}-\bar{e}_{\cdot n})^2 = \frac{b}{n}\sum_{j=1}^{n}(\bar{u}_{\cdot jn}-\bar{u}_{\cdot n})^2 + \frac{b}{n}\sum_{j=1}^{n}(\bar{t}_{\cdot jn}-\bar{t}_{\cdot n})^2 + \frac{2b}{n}\sum_{j=1}^{n}(\bar{u}_{\cdot jn}-\bar{u}_{\cdot n})(\bar{t}_{\cdot jn}-\bar{t}_{\cdot n}),$$

where $\bar{u}_{\cdot jn} = \sum_{i=j}^{j+b-1} u_{in}/b$, and $\bar{u}_{\cdot n} = \sum_{j=1}^{n} u_{jn}/n$. 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 $\varepsilon > 0$

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

Dominated convergence then implies that

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

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

$$Var[\sqrt{n}\hat{\tau}|X,Z] = \frac{1}{n}\sum_{i=1}^{n}\sigma^2(x_{in}, z_{in}) + \frac{1}{n}\sum_{i=n+1}^{N}\frac{Q_K^2(i)}{K^2}\sigma^2(x_{in}, z_{in}), \tag{10}$$

where $Q_K^2(i)$ denotes the number of times unit i (in the reference group) is used as a match given that K matches per unit (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

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

Hence,

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

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

$$\frac{b}{n}\sum_{j=1}^{n}(\bar{u}_{\cdot jn} - \bar{u}_{\cdot n})(\bar{t}_{\cdot jn} - \bar{t}_{\cdot n}) \leq \sqrt{\frac{b}{n}\sum_{j=1}^{n}(\bar{u}_{\cdot jn} - \bar{u}_{\cdot n})^2}\sqrt{\frac{b}{n}\sum_{j=1}^{n}(\bar{t}_{\cdot jn} - \bar{t}_{\cdot n})^2} = o_p(1). \tag{13}$$

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

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

$$\mathcal{L}(\sqrt{2bn}e'^{*}_{\cdot n}|X,Z,Y) \;= \mathcal{L}(\sqrt{2bn}r'^{*}_{\cdot n}|X,Z,Y) + \mathcal{L}(\sqrt{2bn}\delta'^{*}_{\cdot n}|X,Z,Y),$$

where $r'^*_{\cdot n}$ and $\delta'^*_{\cdot n}$ are constructed as $e'^*_{\cdot n}$, while replacing e_{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

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

That

$$\mathcal{L}(\sqrt{2bn}r'^*_{\cdot n}|X,Z,Y) \overset{wa(p)}{\longleftrightarrow} \mathcal{L}(\sqrt{n}(\hat{\tau} - \bar{\tau}_{\cdot n})|X,Z) \quad \text{as } n \to \infty,$$

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,

$$\mathcal{L}(\sqrt{2bn}e'^*_{\cdot n}|X,Z,Y) \overset{wa(p)}{\longleftrightarrow} \mathcal{L}(\sqrt{n}(\hat{\tau} - \bar{\tau}_{\cdot n})|X,Z) \quad \text{as } n \to \infty.$$

■

Proof of Theorem 4. We have that

$$\frac{1}{2bn}\sum_{j=1}^{n}(e'_{\cdot jn})^2 = \frac{1}{2bn}\sum_{j=1}^{n}(u'_{\cdot jn} + t'_{\cdot jn})^2 = \frac{1}{2bn}\sum_{j=1}^{n}(u'_{\cdot jn})^2 + \frac{1}{2bn}\sum_{j=1}^{n}(t'_{\cdot jn})^2 + \frac{1}{bn}\sum_{j=1}^{n}u'_{\cdot jn}t'_{\cdot jn},$$

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

$$\frac{1}{bn}\sum_{j=1}^{n}(t'_{\cdot jn})^2 \le \frac{2b}{n}\sum_{j=1}^{n}(\bar{\tau}_{\cdot jn} - \bar{\tau}_{\cdot i+2b,n})^2 + \frac{2b}{n}\sum_{j=1}^{n}(\bar{\delta}_{\cdot jn} - \bar{\delta}_{\cdot i+2b,n})^2$$

$$\le \frac{2}{n}\sum_{j=1}^{n}\sum_{i=j}^{j+b-1}(\tau_{in} - \tau_{i+2b,n})^2 + \frac{8b}{n}\sum_{j=1}^{n}\bar{\delta}^2_{\cdot jn}.$$

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

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

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

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

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

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

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

By dominated convergence we thus have that

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

From similar arguments as for (11) and (12) it follows that $\sum_{j=1}^{n} (u'_{\cdot jn})^2/(2bn) = O_p(1)$. Now, by the Cauchy-Schwartz inequality, (14) and (15)

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

Hence, the desired result follows. ■