# On the Failure of the Bootstrap for Matching Estimators*

Alberto Abadie[†] Guido W. Imbens[‡]

First Draft: May 2004
This Draft: March 2005

**Abstract**

Matching estimators are widely used for the evaluation of programs or treatments. Often researchers use bootstrapping methods for inference. 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*

---

*We are grateful for comments by Peter Bickel. Financial support for this research was generously provided through NSF grants SES-0350645 (Abadie) and SES 0136789 (Imbens).

†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 for conducting inference.[2] Such methods have not been formally justified, and due to the non-smooth nature of nearest neighbor or matching methods there is reason for concern about their validity. On the other hand, we are not aware of any examples 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] Here we resolve this question. We show in a simple case with a single continuous covariate that the standard bootstrap does indeed fail to provide asymptotically valid confidence intervals and we provide some intuition for this failure. We present theoretical calculations for the difference between the bootstrap and nominal coverage rates. These theoretical calculations are supported by Monte Carlo evidence. We show that the bootstrap confidence intervals can have over– as well as under coverage.

Alternative analytic variance estimators have been proposed by Abadie and Imbens (2004, AI from here on). Since the standard bootstrap is shown to be invalid, together with the subsampling bootstrap these are now the only variance estimators available that are formally justified.

The rest of the paper is organized as follows. Section 2 reviews the basic notation and setting of matching estimators. Section 3 presents theoretical results on the lack of

[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 (2004). Yamani, Lauer, Starling, Pothier, Tuzcu, Ratliff, Cook, Abdo, McNeil, Crowe, Hobbs, Rincon, Bott-Silverman, McCarthy and Young (2004).

[3]Familiar counterexamples where the estimator has no limiting normal distribution include 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 estimators (Beran, 1984). It should be noted that the above conditions do imply that the subsampling (Politis and Romano, 1994; Politis, Romano, and Wolf, 1999) and other 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) are valid.

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

Consider the following standard set up for estimating treatment effects under exogeneity or unconfoundedness (Rubin, 1978; Rosenbaum and Rubin, 1983, Rosenbaum, 2001, Imbens, 2004). We are interested in the evaluation of a treatment on the basis of data on outcomes for treated and control units and covariates. 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. 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 scalar covariate $X_i$.[4] Let $W_i$ indicate whether a unit is from the control sample ($W_i = 0$) or the treatment group ($W_i = 1$). We observe for each unit 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 $\mathbf{X}$ denote the $N$-vector with typical element $X_i$, and similar for $\mathbf{Y}$ and $\mathbf{W}$. Also, let $\mathbf{X}_0$ denote the $N$-vector with typical element $(1 - W_i)\, X_i$, and $\mathbf{X}_1$ the $N$-vector with typical element $W_i\, X_i$. We make the following two assumptions that will justify using matching methods.

**Assumption 2.1:** (UNCONFOUNDEDNESS)

$$\Big(Y_i(0), Y_i(1)\Big) \perp\!\!\!\perp W_i \;\Big|\; X_i,$$

**Assumption 2.2:** (OVERLAP) *For some $c > 0$,*

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

[4]To simplify the calculations we focus on the case with a scalar covariate. With higher dimensional covariates there is the additional complication of biases that may dominate the variance even in large samples. See for a discussion on this Abadie and Imbens (2004).

In this discussion we focus on the average treatment effect for the treated:[5]

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

We estimate this by matching each treated unit to the closest control, and then averaging the within-pair differences. Here, we focus on the case of matching with replacement.

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

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

Then define

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

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 at least one of the covariates is continuously distributed, the set $\mathcal{J}(i)$ will consist of a single index with probability one, but for the bootstrap samples there will often be more than one index in this set. Next, define

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

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

$$K_i = (1 - W_i) \sum_{j=1}^{N} W_j\, 1\{i \in \mathcal{J}(j)\}/\#\mathcal{J}(j).$$

[5]In many cases 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.

Then we can write

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

AI propose two variance estimators.

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

and

$$V^{AI,II} = \frac{1}{N_1} \sum_{i=1}^{N} \Big( Y_i - \hat{Y}_i(0) - \hat{\tau} \Big)^2 + \frac{1}{N_1} \sum_{i=1}^{N} (1 - W_i) K_i \Big( K_i - 1/\#\mathcal{J}(i) \Big)\, \widehat{\sigma}^2_{W_i}(X_i),$$

where $\widehat{\sigma}^2_{W_i}(X_i)$ is an unbiased estimator (although inconsistent) estimator of the conditional variance of $Y_i$ given $W_i$ and $X_i$ based on matching.

Abadie and Imbens show that the first variance estimator, $V^{AI,I}$ is consistent for the normalized conditional variance:

$$\sqrt{N}_1\, \mathbb{V}(\hat{\tau}|\mathbf{X}).$$

The second variance estimator, $V^{AI,II}$, is consistent for the normalized marginal variance:

$$\sqrt{N}_1\, \mathbb{V}(\hat{\tau}).$$

The limiting variances differ by the normalized variance of the conditional average treatment effect: $\sqrt{N}_1\, \mathbb{V}(\hat{\tau}) - \sqrt{N}_1\, \mathbb{V}(\hat{\tau}|\mathbf{X}) = \sqrt{N}_1\, \mathbb{V}(\mathbb{E}[\tau|\mathbf{X}])$.

## 2.2 The Bootstrap

We consider two versions of the bootstrap in this discussion. The first centers the bootstrap distribution around the estimate $\hat{\tau}$ from the original sample, and the second centers it around the mean of the bootstrap distribution of $\hat{\tau}$.

Consider a sample $\mathbf{Z} = (\mathbf{X}, \mathbf{W}, \mathbf{Y})$ with $N_0$ controls and $N_1$ treated units, and matching estimator $\hat{\tau} = t(\mathbf{Z})$. We construct a bootstrap sample with $N_0$ controls and $N_1$ treated by sampling with replacement from the two subsamples. We then calculate the bootstrap estimate $\hat{\tau}_b$. The first way of defining the bootstrap variance is

$$V^I_B = v^I(\mathbf{Z}) = \mathbb{E}\left[ (\hat{\tau}_b - \hat{\tau})^2 \big| \, \mathbf{Z} \right]. \tag{2.4}$$

Second, we consider the population variance of the bootstrap estimator. In other words, we estimate the variance by centering the bootstrap estimator at its mean rather than at the original estimate $\hat{\tau}$:

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

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 $N_1\,\mathbb{V}(\hat{\tau})$ and therefore that this bootstrap estimator for the variance is not valid, in the sense that it does not lead to confidence intervals with large sample coverage equal to nominal coverage. In some cases it will have coverage lower than nominal, and in other cases it will have coverage rates higher than nominal. This will indirectly imply that confidence intervals based on $V_B^{II}$ are not valid 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 $\mathbb{E}[V_B^{II}] \geq \mathbb{E}[V_B^I]$. Thus in the cases where the first bootstrap has actual coverage lower than nominal coverage, it follows that the second bootstrap cannot be valid either.

In most standard settings (i.e., outside of matching) both bootstrap variances would lead to valid confidence intervals. In fact, in most cases $V_B^I$ and $V_B^{II}$ would be identical as typically $\hat{\tau} = \mathbb{E}[\hat{\tau}_b|\mathbf{Z}]$. 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, $\hat{\mu} = \sum_i X_i/N$. In the setting studied in the current paper, however, this is not the case and the two 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 both the exact variance for $\hat{\tau}$ and the approximate (asymptotic) bootstrap variance, and show that these differ.

## 3.1 Data Generating Process

We consider a special case where the following assumptions are satisfied:

**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 $\alpha \in (0, 1)$.*

**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 $x$. Combined with unconfoundedness (Assumption 2.1)($i$) this implies that the average treatment effect for the treated is $\tau$. Under this data generating process $\sum_i W_i Y_i/N_1 = \sum_i W_i Y_i(1)/N_1 = \tau$, and so we can simplify the expression for the estimator given in (2.3) relative to the estimand to

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

Conditional on $\mathbf{X}$ and $\mathbf{W}$ the only stochastic component of $\hat{\tau}$ is $\mathbf{Y}$. By Assumption 3.4 the $Y_i$ 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

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

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

$$\mathbb{V}(\hat{\tau}) = \frac{N_0}{N_1^2}\, \mathbb{E}\left[K_i^2 | W_i = 0\right]. \tag{3.6}$$

**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*

$$\mathbb{V}(\hat{\tau}) = \frac{1}{N_1} + \frac{3}{2} \frac{(N_1 - 1)(N_0 + 8/3)}{N_1(N_0 + 1)(N_0 + 2)}, \tag{3.7}$$

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

$$N_1\, \mathbb{V}(\hat{\tau}) \to 1 + \frac{3}{2}\alpha, \tag{3.8}$$

*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 I Variance

Now we analyze the properties of the bootstrap variance, $V_B^I$ in (2.4). As before, let $\mathbf{Z} = (\mathbf{X}, \mathbf{W}, \mathbf{Y})$ denote the original sample. Also, let $t(\,)$ be the function that defines the estimator, so that $\hat{\tau} = t(\mathbf{Z})$ is the estimate based on the original sample. Let $\mathbf{Z}_b$ denote the $b$-th bootstrap sample, and $\hat{\tau}_b = t(\mathbf{Z}_b)$ the corresponding $b$-th bootstrap estimate, for $b = 1, \ldots, B$. As noted before, we draw bootstrap samples conditional on the two subsample sizes $N_0$ and $N_1$. We will look at the distribution of statistics both conditional on the original sample (denoted by $|\mathbf{Z}$), as well as over replications of the original sample drawn from the same distribution. In this notation,

$$\mathbb{E}\left[V_B^I\right] = \mathbb{E}\left[\mathbb{E}\left[(\hat{\tau}_b - \hat{\tau})^2 \big| \mathbf{Z}\right]\right] = \mathbb{E}\left[(\hat{\tau}_b - \hat{\tau})^2\right], \tag{3.9}$$

is the expected bootstrap variance. We will primarily focus on the normalized variance $N_1 V_B^I$.

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

$$N_1\, \mathbb{E}[V_B^I] \to 1 + \frac{3}{2}\,\alpha\, \frac{5\, \exp(-1) - 2\, \exp(-2)}{3\,(1 - \exp(-1))} + 2\, \exp(-1). \tag{3.10}$$

Recall that the limit of the normalized variance of $\hat{\tau}$ is $1+(3/2)\,\alpha$. For small values of $\alpha$ the bootstrap variance exceeds the true variance by the third term in (3.10), $2\, \exp(-1) \approx 0.74$, or 74%. For large $\alpha$ the second term in (3.10) dominates and the ratio of the bootstrap and true variance is equal to the factor in the second term of (3.10) multiplying $\alpha\,(3/2)$. Since $(5\, \exp(-1) - 2\, \exp(-2))/(3\,(1 - \exp(-1))) \approx 0.83$, it follows that as $\alpha$ increases, the ratio of the bootstrap variance to the actual variance asymptotes to 0.83, suggesting that bootstrap variance can under as well as over estimate the true variance.

So far, we have discussed the relation between the limiting variance of the estimator and the average bootstrap variance. We end this section by 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:

$$N_1\Big(E\left[(\widehat{\tau}_b-\widehat{\tau})^2\big|\,\mathbf{Z}\right]-\mathrm{Var}(\widehat{\tau})\Big)\xrightarrow{a.s.}0.$$

Lemma 3.1 shows that:

$$N_1\mathrm{Var}(\widehat{\tau})\longrightarrow 1+\frac{3}{2}\,\alpha.$$

Lemma 3.2 shows that

$$N_1E\left[(\widehat{\tau}_b-\widehat{\tau})^2\right]\longrightarrow 1+\frac{3}{2}\,\alpha\,\frac{5\exp(-1)-2\exp(-2)}{3(1-\exp(-1))}+2\exp(-1).$$

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

$$N_1E\left[(\widehat{\tau}_b-\widehat{\tau})^2\big|\,\mathbf{Z}\right]\xrightarrow{a.s.}1+\frac{3}{2}\,\alpha.$$

Because $N_1E\left[(\widehat{\tau}_b-\widehat{\tau})^2\big|\,\mathbf{Z}\right]\geq 0$, it follows by Fatou's Lemma, that, as $N\to\infty$

$$\begin{aligned}1+\frac{3}{2}\,\alpha=E\Big[\liminf N_1E\left[(\widehat{\tau}_b-\widehat{\tau})^2\big|\,\mathbf{Z}\right]\Big]&\leq\lim N_1E\left[(\widehat{\tau}_b-\widehat{\tau})^2\right]\\&=1+\frac{3}{2}\,\alpha\,\frac{5\exp(-1)-2\exp(-2)}{3(1-\exp(-1))}+2\exp(-1).\end{aligned}$$

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 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:

$$N_1\Big(E\left[(\widehat{\tau}_b-E[\widehat{\tau}_b|\mathbf{Z}])^2\big|\,\mathbf{Z}\right]-\mathrm{Var}(\widehat{\tau})\Big)\xrightarrow{a.s.}0.$$

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

$$N_1E\left[(\widehat{\tau}_b-E[\widehat{\tau}_b|\mathbf{Z}])^2\big|\,\mathbf{Z}\right]\xrightarrow{a.s.}1+\frac{3}{2}\,\alpha.$$

Notice that $E\left[(\widehat{\tau}_b - E[\widehat{\tau}_b|\mathbf{Z}])^2 \middle| \mathbf{Z}\right] \leq E\left[(\widehat{\tau}_b - \widehat{\tau})^2 \middle| \mathbf{Z}\right]$. By Fatou's Lemma, as $N \rightarrow \infty$

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

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 10000 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 (3.10)) to limit of the actual variance (given in equation (3.8)), for different values of $\alpha$. On the horizontal axis is the log of $\alpha$. As $\alpha$ converges to zero the ratio converges to 1.62. At $\alpha = 0$ the variance ratio is 1.13, and as $\alpha$ goes to infinity the ratio converges to 0.88. The vertical dashed lines indicate the three designs ($\alpha = 0.1$, $\alpha = 1$, and $\alpha = 10$).

The simulation results are reported in Table 1. The first row of the table gives the theoretical (exact) variances, as calculated in equation (3.7), normalized by $N_1$. The second and third rows present the normalized versions of the two Abadie-Imbens variance estimators. The second row is the variance for the conditional average treatment effect for the treated (conditional on the covariates), and the third is the variance for the population average treatment effect. Both are valid in large samples. Since the conditional average treatment effect is zero for all values of the covariates, these two estimate the same object in large samples. In brackets standard errors for these averages are presented. Of most interest is to see the difference between these variance estimates and the theoretical variance. For example, for design I, the AI Var 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, this difference is a finite sample phenomenon.

The fourth row presents the theoretical calculation for the asymptotic bootstrap I variance, as given in (3.10). The fifth and sixth row give the averages of the estimated bootstrap variances. 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 to the theoretical variance, rows 5 and 4. The difference between rows 4 and 5 is small relative to the difference between the theoretical variance and the theoretical bootstrap difference, but the difference is significantly different from zero. These differences are the results of small sample sizes. The limited number of bootstrap replications makes these averages noisier than they would otherwise be, but it does not affect the average difference.

The next two panels of the table gives coverage rates, first for nominal 90% confidence intervals and then for nominal 95% confidence intervals. The first row constructs a 90% confidence interval by adding and subtracting 1.645 times the standard error based on the theoretical exact variance (3.7). Comparison of this coverage rate to its nominal level is informative about the quality of the normal approximation to the sampling distribution. This appears to be good in all three designs and both levels (90%, and 95%). The second row calculates confidence intervals the same way but using the Abadie-Imbens I variance estimator. The third row calculates confidence intervals the same way but using the Abadie-Imbens II variance estimator. Both give good coverage rates, statistically indistinguishable from the nominal levels given the number of replications (10,000).

The fourth row of the second panel calculates the coverage rate one would expect for the bootstrap based on the difference between the theoretical bootstrap variance and the theoretical variance. To be precise, consider Design I. The theoretical variance is 2.480. The theoretical bootstrap variance is 2.977, or 20% larger. The ratio of the variances is $2.977/2.480 = 1.2005$. Hence, if the estimator itself is normally distributed centered around $\tau = 0$ and with variance 2.480, we would expect the coverage rate of a 90% bootstrap confidence interval to be $\Phi(1.645\sqrt{1.2005}) - \Phi(-1.645\sqrt{1.2005}) = 0.929$. The following row gives the actual coverage rate for a 90% confidence interval obtained by adding and subtracting 1.645 times the square root of the theoretical bootstrap variance (2.977 for Design I), calculated by simulation with 10,000 replications. For Design I this number is 0.931. The fact that this is very close to the coverage we expected for the bootstrap (0.929) suggests that both the normal approximation for the estimator is accurate (confirming results in the previous rows), and that the mean of the bootstrap variance is a good indicator of the center of the distribution of the bootstrap variance. The last two rows of this panel give the 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 standard errors for the coverage rates reflect the uncertainty coming from the finite number of replications (10,000). They are equal to $\sqrt{p\,(1-p)/N_s}$ where for the second panel $p = 0.9$ and for the third panel $p = 0.95$, and $N_s = 10,000$ is the number of replications.

The third panel gives the corresponding numbers for 95% confidence intervals.

Clearly the theoretical calculations correspond fairly closely to the numbers from the simulations. The theoretically predicted coverage rates for the bootstrap confidence intervals are very close to the actual coverage rates. They 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. Theoretically one would expect that using a small number of bootstrap replications lowers the coverage rate of the constructed confidence intervals as one uses a noise measure of the variance. Increasing the number of bootstrap replications significantly for all designs was infeasible as matching is already computationally expensive.[6] We therefore investigated the implications of this choice for the first design which is the fastest to run. For the

[6]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.

same 10,000 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 bootstrap I the coverage rate for the 90% confidence interval was 0.002 (s.e. 0.001) higher with 1,000 bootstrap replications than with 100 bootstrap replications, and 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. Since 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 note we show theoretically that the standard bootstrap is generally not valid for matching estimators. This is somewhat surprising 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 suggests that the validity of the bootstrap may be in doubt. We provide details of a set of special cases where it is possible to work out the exact variance of the estimator as well as the approximate bootstrap variance. We show that in this case bootstrap confidence intervals can lead to under as well as over coverage. A small Monte Carlo study supports the theoretical calculations. The implications of the theoretical arguments are that for matching estimators one should use the variance estimators developed by Abadie and Imbens (2004) or the subsampling bootstrap (Politis, Romano and Wolf, 1999).

## Appendix

Before proving Lemma 3.1 we introduce some additional notation. 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 (ignoring ties), $M_j$ 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:

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

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

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

Also, let $P_i$ be the conditional 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 $P_i = 0$.
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.1:** (Conditional Distribution and Moments of $K(i)$)
*Suppose that assumptions 3.1-3.3 hold. Then, the distribution of $K_i$ conditional on $W_i = 0$ 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 $P_i^1$. In addition, the events $1\{M_{j_1} = i\}$ and $1\{M_{j_2} = i\}$ 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 $P_i$. $\square$
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 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 $1/N_0$. For deriving 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 $i^{\text{th}}$ unit among control units:

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

Furthermore, let $X_{0(i)}$ be the $i^{\text{th}}$ order statistic of the covariates among the control units, so that $X_{0(i)} \leq X_{0(i+1)}$ for $i = 1, \ldots, N_0 - 1$, 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 $f_1(x)$ 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 $P_i$:

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

The representation of $P_i$ as a linear function of order statistics facilitates deriving its distribution. In particular with $X_i|W_i = 0$ uniform on $[0,1]$, $P_i$ can be written as a Beta random variable (if $1 < \iota(i) < N$) or as a linear combination of two correlated Beta random variables in the two boundary cases ($\iota(i) = 1$ or $\iota(i) = N$. This leads to the following result:

**Lemma A.2:** (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 $M$th moment of $P_i$ is bounded by*

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

**Proof:** First, consider $(i)$. Since the $X_i$ conditional on $W_i = 0$ come from a uniform distribution on the interval $[0,1]$, it follows that we can write

$$X_{0(i)} = \left(\sum_{l=0}^{i-1} E_l\right) \Big/ \left(\sum_{l=0}^{N_0} E_l\right),$$

where the $E_l$ are independent unit exponential. Hence

$$P_i = \begin{cases} (2E_0 + E_1) \Big/ \left(2\sum_{l=0}^{N_0} E_l\right) & \text{if } \ \iota(i) = 1, \\ \left(E_{\iota(i)} + E_{\iota(i)+1}\right) \Big/ \left(2\sum_{l=0}^{N_0} E_l\right) & \text{if } \ 1 < \iota(i) < N_0, \\ (2E_{N_0} + E_{N_0-1}) \Big/ \left(2\sum_{l=0}^{N_0} E_l\right) & \text{if } \ \iota(i) = N_0. \end{cases} \tag{A.2}$$

The ratio $\sum_{l=L+1}^{L+K} E_l / \sum_{l=0}^{N} E_l$ has a Beta distribution with parameters $K$ and $N+1-K$. Recall that a Beta distribution with parameters $\alpha$ and $\beta$ has a mean equal to $\alpha/(\alpha+\beta)$ and second moment is equal to $(\alpha+1)\alpha)/((\alpha+\beta)(\alpha+\beta+1))$.
So, for interior $i$ ($i$ such that $1 < \iota(i) < N_0$), we have that

$$2\,P_i \sim \text{Beta}\,(2, N_0 - 1)\,.$$

The first and second moment of these $P(i)$ are

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

and

$$\mathbb{E}\left[P_i^2|1 < \iota(i) < N_0, W_i = 0\right] = \frac{1}{(N_0+1)^2} + \frac{2(N_0-1)}{4(N_0+1)^2(N_0+2)} = \frac{3}{2}\,\frac{1}{(N_0+1)(N_0+2)}.$$

For the smallest and largest observation it is a little trickier. For $i$ such that $\iota(i) = 1$ or $\iota(i) = N_0$, the distribution of $P_i$ is as the distribution of $V_1 + V_2/2$, where

$$V_1 \overset{d}{\sim} E_1 \Big/ \sum_{l=0}^{N_0} E_l \sim \text{Beta}\,(1, N_0)\,, \qquad V_2 \overset{d}{\sim} E_2 \Big/ \sum_{l=0}^{N_0} E_l \sim \text{Beta}\,(1, N_0)\,,$$

and

$$V_1 + V_2 \overset{d}{\sim} (E_1 + E_2) \Big/ \sum_{l=0}^{N_0} E_l \sim \text{Beta}\,(2, N_0 - 1)\,,$$

with all independent unit exponential $E_l$. To get the first and second moment of $V_1 + V_2/2$ we need the first and second moment of $V_1$ and $V_2$ (which are the identical) and the second moment of $V_1 + V_2$:

$$\mathbb{E}[V_1] = \mathbb{E}[V_2] = \frac{1}{N_0+1}, \qquad \mathbb{E}[V_1^2] = \mathbb{E}[V_2^2] = \frac{2}{(N_0+1)(N_0+2)},$$

$$\mathbb{E}[V_1 + V_2] = \frac{2}{N_0+1}, \qquad \mathbb{E}[(V_1+V_2)^2] = \frac{6}{(N_0+1)(N_0+2)}.$$

Then we can back out the expectation of $V_1\,V_2$:

$$\mathbb{E}[V_1\,V_2] = \frac{1}{2}\left(\mathbb{E}[(V_1+V_2)^2] - \mathbb{E}[V_1^2] - \mathbb{E}[V_2^2]\right) = \frac{1}{(N_0+1)(N_0+2)}.$$

Then the first two moments of $V_1 + V_2/2$ are

$$\mathbb{E}[P_i|\iota(i) \in \{1, N_0\}, W_i = 0] = \mathbb{E}[V_1 + V_2/2] = \frac{3}{2}\,\frac{1}{N_0+1},$$

and

$$\mathbb{E}[P_i^2|\iota(i) \in \{1, N_0\}, W_i = 0] = \mathbb{E}[(V_1 + V_2/2)^2] = \mathbb{E}[V_1^2] + \mathbb{E}[V_1\,V_2] + \frac{1}{4}\,\mathbb{E}[V_2^2] = \frac{7}{2(N_0+1)(N_0+2)},$$

for the two boundary units $i$ where $\iota(i) \in \{1, N_0\}$.
Averaging over all units includes two units at the boundary and $N_0 - 2$ interior values. Hence:

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

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 $(ii)$ note that by (A.2) $P_i$ is less than $(E_{i-1} + E_i)/\sum_{l=0}^{N_0} E_l$, which has a Beta distribution with parameters 2 and $N_0 - 1$. Hence the moments of $P_i$ are bounded by those of a Beta distribution with parameters 2 and $N_0 - 1$. The $M$th moment of a Beta distribution with parameters $\alpha$ and $\beta$ 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.1 and A.2,

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

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

Substituting this into (3.6) 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. The $M$th moment of $K_i$ given $\mathbf{W}$ and $\mathbf{X}_0$ is (Johnson, Kotz, and Kemp, 1993):

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

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

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

Finally, we write

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

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

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

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

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

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

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 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:

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

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.

$$K_{b,i} = D_{b,i} \sum_{j=1}^{N} W_j\, B_i(X_j)\, R_{b,j}. \tag{A.4}$$

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

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

Because $Y_i(1) = 0$ by Assumption 3.4, and with $K_{b,i} = 0$ if $W_i = 1$, this reduces to

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

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

$$N_1\, \mathbb{E}[V_B^I] = N_1 \cdot \mathbb{E}[(\hat{\tau}_b - \hat{\tau})^2]$$

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

Using the facts that $\mathbb{E}[Y_i^2|W_i=0]=1$, and $\mathbb{E}[Y_i Y_j|\mathbf{X},\mathbf{W}]=0$ 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 collect and establish some properties of $D_{b,i}$, $R_{b,i}$, $N_{b,0}$, and $B_i(x)$.

**Lemma A.3:** (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)*

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

*(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)*

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

**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 $\mathbf{W}_b$, and suppose that $D_{b,i}=1$. The event that a randomly choosen treated unit will be matched to control unit $i$ conditional on 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 choosen 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). □

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.4:** (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.3 we obtain that $N_b/N \xrightarrow{p} 1-\exp(-1)$. Convergence of moments for the first results follows from the fact that $(N-N_b)/N \leq 1$. For the second result, convergence of moments follows easily from the known fact that the tails of the occupancy distribution in Lemma A.3 (iii) have an exponential bound (see, e.g., Kamath, Motwani, Palem, and Spirakis, 1995). $\square$

**Lemma A.5:** (APPROXIMATE BOOTSTRAP $K$ MOMENTS)
*Suppose that Assumption 3.1 hold. Then,*
$(i)$

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

*and* $(ii)$,

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

**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, for $W_j = W_l = 1$:

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

As a result,

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

Using the results from the two previous lemmas:

$$N_0 \,\Pr\left(B_i(X_j) = 1 \,|\, D_{b,i} = 1, W_j = 1, W_i = 0\right) \longrightarrow \frac{1}{1-\exp(-1)},$$

and

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

Then,

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

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

$$\begin{aligned}
&\mathbb{E}[K_{b,i} \,|\, \mathbf{X}, \mathbf{W}, D_{b,i} = 1, W_i = 0, N_{b,0}] = \mathbb{E}\left[ \sum_{i=1}^{N} W_j \, R_{b,j} \, B_i(X_j) \,\middle|\, \mathbf{X}, \mathbf{W}, D_{b,i} = 1, W_i = 0, N_{b,0} \right] \\
&= \mathbb{E}\left[ \sum_{i=1}^{N} W_j \, \mathbb{E}[R_{b,j} \,|\, \mathbf{X}, \mathbf{W}, D_{b,i} = 1, W_i = 0, W_j = 1, B_i(X_j)] \, B_i(X_j) \,\middle|\, \mathbf{X}, \mathbf{W}, D_{b,i} = 1, W_i = 0, N_{b,0} \right] \\
&= \mathbb{E}\left[ \sum_{i=1}^{N} W_j \, B_i(X_j) \,\middle|\, \mathbf{X}, \mathbf{W}, D_{b,i} = 1, W_i = 0, N_{b,0} \right] \\
&\qquad\qquad = K_i + (N_1 - K_i) \mathbb{E}[B_i(X_j) \,|\, \mathbf{X}, \mathbf{W}, D_{b,i} = 1, W_i = 0, W_j = 1, M_j \neq i, N_{b,0}].
\end{aligned}$$

For some $0 < \delta < 1$, let $c_L(N_0) = N_0^\delta$ and $c_U(N_0) = N_0 - N_0^\delta$. For $c_L(N_0) \leq i \leq c_U(N_0)$ and large enough $N_0$:

$$\begin{aligned}
&\mathbb{E}[B_i(X_j) \,|\, \mathbf{X}, \mathbf{W}, D_{b,i} = 1, W_i = 0, W_j = 1, M_j \neq i, N_{b,0}] \\
&= \frac{X_{0(\iota(i)+2)} - X_{0(\iota(i)+1)}}{2} \frac{N_0 - N_{b,0}}{N_0 - 1} + \frac{X_{0(\iota(i)+3)} - X_{0(\iota(i)+2)}}{2} \frac{N_0 - N_{b,0}}{N_0 - 1} \frac{N_0 - N_{b,0} - 1}{N_0 - 2} + \cdots \\
&\quad + \frac{X_{0(\iota(i)-1)} - X_{0(\iota(i)-2)}}{2} \frac{N_0 - N_{b,0}}{N_0 - 1} + \frac{X_{0(\iota(i)-2)} - X_{0(\iota(i)-1)}}{2} \frac{N_0 - N_{b,0}}{N_0 - 1} \frac{N_0 - N_{b,0} - 1}{N_0 - 2} + \cdots
\end{aligned}$$

Using the results in Johnson, Kotz, and Balakrishnan (1995), page 280, for $l \geq 1$:

$$\mathbb{E}\left[ \left( \frac{X_{0(\iota(i)+l+1)} - X_{0(\iota(i)+l)}}{2} \right) \left( \frac{X_{0(\iota(i)+1)} - X_{0(\iota(i)-1)}}{2} \right) \right] = \frac{1}{2(N_0 + 1)(N_0 + 2)},$$

and

$$\mathbb{E}\left[ \left( \frac{X_{0(\iota(i)-l)} - X_{0(\iota(i)-l-1)}}{2} \right) \left( \frac{X_{0(\iota(i)+1)} - X_{0(\iota(i)-1)}}{2} \right) \right] = \frac{1}{2(N_0 + 1)(N_0 + 2)}.$$

Therefore,

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

and

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

$\square$

**Proof of Lemma 3.2:** From previous results:

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

$\square$

## References


Abadie, A., and G. Imbens, (2004), "Large Sample Properties of Matching Estimators for Average Treatment Effects," unpublished manuscript, Kennedy School of Government, Harvard University.

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*, v 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*, v 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*, v. 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*

Dehejia, R., and S. Wahba, (2002) "Propensity Score-Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics*.

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. 20.

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*, (64(4).

Ichino, A. and S. Becker, (2002), "Estimation of Average Treatment Effects Based on Propensity Scores" *The Stata Journal* Vol.2, No.4, pp. 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.

Politis, N., and J. Romano, (1994), "Large Sample Confidence Regions Based on Subsamples Under Minimal Assumptions," *The Annals of Statistics*, v 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, No. 2, 275-295.

Puhani, P. (2002) "An Advantage Through Training In Poland," *Labour* 2002. 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*, 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*, 86(1).

Smith, J., and P. Todd (2004) "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" forthcoming, *Journal of Econometrics*.

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, Volume 4, Issue 2, Page 257.

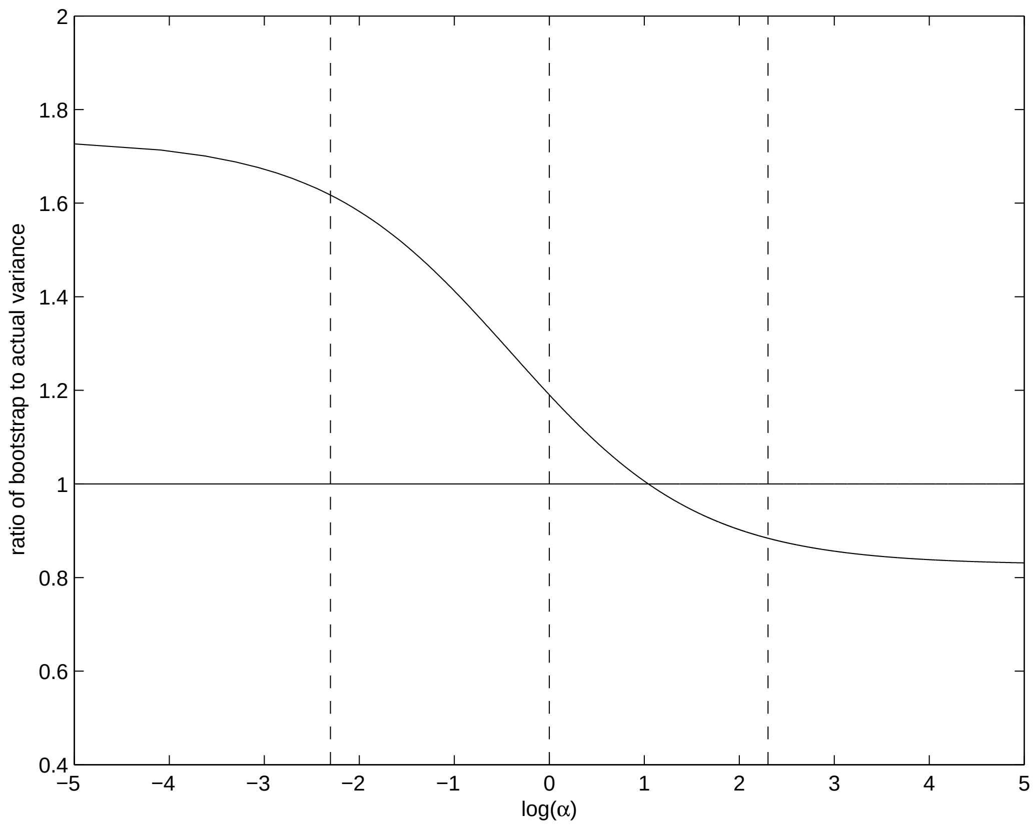


Figure 1. Ratio of Bootstrap to Actual Variance

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

| | Design I | | Design II | | Design III | |
|---|---|---|---|---|---|---|
| Sample Size | $N_0 = 100$, $N_1 = 100$ | | $N_0 = 100$, $N_1 = 1000$ | | $N_0 = 1000$, $N_1 = 100$ | |
| | est. | (s.e.) | est. | (s.e.) | est. | (s.e.) |
| Theoretical Var | 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) |
| Theor 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 | | | | | | |
| Theoretical 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) |
| Theor Boots I Coverage Rate | 0.929 | | 0.879 | | 0.964 | |
| Theor Boots Var I | 0.931 | (0.003) | 0.879 | (0.003) | 0.967 | (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 | | | | | | |
| Theoretical 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) |
| Theor Boots I Coverage Rate | 0.968 | | 0.935 | | 0.987 | |
| Theor Boots Var I | 0.969 | (0.002) | 0.937 | (0.002) | 0.989 | (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) |