DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD

On the Failure of the Bootstrap for Matching Estimators

(Article begins on next page)

On the Failure of the Bootstrap for Matching Estimators*

Alberto Abadie†   Guido W. Imbens‡

May 2006

Abstract

Matching estimators are widely used for the evaluation of programs or treatments. Often researchers use bootstrapping methods for inference. However, no formal justification for the use of the bootstrap has been provided. Here we show that the bootstrap is in general not valid, even in the simple case with a single continuous covariate when the estimator is root-N consistent and asymptotically normally distributed with zero asymptotic bias. Due to the extreme non-smoothness of nearest neighbor matching, the standard conditions for the bootstrap are not satisfied, leading the bootstrap variance to diverge from the actual variance. Simulations confirm the difference between actual and nominal coverage rates for bootstrap confidence intervals predicted by the theoretical calculations. To our knowledge, this is the first example of a root-N consistent and asymptotically normal estimator for which the bootstrap fails to work.

JEL Classification: C14, C21, C52
Keywords: *Average Treatment Effects, Bootstrap, Matching, Confidence Intervals*

†John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138, and NBER. Electronic correspondence: alberto_abadie@harvard.edu, http://www.ksg.harvard.edu/fs/aabadie/

‡Department of Economics, and Department of Agricultural and Resource Economics, University of California at Berkeley, 330 Giannini Hall, Berkeley, CA 94720-3880, and NBER. Electronic correspondence: imbens@econ.berkeley.edu, http://elsa.berkeley.edu/users/imbens/

1   Introduction

Matching methods have become very popular for the estimation of treatment effects.[1] Often researchers use bootstrap methods to calculate the standard errors of matching estimators.[2] Bootstrap inference for matching estimators has not been formally justified. Because of the non-smooth nature of some matching methods and the lack of evidence that the resulting estimators are asymptotically linear (e.g., nearest neighbor matching with a fixed number of neighbors), there is reason for concern about their validity of the bootstrap in this context.

At the same time, we are not aware of any example where an estimator is root-N consistent, as well as asymptotically normally distributed with zero asymptotic bias and yet where the standard bootstrap fails to deliver valid confidence intervals.[3] This article addresses the question of the validity of the bootstrap for nearest-neighbor matching estimators with a fixed number of neighbors. We show in a simple case with a single continuous covariate that the standard bootstrap does indeed fail to provide asymptotically valid confidence intervals, in spite of the fact that the estimator is root-N consistent and asymptotically normal with no asymptotic bias. We provide some intuition for this failure. We present theoretical calculations for the asymptotic behavior of the difference between the variance of the matching estimator and the average of the bootstrap variance. These theoretical calculations are supported by Monte Carlo evidence. We show that the bootstrap confidence intervals can have over-coverage as well as under-coverage. The

---

[1] E.g., Dehejia and Wahba, (1999). See Rosenbaum (2001) and Imbens (2004) for surveys.

[2] A partial list of recent papers using matching with bootstrapped standard errors includes Agodini and Dynarski (2004), Dehejia and Wahba (1999, 2002), Galasso and Ravaillon (2003), Guarcello, Mealli, and Rosati (2003), Heckman, Ichimura and Todd (1997), Ichino and Becker (2002), Imai (2005), Jalan and Ravallion (1999), Lechner (2002), Myers, Olsen, Seftor, Young, and Tuttle (2002) Pradhan, and Rawlings (2003), Puhani (2002), Sianesi (2004), Smith and Todd (2005), Yamani, Lauer, Starling, Pothier, Tuzcu, Ratliff, Cook, Abdo, McNeil, Crowe, Hobbs, Rincon, Bott-Silverman, McCarthy and Young (2004).

[3] Familiar examples of failure of the bootstrap for estimators with non-normal limiting distributions arise in the contexts of estimating the maximum of the support of a random variable (Bickel and Freedman, 1981), estimating the average of a variable with infinite variance (Arthreya, 1987), and super-efficient estimation (Beran, 1984). Resampling inference in these contexts can be conducted using alternative methods such as subsampling (Politis and Romano, 1994; Politis, Romano, and Wolf, 1999) and versions of the bootstrap where the size of the bootstrap sample is smaller than the sample size (e.g., Bickel, Götze and Van Zwet, 1997). See Hall (1992) and Horowitz (2003) for general discussions.

results do not address whether nearest neighbor estimators with the number of neighbors increasing with the sample size do satisfy asymptotic linearity or whether the bootstrap is valid for such estimators as in practice many researchers have used estimators with very few (e.g., one) nearest neighbor(s).

In Abadie and Imbens (2006) we have proposed analytical estimators of the asymptotic variance of matching estimators. Because the standard bootstrap is shown to be invalid, together with subsampling (Politis, Romano, and Wolf, 1999) these are now the only available methods of inference that are formally justified.[4]

The rest of the article is organized as follows. Section 2 reviews the basic notation and setting of matching estimators. Section 3 presents theoretical results on the lack of validity of the bootstrap for matching estimators, along with simulations that confirm the formal results. Section 4 concludes. The appendix contains proofs.

2 Set up

2.1 Basic Model

In this article we adopt the standard model of treatment effects under unconfoundedness (Rubin, 1978; Rosenbaum and Rubin, 1983, Heckman, Ichimura and Todd, 1997, Rosenbaum, 2001, Imbens, 2004). The goal is to evaluate the effect of a treatment on the basis of data on outcomes and covariates for treated and control units. We have a random sample of N_0 units from the control population, and a random sample of N_1 units from the treated population, with $N = N_0 + N_1$. Each unit is characterized by a pair of potential outcomes, $Y_i(0)$ and $Y_i(1)$, denoting the outcomes under the control and active treatment respectively. We observe $Y_i(0)$ for units in the control sample, and $Y_i(1)$ for units in the treated sample. For all units we observe a covariate vector, X_i.[5] Let W_i indicate whether a unit is from the control sample ($W_i = 0$) or the treatment sample ($W_i = 1$). For each unit we observe the triple (X_i, W_i, Y_i) where $Y_i = W_i\,Y_i(1) + (1 - W_i)\,Y_i(0)$ is the observed

---

[4]Politis, Romano, and Wolf (1999) show that subsampling produces valid inference for statistics with stable asymptotic distributions.

[5]To simplify our proof of lack of validity of the bootstrap we will consider in our calculations the case with a scalar covariate. With higher dimensional covariates there is the additional complication of biases that may dominate the asymptotic distribution of matching estimators (Abadie and Imbens, 2006).

outcome. Let $\mathbf{X}$ an N-column matrix with column i equal to X_i, and similar for $\mathbf{Y}$ and $\mathbf{W}$. Also, let $\mathbf{X}_0$ denote the N-column matrix with column i equal to $(1 - W_i) X_i$, and $\mathbf{X}_1$ the N-column matrix with column i equal to $W_i X_i$. The following two assumptions are the identification conditions behind matching estimators.

Assumption 2.1: (UNCONFOUNDEDNESS) *For almost all x, $(Y_i(1), Y_i(0))$ is indepen-dent of W_i conditional on $X_i = x$, or*

$$\Big(Y_i(0), Y_i(1)\Big) \perp\!\!\!\perp W_i \mid X_i = x, \quad (a.s.)$$

Assumption 2.2: (OVERLAP) *For some $c > 0$, and almost all x*

$$c \leq \Pr(W_i = 1 | X_i = x) \leq 1 - c.$$

In this article we focus on matching estimation of the average treatment effect for the treated:[6]

$$\tau = \mathbb{E}[Y_i(1) - Y_i(0)|W_i = 1]. \tag{2.1}$$

A nearest neighbor matching estimator of τ matches each treated unit i to the control unit j with the closest value for the covariate, and then averages the within-pair outcome differences, $Y_i - Y_j$, over the N_1 matched pairs. Here we focus on the case of matching with replacement, so each control unit can be used as a match for more than one treated units.

Formally, for all treated units i (that is, units with $W_i = 1$) let D_i be the distance between the covariate value for observation i and the covariate value for the closest (control) match:

$$D_i = \min_{j=1,\ldots,N:W_j=0} \|X_i - X_j\|.$$

Then let

$$\mathcal{J}(i) = \{j \in \{1, 2, \ldots, N\} : W_j = 0, \|X_i - X_j\| = D_i\}$$

---

[6]In many cases, the interest is in the average effect for the entire population. We focus here on the average effect for the treated because it simplifies the calculations below. Since the overall average effect is the weighted sum of the average effect for the treated and the average effect for the controls it suffices to show that the bootstrap is not valid for one of the components.

be the set of closest matches for treated unit i. If unit i is a control unit, then $\mathcal{J}(i)$ is defined to be the empty set. When X_i is continuously distributed, the set $\mathcal{J}(i)$ will consist of a single index with probability one, but for bootstrap samples there will often be more than one index in this set (because an observation from the original sample may appear multiple times in the bootstrap sample). For each treated unit, i, let

$$\hat{Y}_i(0) = \frac{1}{\#\mathcal{J}(i)} \sum_{j \in \mathcal{J}(i)} Y_j$$

be the average outcome in the set of the closest matches for observation i, where $\#\mathcal{J}(i)$ is the number of elements of the set $\mathcal{J}(i)$. The matching estimator of τ is then

$$\hat{\tau} = \frac{1}{N_1} \sum_{i:W_i=1} \left(Y_i - \hat{Y}_i(0) \right). \tag{2.2}$$

For the subsequent discussion it is useful to write the estimator in a different way. Let K_i denote the weighted number of times unit i is used as a match (if unit i is a control unit, with $K_i = 0$ if unit i is a treated unit):

$$K_i = \begin{cases} 0 & \text{if } W_i = 1, \\ \sum_{W_j=1} 1\{i \in \mathcal{J}(j)\} \frac{1}{\#\mathcal{J}(j)} & \text{if } W_i = 0. \end{cases}$$

Then we can write

$$\hat{\tau} = \frac{1}{N_1} \sum_{i=1}^{N} (W_i - K_i) Y_i. \tag{2.3}$$

Let

$$K_i' = \begin{cases} 0 & \text{if } W_i = 1, \\ \sum_{W_j=1} 1\{i \in \mathcal{J}(j)\} \left(\frac{1}{\#\mathcal{J}(j)} \right)^2 & \text{if } W_i = 0. \end{cases}$$

Abadie and Imbens (2006) prove that under certain conditions (for example, when X is a scalar variable) the nearest-neighbor matching estimator in (2.2) is root-N consistent and asymptotically normal with zero asymptotic bias.[7] Abadie and Imbens propose two

---

[7]More generally, Abadie and Imbens (2002) propose a bias correction that makes matching estimators root-N consistent and asymptotically normal regardless of the dimension of X.

variance estimators:

$$\widehat{V}^{AI,I} = \frac{1}{N_1^2} \sum_{i=1}^{N} \left(W_i - K_i\right)^2 \widehat{\sigma}^2(X_i, W_i),$$

and

$$\widehat{V}^{AI,II} = \frac{1}{N_1^2} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i(0) - \hat{\tau}\right)^2 + \frac{1}{N_1^2} \sum_{i=1}^{N} (K_i^2 - K_i') \widehat{\sigma}^2(X_i, W_i),$$

where $\widehat{\sigma}^2(X_i, W_i)$ is an estimator of the conditional variance of Y_i given W_i and X_i based on matching. Let $l_j(i)$ be the j-th closest match to unit i, in terms of the covariates, among the units with the same value for the treatment (that is, units in the treatment groups are matched to units in the treatment group, and units in the control group are matched to units in the control group).[8] Define

$$\widehat{\sigma}^2(X_i, W_i) = \frac{J}{J+1} \left(Y_i - \frac{1}{J} \sum_{j=1}^{J} Y_{l_j(i)}\right)^2. \tag{2.4}$$

Let $\mathbb{V}(\hat{\tau})$ be the variance of $\hat{\tau}$, and let $\mathbb{V}(\hat{\tau}|\mathbf{X}, \mathbf{W})$ the variance of $\hat{\tau}$ conditional on $\mathbf{X}$ and $\mathbf{W}$. Abadie and Imbens (2006) show that (under weak regularity conditions) the normalized version of first variance estimator, $N_1 \widehat{V}^{AI,I}$ is consistent for the normalized conditional variance, $N_1 \mathbb{V}(\hat{\tau}|\mathbf{X}, \mathbf{W})$:

$$N_1 \left(\mathbb{V}(\hat{\tau}|\mathbf{X}, \mathbf{W}) - \widehat{V}^{AI,I}\right) \overset{p}{\longrightarrow} 0,$$

for fixed J as $N \to \infty$. The normalized version of the second variance estimator, $N_1 \widehat{V}^{AI,II}$, is consistent for the normalized marginal variance, $N_1 \mathbb{V}(\hat{\tau})$:

$$N_1 \left(\mathbb{V}(\hat{\tau}) - \widehat{V}^{AI,II}\right) \overset{p}{\longrightarrow} 0,$$

for fixed J as $N \to \infty$.

---

[8]To simplify the notation, here we consider only the case without matching ties. The extension to accommodate ties is immediate (see Abadie, Drukker, Herr, and Imbens, 2004), but it is not required for the purpose of the analysis in this article.

2.2 The Bootstrap

We consider two versions of the bootstrap in this discussion. The first version centers the bootstrap variance at the matching estimate in the original sample. The second version centers the bootstrap variance at the mean of the bootstrap distribution of the matching estimator.

Consider a random sample $\mathbf{Z} = (\mathbf{X}, \mathbf{W}, \mathbf{Y})$ with N_0 controls and N_1 treated units. The matching estimator, $\hat{\tau}$, is a functional $t(\cdot)$ of the original sample: $\hat{\tau} = t(\mathbf{Z})$. We construct a bootstrap sample, $\mathbf{Z_b}$, with N_0 controls and N_1 treated by sampling with replacement from the two subsamples. We then calculate the bootstrap estimator, $\hat{\tau}_b$, applying the functional $t(\cdot)$ to the bootstrap sample: $\hat{\tau}_b = t(\mathbf{Z_b})$. The first version of the bootstrap variance is the second moment of $(\hat{\tau}_b - \hat{\tau})$ conditional on the sample, $\mathbf{Z}$:

$$V^{B,I} = v^I(\mathbf{Z}) = \mathbb{E}\left[\left(\hat{\tau}_b - \hat{\tau}\right)^2 \middle| \mathbf{Z}\right]. \tag{2.5}$$

The second version of the bootstrap variance centers the bootstrap variance at the bootstrap mean, $\mathbb{E}[\hat{\tau}_b|\mathbf{Z}]$, rather than at the original estimate, $\hat{\tau}$:

$$V^{B,II} = v^{II}(\mathbf{Z}) = \mathbb{E}\left[\left(\hat{\tau}_b - \mathbb{E}\left[\hat{\tau}_b|\mathbf{Z}\right]\right)^2 \middle| \mathbf{Z}\right]. \tag{2.6}$$

Although these bootstrap variances are defined in terms of the original sample $\mathbf{Z}$, in practice an easier way to calculate them is by drawing B bootstrap samples. Given B bootstrap samples with bootstrap estimates $\hat{\tau}_b$, for $b = 1, \ldots, B$, we can obtain unbiased estimators for these two variances as

$$\hat{V}^{B,I} = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\tau}_b - \hat{\tau}\right)^2,$$

and

$$\hat{V}^{B,II} = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\tau}_b - \left(\frac{1}{B}\sum_{b=1}^{B}\hat{\tau}_b\right)\right)^2.$$

We will focus on the first bootstrap variance, $V^{B,I}$, and its unconditional expectation, $\mathbb{E}[V^{B,I}]$. We shall show that in general $N_1 \mathbb{E}[V^{B,I}]$ does not converge to $N_1 \mathbb{V}(\hat{\tau})$. We will show that in some cases the limit of $N_1(\mathbb{E}[V^{B,I}] - \mathbb{V}(\hat{\tau}))$ is positive and that in other cases

this limit is negative. As a result, we will show that $N_1 V^{B,I}$ is not a consistent estimator of the limit of $N_1 \mathbb{V}(\hat{\tau})$. This will indirectly imply that $N_1 V^{B,II}$ is not consistent either. Because

$$\mathbb{E}\left[\left(\hat{\tau}_b - \hat{\tau}\right)^2 \middle| \mathbf{Z}\right] \geq \mathbb{E}\left[\left(\hat{\tau}_b - \mathbb{E}\left[\hat{\tau}_b | \mathbf{Z}\right]\right)^2 \middle| \mathbf{Z}\right],$$

it follows that $\mathbb{E}[V^{B,I}] \geq \mathbb{E}[V^{B,II}]$. Thus in the cases where the limit of $N_1(\mathbb{E}[V^{B,I}] - \mathbb{V}(\hat{\tau}))$ is smaller than zero, it follows that the limit of $N_1(\mathbb{E}[V^{B,II}] - \mathbb{V}(\hat{\tau}))$ is also smaller than zero.

In most standard settings, both centering the bootstrap variance at the estimate in the original sample or at the average of the bootstrap distribution of the estimator lead to valid confidence intervals. In fact, in many settings the average of the bootstrap distribution of an estimator is identical to the estimate in the original sample. For example, if we are interested in constructing a confidence interval for the population mean $\mu = \mathbb{E}[X]$ given a random sample $X_1, \ldots, X_N$, the expected value of the bootstrap statistic, $\mathbb{E}[\hat{\mu}_b | X_1, \ldots, X_N]$, is equal to the sample average for the original sample, $\hat{\mu} = \sum_i X_i / N$. For matching estimators, however, it is easy to construct examples where the average of the bootstrap distribution of the estimator differs from the estimate in the original sample. As a result, the two bootstrap variance estimators will lead to different confidence intervals with potentially different coverage rates.

3 An Example where the Bootstrap Fails

In this section we discuss in detail a specific example where we can calculate the limits of $N_1 \mathbb{V}(\hat{\tau})$ and $N_1 \mathbb{E}[V^{B,I}]$ and show that they differ.

3.1 Data Generating Process

We consider the following data generating process:

Assumption 3.1: *The marginal distribution of the covariate X is uniform on the interval $[0, 1]$*

Assumption 3.2: *The ratio of treated and control units is $N_1/N_0 = \alpha$ for some $\alpha > 0$.*

Assumption 3.3: *The propensity score $e(x) = \Pr(W_i = 1 | X_i = x)$ is constant as a function of x.*

Assumption 3.4: *The distribution of $Y_i(1)$ is degenerate with $\Pr(Y_i(1) = \tau) = 1$, and the conditional distribution of $Y_i(0)$ given $X_i = x$ is normal with mean zero and variance one.*

The implication of Assumptions 3.2 and 3.3 is that the propensity score is $e(x) = \alpha/(1 + \alpha)$.

3.2 Exact Variance and Large Sample Distribution

The data generating process implies that conditional on $X = x$ the treatment effect is equal to $\mathbb{E}[Y(1) - Y(0)|X = x] = \tau$ for all x. Therefore, the average treatment effect for the treated is equal to τ. Under this data generating process $\sum_i W_i Y_i / N_1 = \sum_i W_i Y_i(1)/N_1 = \tau$, which along with equation (2.3) implies:

$$\hat{\tau} - \tau = -\frac{1}{N_1} \sum_{i=1}^{N} K_i Y_i.$$

Conditional on $\mathbf{X}$ and $\mathbf{W}$ the only stochastic component of $\hat{\tau}$ is $\mathbf{Y}$. By Assumption 3.4 the Y_i-s are mean zero, unit variance, and independent of $\mathbf{X}$. Thus $\mathbb{E}[\hat{\tau} - \tau | \mathbf{X}, \mathbf{W}] = 0$. Because (i) $\mathbb{E}[Y_i Y_j | W_i = 0, \mathbf{X}, \mathbf{W}] = 0$ for $i \neq j$, (ii) $\mathbb{E}[Y_i^2 | W_i = 0, \mathbf{X}, \mathbf{W}] = 1$ and (iii) K_i is a deterministic function of $\mathbf{X}$ and $\mathbf{W}$, it also follows that the conditional variance of $\hat{\tau}$ given $\mathbf{X}$ and $\mathbf{W}$ is

$$\mathbb{V}(\hat{\tau} | \mathbf{X}, \mathbf{W}) = \frac{1}{N_1^2} \sum_{i=1}^{N} K_i^2.$$

Because $\mathbb{V}(\mathbb{E}[\hat{\tau} | \mathbf{X}, \mathbf{W}]) = \mathbb{V}(\tau) = 0$, the (exact) unconditional variance of the matching estimator is therefore equal to the expected value of the conditional variance:

$$\mathbb{V}(\hat{\tau}) = \frac{N_0}{N_1^2} \mathbb{E}\left[K_i^2 | W_i = 0\right]. \tag{3.7}$$

Lemma 3.1: (Exact Variance of Matching Estimator)
Suppose that Assumptions 2.1, 2.2, and 3.1-3.4 hold. Then

(i) the exact variance of the matching estimator is

$$\mathbb{V}(\hat{\tau}) = \frac{1}{N_1} + \frac{3}{2}\frac{(N_1 - 1)(N_0 + 8/3)}{N_1(N_0 + 1)(N_0 + 2)}, \tag{3.8}$$

(ii) as $N \to \infty$,

$$N_1\,\mathbb{V}(\hat{\tau}) \to 1 + \frac{3}{2}\,\alpha, \tag{3.9}$$

and (iii),

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

All proofs are given in the Appendix.

3.3 The Bootstrap Variance

Now we analyze the properties of the bootstrap variance, $V^{B,I}$ in (2.5). As before, let $\mathbf{Z} = (\mathbf{X}, \mathbf{W}, \mathbf{Y})$ denote the original sample. We will look at the distribution of statistics both conditional on the original sample, as well as over replications of the original sample drawn from the same distribution. Notice that

$$\mathbb{E}\left[V^{B,I}\right] = \mathbb{E}\left[\mathbb{E}\left[(\hat{\tau}_b - \hat{\tau})^2 \middle| \mathbf{Z}\right]\right] = \mathbb{E}\left[(\hat{\tau}_b - \hat{\tau})^2\right] \tag{3.10}$$

is the expected bootstrap variance. The following lemma establishes the limit of $N_1\,\mathbb{E}[V^{B,I}]$ under our data generating process.

Lemma 3.2: (BOOTSTRAP VARIANCE I) *Suppose that Assumptions 3.1-3.4 hold. Then, as $N \to \infty$:*

$$N_1\,\mathbb{E}[V^{B,I}] \to 1 + \frac{3}{2}\,\alpha\,\frac{5\exp(-1) - 2\exp(-2)}{3\,(1 - \exp(-1))} + 2\exp(-1). \tag{3.11}$$

Recall that the limit of the normalized variance of $\hat{\tau}$ is $1 + (3/2)\,\alpha$. For small values of α the limit of the expected bootstrap variance exceeds the limit variance by the third term in (3.11), $2\exp(-1) \simeq 0.74$, or 74%. For large values of α the second term in (3.11) dominates and the ratio of the limit expected bootstrap and limit variance is equal to the factor in the second term of (3.11) multiplying $(3/2)\alpha$. Since $(5\exp(-1) -$

$2 \exp(-2))/(3(1 - \exp(-1))) \simeq 0.83$, it follows that as α increases, the ratio of the limit expected bootstrap variance to the limit variance asymptotes to 0.83, suggesting that in large samples the bootstrap variance can under as well as over estimate the true variance.

So far, we have established the relation between the limiting variance of the estimator and the limit of the average bootstrap variance. We end this section with a discussion of the implications of the previous two lemmas for the validity of the bootstrap. The first version of the bootstrap provides a valid estimator of the asymptotic variance of the simple matching estimator if:

$$N_1 \left(\mathbb{E}\left[(\widehat{\tau}_b - \widehat{\tau})^2 \,\middle|\, \mathbf{Z} \right] - \mathbb{V}(\widehat{\tau}) \right) \xrightarrow{p} 0.$$

Lemma 3.1 shows that:

$$N_1 \mathbb{V}(\widehat{\tau}) \longrightarrow 1 + \frac{3}{2}\alpha.$$

Lemma 3.2 shows that

$$N_1 \mathbb{E}\left[(\widehat{\tau}_b - \widehat{\tau})^2 \right] \longrightarrow 1 + \frac{3}{2}\alpha \frac{5\exp(-1) - 2\exp(-2)}{3(1 - \exp(-1))} + 2\exp(-1).$$

Assume that the first version of the bootstrap provides a valid estimator of the asymptotic variance of the simple matching estimator. Then,

$$N_1 \mathbb{E}\left[(\widehat{\tau}_b - \widehat{\tau})^2 \,\middle|\, \mathbf{Z} \right] \xrightarrow{p} 1 + \frac{3}{2}\alpha.$$

Because $N_1 \mathbb{E}\left[(\widehat{\tau}_b - \widehat{\tau})^2 \,\middle|\, \mathbf{Z} \right] \geq 0$, it follows by Portmanteau Lemma (see, e.g., van der Vaart, 1998, page 6) that, as $N \to \infty$,

$$1 + \frac{3}{2}\alpha \leq \lim \mathbb{E}\left[N_1 \mathbb{E}\left[(\widehat{\tau}_b - \widehat{\tau})^2 \,\middle|\, \mathbf{Z} \right] \right] = \lim N_1 \mathbb{E}\left[(\widehat{\tau}_b - \widehat{\tau})^2 \right]$$
$$= 1 + \frac{3}{2}\alpha \frac{5\exp(-1) - 2\exp(-2)}{3(1 - \exp(-1))} + 2\exp(-1).$$

However, the algebraic inequality

$$1 + \frac{3}{2}\alpha \leq 1 + \frac{3}{2}\alpha \frac{5\exp(-1) - 2\exp(-2)}{3(1 - \exp(-1))} + 2\exp(-1),$$

does not hold for large enough α. As a result, the first version of the bootstrap does not provide a valid estimator of the asymptotic variance of the simple matching estimator.

The second version of the bootstrap provides a valid estimator of the asymptotic variance of the simple matching estimator if:

$$N_1\left(\mathbb{E}\left[(\widehat{\tau}_b - \mathbb{E}[\widehat{\tau}_b|\mathbf{Z}])^2\big|\,\mathbf{Z}\right] - \mathbb{V}(\widehat{\tau})\right) \xrightarrow{p} 0.$$

Assume that the second version of the bootstrap provides a valid estimator of the asymptotic variance of the simple matching estimator. Then,

$$N_1\mathbb{E}\left[(\widehat{\tau}_b - \mathbb{E}[\widehat{\tau}_b|\mathbf{Z}])^2\big|\,\mathbf{Z}\right] \xrightarrow{p} 1 + \frac{3}{2}\,\alpha.$$

Notice that $\mathbb{E}\left[(\widehat{\tau}_b - \mathbb{E}[\widehat{\tau}_b|\mathbf{Z}])^2\big|\,\mathbf{Z}\right] \leq \mathbb{E}\left[(\widehat{\tau}_b - \widehat{\tau})^2\big|\,\mathbf{Z}\right]$. By Portmanteau Lemma, as $N \to \infty$

$$1 + \frac{3}{2}\,\alpha \leq \liminf \mathbb{E}\left[N_1\mathbb{E}\left[(\widehat{\tau}_b - \mathbb{E}[\widehat{\tau}_b|\mathbf{Z}])^2\big|\,\mathbf{Z}\right]\right] \leq \lim E\left[N_1\mathbb{E}\left[(\widehat{\tau}_b - \widehat{\tau})^2\big|\,\mathbf{Z}\right]\right]$$

$$= \lim N_1\mathbb{E}\left[(\widehat{\tau}_b - \widehat{\tau})^2\right] = 1 + \frac{3}{2}\,\alpha\,\frac{5\exp(-1) - 2\exp(-2)}{3(1 - \exp(-1))} + 2\exp(-1).$$

Again, this inequality does not hold for large enough α. As a result, the second version of the bootstrap does not provide a valid estimator of the asymptotic variance of the simple matching estimator.

3.4 Simulations

We consider three designs: $N_0 = N_1 = 100$ (Design I), $N_0 = 100$, $N_1 = 1000$ (Design II), and $N_0 = 1000$, $N_1 = 100$ (Design III), We use 10,000 replications, and 100 bootstrap samples in each replication. These designs are partially motivated by Figure 1, which gives the ratio of the limit of the expectation of the bootstrap variance (given in equation (3.11)) to limit of the actual variance (given in equation (3.9)), for different values of α. On the horizontal axis is the log of α. As α converges to zero the variance ratio converges to 1.74; at $\alpha = 1$ the variance ratio is 1.19; and as α goes to infinity the variance ratio converges to 0.83. The vertical dashed lines indicate the three designs that we adopt in our simulations: $\alpha = 0.1$, $\alpha = 1$, and $\alpha = 10$.

The simulation results are reported in Table 1. The first row of the table gives normalized exact variances, $N_1\mathbb{V}(\widehat{\tau})$, calculated from equation (3.8). The second and third rows present averages (over the 10,000 simulation replications) of the normalized

variance estimators from Abadie and Imbens (2006). The second row reports averages of $N_1\widehat{V}^{AI,I}$ and the third row reports averages of $N_1\widehat{V}^{AI,II}$. In large samples, $N_1\widehat{V}^{AI,I}$ and $N_1\widehat{V}^{AI,II}$ are consistent for $N_1\mathbb{V}(\hat{\tau}|\mathbf{X},\mathbf{W})$ and $N_1\mathbb{V}(\hat{\tau})$, respectively. Because, for our data generating process, the conditional average treatment effect is zero for all values of the covariates, $N_1\widehat{V}^{AI,I}$ and $N_1\widehat{V}^{AI,II}$ converge to the same parameter. Standard errors (for the averages over 10,000 replications) are reported in parentheses. The first three rows of Table 1 allow us to assess the difference between the averages of the Abadie-Imbens (AI) variance estimators and the theoretical variances. For example, for Design I, the normalized AI Var I estimator $(N_1\widehat{V}^{AI,I})$ is on average 2.449, with a standard error of 0.006. The theoretical variance is 2.480, so the difference between the theoretical and AI variance I is approximately 1%, although it is statistically significant at about 5 standard errors. Given the theoretical justification of the variance estimator, this difference is a finite sample phenomenon.

The fourth row reports the limit of normalized expected bootstrap variance, $N_1\mathbb{E}[V^{B,I}]$, calculated as in (3.11). The fifth and sixth rows give normalized averages of the estimated bootstrap variances, $N_1\widehat{V}^{B,I}$ and $N_1\widehat{V}^{B,II}$, over the 10,000 replications. These variances are estimated for each replication using 100 bootstrap samples, and then averaged over all replications. Again it is interesting to compare the average of the estimated bootstrap variance in the fifth row to the limit of the expected bootstrap variance in the fourth row. The differences between the fourth and fifth rows are small (although significantly different from zero as a result of the small sample size). The limited number of bootstrap replications makes these averages noisier than they would otherwise be, but it does not affect the average difference. The results in the fifth row illustrate our theoretical calculations in Lemma 3.2: the average bootstrap variance can over-estimate or under-estimate the variance of the matching estimator.

The next two panels of the table report coverage rates, first for nominal 90% confidence intervals and then for nominal 95% confidence intervals. The standard errors for the coverage rates reflect the uncertainty coming from the finite number of replications (10,000). They are equal to $\sqrt{p(1-p)/R}$ where for the second panel $p = 0.9$ and for

the third panel $p = 0.95$, and $R = 10,000$ is the number of replications.

The first rows of the last two panels of Table 1 report coverage rates of 90% and 95% confidence intervals constructed in each replication as the point estimate, $\hat{\tau}$, plus/minus 1.645 and 1.96 times the square root of the variance in (3.8). The results show coverage rates which are statistically indistinguishable from the nominal levels, for all three designs and both levels (90% and 95%). The second row of the second and third panels of Table 1 report coverage rates for confidence intervals calculated as in the preceding row but using the estimator $\widehat{V}^{AI,I}$ in (2.5). The third row report coverage rates for confidence intervals constructed with the estimator $\widehat{V}^{AI,II}$ in (2.6). Both $\widehat{V}^{AI,I}$ and $\widehat{V}^{AI,II}$ produce confidence intervals with coverage rates that are statistically indistinguishable from the nominal levels.

The last two rows of the second panel of Table 1 report coverage rates for bootstrap confidence intervals obtained by adding and subtracting 1.645 times the square root of the estimated bootstrap variance in each replication, again over the 10,000 replications. The third panel gives the corresponding numbers for 95% confidence intervals.

Our simulations reflect the lack of validity of the bootstrap found in the theoretical calculations. Coverage rates of confidence intervals constructed with the bootstrap estimators of the variance are different from nominal levels in substantially important and statistically highly significant magnitudes. In Designs I and III the bootstrap has coverage larger the nominal coverage. In Design II the bootstrap has coverage smaller than nominal. In neither case the difference is huge, but it is important to stress that this difference will not disappear with a larger sample size, and that it may be more substantial for different data generating processes.

The bootstrap calculations in this table are based on 100 bootstrap replications. Increasing the number of bootstrap replications significantly for all designs was infeasible as matching is already computationally expensive.[9] We therefore investigated the implications of this choice for Design I, which is the fastest to run. For the same 10,000

---

[9]Each calculation of the matching estimator requires N_1 searches for the minimum of an array of length N_0, so that with B bootstrap replications and R simulations one quickly requires large amounts of computer time.

14

replications we calculated both the coverage rates for the 90% and 95% confidence intervals based on 100 bootstrap replications and based on 1,000 bootstrap replications. For the confidence intervals based on $\widehat{V}^{B,I}$ the coverage rate for a 90% nominal level was 0.002 (s.e. 0.001) higher with 1,000 bootstrap replications than with 100 bootstrap replications. The coverage rate for the 95% confidence interval was 0.003 (s.e., 0.001) higher with 1,000 bootstrap replications than with 100 bootstrap replications. Because the difference between the bootstrap coverage rates and the nominal coverage rates for this design are 0.031 and 0.022 for the 90% and 95% confidence intervals respectively, the number of bootstrap replications can only explain approximately 6-15% of the difference between the bootstrap and nominal coverage rates. We therefore conclude that using more bootstrap replications would not substantially change the results in Table 1.

4   Conclusion

In this article we prove that the bootstrap is not valid for the standard nearest-neighbor matching estimator with replacement. This is a somewhat surprising discovery, because in the case with a scalar covariate the matching estimator is root-N consistent and asymptotically normally distributed with zero asymptotic bias. However, the extreme non-smooth nature of matching estimators and the lack of evidence that the estimator is asymptotically linear explain the lack of validity of the bootstrap. We investigate a special case where it is possible to work out the exact variance of the estimator as well as the limit of the average bootstrap variance. We show that in this case the limit of the average bootstrap variance can be greater or smaller than the limit variance of the matching estimator. This implies that the standard bootstrap fails to provide valid inference for the matching estimator studied in this article. A small Monte Carlo study supports the theoretical calculations. The implication for empirical practice of these results is that for nearest-neighbor matching estimators with replacement one should use the variance estimators developed by Abadie and Imbens (2006) or the subsampling bootstrap (Politis, Romano and Wolf, 1999). It may well be that if the number of neighbors increases with the sample size the matching estimator does become asymptotically linear and sufficiently

regular for the bootstrap to be valid. However, the increased risk of a substantial bias has led many researchers to focus on estimators where the number of matches is very small, often just one, and the asymptotics based on an increasing number of matches may not provide a good approximation in such cases.

Finally, our results cast doubts on the validity of the standard bootstrap for other estimators that are asymptotically normal but not asymptotically linear (see, e.g., Newey and Windmeijer, 2005).

Appendix

Before proving Lemma 3.1 we introduce some notation and preliminary results. Let $X_1, \ldots, X_N$ be a random sample from a continuous distribution. Let M_j be the index of the closest match for unit j. That is, if $W_j = 1$, then M_j is the unique index (ties happen with probability zero), with $W_{M_j} = 0$, such that $\|X_j - X_{M_j}\| \leq \|X_j - X_i\|$, for all i such that $W_i = 0$. If $W_j = 0$, then $M_j = 0$. Let K_i be the number of times unit i is the closest match for a treated observation:

$$K_i = (1 - W_i) \sum_{j=1}^{N} W_j \, 1\{M_j = i\}.$$

Following this definition K_i is zero for treated units. Using this notation, we can write the estimator for the average treatment effect on the treated as:

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

Also, let P_i be the probability that the closest match for a randomly chosen treated unit j is unit i, conditional on both the vector of treatment indicators $\mathbf{W}$ and on vector of covariates for the control units $\mathbf{X}_0$:

$$P_i = \Pr(M_j = i | W_j = 1, \mathbf{W}, \mathbf{X}_0).$$

For treated units we define $P_i = 0$.

The following lemma provides some properties of the order statistics of a sample from the standard uniform distribution.

Lemma A.1: *Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(N)}$ be the order statistics of a random sample of size N from a standard uniform distribution, $U(0,1)$. Then, for $1 \leq i \leq j \leq N$,*

$$E[X_{(i)}^r (1 - X_{(j)})^s] = \frac{i^{[r]}(N - j + 1)^{[s]}}{(N + 1)^{[r+s]}},$$

where for a positive integer, a, and a non-negative integer, b: $a^{[b]} = (a + b - 1)!/(a - 1)!$. Moreover, for $1 \leq i \leq N$, $X_{(i)}$ has a Beta distribution with parameters $(i, N - i + 1)$; for $1 \leq i \leq j \leq N$, $(X_{(j)} - X_{(i)})$ has a Beta distribution with parameters $(j - i, N - (j - i) + 1)$.

Proof: All the results of this lemma, with the exception of the distribution of differences of order statistics, appear in Johnson, Kotz, and Balakrishnan (1994). The distribution of differences of order statistics can be easily derived from the joint distribution of order statistics provided in Johnson, Kotz, and Balakrishnan (1994). □

Notice that the lemma implies the following results:

$$
\begin{aligned}
E[X_{(i)}] &= \frac{i}{N + 1} && \text{for } 1 \leq i \leq N, \\
E[X_{(i)}^2] &= \frac{i(i + 1)}{(N + 1)(N + 2)} && \text{for } 1 \leq i \leq N, \\
E[X_{(i)} X_{(j)}] &= \frac{i(j + 1)}{(N + 1)(N + 2)} && \text{for } 1 \leq i \leq j \leq N.
\end{aligned}
$$

First we investigate the first two moments of K_i, starting by studying the conditional distribution of K_i given $\mathbf{X}_0$ and $\mathbf{W}$.

17

Lemma A.2: (CONDITIONAL DISTRIBUTION AND MOMENTS OF K_i)
Suppose that assumptions 3.1-3.3 hold. Then, the distribution of K_i conditional on $W_i = 0$, $\mathbf{W}$, and $\mathbf{X}_0$ is binomial with parameters (N_1, P_i):

$$K_i | W_i = 0, \mathbf{W}, \mathbf{X}_0 \sim \mathcal{B}(N_1, P_i).$$

Proof: By definition $K_i = (1 - W_i) \sum_{j=1}^{N} W_j \, 1\{M_j = i\}$. The indicator $1\{M_j = i\}$ is equal to one if the closest control unit for X_j is i. This event has probability P_i. In addition, the events $1\{M_{j_1} = i\}$ and $1\{M_{j_2} = i\}$, for $W_{j_1} = W_{j_2} = 1$ and $j_1 \neq j_2$, are independent conditional on $\mathbf{W}$ and $\mathbf{X}_0$. Because there are N_1 treated units the sum of these indicators follows a binomial distribution with parameters N_1 and P_i. □

This implies the following conditional moments for K_i:

$$\mathbb{E}[K_i | \mathbf{W}, \mathbf{X}_0] = (1 - W_i) \, N_1 \, P_i,$$

$$\mathbb{E}[K_i^2 | \mathbf{W}, \mathbf{X}_0] = (1 - W_i) \left(N_1 \, P_i + N_1 \, (N_1 - 1) \, P_i^2 \right).$$

To derive the marginal moments of K_i we need first to analyze the properties of the random variable P_i. Exchangeability of the units implies that the marginal expectation of P_i given N_0, N_1 and $W_i = 0$ is equal to $1/N_0$. To derive the second moment of P_i it is helpful to express P_i in terms of the order statistics of the covariates for the control group. For control unit i let $\iota(i)$ be the order of the covariate for the ith unit among control units:

$$\iota(i) = \sum_{j=1}^{N} (1 - W_j) \, 1\{X_j \leq X_i\}.$$

Furthermore, let $X_{0(k)}$ be the kth order statistic of the covariates among the control units, so that $X_{0(1)} \leq X_{0(2)} \leq \ldots X_{0(N_0)}$, and for control units $X_{0(\iota(i))} = X_i$. Ignoring ties, a treated unit with covariate value x will be matched to control unit i if

$$\frac{X_{0(\iota(i)-1)} + X_{0(\iota(i))}}{2} \leq x \leq \frac{X_{0(\iota(i)+1)} + X_{0(\iota(i))}}{2},$$

if $1 < \iota(i) < N_0$. If $\iota(i) = 1$, then x will be matched to unit i if

$$x \leq \frac{X_{0(2)} + X_{0(1)}}{2},$$

and if $\iota(i) = N_0$, x will be matched to unit i if

$$\frac{X_{0(N_0-1)} + X_{0(N_0)}}{2} < x.$$

To get the value of P_i we need to integrate the density of X conditional on $W = 1$, $f_1(x)$, over these sets. With a uniform distribution for the covariates in the treatment group ($f_1(x) = 1$, for $x \in [0, 1]$), we get the following representation for P_i:

$$P_i = \begin{cases} (X_{0(2)} + X_{0(1)})/2 & \text{if } \iota(i) = 1, \\ \left(X_{0(\iota(i)+1)} - X_{0(\iota(i)-1)} \right)/2 & \text{if } 1 < \iota(i) < N_0, \\ 1 - (X_{0(N_0-1)} + X_{0(N_0)})/2 & \text{if } \iota(i) = N_0. \end{cases} \tag{A.2}$$

Lemma A.3: (MOMENTS OF P_i)
Suppose that Assumptions 3.1–3.3 hold. Then
(i), the second moment of P_i conditional on $W_i = 0$ is

$$\mathbb{E}[P_i^2 | W_i = 0] = \frac{3N_0 + 8}{2N_0(N_0 + 1)(N_0 + 2)},$$

and (ii), the Mth moment of P_i is bounded by

$$\mathbb{E}[P_i^M|W_i = 0] \leq \left(\frac{1+M}{N_0+1}\right)^M.$$

Proof: First, consider (i). Conditional on $W_i = 0$, X_i has a uniform distribution on the interval $[0, 1]$. Using Lemma A.1 and equation (A.2), for interior i (i such that $1 < \iota(i) < N_0$), we have that

$$\mathbb{E}[P_i|1 < \iota(i) < N_0, W_i = 0] = \frac{1}{N_0+1},$$

and

$$\mathbb{E}\left[P_i^2|1 < \iota(i) < N_0, W_i = 0\right] = \frac{3}{2}\frac{1}{(N_0+1)(N_0+2)}.$$

For the smallest and largest observations:

$$\mathbb{E}[P_i|\iota(i) \in \{1, N_0\}, W_i = 0] = \frac{3}{2}\frac{1}{N_0+1},$$

and

$$\mathbb{E}[P_i^2|\iota(i) \in \{1, N_0\}, W_i = 0] = \frac{7}{2(N_0+1)(N_0+2)}.$$

Averaging over all units includes two units at the boundary and $N_0 - 2$ interior values, we obtain:

$$\mathbb{E}[P_i|W_i = 0] = \frac{N_0-2}{N_0}\frac{1}{(N_0+1)} + \frac{2}{N_0}\frac{3}{2}\frac{1}{(N_0+1)} = \frac{1}{N_0},$$

and

$$\mathbb{E}[P_i^2|W_i = 0] = \frac{N_0-2}{N_0}\frac{3}{2}\frac{1}{(N_0+1)(N_0+2)} + \frac{2}{N_0}\frac{7}{2}\frac{1}{(N_0+1)(N_0+2)} = \frac{3N_0+8}{2N_0(N_0+1)(N_0+2)}.$$

For (ii) notice that

$$P_i = \begin{cases} (X_{0(2)} + X_{0(1)})/2 \leq X_{0(2)} & \text{if } \iota(i) = 1, \\ \left(X_{0(\iota(i)+1)} - X_{0(\iota(i)-1)}\right)/2 \leq \left(X_{0(\iota(i)+1)} - X_{0(\iota(i)-1)}\right) & \text{if } 1 < \iota(i) < N_0, \\ 1 - (X_{0(N_0-1)} + X_{0(N_0)})/2 \leq 1 - X_{0(N_0-1)} & \text{if } \iota(i) = N_0. \end{cases} \tag{A.3}$$

Because the right-hand sides of the inequalities in equation (A.3) all have a Beta distribution with parameters $(2, N_0 - 1)$, the moments of P_i are bounded by those of a Beta distribution with parameters 2 and N_0-1. The Mth moment of a Beta distribution with parameters α and β is $\prod_{j=0}^{M-1}(\alpha+j)/(\alpha+\beta+j)$. This is bounded by $(\alpha+M-1)^M/(\alpha+\beta)^M$, which completes the proof of the second part of the Lemma. $\square$

Proof of Lemma 3.1:

First we prove (i). The first step is to calculate $\mathbb{E}[K_i^2|W_i = 0]$. Using Lemmas A.2 and A.3,

$$\mathbb{E}[K_i^2|W_i = 0] = N_1 \mathbb{E}[P_i|W_i = 0] + N_1(N_1 - 1)\mathbb{E}[P_i^2|W_i = 0]$$

$$= \frac{N_1}{N_0} + \frac{3}{2}\frac{N_1(N_1-1)(N_0+8/3)}{N_0(N_0+1)(N_0+2)}.$$

Substituting this into (3.7) we get:

$$\mathbb{V}(\hat{\tau}) = \frac{N_0}{N_1^2}\mathbb{E}[K_i^2|W_i = 0] = \frac{1}{N_1} + \frac{3}{2}\frac{(N_1-1)(N_0+8/3)}{N_1(N_0+1)(N_0+2)},$$

19

proving part (i).

Next, consider part (ii). Multiply the exact variance of $\hat{\tau}$ by N_1 and substitute $N_1 = \alpha N_0$ to get

$$N_1 \, \mathbb{V}(\hat{\tau}) = 1 + \frac{3}{2} \frac{(\alpha N_0 - 1)(N_0 + 8/3)}{(N_0 + 1)(N_0 + 2)}.$$

Then take the limit as $N_0 \to \infty$ to get:

$$\lim_{N \to \infty} N_1 \, \mathbb{V}(\hat{\tau}) = 1 + \frac{3}{2} \alpha.$$

Finally, consider part (iii). Let $S(r, j)$ be a Stirling number of the second kind. The Mth moment of K_i given $\mathbf{W}$ and $\mathbf{X}_0$ is (Johnson, Kotz, and Kemp, 1993):

$$\mathbb{E}[K_i^M | \mathbf{X}_0, W_i = 0] = \sum_{j=0}^{M} \frac{S(M, j) N_0! \, P_i^j}{(N_0 - j)!}.$$

Therefore, applying Lemma A.3 (ii), we obtain that the moments of K_i are uniformly bounded:

$$\begin{aligned}
\mathbb{E}[K_i^M | W_i = 0] &= \sum_{j=0}^{M} \frac{S(M, j) N_0!}{(N_0 - j)!} \mathbb{E}[P_i^j | W_i = 0] \le \sum_{j=0}^{M} \frac{S(M, j) N_0!}{(N_0 - j)!} \left(\frac{1 + M}{N_0 + 1} \right)^j \\
&\le \sum_{j=0}^{M} S(M, j)(1 + M)^j.
\end{aligned}$$

Notice that

$$\mathbb{E}\left[\frac{1}{N_1} \sum_{i=1}^{N} K_i^2 \right] = \frac{N_0}{N_1} \mathbb{E}[K_i^2 | W_i = 0] \to 1 + \frac{3}{2} \alpha,$$

$$\mathbb{V}\left(\frac{1}{N_1} \sum_{i=1}^{N} K_i^2 \right) \le \frac{N_0}{N_1^2} \mathbb{V}(K_i^2 | W_i = 0) \to 0,$$

because $\mathrm{cov}(K_i^2, K_j^2 | W_i = W_j = 0, i \ne j) \le 0$ (see Joag-Dev and Proschan, 1983). Therefore:

$$\frac{1}{N_1} \sum_{i=1}^{N} K_i^2 \xrightarrow{p} 1 + \frac{3}{2} \alpha. \tag{A.4}$$

Finally, we write

$$\hat{\tau} - \tau = \frac{1}{N_1} \sum_{i=1}^{N} \xi_i,$$

where $\xi_i = -K_i Y_i$. Conditional on $\mathbf{X}$ and $\mathbf{W}$ the ξ_i are independent, and the distribution of ξ_i is degenerate at zero for $W_i = 1$ and normal $\mathcal{N}(0, K_i^2)$ for $W_i = 0$. Hence, for any $c \in \mathbb{R}$:

$$\Pr\left(\left(\frac{1}{N_1} \sum_{i=1}^{N} K_i^2 \right)^{-1/2} \sqrt{N_1}(\hat{\tau} - \tau) \le c \,\Big|\, \mathbf{X}, \mathbf{W} \right) = \Phi(c),$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal variable. Integrating over the distribution of $\mathbf{X}$ and $\mathbf{W}$ yields:

$$\Pr\left(\left(\frac{1}{N_1} \sum_{i=1}^{N} K_i^2 \right)^{-1/2} \sqrt{N_1}(\hat{\tau} - \tau) \le c \right) = \Phi(c).$$

Now, Slustky's Theorem implies (iii). $\qquad\square$

Next we introduce some additional notation. Let $R_{b,i}$ be the number of times unit i is in the bootstrap sample. In addition, let $D_{b,i}$ be an indicator for inclusion of unit i in the bootstrap sample, so that $D_{b,i} = 1\{R_{b,i} > 0\}$. Let $N_{b,0} = \sum_{i=1}^{N}(1 - W_i)\,D_{b,i}$ be the number of distinct control units in the bootstrap sample. Finally, define the binary indicator $B_i(x)$, for $i = 1\ldots, N$ to be the indicator for the event that in the bootstrap sample a treated unit with covariate value x would be matched to unit i. That is, for this indicator to be equal to one the following three conditions need to be satisfied: (i) unit i is a control unit, (ii) unit i is in the bootstrap sample, and (iii) the distance between X_i and x is less than or equal to the distance between x and any other control unit in the bootstrap sample. Formally:

$$B_i(x) = \begin{cases} 1 & \text{if } |x - X_i| = \min_{k:W_k=0, D_{b,k}=1} |x - X_k|, \text{ and } D_{b,i} = 1, W_i = 0, \\ 0 & \text{otherwise.} \end{cases}$$

For the N units in the original sample, let $K_{b,i}$ be the number of times unit i is used as a match in the bootstrap sample.

$$K_{b,i} = \sum_{j=1}^{N} W_j\, B_i(X_j)\, R_{b,j}. \qquad (A.5)$$

We can write the estimated treatment effect in the bootstrap sample as

$$\hat{\tau}_b = \frac{1}{N_1} \sum_{i=1}^{N} W_i\, R_{b,i}\, Y_i - K_{b,i}\, Y_i.$$

Because $Y_i(1) = \tau$ by Assumption 3.4, and $\sum_{i=1}^{N} W_i R_{b,i} = N_1$, then

$$\hat{\tau}_b - \tau = -\frac{1}{N_1} \sum_{i=1}^{N} K_{b,i}\, Y_i.$$

The difference between the original estimate $\hat{\tau}$ and the bootstrap estimate $\hat{\tau}_b$ is

$$\hat{\tau}_b - \hat{\tau} = \frac{1}{N_1} \sum_{i=1}^{N} (K_i - K_{b,i})\, Y_i = \frac{1}{\alpha\, N_0} \sum_{i=1}^{N} (K_i - K_{b,i})\, Y_i.$$

We will calculate the expectation

$$N_1\, \mathbb{E}[V^{B,I}] = N_1 \cdot \mathbb{E}[(\hat{\tau}_b - \hat{\tau})^2] = \frac{N_1}{\alpha^2\, N_0^2} \mathbb{E}\left[\sum_{i=1}^{N} \sum_{j=1}^{N} (K_i - K_{b,i})\, Y_i (K_j - K_{b,j})\, Y_j \right].$$

Using the facts that $\mathbb{E}[Y_i^2|\mathbf{X}, \mathbf{W}, W_i = 0] = 1$, and $\mathbb{E}[Y_i Y_j|\mathbf{X}, \mathbf{W}, W_i = W_j = 0] = 0$ if $i \neq j$, this is equal to

$$N_1\, \mathbb{E}[V^{B,I}] = \frac{1}{\alpha} \mathbb{E}\left[(K_{b,i} - K_i)^2 | W_i = 0 \right].$$

The first step in deriving this expectation is to establish some properties of $D_{b,i}$, $R_{b,i}$, $N_{b,0}$, and $B_i(x)$.

Lemma A.4: (PROPERTIES OF $D_{b,i}$, $R_{b,i}$, $N_{b,0}$, AND $B_i(x)$)
Suppose that Assumptions 3.1-3.3 hold. Then, for $w \in \{0,1\}$, and $n \in \{1,\ldots,N_0\}$
(i)

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

21

(ii)

$$D_{b,i}|W_i = w, \mathbf{Z} \sim \mathcal{B}\left(1, 1 - (1 - 1/N_w)^{N_w}\right),$$

(iii)

$$\Pr(N_{b,0} = n) = \binom{N_0}{N_0 - n} \frac{n!}{N_0^{N_0}} S(N_0, n),$$

(iv)

$$\Pr(B_i(X_j) = 1|W_j = 1, W_i = 0, D_{b,i} = 1, N_{b,0}) = \frac{1}{N_{b,0}},$$

(v) for $l \neq j$

$$\Pr(B_i(X_l) B_i(X_j) = 1|W_j = W_l = 1, W_i = 0, D_{b,i} = 1, N_{b,0}) = \frac{3N_{b,0} + 8}{2N_{b,0}(N_{b,0} + 1)(N_{b,0} + 2)},$$

(vi)

$$\mathbb{E}[N_{b,0}/N_0] = 1 - (1 - 1/N_0)^{N_0} \to 1 - \exp(-1),$$

(vii)

$$\frac{1}{N_0} \mathbb{V}(N_{b,0}) = (N_0 - 1)(1 - 2/N_0)^{N_0} + (1 - 1/N_0)^{N_0} - N_0(1 - 1/N_0)^{2N_0} \to \exp(-1)(1 - 2\exp(-1)).$$

Proof: Parts *(i)*, *(ii)*, and *(iv)* are trivial. Part *(iii)* follows easily from equation (3.6) in page 110 of Johnson and Kotz (1977). Next, consider part *(v)*. First condition on $\mathbf{X}_{0b}$ and $\mathbf{W}_b$ (the counterparts of $\mathbf{X}_0$ and $\mathbf{W}$ in the b-th bootstrap sample), and suppose that $D_{b,i} = 1$. The event that a randomly chosen treated unit will be matched to control unit i conditional on $\mathbf{X}_{0b}$ and $\mathbf{W}_b$ depends on the difference in order statistics of the control units in the bootstrap sample. The equivalent in the original sample is P_i. The only difference is that the bootstrap control sample is of size $N_{0,b}$. The conditional probability that two randomly chosen treated units are both matched to control unit i is the square of the difference in order statistics. It marginal expectation is the equivalent in the bootstrap sample of $\mathbb{E}[P_i^2|W_i = 0]$, again with the sample size scaled back to $N_{b,0}$. Parts *(vi)* and *(vii)* can be derived by making use of equation (3.13) on page 114 in Johnson and Kotz (1977). $\square$

Next, we prove a general result for the bootstrap. Consider a sample of size N, indexed by $i = 1, \ldots, N$. Let $D_{b,i}$ be an indicator whether observation i is in bootstrap sample b. Let $N_b = \sum_{i=1}^N D_{b,i}$ be the number of distinct observations in bootstrap sample b.

Lemma A.5: (BOOTSTRAP) *For all $m \geq 0$:*

$$\mathbb{E}\left[\left(\frac{N - N_b}{N}\right)^m\right] \to \exp(-m),$$

and

$$\mathbb{E}\left[\left(\frac{N}{N_b}\right)^m\right] \to \left(\frac{1}{1 - \exp(-1)}\right)^m.$$

Proof: From parts *(vi)* and *(vii)* of Lemma A.4 we obtain that $(N - N_b)/N \xrightarrow{p} \exp(-1)$. By the Continuous Mapping Theorem, $N/N_b \xrightarrow{p} 1/(1 - \exp(-1))$. To obtain convergence of moments it suffices that, for any $m \geq 0$, $\mathbb{E}[((N - N_b)/N)^m]$ and $\mathbb{E}[(N/N_b)^m]$ are uniformly bounded in N (see, e.g., van der Vaart, 1998). For $\mathbb{E}[((N - N_b)/N)^m]$ uniform boundedness follows from the fact that, for any $m \geq 0$,

$((N - N_b)/N)^m \leq 1$. For $\mathbb{E}[(N/N_b)^m]$ the proof is a little bit more complicated. As before, it is enough to show that $\mathbb{E}[(N/N_b)^m]$ is uniformly bounded for $N \geq 1$. Let $\lambda(N) = (1 - 1/N)^N$. Notice that $\lambda(N) < \exp(-1)$ for all $N \geq 1$. Let $0 < \theta < \exp(1) - 1$. Notice that, $1 - (1 + \theta)\lambda(N) > 0$. Therefore:

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{N}{N_b}\right)^m\right] &= \mathbb{E}\left[\left(\frac{N}{N_b}\right)^m \left(1\left\{\frac{N}{N_b} < \frac{1}{1 - (1 + \theta)\lambda(N)}\right\} + 1\left\{\frac{N}{N_b} \geq \frac{1}{1 - (1 + \theta)\lambda(N)}\right\}\right)\right] \\
&\leq \left(\frac{1}{1 - (1 + \theta)\exp(-1)}\right)^m + \mathbb{E}\left[\left(\frac{N}{N_b}\right)^m \Big| \frac{N}{N_b} \geq \frac{1}{1 - (1 + \theta)\lambda(N)}\right] \Pr\left(\frac{N}{N_b} \geq \frac{1}{1 - (1 + \theta)\lambda(N)}\right) \\
&\leq \left(\frac{1}{1 - (1 + \theta)\exp(-1)}\right)^m + N^m \Pr\left(\frac{N}{N_b} \geq \frac{1}{1 - (1 + \theta)\lambda(N)}\right).
\end{aligned}
$$

Therefore, for the expectation $\mathbb{E}[(N/N_b)^m]$ to be uniformly bounded, it is sufficient that the probability $\Pr(N/N_b \geq (1 - (1 + \theta)\lambda(N))^{-1})$ converges to zero at an exponential rate as $N \to \infty$. Notice that

$$
\begin{aligned}
\Pr\left(\frac{N}{N_b} \geq \frac{1}{1 - (1 + \theta)\lambda(N)}\right) &= \Pr\left(N - N_b - N\lambda(N) \geq \theta N\lambda(N)\right) \\
&\leq \Pr\left(|N - N_b - N\lambda(N)| \geq \theta N\lambda(N)\right).
\end{aligned}
$$

Theorem 2 in Kamath, Motwani, Palem, and Spirakis (1995) implies:

$$
\Pr\left(|N - N_b - N\lambda(N)| \geq \theta N\lambda(N)\right) \leq 2\exp\left(-\frac{\theta^2 \lambda(N)^2 (N - 1/2)}{1 - \lambda(N)^2}\right).
$$

Because for $N \geq 1$, $\lambda(N)^2/(1 - \lambda(N)^2)$ is increasing in N (converging to $(\exp(2) - 1)^{-1} > 0$ as $N \to \infty$), the last equation establishes an exponential bound on the tail probability of N/N_b. $\qquad \square$

Lemma A.6: (APPROXIMATE BOOTSTRAP K MOMENTS)
Suppose that assumptions 3.1 to 3.3 hold. Then,
(i)

$$
\mathbb{E}[K_{b,i}^2 \mid W_i = 0] \to 2\alpha + \frac{3}{2}\frac{\alpha^2}{(1 - \exp(-1))},
$$

and (ii),

$$
\mathbb{E}[K_{b,i} K_i | W_i = 0] \to (1 - \exp(-1))\left(\alpha + \frac{3}{2}\alpha^2\right) + \alpha^2 \exp(-1).
$$

Proof: First we prove part (i). Notice that for i, j, l, such that $W_i = 0$, $W_j = W_l = 1$

$$
(R_{b,j}, R_{b,l}) \perp\!\!\!\perp D_{b,i}, B_i(X_j), B_i(X_l).
$$

Notice also that $\{R_{b,j} : W_j = 1\}$ are exchangeable with:

$$
\sum_{W_j = 1} R_{b,j} = N_1.
$$

Therefore, applying Lemma A.4(i), for $W_j = W_l = 1$:

$$
\mathrm{cov}(R_{b,j}, R_{b,l}) = -\frac{\mathbb{V}(R_{b,j})}{(N_1 - 1)} = -\frac{1 - 1/N_1}{(N_1 - 1)} \longrightarrow 0.
$$

As a result,

$$\mathbb{E}[R_{b,j}\,R_{b,l}\mid D_{b,i}=1, B_i(X_j)=B_i(X_l)=1, W_i=0, W_j=W_l=1, j\neq l]$$
$$-\left(\mathbb{E}[R_{b,j}\mid D_{b,i}=1, B_i(X_j)=B_i(X_l)=1, W_i=0, W_j=W_l=1, j\neq l]\right)^2 \longrightarrow 0.$$

By Lemma A.4(i),

$$\mathbb{E}[R_{b,j}\mid D_{b,i}=1, B_i(X_j)=B_i(X_l)=1, W_i=0, W_j=W_l=1, j\neq l]=1.$$

Therefore,

$$E[R_{b,j}\,R_{b,l}\mid D_{b,i}=1, B_i(X_j)=B_i(X_l)=1, W_i=0, W_j=W_l=1, j\neq l]\longrightarrow 1.$$

In addition,

$$\mathbb{E}\left[R_{b,j}^2\mid D_{b,i}=1, B_i(X_j)=1, W_j=1, W_i=0\right]=N_1(1/N_1)+N_1(N_1-1)(1/N_1^2)\longrightarrow 2.$$

Notice that

$$\Pr(D_{b,i}=1|W_i=0, W_j=W_l=1, j\neq l, N_{b,0})=\Pr(D_{b,i}=1|W_i=0, N_{b,0})=\frac{N_{b,0}}{N_0}.$$

Using Bayes' Rule:

$$\Pr(N_{b,0}=n|D_{b,i}=1, W_i=0, W_j=W_l=1, j\neq l)=\Pr(N_{b,0}=n|D_{b,i}=1, W_i=0)$$
$$=\frac{\Pr(D_{b,i}=1|W_i=0, N_{b,0}=n)\Pr(N_{b,0}=n)}{\Pr(D_{b,i}=1|W_i=0)}=\frac{\dfrac{n}{N_0}\Pr(N_{b,0}=n)}{1-(1-1/N_0)^{N_0}}.$$

Therefore,

$$N_0\Pr(B_i(X_j)=1|D_{b,i}=1, W_i=0, W_j=1)$$
$$=N_0\sum_{n=1}^{N_0}\Pr(B_i(X_j)=1|D_{b,i}=1, W_i=0, W_j=1, N_{b,0}=n)$$
$$\times\Pr(N_{b,0}=n|D_{b,i}=1, W_i=0, W_j=1)$$
$$=N_0\sum_{n=1}^{N_0}\frac{1}{n}\left(\frac{n}{N_0}\right)\frac{\Pr(N_{b,0}=n)}{1-(1-1/N_0)^{N_0}}=\frac{1}{1-(1-1/N_0)^{N_0}}\longrightarrow\frac{1}{1-\exp(-1)}.$$

In addition,

$$N_0^2\Pr(B_i(X_j)B_i(X_l)|D_{b,i}=1, W_i=0, W_j=W_l=1, j\neq l, N_{b,0})$$
$$=\frac{3}{2}\frac{N_0^2(N_{b,0}+8/3)}{N_{b,0}(N_{b,0}+1)(N_{b,0}+2)}\xrightarrow{p}\frac{3}{2}\left(\frac{1}{1-\exp(-1)}\right)^2.$$

Therefore

$$\sum_{n=1}^{N_0}\left(N_0^2\Pr(B_i(X_j)B_i(X_l)|D_{b,i}=1, W_i=0, W_j=W_l=1, j\neq l, N_{b,0})\right)^2$$
$$\times\Pr(N_{b,0}=n|D_{b,i}=1, W_i=0, W_j=W_l=1, j\neq l)$$
$$=\sum_{n=1}^{N_0}\left(\frac{3}{2}\frac{N_0^2(n+8/3)}{n(n+1)(n+2)}\right)^2\frac{\dfrac{n}{N_0}\Pr(N_{b,0}=n)}{1-(1-1/N_0)^{N_0}}$$
$$\leq\frac{9}{4}\left(\frac{1}{1-\exp(-1)}\right)\sum_{n=1}^{N_0}\frac{N_0^4(n+8/3)^2}{n^6}\Pr(N_{b,0}=n).$$

Notice that

$$\sum_{n=1}^{N_0} \frac{N_0^4(n+8/3)^2}{n^6} \Pr(N_{b,0}=n) \le \left(1+\frac{16}{3}+\frac{64}{9}\right) \sum_{n=1}^{N_0} \left(\frac{N_0}{n}\right)^4 \Pr(N_{b,0}=n),$$

which is bounded away from infinity (as shown in the proof of Lemma A.5). Convergence in probability of a random variable along with boundedness of its second moment implies convergence of the first moment (see, e.g., van der Vaart, 1998). As a result,

$$N_0^2 \Pr(B_i(X_j)B_i(X_l)|D_{b,i}=1, W_i=0, W_j=W_l=1, j \ne l) \longrightarrow \frac{3}{2}\left(\frac{1}{1-\exp(-1)}\right)^2.$$

Then, using these preliminary results, we obtain:

$$
\begin{aligned}
\mathbb{E}[K_{b,i}^2 \,|\, W_i=0] \;\; = \;\; & \mathbb{E}\left[\sum_{j=1}^N \sum_{l=1}^N W_j W_l \, B_i(X_j)\, B_i(X_l)\, R_{b,j}\, R_{b,l} \;\Big|\; W_i=0\right] \\
= \;\; & \mathbb{E}\left[\sum_{j=1}^N W_j \, B_i(X_j)\, R_{b,j}^2 \;\Big|\; W_i=0\right] \\
+ \;\; & \mathbb{E}\left[\sum_{j=1}^N \sum_{l \ne j} W_j\, W_l\, B_i(X_j)\, B_i(X_l)\, R_{b,j}\, R_{b,l} \;\Big|\; W_i=0\right] \\
= \;\; & N_1 \mathbb{E}\left[R_{b,j}^2 \,|\, D_{b,i}=1, B_i(X_j)=1, W_j=1, W_i=0\right] \\
& \times \Pr\left(B_i(X_j)=1 \,|\, D_{b,i}=1, W_j=1, W_i=0\right) \Pr(D_{b,i}=1 \,|\, W_j=1, W_i=0) \\
+ \;\; & N_1(N_1-1) \mathbb{E}\left[R_{b,j}\, R_{b,l} \,|\, D_{b,i}=1, B_i(X_j)=B_i(X_l)=1, W_j=W_l=1, j \ne l, W_i=0\right] \\
& \times \Pr\left(B_i(X_j)B_i(X_l)=1 \,|\, D_{b,i}=1, W_j=W_l=1, j \ne l, W_i=0\right) \\
& \times \Pr(D_{b,i}=1 \,|\, W_j=W_l=1, j \ne l, W_i=0) \\
\longrightarrow \;\; & 2\alpha + \frac{3}{2}\frac{\alpha^2}{(1-\exp(-1))}.
\end{aligned}
$$

This finishes the proof of part (i). Next, we prove part (ii).

$$\mathbb{E}[K_i K_{b,i} | \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0]$$

$$= \mathbb{E}\left[\sum_{j=1}^{N} W_j 1\{M_j = i\} \sum_{l=1}^{N} W_l B_i(X_l) R_{b,l} \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$= \mathbb{E}\left[\sum_{j=1}^{N} \sum_{l=1}^{N} W_j W_l 1\{M_j = i\} B_i(X_l) R_{b,l} \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$= \mathbb{E}\left[\sum_{j=1}^{N} \sum_{l=1}^{N} W_j W_l 1\{M_j = i\} B_i(X_l) \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$= \mathbb{E}\left[\sum_{j=1}^{N} \sum_{l=1}^{N} W_j W_l 1\{M_j = i\} 1\{M_l = i\} B_i(X_l) \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$+ \mathbb{E}\left[\sum_{j=1}^{N} \sum_{l=1}^{N} W_j W_l 1\{M_j = i\} 1\{M_l \neq i\} B_i(X_l) \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$= \mathbb{E}\left[\sum_{j=1}^{N} \sum_{l=1}^{N} W_j W_l 1\{M_j = i\} 1\{M_l = i\} \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$+ \mathbb{E}\left[\sum_{j=1}^{N} \sum_{l=1}^{N} W_j W_l 1\{M_j = i\} 1\{M_l \neq i\} B_i(X_l) \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$= \mathbb{E}\left[K_i^2 \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$+ \mathbb{E}\left[\sum_{j=1}^{N} \sum_{l \neq j} W_j W_l 1\{M_j = i\} 1\{M_l \neq i\} B_i(X_l) \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right].$$

Conditional on $\mathbf{X}_0$, $W_i = 0$, and $D_{b,i} = 1$, the probability that a treated observation, l, that was not matched to i in the original sample, is matched to i in a bootstrap sample does not depend on the covariate values of the other treated observations (or on $\mathbf{W}$). Therefore:

$$
\begin{aligned}
B_i^0 &= \mathbb{E}[B_i(X_l) | \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0, W_j = W_l = 1, M_j = i, M_l \neq i] \\
&= \mathbb{E}[B_i(X_l) | \mathbf{X}_0, D_{b,i} = 1, W_i = 0, W_l = 1, M_l \neq i].
\end{aligned}
$$

As a result:

$$\mathbb{E}\left[\sum_{j=1}^{N} \sum_{l \neq j} W_j W_l 1\{M_j = i\} 1\{M_l \neq i\} B_i(X_l) \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$= B_i^0 \, \mathbb{E}\left[\sum_{j=1}^{N} \sum_{l \neq j} W_j W_l 1\{M_j = i\} 1\{M_l \neq i\} \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right]$$

$$= B_i^0 \, \mathbb{E}\left[K_i(N_1 - K_i) \,\Big|\, \mathbf{X}_0, \mathbf{W}, D_{b,i} = 1, W_i = 0 \right] = B_i^0 \, \mathbb{E}\left[K_i(N_1 - K_i) \,\Big|\, \mathbf{X}_0, W_i = 0 \right].$$

Conditional on $\mathbf{X}_0$ and $W_i = 0$, K_i has a Binomial distribution with parameters (N_1, P_i). Therefore:

$$
\begin{aligned}
\mathbb{E}[K_i(N_1 - K_i) \,|\, \mathbf{X}_0, W_i = 0] &= N_1^2 P_i - N_1 P_i - N_1(N_1 - 1)P_i^2 \\
&= N_1(N_1 - 1)P_i(1 - P_i).
\end{aligned}
$$

Therefore:

$$\mathbb{E}\left[\sum_{j=1}^{N}\sum_{l\neq j}W_jW_l1\{M_j=i\}1\{M_l\neq i\}B_i(X_l)\,\bigg|\,\iota(i),P_i,D_{b,i}=1,W_i=0\right]$$

$$=\mathbb{E}[B_i^0\,|\,\iota(i),P_i,D_{b,i}=1,W_i=0]N_1(N_1-1)P_i(1-P_i).$$

In addition, the probability that r specified observations do not appear in a bootstrap sample conditional on that another specified observation appears in the sample is (apply Bayes' theorem):

$$\frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}1\{r\leq N_0-1\}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}}.$$

Notice that for a fixed r this probability converges to $\exp(-r)$, as $N_0\to\infty$. Notice also that this probability is bounded by $\exp(-r)/(1-\exp(-1))$, which is integrable:

$$\sum_{r=1}^{\infty}\frac{\exp(-r)}{1-\exp(-1)}=\frac{\exp(-1)}{(1-\exp(-1))^2}.$$

As a result, by the dominated convergence theorem for infinite sums:

$$\lim_{N_0\to\infty}\sum_{r=1}^{\infty}\frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}1\{r\leq N_0-1\}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}}$$

$$=\sum_{r=1}^{\infty}\exp(-r)=\frac{\exp(-1)}{1-\exp(-1)}.$$

For $k,d\in\{1,\dots,N_0-1\}$ and $k+d\leq N_0$, let $\Delta_{d(k)}=X_{0(k+d)}-X_{0(k)}$. In addition, let $\Delta_{d(0)}=X_{0(d)}$, and for $k+d=N_0+1$ let $\Delta_{d(k)}=1-X_{0(k)}$. Notice that:

$$B_i^0=\left(\frac{1}{1-P_i}\right)$$

$$\times\left\{\sum_{r=1}^{N_0-\iota(i)}\left(\frac{\Delta_{1(\iota(i)+r)}}{2}+1\{r=N_0-\iota(i)\}\frac{\Delta_{N_0-\iota(i)+1(\iota(i))}}{2}\right)\frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}}\right.$$

$$\left.+\sum_{r=1}^{\iota(i)-1}\left(\frac{\Delta_{1(\iota(i)-1-r)}}{2}+1\{r=\iota(i)-1\}\frac{\Delta_{\iota(i)(0)}}{2}\right)\frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}}\right\}.$$

In addition, using the results in Lemma A.1 we obtain that, for $1<\iota(i)<N_0$, and $1\leq r\leq N_0-\iota(i)$, we have:

$$\mathbb{E}\left[\Delta_{1(\iota(i)+r)}P_i\,|\,\iota(i),D_{b,i}=1,W_i=0\right]=\frac{1}{(N_0+1)(N_0+2)},$$

$$\mathbb{E}\left[\Delta_{N_0-\iota(i)+1(\iota(i))}P_i \mid \iota(i), D_{b,i}=1, W_i=0\right] = \frac{(N_0-\iota(i))+3/2}{(N_0+1)(N_0+2)}.$$

For $1 < \iota(i) < N_0$, and $1 \le r \le \iota(i)-1$, we have:

$$\mathbb{E}\left[\Delta_{1(\iota(i)-1-r)}P_i \mid \iota(i), D_{b,i}=1, W_i=0\right] = \frac{1}{(N_0+1)(N_0+2)},$$

$$\mathbb{E}\left[\Delta_{\iota(i)(0)}P_i \mid \iota(i), D_{b,i}=1, W_i=0\right] = \frac{(\iota(i)-1)+3/2}{(N_0+1)(N_0+2)}.$$

Therefore, for $1 < \iota(i) < N_0$:

$$\mathbb{E}\left[\sum_{j=1}^{N}\sum_{l\ne j} W_j W_l 1\{M_j=i\}1\{M_l\ne i\}B_i(X_l) \,\Big|\, \iota(i), D_{b,i}=1, W_i=0\right] = \frac{N_1(N_1-1)}{2(N_0+1)(N_0+2)}$$

$$\times\left\{\sum_{r=1}^{N_0-\iota(i)}\left(1+1\{r=N_0-\iota(i)\}(N_0-\iota(i)+3/2)\right)\frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}}\right.$$

$$\left.+\sum_{r=1}^{\iota(i)-1}\left(1+1\{r=\iota(i)-1\}(\iota(i)+1/2)\right)\frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}}\right\}.$$

For $\iota(i)=1$ and $1 \le r \le N_0-1$, we obtain:

$$\mathbb{E}\left[\Delta_{1(1+r)}P_i \mid \iota(i)=1, D_{b,i}=1, W_i=0\right] = \frac{3}{2}\frac{1}{(N_0+1)(N_0+2)},$$

$$\mathbb{E}\left[\Delta_{N_0(1)}P_i \mid \iota(i)=1, D_{b,i}=1, W_i=0\right] = \frac{3}{2}\frac{N_0+1/3}{(N_0+1)(N_0+2)},$$

$$\mathbb{E}\left[\sum_{j=1}^{N}\sum_{l\ne j} W_j W_l 1\{M_j=i\}1\{M_l\ne i\}B_i(X_l) \,\Big|\, \iota(i)=1, D_{b,i}=1, W_i=0\right] = \frac{3N_1(N_1-1)}{4(N_0+1)(N_0+2)}$$

$$\times\sum_{r=1}^{N_0-1}\left(1+1\{r=N_0-1\}(N_0+1/3)\right)\frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}}.$$

with analogous results for the case $\iota(i)=N_0$. Let

$$T(N_0, N_1, n) = \mathbb{E}\left[\sum_{j=1}^{N}\sum_{l\ne j} W_j W_l 1\{M_j=i\}1\{M_l\ne i\}B_i(X_l) \,\Big|\, \iota(i)=n, D_{b,i}=1, W_i=0\right],$$

Then,

$$T(N_0, N_1, n) = \frac{N_1(N_1-1)}{(N_0+1)(N_0+2)}\left(R_{N_0}(n) + U_{N_0}(n)\right),$$

28

where

$$
R_{N_0}(n) = \frac{1}{2}\Bigg\{ \sum_{r=1}^{N_0-n} \frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}}
$$

$$
+ \sum_{r=1}^{n-1} \frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}} \Bigg\},
$$

$$
U_{N_0}(n) = \frac{1}{2}\Bigg\{ (N_0-n+3/2)\frac{\left(1-\left(1-\frac{1}{n}\right)^{N_0}\right)\left(1-\frac{N_0-n}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}}
$$

$$
+ (n-1+3/2)\frac{\left(1-\left(1-\frac{1}{N_0-n+1}\right)^{N_0}\right)\left(1-\frac{n-1}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}} \Bigg\},
$$

for $1 < n < N_0$. For $n = 1$:

$$
R_{N_0}(1) = \frac{3}{4}\sum_{r=1}^{N_0-1} \frac{\left(1-\left(1-\frac{1}{N_0-r}\right)^{N_0}\right)\left(1-\frac{r}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}},
$$

$$
U_{N_0}(1) = \frac{3}{4}(N_0+1/3)\frac{\left(1-\frac{N_0-1}{N_0}\right)^{N_0}}{1-\left(1-\frac{1}{N_0}\right)^{N_0}},
$$

with analogous expressions for $n = N_0$.

Let $T = \alpha^2 \exp(-1)/(1-\exp(-1))$. Then,

$$
\begin{aligned}
T - T(N_0, N_1, n) &= \alpha^2 \left(\frac{\exp(-1)}{1-\exp(-1)} - R_{N_0}(n) - U_{N_0}(n) \right) \\
&+ \left(\alpha^2 - \frac{N_1(N_1-1)}{(N_0+1)(N_0+2)} \right)\Big(R_{N_0}(n) + U_{N_0}(n) \Big).
\end{aligned}
$$

Notice that, for $0 < n < N_0$,

$$
\begin{aligned}
R_{N_0}(n) &\leq \frac{1}{1-\exp(-1)}\sum_{r=1}^{\infty}\left(1-\frac{r}{N_0}\right)^{N_0} \leq \frac{1}{1-\exp(-1)}\sum_{r=1}^{\infty}\exp(-r) \\
&= \frac{\exp(-1)}{(1-\exp(-1))^2}.
\end{aligned}
$$

Notice also that, because $\log \lambda \le \lambda - 1$ for any $\lambda > 0$, we have that for all n such that $0 < n < N_0$, $\log(n(1 - n/N_0)^{N_0}) = \log(n) + N_0 \log(1 - n/N_0) \le \log(n) - n \le -1$, and therefore $n(1 - n/N_0)^{N_0} \le \exp(-1)$. This implies that the quantities $(N_0 - n + 3/2)(1 - (N_0 - n)/N_0)^{N_0}$ and $(n + 1/2)(1 - (n-1)/N_0)^{N_0}$ are bounded by $\exp(-1) + 3/2$. Therefore, for $0 < n < N_0$:

$$U_{N_0}(n) \le \frac{\exp(-1) + 3/2}{1 - \exp(-1)}.$$

Similarly, for $n \in \{1, N_0\}$, we obtain

$$R_{N_0}(n) \le \frac{3}{4} \frac{\exp(-1)}{(1 - \exp(-1))^2},$$

and

$$U_{N_0}(n) \le \frac{3}{4} \frac{\exp(-1) + 4/3}{1 - \exp(-1)}.$$

Consequently, $R_{N_0}(n)$ and $U_{N_0}(n)$ are uniformly bounded by some constants $\bar{R}$ and $\bar{U}$ for all $N_0, n \in \mathbb{N}$ (the set of positive integers) with $n \le N_0$. Therefore, for all $N_0, n \in \mathbb{N}$ with $n \le N_0$

$$|T - T(N_0, N_1, n)| \le \alpha^2 \left(\frac{\exp(-1)}{1 - \exp(-1)} + \bar{R} + \bar{U} \right) + \left(\alpha^2 + \frac{N_1(N_1 - 1)}{(N_0 + 1)(N_0 + 2)} \right) (\bar{R} + \bar{U}).$$

Because every convergent sequence in $\mathbb{R}$ is bounded, $(N_1(N_1 - 1))/((N_0 + 1)(N_0 + 2))$ is bounded by some constant $\bar{\alpha}^2$. As a result, there exist some constant $\bar{D}$ such that $|T - T(N_0, N_1, n)| \le \bar{D}$.

Notice that for all n such that $0 < n < N_0$, $R_{N_0}(n) = R^*_{N_0}(n) + V_{N_0}(n)$, where

$$R^*_{N_0}(n) = \frac{1}{2} \left\{ \sum_{r=1}^{N_0 - n} \left(1 - \frac{r}{N_0} \right)^{N_0} + \sum_{r=1}^{n-1} \left(1 - \frac{r}{N_0} \right)^{N_0} \right\},$$

and

$$V_{N_0}(n) = \frac{1}{2} \left\{ \sum_{r=1}^{N_0 - n} \frac{\left[\left(1 - \frac{1}{N_0}\right)^{N_0} - \left(1 - \frac{1}{N_0 - r}\right)^{N_0} \right] \left(1 - \frac{r}{N_0}\right)^{N_0}}{1 - \left(1 - \frac{1}{N_0}\right)^{N_0}} \right. $$
$$\left. + \sum_{r=1}^{n-1} \frac{\left[\left(1 - \frac{1}{N_0}\right)^{N_0} - \left(\frac{1}{N_0 - r}\right)^{N_0} \right] \left(1 - \frac{r}{N_0}\right)^{N_0}}{1 - \left(1 - \frac{1}{N_0}\right)^{N_0}} \right\},$$

Notice that for all n such that $0 < n < N_0$, $0 \le V_{N_0}(n) \le \bar{V}_{N_0}$, where

$$\bar{V}_{N_0} = \sum_{r=1}^{\infty} \frac{\left[\left(1 - \frac{1}{N_0}\right)^{N_0} - \left(1 - \frac{1}{N_0 - r}\right)^{N_0} \right] \left(1 - \frac{r}{N_0}\right)^{N_0} 1\{r \le N_0 - 1\}}{1 - \left(1 - \frac{1}{N_0}\right)^{N_0}},$$

which does not depend on n. Applying the dominated convergence theorem, it is easy to show that $\bar{V}_{N_0} \to 0$, as $N_0 \to \infty$.

For some δ such that $0 < \delta < 1$, let $\mathbb{I}_{\delta,N_0} = \{n \in \mathbb{N} : 1 + N_0^\delta < n < N_0 - N_0^\delta\}$ and $\bar{\mathbb{I}}_{\delta,N_0} = \{n \in \mathbb{N} : n \leq 1 + N_0^\delta\} \cup \{n \in \mathbb{N} : N_0 - N_0^\delta \leq n \leq N_0\}$.

Notice that

$$
\begin{aligned}
R_{N_0}^*(n) &< \frac{1}{2} \left\{ \sum_{r=1}^{N_0-n} \exp(-r) + \sum_{r=1}^{n-1} \exp(-r) \right\} \\
&< \sum_{r=1}^{\infty} \exp(-r) = \frac{\exp(-1)}{1 - \exp(-1)}.
\end{aligned}
$$

For $n \in \mathbb{I}_{\delta,N_0}$, $N_0 - n > N_0^\delta$ and $n - 1 > N_0^\delta$. Therefore,

$$
R_{N_0}^*(n) > \sum_{r=1}^{\infty} \left(1 - \frac{r}{N_0}\right)^{N_0} 1\{r \leq N_0^\delta\},
$$

Let

$$
\bar{D}_{N_0} = \frac{\exp(-1)}{1 - \exp(-1)} - \sum_{r=1}^{\infty} \left(1 - \frac{r}{N_0}\right)^{N_0} 1\{r \leq N_0^\delta\}.
$$

It follows that

$$
\left| \frac{\exp(-1)}{1 - \exp(-1)} - R_{N_0}^*(n) \right| < \bar{D}_{N_0}.
$$

Notice that $\bar{D}_{N_0}$ does not depend on n. Also, applying the dominated convergence theorem, it is easy to show that $\bar{D}_{N_0} \to 0$, as $N_0 \to \infty$.

In addition, for $n \in \mathbb{I}_{N_0,\delta}$, it has to be the case that $n \geq 2$ and $n \leq N_0 - 1$. As a result,

$$
\left((N_0 - n) + 3/2\right) \left(1 - \frac{N_0 - n}{N_0}\right)^{N_0} < N_0 \left(1 - \frac{N_0^\delta}{N_0}\right)^{N_0} < N_0 \exp(-N_0^\delta),
$$

and

$$
\left((n - 1) + 3/2\right) \left(1 - \frac{n - 1}{N_0}\right)^{N_0} < N_0 \left(1 - \frac{N_0^\delta}{N_0}\right)^{N_0} < N_0 \exp(-N_0^\delta).
$$

Therefore, for $n \in \mathbb{I}_{N_0,\delta}$, $U_{N_0}(n) < \bar{U}_{N_0}$, where

$$
\bar{U}_{N_0} = \frac{1}{1 - \exp(-1)} N_0 \exp(-N_0^\delta).
$$

Notice that $\bar{U}_{N_0} \to 0$, as $N_0 \to \infty$.

The last set of results imply that for $n \in \mathbb{I}_{\delta,N_0}$,

$$
\begin{aligned}
|T - T(N_0, N_1, n)| &\leq \alpha^2 \left(\bar{D}_{N_0} + \bar{V}_{N_0} + \bar{U}_{N_0}\right) \\
&+ \left|\alpha^2 - \frac{N_1(N_1 - 1)}{(N_0 + 1)(N_0 + 2)}\right| (\bar{R} + \bar{U}).
\end{aligned}
$$

Let $\#\mathbb{I}_{\delta,N_0}$ and $\#\bar{\mathbb{I}}_{\delta,N_0}$ be the cardinalities of the sets $\mathbb{I}_{\delta,N_0}$ and $\bar{\mathbb{I}}_{\delta,N_0}$ respectively. Notice that $\#\mathbb{I}_{\delta,N_0} < N_0$, $\#\mathbb{I}_{\delta,N_0}/N_0 \to 1$ and $\#\bar{\mathbb{I}}_{\delta,N_0}/N_0 \to 0$.

$$\left| \mathbb{E}\left[\sum_{j=1}^{N} \sum_{l \neq j} W_j W_l 1\{M_j = i\} 1\{M_l \neq i\} B_i(X_l) \,\Big|\, D_{b,i} = 1, W_i = 0 \right] - T \right|$$

$$= \sum_{n=1}^{\infty} |T - T(N_0, N_1, n)| \, 1\{1 \leq n \leq N_0\}/N_0$$

$$= \sum_{n=1}^{\infty} |T - T(N_0, N_1, n)| \, 1\{n \in \mathbb{I}_{\delta, N_0}\}/N_0$$

$$+ \sum_{n=1}^{\infty} |T - T(N_0, N_1, n)| \, 1\{n \in \bar{\mathbb{I}}_{\delta, N_0}\}/N_0.$$

Using the bounds established above and the fact that $\#\mathbb{I}_{\delta, N_0} < N_0$, we obtain:

$$
\begin{aligned}
\sum_{n=1}^{\infty} |T - T(N_0, N_1, n)| \, 1\{n \in \mathbb{I}_{\delta, N_0}\}/N_0 \quad \leq \quad & \alpha^2 \left(\bar{D}_{N_0} + \bar{V}_{N_0} + \bar{U}_{N_0} \right) \sum_{n=1}^{\infty} 1\{n \in \mathbb{I}_{\delta, N_0}\}/N_0 \\
+ \quad & \left| \alpha^2 - \frac{N_1(N_1 - 1)}{(N_0 + 1)(N_0 + 2)} \right| (\bar{R} + \bar{U}) \sum_{n=1}^{\infty} 1\{n \in \mathbb{I}_{\delta, N_0}\}/N_0 \\
\leq \quad & \alpha^2 \left(\bar{D}_{N_0} + \bar{V}_{N_0} + \bar{U}_{N_0} \right) \\
+ \quad & \left| \alpha^2 - \frac{N_1(N_1 - 1)}{(N_0 + 1)(N_0 + 2)} \right| (\bar{R} + \bar{U}) \longrightarrow 0.
\end{aligned}
$$

Notice also that:

$$\sum_{n=1}^{\infty} |T - T(N_0, N_1, n)| \, 1\{n \in \bar{\mathbb{I}}_{\delta, N_0}\}/N_0 \leq \bar{D} \frac{1}{N_0} \sum_{n=1}^{\infty} 1\{n \in \bar{\mathbb{I}}_{\delta, N_0}\} = \bar{D} \frac{\#\bar{\mathbb{I}}_{\delta, N_0}}{N_0} \longrightarrow 0.$$

As a result, we obtain:

$$\mathbb{E}\left[\sum_{j=1}^{N} \sum_{l \neq j} W_j W_l 1\{M_j = i\} 1\{M_l \neq i\} B_i(X_l) \,\Big|\, D_{b,i} = 1, W_i = 0 \right] \longrightarrow \alpha^2 \frac{\exp(-1)}{1 - \exp(-1)}.$$

Now, because

$$\mathbb{E}[K_i^2 \,|\, D_{b,i} = 1, W_i = 0] = \mathbb{E}[K_i^2 \,|\, W_i = 0] \to \alpha + \frac{3}{2}\alpha^2,$$

we obtain

$$\mathbb{E}[K_i K_{b,i} \,|\, D_{b,i} = 1, W_i = 0] \to \alpha + \frac{3}{2}\alpha^2 + \alpha^2 \frac{\exp(-1)}{1 - \exp(-1)}.$$

Therefore, because $\mathbb{E}[K_i K_{b,i} \,|\, D_{b,i} = 0, W_i = 0] = 0$, we obtain

$$\mathbb{E}[K_i K_{b,i} \,|\, W_i = 0] \to \left(\alpha + \frac{3}{2}\alpha^2 + \alpha^2 \frac{\exp(-1)}{1 - \exp(-1)} \right) (1 - \exp(-1)).$$

$\square$

Proof of Lemma 3.2: From previous results:

$$
\begin{aligned}
N_1 \mathbb{E}[V^{B,I}] &= \frac{1}{\alpha} \left(\mathbb{E}[K_{b,i}^2 | W_i = 0] - 2 \mathbb{E}[K_{b,i} K_i | W_i = 0] + \mathbb{E}[K_i^2 | W_i = 0] \right) \\
&\rightarrow \frac{1}{\alpha} \left[2\alpha + \frac{3}{2} \frac{\alpha^2}{(1 - \exp(-1))} - 2(1 - \exp(-1)) \left(\alpha + \frac{3}{2} \alpha^2 + \frac{\exp(-1)}{1 - \exp(-1)} \alpha^2 \right) + \alpha + \frac{3}{2} \alpha^2 \right] \\
&= \alpha \left(\frac{3}{2(1 - \exp(-1)} - 3(1 - \exp(-1)) - 2\exp(-1) + \frac{3}{2} \right) + 2 - 2 + 2\exp(-1) + 1 \\
&= 1 + \frac{3}{2} \alpha \frac{5\exp(-1) - 2\exp(-2)}{3(1 - \exp(-1))} + 2\exp(-1).
\end{aligned}
$$

$\square$

References

ABADIE, A., D. DRUKKER, J. LEBER HERR, AND G.W. IMBENS, (2004), "Implementing Matching Estimators for Average Treatment Effects in Stata." *Stata Journal*, vol. 4(3), 290-311.

ABADIE, A., AND G. IMBENS, (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, vol. 74(1), 235-267.

ABADIE, A., AND G. IMBENS, (2002), "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects," Technical Working Paper T0283, NBER.

AGODINI, R., AND M. DYNARSKI, (2004), "Are Experiments the Only Option? A Look at Dropout Prevention Programs," *The Review of Economics and Statistics*, vol. 86(1), 180-194.

ARTHREYA, K., (1987), "Bootstrap of the Mean in the Infinite Variance Case," *The Annals of Statistics*, vol. 15(2), 724-731.

BERAN, R., (1984), "Bootstrap Methods in Statistics," *Jahresberichte des Deutschen Mathematischen Vereins.* vol. 86, 14-30.

BICKEL, P., AND D. FREEDMAN, (1981), "Some Asymptotic Theory for the Bootstrap," *The Annals of Statistics*, vol. 9(6), 1196-1217.

BICKEL, P., F. GÖTZE, AND W. VAN ZWET, (1997), "Resampling Fewer than n Observations: Gains, Losses and Remedies for Losses," *Statistica Sinica*, vol. 7, 1-31.

DEHEJIA, R., AND S. WAHBA (1999) "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, vol. 94, 1053-1062.

DEHEJIA, R., AND S. WAHBA, (2002) "Propensity Score-Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics*, vol. 84, 151-161.

GALASSO, AND RAVALLION (2003) "Social Protection in a Crisis: Argentina's Plan "Jefes y Jefes"," World Bank Policy Research Paper 3165.

GUARCELLO, L., F. MEALLI, AND F. ROSATI, (2003) "Household Vulnerability and Child Labor: the Effect of Shocks, Credit Rationing." Working Paper, Centre for Household Income, Labour, and Demographic Economics.

HALL, P. (1992), *The Bootstrap and Edgeworth Expansions*, New York: Springer.

HECKMAN, J., H. ICHIMURA, AND P. TODD (1997), "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, vol. 64(4), 605-654.

HOROWITZ, J. (2003), "The Bootstrap," in J. Heckman and E. Leamer eds. Handbook of Econometrics, vol. 5. New York: Elsevier Science.

ICHINO, A. AND S. BECKER, (2002), "Estimation of Average Treatment Effects Based on Propensity Scores" *The Stata Journal* vol. 2(4), 358-377.

IMAI, K. (2005), "Do Get-Out-The-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." *American Political Science Review*, Forthcoming.

IMBENS, G. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity," *Review of Economics and Statistics*, vol. 86(1), 4-29.

JALAN, J. AND MARTIN RAVALLION, (1999) "Income Gains to the Poor from Workfare: Estimates for Argentina's Trabajar Program". Working Paper. The World Bank, Policy Research Working Paper Series, # 2149.

JOHNSON, N. AND S. KOTZ (1977), *Urn Models and Their Applications*, John Wiley & Sons, New York.

JOHNSON, N., S. KOTZ, AND N. BALAKRISHNAN (1995), *Continuous Univariate Distributions, volume 2*, Wiley, New York.

JOHNSON, N., S. KOTZ, AND A. KEMP (1993), *Univariate Discrete Distributions*, 2nd Edition, Wiley, New York.

JOAG-DEV, AND F. PROSCHAN (1983), "Negative Association of Random Variables, with Applications," *The Annals of Statistics*, vol. 11(1), 286-295.

KAMATH, A., R. MOTWANI, K. PALEM, AND P. SPIRAKIS, (1995) "Tail Bounds for Occupancy and the Satisfiability Threshold Conjecture." *Randomized Structure and Algorithms*, vol. 7, 59-80.

LECHNER, M. (2002), "Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programmes by Matching Methods", *Journal of the Royal Statistical Society* Series A (Statistics in Society). v 165, issue 1, p 59.

MYERS, D., R. OLSEN, N. SEFTOR, J. YOUNG, C. TUTTLE, (2004) "The Impacts of Regular Upward Bound:? Results from the Third Follow-Up Data." 2004. US Dept of Education, Office of the Undersecretary, Doc # 2004-13.

NEWEY, W., AND F. WINDMEIJER, (2005), "GMM with Many Weak Moment Conditions," unpublished manuscript.

POLITIS, N., AND J. ROMANO, (1994), "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *The Annals of Statistics*, vol. 22(4), 2031-2050.

POLITIS, N., J. ROMANO, AND M. WOLF (1999) *Subsampling*, Springer Verlag, New York.

PRADHAN, M., L. RAWLINGS, (2002), "The Impact and Targeting of Social Infrastructure Investments: Lessons from the Nicaraguan Social Fund," *World Bank Economic Review,* vol. 16(2), 275-295.

PUHANI, P. (2002) "An Advantage Through Training In Poland," *Labour*, vol. 16(3), 569-608.

ROSENBAUM, P., (1989), "Optimal Matching in Observational Studies", *Journal of the American Statistical Association*, 84, 1024-1032.

ROSENBAUM, P., (2001), *Observational Studies*, second edition, Springer Verlag, New York.

ROSENBAUM, P., AND D. RUBIN, (1983a), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, vol. 70, 41-55.

SIANESI, B (2004), "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990's," *Review of Economics and Statistics*, vol. 86(1), 133-155.

SMITH, J., AND P. TODD (2005) "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* vol. 125, 305-353.

VAN DER VAART, A., (1998), *Asymptotic Statistics*, Cambridge University Press, New York.

YAMANI, M., M. LAUER, R. STARLING, C. POTHIER, E. TUZCU, N. RATLIFF, D. COOK, A. ABDO, A. MCNEIL, T. CROWE, R. HOBBS, G. RINCON, C. BOTT-SILVERMAN, P. MCCARTHY AND J. YOUNG, (2004) "Impact of Donor Spontaneous Intracranial Hemorrhage on Outcome after Heart Transplantation" *American Journal of Transplantation* February, vol. 4(2), 257.
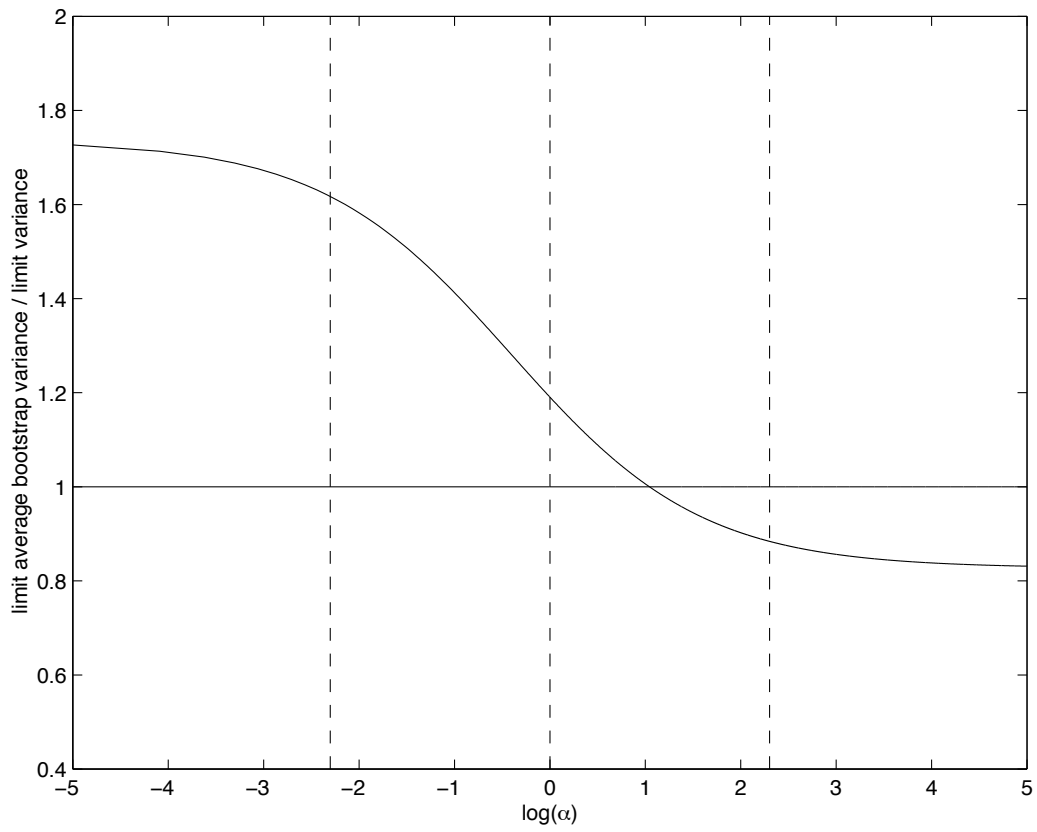
FIGURE 1. RATIO OF LIMIT AVERAGE BOOTSTRAP VARIANCE TO LIMIT VARIANCE

37

Table 1: SIMULATION RESULTS, 10,000 REPLICATIONS, 100 BOOTSTRAP DRAWS

| | Design I | | Design II | | Design III | |
| | $N_0 = 100, N_1 = 100$ | | $N_0 = 100, N_1 = 1000$ | | $N_0 = 1000, N_1 = 100$ | |
Sample Size	est.	(s.e.)	est.	(s.e.)	est.	(s.e.)
Exact Variance	2.480		15.930		1.148	
AI Var I	2.449	(0.006)	15.871	(0.033)	1.130	(0.002)
AI Var II	2.476	(0.006)	15.887	(0.033)	1.144	(0.001)
Limit Expected Boots Var I	2.977		14.144		1.860	
Boots Var I	2.914	(0.009)	14.008	(0.042)	1.836	(0.004)
Boots Var II	2.614	(0.007)	12.774	(0.034)	1.632	(0.003)
Coverage Rate 90% Confidence Interval						
Exact Variance	0.901	(0.003)	0.901	(0.003)	0.901	(0.003)
AI Var I	0.896	(0.003)	0.897	(0.003)	0.900	(0.003)
AI Var II	0.895	(0.003)	0.897	(0.003)	0.897	(0.003)
Boots Var I	0.931	(0.003)	0.870	(0.003)	0.980	(0.003)
Boots Var II	0.903	(0.003)	0.847	(0.003)	0.948	(0.003)
Coverage Rate 95% Confidence Interval						
Exact Variance	0.951	(0.002)	0.951	(0.002)	0.952	(0.002)
AI Var I	0.949	(0.002)	0.950	(0.002)	0.950	(0.002)
AI Var II	0.948	(0.002)	0.950	(0.002)	0.947	(0.002)
Boots Var I	0.972	(0.002)	0.934	(0.002)	0.995	(0.002)
Boots Var II	0.953	(0.002)	0.915	(0.002)	0.981	(0.002)