*Econometrica*, Vol. 76, No. 6 (November, 2008), 1537–1557

## NOTES AND COMMENTS

# ON THE FAILURE OF THE BOOTSTRAP FOR MATCHING ESTIMATORS

BY ALBERTO ABADIE AND GUIDO W. IMBENS [1]

Matching estimators are widely used in empirical economics for the evaluation of programs or treatments. Researchers using matching methods often apply the bootstrap to calculate the standard errors. However, no formal justification has been provided for the use of the bootstrap in this setting. In this article, we show that the standard bootstrap is, in general, not valid for matching estimators, even in the simple case with a single continuous covariate where the estimator is root-$N$ consistent and asymptotically normally distributed with zero asymptotic bias. Valid inferential methods in this setting are the analytic asymptotic variance estimator of Abadie and Imbens (2006a) as well as certain modifications of the standard bootstrap, like the subsampling methods in Politis and Romano (1994).

KEYWORDS: Average treatment effects, bootstrap, matching.

## 1. INTRODUCTION

MATCHING METHODS have become very popular for the estimation of treatment effects in the absence of experimental data.[2] Researchers using matching methods often apply the bootstrap to calculate the standard errors. However, bootstrap inference for matching estimators has not been formally justified.

This article addresses the question of the validity of the standard bootstrap for nearest-neighbor matching estimators with replacement and a fixed number of neighbors. We focus on the case of a fixed number of neighbors because it conforms to the usual practice in empirical economics, where researchers applying matching estimators typically employ nearest-neighbor matching with a very limited number of neighbors (e.g., one). We show in a simple case, with a single continuous covariate, that the standard bootstrap fails to provide asymptotically valid standard errors, in spite of the fact that the matching estimator is root-$N$ consistent and asymptotically normal with no asymptotic bias. We show that the average bootstrap variance can overestimate as well as underestimate the asymptotic variance of matching estimators. We provide some intuition for the failure of the bootstrap in this context.[3]

[1]We are grateful for comments by Peter Bickel, Stéphane Bonhomme, Joel Horowitz, Francis Kramarz, Whitney Newey, seminar participants at Princeton, CEMFI, CREST, and Harvard/MIT, and two anonymous reviewers. Financial support for this research was generously provided through NSF Grants SES-0350645 (Abadie), SES-0136789, and SES-0452590 (Imbens).

[2]For example, Dehejia and Wahba (1999). See Rosenbaum (2001) and Imbens (2004) for surveys.

[3]Other examples of failure of the bootstrap arise in the contexts of estimating the maximum of the support of a random variable (Bickel and Freedman (1981)), estimating the average of a

 DOI: 10.3982/ECTA6474

There are valid alternatives to the bootstrap for inference with matching estimators. In Abadie and Imbens (2006a), we derived conditions under which the nearest-neighbor matching estimator with replacement and a fixed number of matches is root-$N$ consistent and asymptotically normal, and we proposed an analytic estimator of the asymptotic variance. Under those conditions, the validity of certain alternatives to the bootstrap, such as subsampling (Politis and Romano (1994)) or the $M$-out-of-$N$ bootstrap (Bickel, Götze, and van Zwet (1997)), can be established from general results.[4]

## 2. SETUP

### 2.1. *Basic Model*

In this article we adopt the standard model of treatment effects under unconfoundedness (Rosenbaum and Rubin (1983), Heckman, Ichimura, and Todd (1998), Rosenbaum (2001), Imbens (2004)). The goal is to evaluate the effect of a treatment on the basis of data on outcomes, treatments, and covariates for treated and untreated units. We have a random sample of $N_0$ units from the control (untreated) population and a random sample of $N_1$ units from the treated population, with total sample size $N = N_0 + N_1$. 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 treatments, 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 outcome. Let $\mathbf{X}$ be an $N$-column matrix with column $i$ equal to $X_i$ and assume analogous notation for $\mathbf{Y}$ and $\mathbf{W}$.

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

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

We make the following two identifying assumptions:

variable with infinite variance (Arthreya (1987)), and superefficient estimation (Beran (1984)). Resampling inference in these contexts can be conducted using alternative methods such as subsampling (Politis and Romano (1994)) 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.

[4]See, for example, Politis, Romano, and Wolf (1999).

[5]To simplify our proof of lack of validity of the bootstrap, we will consider in our calculations the case of 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 (2006a)).

ASSUMPTION 2.1—Unconfoundedness: *For almost all* $x$, $(Y_i(1), Y_i(0))$ *is independent of* $W_i$ *conditional on* $X_i = x$ *or*

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

ASSUMPTION 2.2—Overlap: *For some* $c > 0$ *and almost all* $x$,

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

A nearest-neighbor matching estimator of $\tau$ 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. In this article, we focus on the case of matching with replacement, so each control unit may be used as a match for more than one treated unit.

For each treated unit $i$, let $D_i$ be the distance between the covariate value for observation $i$ and the covariate value for the closest untreated match:

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

Then let

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

be the set of closest matches for treated unit $i$. If unit $i$ is an untreated 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 1, 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 $\tau$ is then

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

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 an untreated unit, with $K_i = 0$ if unit $i$ is a treated unit):

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

Then we can write

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

Let

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

In Abadie and Imbens (2006a) we proved that under certain conditions (for example, when $X_i$ is a scalar variable) the nearest-neighbor matching estimator in (1) is root-$N$ consistent and asymptotically normal with zero asymptotic bias.[6] We also proposed a consistent estimator for the asymptotic variance of $\hat{\tau}$:

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

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 $l(i)$, be the 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).[7] Then

$$\widehat{\sigma}^2(X_i, W_i) = \frac{1}{2}(Y_i - Y_{l(i)})^2.$$

[6]More generally, in Abadie and Imbens (2007), we proposed a bias correction that makes matching estimators root-$N$ consistent and asymptotically normal regardless of the dimension of $X_i$.

[7]To simplify the notation, here we consider only the case without matching ties. The extension to accommodate ties is immediate, but it is not required for the purpose of the analysis in this article.

Let $\mathbb{V}(\hat{\tau})$ be the variance of $\hat{\tau}$. In Abadie and Imbens (2006a), we showed that (under regularity conditions) the normalized version of the variance estimator, $N_1\widehat{V}^{\mathrm{AI}}$ is consistent for the normalized variance, $N_1\mathbb{V}(\hat{\tau})$:

$$N_1(\mathbb{V}(\hat{\tau}) - \widehat{V}^{\mathrm{AI}}) \xrightarrow{p} 0.$$

## 2.2. *The Bootstrap*

In this article we consider two versions of the bootstrap variance commonly used in empirical research. 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: $\hat{\tau}_b = t(\mathbf{Z}_b)$. We denote expectations over the bootstrap distribution (conditional on the sample) as $\mathbb{E}[\cdot|\mathbf{Z}]$. The first version of the bootstrap variance is the second moment of $(\hat{\tau}_b - \hat{\tau})$ conditional on $\mathbf{Z}$:

$$V^{B,\mathrm{I}} = \mathbb{E}[(\hat{\tau}_b - \hat{\tau})^2|\mathbf{Z}].$$

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, $\hat{\tau}$:

$$V^{B,\mathrm{II}} = \mathbb{E}\big[(\hat{\tau}_b - \mathbb{E}[\hat{\tau}_b|\mathbf{Z}])^2|\mathbf{Z}\big].$$

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

$$\hat{V}^{B,\mathrm{I}} = \frac{1}{B}\sum_{b=1}^{B}(\hat{\tau}_b - \hat{\tau})^2, \quad \text{and}$$

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

We will focus on the first bootstrap variance, $V^{B,\mathrm{I}}$, and its expectation, $\mathbb{E}[V^{B,\mathrm{I}}]$. We shall show that, in general, $N_1(\mathbb{E}[V^{B,\mathrm{I}}] - \mathbb{V}(\hat{\tau}))$ does not converge to zero. We will show that in some cases the limit of $N_1(\mathbb{E}[V^{B,\mathrm{I}}] - \mathbb{V}(\hat{\tau}))$ is positive and that in other cases this limit is negative. As a result, we will show that $N_1 V^{B,\mathrm{I}}$ is not a consistent estimator of the limit of $N_1 \mathbb{V}(\hat{\tau})$. This will indirectly imply that $N_1 V^{B,\mathrm{II}}$ is not consistent either. Because $\mathbb{E}[(\hat{\tau}_b - \hat{\tau})^2|\mathbf{Z}] \geq \mathbb{E}[(\hat{\tau}_b - \mathbb{E}[\hat{\tau}_b|\mathbf{Z}])^2|\mathbf{Z}]$, it follows that $\mathbb{E}[V^{B,\mathrm{I}}] \geq \mathbb{E}[V^{B,\mathrm{II}}]$. Thus in the cases where the limit of $N_1(\mathbb{E}[V^{B,\mathrm{I}}] - \mathbb{V}(\hat{\tau}))$ is smaller than zero, it follows that the limit of $N_1(\mathbb{E}[V^{B,\mathrm{II}}] - \mathbb{V}(\hat{\tau}))$ is also smaller than zero.

## 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 $N_1 \mathbb{E}[V^{B,\mathrm{I}}]$.

### 3.1. *Data Generating Process*

We consider the following data generating process (DGP):

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*.

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* 0 *and variance* 1.

It follows from Assumptions 3.2 and 3.3 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 average treatment effect is equal to $\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x] = \tau$ for all $x$. Therefore, the average treatment effect for the treated is equal to $\tau$. 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) 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, given $W_i = 0$, the $Y_i$'s are mean zero, unit variant, and independent of $\mathbf{X}$. Thus $\mathbb{E}[\hat{\tau}|\mathbf{X},\mathbf{W}]=\tau$. Because (i) $\mathbb{E}[Y_iY_j|W_i=0,\mathbf{X},\mathbf{W}]=0$ 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

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

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

$$\mathbb{V}(\hat{\tau})=\mathbb{E}(\mathbb{V}(\hat{\tau}|\mathbf{X},\mathbf{W}))=\frac{N_0}{N_1^2}\mathbb{E}[K_i^2|W_i=0]. \tag{3}$$

LEMMA 3.1—Exact Variance: *Suppose that Assumptions 2.1, 2.2, and 3.1–3.4 hold. Then*:

(i) *The exact variance of the matching estimator is*

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

(ii) *As* $N\to\infty$,

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

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

See the Appendix for proofs.

### 3.3. *The Bootstrap Variance*

Now we analyze the properties of the bootstrap variance, $V^{B,\mathrm{I}}$. As before, let $\mathbf{Z}=(\mathbf{X},\mathbf{W},\mathbf{Y})$ denote the original sample. Notice that

$$\mathbb{E}[V^{B,\mathrm{I}}]=\mathbb{E}\big[\mathbb{E}[(\hat{\tau}_b-\hat{\tau})^2|\mathbf{Z}]\big]=\mathbb{E}[(\hat{\tau}_b-\hat{\tau})^2] \tag{4}$$

is the expected bootstrap variance. Notice also that the expectation $\mathbb{E}[(\hat{\tau}_b-\hat{\tau})^2|\mathbf{Z}]$ is taken over the bootstrap distribution (conditional on $\mathbf{Z}$). The expectation $\mathbb{E}[(\hat{\tau}_b-\hat{\tau})^2]$ averages $\mathbb{E}[(\hat{\tau}_b-\hat{\tau})^2|\mathbf{Z}]$ over the population distribution of $\mathbf{Z}$. Let $K_{b,i}$ be the number of times that unit $i$ in the original sample is used as a match in bootstrap sample $b$. For the DGP of Section 3.1,

$$\hat{\tau}_b=\tau-\frac{1}{N_1}\sum_{i=1}^{N}K_{b,i}Y_i. \tag{5}$$

From equations (2) and (5) we obtain

$$\begin{aligned}
(6)\qquad \mathbb{E}[(\hat{\tau}_b-\hat{\tau})^2] &= \mathbb{E}\left[\left(\frac{1}{N_1}\sum_{i=1}^{N}(K_{b,i}-K_i)Y_i\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{N_1^2}\sum_{i=1}^{N}(K_{b,i}-K_i)^2\right] = \frac{N_0}{N_1^2}\mathbb{E}\big[(K_{b,i}-K_i)^2|W_i=0\big].
\end{aligned}$$

The following lemma establishes the limit of $N_1\mathbb{E}[V^{B,\mathrm{I}}]$ under our DGP.

LEMMA 3.2—Bootstrap Variance: *Suppose that Assumptions* 3.1–3.4 *hold*. *Then*, *as* $N\to\infty$,

$$(7)\qquad N_1\mathbb{E}[V^{B,\mathrm{I}}]\to 1+\frac{3}{2}\alpha\frac{5\exp(-1)-2\exp(-2)}{3(1-\exp(-1))}+2\exp(-1).$$

Recall that the limit of the normalized variance of $\hat{\tau}$ is $1+(3/2)\alpha$. For small values of $\alpha$ the limit of the expected bootstrap variance exceeds the limit variance by the third term in (7), $2\exp(-1)\simeq 0.74$, or 74%. For large values of $\alpha$, the second term in (7) dominates and the ratio of the limit expected bootstrap and limit variance is equal to the factor in the second term of (7) multiplying $(3/2)\alpha$. Since $(5\exp(-1)-2\exp(-2))/(3(1-\exp(-1)))\simeq 0.83$, it follows that as $\alpha$ increases, the ratio of the limit expected bootstrap variance to the limit variance converges to 0.83, suggesting that in large samples the bootstrap variance can under- as well as overestimate the true variance.

### 3.4. *Failure of the Bootstrap*

So far, we have established the relationship 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 bootstrap provides a valid estimator of the asymptotic variance of the simple matching estimator if

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

Lemmas 3.1 and 3.2 show that

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

and

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

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

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

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

$$\begin{aligned} 1 + \frac{3}{2}\alpha &\leq \lim \mathbb{E}\big[N_1 \mathbb{E}[(\widehat{\tau}_b - \widehat{\tau})^2 | \mathbf{Z}]\big] = \lim N_1 \mathbb{E}[(\widehat{\tau}_b - \widehat{\tau})^2] \\ &= 1 + \frac{3}{2}\alpha \frac{5\exp(-1) - 2\exp(-2)}{3(1 - \exp(-1))} + 2\exp(-1). \end{aligned}$$

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 $\alpha$. As a result, 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

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

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

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

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

$$\begin{aligned} 1 + \frac{3}{2}\alpha &\leq \liminf \mathbb{E}\big[N_1 \mathbb{E}\big[(\widehat{\tau}_b - \mathbb{E}[\widehat{\tau}_b | \mathbf{Z}])^2 | \mathbf{Z}\big]\big] \\ &\leq \lim \mathbb{E}\big[N_1 \mathbb{E}[(\widehat{\tau}_b - \widehat{\tau})^2 | \mathbf{Z}]\big] = \lim N_1 \mathbb{E}[(\widehat{\tau}_b - \widehat{\tau})^2] \\ &= 1 + \frac{3}{2}\alpha \frac{5\exp(-1) - 2\exp(-2)}{3(1 - \exp(-1))} + 2\exp(-1). \end{aligned}$$

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

Because the variance is an unbounded functional, inconsistency of the bootstrap variance does not necessarily imply inconsistency of the bootstrap estimator of the asymptotic distribution of matching estimators. However, using an argument similar to the one applied above, it is easy to see that the bootstrap distribution of $\sqrt{N_1}(\widehat{\tau}_b - \widehat{\tau})$ is not consistent, in general, for the asymptotic distribution of $\sqrt{N_1}(\widehat{\tau} - \tau)$. The reason is that if the bootstrap is consistent for the asymptotic distribution of $\sqrt{N_1}(\widehat{\tau} - \tau)$, then the limit inferior of the variance of $\sqrt{N_1}(\widehat{\tau}_b - \widehat{\tau})$ should not be smaller than $1 + (3/2)\alpha$, which we have shown happens for large enough $\alpha$.

As is apparent from equation (2), the matching estimator becomes linear after conditioning on $\mathbf{X}$ and $\mathbf{W}$. The reason is that $K_1, \ldots, K_N$ are fixed once we condition on $\mathbf{X}$ and $\mathbf{W}$. This implies that the wild bootstrap of Härdle and Mammen (1993) can be used to estimate the conditional distribution of matching estimators.[8] The reason why the bootstrap fails to reproduce the unconditional distribution of $\hat{\tau}$ is that the bootstrap fails to reproduce the distribution of $K_i$, even in large samples. To gain some intuition about this, consider the DGP of Section 3.1. Equations (3), (4), and (6) imply that $N_1(\mathbb{E}[V^{B,\mathrm{I}}] - \mathbb{V}(\hat{\tau})) \to 0$ if and only if $\mathbb{E}[(K_{b,i} - K_i)^2 | W_i = 0] - \mathbb{E}[K_i^2 | W_i = 0] \to 0$. Consider the situation when $\alpha = N_1/N_0$ is small. Then, because the number of control units is large relative to the number of treated units, most observations in the control group are used as a match no more than once, so $\Pr(K_i > 1 | W_i = 0)$ is small. In a bootstrap sample, however, treated units can appear multiple times. Every time that a treated unit appears in the bootstrap sample, this unit is matched to the same control unit, creating instances in which $K_{b,i} - K_i > 1$. This problem does not disappear by increasing the sample size. As a result, even in large samples, the bootstrap fails to reproduce the distribution of $K_i$ and, in particular, it fails to reproduce $\mathbb{E}[K_i^2 | W_i = 0]$.

## 4. CONCLUSION

The results in this article have an immediate implication for empirical practice: bootstrap standard errors are not valid as the basis for inference with simple nearest-neighbor matching estimators with replacement and a fixed number of neighbors. In Abadie and Imbens (2006a), we proposed a valid estimator of the variance of matching estimators that is based on a normal approximation to the asymptotic distribution of these estimators. Simulation results in Abadie and Imbens (2006b) suggest that the analytic standard errors proposed in Abadie and Imbens (2006a) work well even in fairly small samples. Alternative inferential methods for matching estimators are the subsampling method of Politis and Romano (1994) and the $M$-out-of-$N$ bootstrap of Bickel, Götze, and van Zwet (1997).

[8]We are grateful to a referee for suggesting this.

In this article we consider only simple nearest-neighbor matching estimators with a fixed number of matches. Heckman, Ichimura, and Todd (1998) have proposed kernel-based matching methods for which the number of matches increases with the sample size. Because these estimators are asymptotically linear, we anticipate that the bootstrap provides valid inference. The same conjecture applies to other asymptotically linear estimators of average treatment effects, such as the propensity score weighting estimator proposed by Hirano, Imbens, and Ridder (2003). In addition, if $X_i$ includes only discrete covariates with a finite number of possible values, then a simple matching estimator can be constructed to match each observation in the treatment group to all untreated observations with the same value of $X_i$. This matching estimator is just a weighted average of differences in means across groups defined by the values of the covariates. As a result, the standard bootstrap provides valid inference in this context.

## APPENDIX

Before proving Lemma 3.1, we introduce some notation and preliminary results. Let $X_1, \dots, X_N$ 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 0), 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

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

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 the 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 $1 \leq i \leq j \leq N$,*

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

*where for a positive integer $a$ and a nonnegative 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 $(j - i, N - (j - i) + 1)$.*

The proof of this lemma and of other preliminary lemmas in this appendix are available in the working paper version of this article (Abadie and Imbens (2006b)).

Notice that the lemma implies the following results:

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

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

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 $(N_1, P_i)$:*

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

This implies the following conditional moments for $K_i$:

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

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

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

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

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

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 obtain $P_i$, 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 obtain the following representation for $P_i$:

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

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*

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

(ii) *The $M$th moment of $P_i$ is bounded by*

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

The proof of this lemma follows from equation (A.1) and Lemma A.1 (see Abadie and Imbens (2006b)).

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), we get

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

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. For any nonnegative integer $M$, the $M$th moment of $K_i$ given $\mathbf{W}$ and $\mathbf{X}_0$ is (Johnson, Kotz, and Kemp (1993))

$$\mathbb{E}[K_i^M|\mathbf{X}_0, W_i = 0] = \sum_{j=0}^{M}\frac{S(M, j)N_1!P_i^j}{(N_1 - j)!}.$$

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

$$\mathbb{E}[K_i^M|W_i = 0] = \sum_{j=0}^{M}\frac{S(M, j)N_1!}{(N_1 - j)!}\mathbb{E}[P_i^j|W_i = 0]$$
$$\leq \sum_{j=0}^{M}\frac{S(M, j)N_1!}{(N_1 - j)!}\left(\frac{1 + M}{N_0 + 1}\right)^j$$
$$\leq \sum_{j=0}^{M} S(M, j)\alpha^j(1 + M)^j.$$

Notice that

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

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

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

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

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 $\xi_i$ are independent, and the distribution of $\xi_i$ is degenerate at zero for $W_i = 1$ and normal $\mathcal{N}(0, K_i^2)$ for $W_i = 0$. Hence, for any $c \in \mathbb{R}$,

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

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

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

Now, Slustky's theorem implies (iii). *Q.E.D.*

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 variable $B_i(x)$ for $i = 1, \dots, N$ 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 1, 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:

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

Equation (6) implies

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

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* $n \in \{1, \dots, N_0\}$:

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

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

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

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

(v) *For* $l \neq j$,

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

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

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

Next, we prove a general result for the bootstrap. Consider a sample of size $N$, indexed by $i = 1, \dots, N$. Let $D_{b,i}$ indicate 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*

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

LEMMA A.6—Approximate Bootstrap $K$ Moments: *Suppose that Assumptions 3.1–3.3 hold. Then*:

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

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

PROOF: Here we prove (i). The proof of part (ii) is similar in spirit, but much longer (see Abadie and Imbens (2006b)). Notice that for $i$, $j$, $l$, such that $W_i = 0$ and $W_j = W_l = 1$,

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

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 $W_j = W_l = 1$,

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

As a result,

$$\begin{aligned}
&\mathbb{E}[R_{b,j} R_{b,l} | D_{b,i} = 1, B_i(X_j) = B_i(X_l) = 1, W_i = 0, W_j = W_l = 1, j \neq l] \\
&\quad - \big(\mathbb{E}[R_{b,j} | D_{b,i} = 1, B_i(X_j) = B_i(X_l) = 1, W_i = 0, \\
&\quad W_j = W_l = 1, j \neq l]\big)^2 \longrightarrow 0.
\end{aligned}$$

By Lemma A.4(i),

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

Therefore,

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

In addition,

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

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,

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

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,

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

Therefore,

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

Notice that

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

which is bounded away from infinity (this is 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,

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

Then, using these preliminary results, we obtain

$$\mathbb{E}[K_{b,i}^2|W_i=0]$$
$$=\mathbb{E}\left[\sum_{j=1}^{N}\sum_{l=1}^{N}W_jW_lB_i(X_j)B_i(X_l)R_{b,j}R_{b,l}\Big|W_i=0\right]$$
$$=\mathbb{E}\left[\sum_{j=1}^{N}W_jB_i(X_j)R_{b,j}^2\Big|W_i=0\right]$$
$$+\mathbb{E}\left[\sum_{j=1}^{N}\sum_{l\neq j}W_jW_lB_i(X_j)B_i(X_l)R_{b,j}R_{b,l}\Big|W_i=0\right]$$

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

*Q.E.D.*

PROOF OF LEMMA 3.2: From preliminary results,

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

*Q.E.D.*

## REFERENCES


ABADIE, A., AND G. IMBENS (2006a): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267. [1538,1540,1541,1546]

——— (2006b): "On the Failure of the Bootstrap for Matching Estimators," Working Paper, available at http://www.ksg.harvard.edu/fs/aabadie. [1546,1548,1549,1553]

——— (2007): "Bias-Corrected Matching Estimators for Average Treatment Effects," Unpublished Manuscript, John F. Kennedy School of Government, Harvard University. [1540]

ARTHREYA, K. (1987): "Bootstrap of the Mean in the Infinite Variance Case," *Annals of Statistics*, 15, 724–731. [1538]

BERAN, R. (1984): "Bootstrap Methods in Statistics," *Jahresberichte des Deutschen Mathematischen Vereins*, 86, 14–30. [1538]

BICKEL, P., AND D. FREEDMAN (1981): "Some Asymptotic Theory for the Bootstrap," *Annals of Statistics*, 9, 1196–1217. [1537]

BICKEL, P., F. GÖTZE, AND W. VAN ZWET (1997): "Resampling Fewer Than $n$ Observations: Gains, Losses and Remedies for Losses," *Statistica Sinica*, 7, 1–31. [1538,1546]
DEHEJIA, R., AND S. WAHBA (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062. [1537]
HALL, P. (1992): *The Bootstrap and Edgeworth Expansions*. New York: Springer Verlag. [1538]
HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, 21, 1926–1947. [1546]
HECKMAN, J., H. ICHIMURA, AND P. TODD (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294. [1538,1547]
HIRANO, K., G. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189. [1547]
HOROWITZ, J. (2003): "The Bootstrap," in *Handbook of Econometrics*, Vol. 5, ed. by J. Heckman and E. Leamer. New York: Elsevier Science. [1538]
IMBENS, G. (2004): "Nonparametric Estimation of Average Treatment Effects Under Exogeneity," *Review of Economics and Statistics*, 86, 4–29. [1537,1538]
JOAG-DEV, AND F. PROSCHAN (1983): "Negative Association of Random Variables, With Applications," *The Annals of Statistics*, 11, 286–295. [1551]
JOHNSON, N., S. KOTZ, AND A. KEMP (1993): *Univariate Discrete Distributions* (Second Ed.). New York: Wiley. [1550]
POLITIS, N., AND J. ROMANO (1994): "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *The Annals of Statistics*, 22, 2031–2050. [1538,1546]
POLITIS, N., J. ROMANO, AND M. WOLF (1999): *Subsampling*. New York: Springer Verlag. [1538]
ROSENBAUM, P. (2001): *Observational Studies* (Second Ed.). New York: Springer Verlag. [1537, 1538]
ROSENBAUM, P., AND D. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55. [1538]
VAN DER VAART, A. (1998): *Asymptotic Statistics*. New York: Cambridge University Press. [1545, 1555]

*John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138, U.S.A. and NBER; alberto_abadie@harvard.edu, http://www.ksg.harvard.edu/fs/aabadie/*

*and*

*Dept. of Economics, Harvard University, M-24 Litauer Center, 1805 Cambridge Street, Cambridge, MA 02138, U.S.A. and NBER; imbens@harvard.edu, http://www.economics.harvard.edu/faculty/imbens/imbens.html.*

*Manuscript received May, 2006; final revision received May, 2008.*