BOOTSTRAP INFERENCE OF MATCHING ESTIMATORS FOR AVERAGE TREATMENT EFFECTS

TAISUKE OTSU AND YOSHIYASU RAI

Abstract. Abadie and Imbens (2008) showed that the standard naive bootstrap is inconsistent to estimate the distribution of the matching estimator for treatment effects with a fixed number of matches. This article proposes an asymptotically valid inference method for the matching estimators based on the wild bootstrap. The key idea is to resample not only the regression residuals of treated and untreated observations but also the ones to estimate the average treatment effects. The proposed method is valid even for the case of vector covariates by incorporating the bias correction method in Abadie and Imbens (2011), and is applicable to estimate the average treatment effect and the counterpart for the treated population. A simulation study indicates that our wild bootstrap method is favorably comparable to the asymptotic normal approximation. As an empirical illustration, we apply our bootstrap method to the National Supported Work data.

1. Introduction

The method of matching is applied widely in empirical research for program evaluations. In the series of papers, Abadie and Imbens (2006, 2008, 2011, 2012) studied various properties of the matching estimator for average treatment effects with a fixed number of matches. In contrast to nonparametric estimation methods for treatment effects (e.g., Heckman, Ichimura and Todd, 1998, and Hirano, Imbens and Ridder, 2003), Abadie and Imbens' estimator shows rather nonstandard behaviors due to the highly nonsmooth functional form caused by the fixed number of matches. Abadie and Imbens (2006) showed that the matching estimator is not $\sqrt{N}$-consistent in general. Abadie and Imbens (2011) proposed a bias correction method based on nonparametric series regression. In addition to the nonstandard asymptotic behavior of the point estimator, Abadie and Imbens (2008) provided an example to show that the standard naive bootstrap (i.e., resample the observation vector with equal weights) is inconsistent to estimate the distribution of the matching estimator for treatment effects. As Abadie and Imbens (2008) argued, the main reason for the failure of the naive bootstrap is that it fails to reproduce the distribution of the number of times each unit is used as a match. Therefore, Abadie and Imbens (2008) recommended to use the asymptotic standard error in Abadie and Imbens (2006) or subsampling in Politis and Romano (1994) for inference.

In this paper, we propose an alternative inference method for the matching estimator based on the wild bootstrap (Wu, 1986, and Mammen, 1993). We show that even though the naive bootstrap is inconsistent to estimate the distribution of the matching estimator, the wild bootstrap

We thank Joachim Freyberger, Jack Porter, Xiaoxia Shi and Christopher Taber for thoughtful comments. We also thank Martin Huber for a computer program. Financial support from the ERC Consolidator Grant (SES-0720961) is gratefully acknowledged (Otsu).

approach can be deliberately applied to estimate consistently the distribution of the estimator. Indeed, Abadie and Imbens (2008, p. 1546) mentioned possibility of the wild bootstrap for valid inference. This paper formally confirms this possibility. However, we emphasize that it is not a trivial task to construct a valid bootstrap counterpart for the matching estimator due to its nonstandard nature. There are two key features in our wild bootstrap method. First, in the wild bootstrap, the treatment assignment and covariates are treated as fixed and thus the distribution of the number of times each unit is used as a match is preserved. Second, we resample not only the regression residuals of treated and untreated observations but also the ones to estimate the average treatment effects. By this construction, we can recover the distribution of the matching estimator. We show the asymptotic validity of the wild bootstrap methods for the average treatment effect and counterpart on the treated population. Also the proposed methods are valid even for the case of vector covariates by incorporating the bias correction terms in Abadie and Imbens (2011). A small simulation study indicates that our wild bootstrap method is favorably comparable to the asymptotic normal approximation. Finally, the proposed method is illustrated by an empirical analysis using the National Supported Work data.

The paper is organized as follows. Section 2 introduces the basic setup and notation. In Section 3, we present our wild bootstrap method and show its asymptotic validity. In Section 4, some simulation evidences are provided. In Section 5, we apply our wild bootstrap method to the National Supported Work data. Section 6 concludes. All proofs are contained in Appendix A. Tables and figures are presented in Appendix B.

2. Setup

Let us introduce the basic setup. Our notation closely follows that of Abadie and Imbens (2006). For each unit $i = 1, \ldots, N$, we observe an indicator variable W_i for a treatment ($W_i = 1$ if treated and $W_i = 0$ otherwise), and outcome

$$Y_i = \begin{cases} Y_i(0) & \text{if } W_i = 0, \\ Y_i(1) & \text{if } W_i = 1, \end{cases}$$

where $Y_i(0)$ and $Y_i(1)$ are potential outcomes for $W_i = 0$ and $W_i = 1$, respectively. Also we observe a vector of covariates X_i for each unit. Based on the non-experimental observations $\{Y_i, W_i, X_i\}_{i=1}^N$ of size N, we wish to conduct inference on the average treatment effect and the one for the treated population

$$\begin{aligned} \tau &= E[Y_i(1) - Y_i(0)], \\ \tau^t &= E[Y_i(1) - Y_i(0)|W_i = 1], \end{aligned}$$

respectively. Let $\mathbb{I}\{A\}$ be the indicator function for an event A and $|x|$ be the Euclidean norm. To estimate the treatment effects τ and τ^t, we consider the matching estimators based on the distance measured by the covariates,

$$\begin{aligned}\hat{\tau} &= \frac{1}{N}\sum_{i=1}^{N}\{\hat{Y}_i(1) - \hat{Y}_i(0)\}, \\ \hat{\tau}^t &= \frac{1}{N_1}\sum_{i:W_i=1}\{Y_i - \hat{Y}_i(0)\},\end{aligned}$$

respectively, where $N_1 = \sum_{i=1}^{N} \mathbb{I}\{W_i = 1\}$ is the number of treated observations, $\hat{Y}_i(0)$ and $\hat{Y}_i(1)$ are estimates of the potential outcomes by imputations defined as

$$\hat{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0, \\ \frac{1}{M}\sum_{j\in\mathcal{J}_M(i)} Y_j & \text{if } W_i = 1, \end{cases} \qquad \hat{Y}_i(1) = \begin{cases} \frac{1}{M}\sum_{j\in\mathcal{J}_M(i)} Y_j & \text{if } W_i = 0, \\ Y_i & \text{if } W_i = 1, \end{cases}$$

and $\mathcal{J}_M(i)$ is the set of indices of the first M matches for unit i,

$$\mathcal{J}_M(i) = \left\{ j \in \{1,\dots,N\} : W_j = 1 - W_i, \sum_{l:W_l=1-W_i} \mathbb{I}\{|X_l - X_i| \le |X_j - X_i|\} \le M \right\}.$$

In the estimators $\hat{\tau}$ and $\hat{\tau}^t$, each unit may be used as a match more than once (matching with replacement). The Euclidean distance $|\cdot|$ for matching may be replaced with the weighted Euclidean.

In practice, it is common that the number of matches M is small (could be one) even though the sample size N is large. To characterize the behaviors of the matching estimators $\hat{\tau}$ and $\hat{\tau}^t$ in such a practical scenario, Abadie and Imbens (2006) analyzed the asymptotic properties of $\hat{\tau}$ and $\hat{\tau}^t$ as the sample size N increases to infinity with fixed M (we call it as the fixed-M asymptotics). Following Abadie and Imbens (2006), we impose the following assumptions. Let $\mu(w,x) = E[Y|W = w, X = x]$, $\sigma^2(w,x) = Var(Y|W = w, X = x)$, and $N_0 = N - N_1$.

Assumption M. (Conditions for $\hat{\tau}$)

- **(i):** $\{Y_i, W_i, X_i\}_{i=1}^{N}$ *is an i.i.d. sample of* (Y, W, X).
- **(ii):** X *is continuously distributed on compact and convex support* $\mathbb{X} \subset \mathbb{R}^k$. *The density of* X *is bounded and bounded away from zero on* $\mathbb{X}$.
- **(iii):** W *is independent of* $(Y(0), Y(1))$ *conditional on* $X = x$ *for almost every* x. *There exists a positive constant* c *such that* $\Pr\{W = 1|X = x\} \in (c, 1-c)$ *for almost every* x.
- **(iv):** *For* $w = 0$ *and* 1, $\mu(w,x)$ *and* $\sigma^2(w,x)$ *are Lipschitz in* $\mathbb{X}$, $\sigma^2(w,x)$ *is bounded away from zero on* $\mathbb{X}$, *and* $E[Y^4|W = w, X = x]$ *is bounded uniformly on* $\mathbb{X}$.

Assumption Mt. (Conditions for $\hat{\tau}^t$)

- **(i):** *Conditional on* $W_i = w$, *the sample consists of independent draws from* $Y, X|W = w$ *for* $w = 0, 1$. *For some* $r \le 1$, $N_1^r/N_0 \to \theta \in (0,\infty)$.
- **(ii):** X *is continuously distributed on compact and convex support* $\mathbb{X} \subset \mathbb{R}^k$. *The density of* X *is bounded and bounded away from zero on* $\mathbb{X}$.
- **(iii):** W *is independent of* $Y(0)$ *conditional on* $X = x$ *for almost every* x. *There exists a positive constant* c *such that* $\Pr\{W = 1|X = x\} < 1 - c$ *for almost every* x.

(iv): *For $w = 0, 1$, $\mu(w,x)$ and $\sigma^2(w,x)$ are Lipschitz in $\mathbb{X}$, $\sigma^2(w,x)$ is bounded away from zero on $\mathbb{X}$, and $E[Y^4|W = w, X = x]$ is bounded uniformly on $\mathbb{X}$.*

The same comments to Abadie and Imbens (2006) apply. Assumption M is used for the estimator $\hat{\tau}$ of the average treatment effect. Assumption M (i) is on the sampling process. Assumption M (ii) is on the distribution form of the covariates X. The assumption that X is continuously distributed can be relaxed. Discrete covariates with finite support can be accommodated by using the subsamples. Such discrete covariates do not change the asymptotic results below. Assumption M (iii) contains standard unconfoundedness and overlap conditions to identify the average treatment effect τ. Assumption M (iv) lists boundedness and smoothness conditions for the conditional mean and variance functions. Assumption Mt is used for the estimator $\hat{\tau}^t$ on the treated population and similar comments to Assumption M apply. Assumption Mt (iii) is weaker than Assumption M (iii) to allow conditional independence instead of full independence and overlap over a subset of support $\mathbb{X}$.

Under Assumption M, Abadie and Imbens (2006, Theorems 3 and 4) showed that $\hat{\tau} \xrightarrow{p} \tau$ and

$$\frac{\sqrt{N}(\hat{\tau} - B_N - \tau)}{\sigma_N} \xrightarrow{d} N(0,1), \tag{1}$$

as $N \to \infty$ (but M is fixed), where B_N and σ_N^2 are asymptotic bias and variance terms, respectively, defined as

$$\begin{aligned}
B_N &= \frac{1}{N}\sum_{i=1}^{N}(2W_i - 1)\left(\frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}\{\mu(1-W_i, X_i) - \mu(1-W_i, X_j)\}\right), \\
\sigma_N^2 &= \sigma_{1N}^2 + \sigma_2^2, \\
\sigma_{1N}^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(1 + \frac{1}{M}\sum_{l=1}^{N}\mathbb{I}\{i \in \mathcal{J}_M(l)\}\right)^2 \sigma^2(W_i, X_i), \\
\sigma_2^2 &= E[\{(\mu(1,X_i) - \mu(0,X_i)) - \tau\}^2].
\end{aligned} \tag{2}$$

Also, under Assumption Mt, Abadie and Imbens (2006) showed $\hat{\tau}^t \xrightarrow{p} \tau$ and

$$\frac{\sqrt{N_1}(\hat{\tau}^t - B_N^t - \tau^t)}{\sigma_N^t} \xrightarrow{d} N(0,1), \tag{3}$$

where

$$\begin{aligned}
B_N^t &= \sum_{i=1}^{N} W_i\left(\frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}\{\mu(0, X_i) - \mu(0, X_j)\}\right), \\
(\sigma_N^t)^2 &= (\sigma_{1N}^t)^2 + (\sigma_2^t)^2, \\
(\sigma_{1N}^t)^2 &= \frac{1}{N_1}\sum_{i=1}^{N}\left(W_i + (1-W_i)\frac{1}{M}\sum_{l=1}^{N}\mathbb{I}\{i \in \mathcal{J}_M(l)\}\right)^2 \sigma^2(W_i, X_i), \\
(\sigma_2^t)^2 &= E[\{(\mu(1,X_i) - \mu(0,X_i)) - \tau^t\}^2|W_i = 1].
\end{aligned} \tag{4}$$

In empirical applications, the number of matches M is typically small (could be one) even for large sample sizes. The fixed-M asymptotics in (1) and (3) provide useful approximations for the distributions of the matching estimators $\hat{\tau}$ and $\hat{\tau}^t$ in such a situation. A key feature of the asymptotic distributions in (1) and (3) is the presence of the bias terms B_N and B_N^t that depend on M. As shown in Abadie and Imbens (2006, Theorems 1 and 2), these bias terms satisfy $B_N = O_p(N^{-1/k})$ and $B_N^t = O_p(N_1^{-r/k})$, where k is the dimension of X and r appears in Assumption Mt (i). Therefore, if $k \geq 2$, the matching estimator $\hat{\tau}$ is not $\sqrt{N}$-consistent for τ. Also, if $k \geq 2r$, the estimator $\hat{\tau}^t$ is not $\sqrt{N_1}$-consistent for τ^t.

In order to deal with these problems, Abadie and Imbens (2011) estimated the bias terms B_N and B_N^t by

$$\hat{B}_N = \frac{1}{N}\sum_{i=1}^{N}(2W_i - 1)\left(\frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}\{\hat{\mu}(1-W_i, X_i) - \hat{\mu}(1-W_i, X_j)\}\right), \tag{5}$$

$$\hat{B}_N^t = \frac{1}{N_1}\sum_{i=1}^{N}W_i\left(\frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}\{\hat{\mu}(0, X_i) - \hat{\mu}(0, X_j)\}\right),$$

respectively, where $\hat{\mu}(w,x)$ is a nonparametric estimator of $\mu(w,x)$. The requirements on $\hat{\mu}(w,x)$ are presented in Theorem 1 below. Abadie and Imbens (2011, Theorem 2) employed a series estimator for $\hat{\mu}(w,x)$ with the series length increasing to infinity, and showed $\sqrt{N}(\hat{B}_N - B_N) \xrightarrow{p} 0$ and $\sqrt{N_1}(\hat{B}_N^t - B_N^t) \xrightarrow{p} 0$ under suitable conditions allowing X to be a vector. These surprisingly fast convergence rates follow from the fact that $\hat{B}_N$ and $\hat{B}_N^t$ basically estimate the contrast $\mu(w,\dot{x}) - \mu(w,x)$ with $\dot{x} - x \to 0$. However, in contrast to the number of matches M, the series length should increase to infinity to guarantee the fast convergence rate. Based on (1) and (3), the bias corrected estimators are obtained as $\tilde{\tau} = \hat{\tau} - \hat{B}_N$ and $\tilde{\tau}^t = \hat{\tau}^t - \hat{B}_N^t$, and these estimators achieve the $\sqrt{N}$-consistency $\sqrt{N}(\tilde{\tau} - \tau)/\sigma_N \xrightarrow{d} N(0,1)$ and $\sqrt{N_1}$-consistency $\sqrt{N_1}(\tilde{\tau}^t - \tau^t)/\sigma_N^t \xrightarrow{d} N(0,1)$ (Abadie and Imbens, 2011, Theorem 2). Since the asymptotic variances σ_N^2 and $(\sigma_N^t)^2$ can be consistently estimated (see, Theorem 6 of Abadie and Imbens, 2006), these asymptotic normality results on $\tilde{\tau}$ and $\tilde{\tau}^t$ yield the asymptotic t confidence intervals for τ and τ^t.

Alternatively, we may consider bootstrap inference based on some resampling scheme. However, Abadie and Imbens (2008) provided an example showing that the naive bootstrap method (i.e., resampling from the observations $\{Y_i, W_i, X_i\}_{i=1}^N$ with uniform weights) cannot consistently estimate the distributions of $\sqrt{N}(\tilde{\tau} - \tau)$ and $\sqrt{N_1}(\tilde{\tau}^t - \tau^t)$ under the fixed-M asymptotics. The main reason is that the naive bootstrap fails to reproduce the distribution of the number of times each unit is used as a match. To best of our knowledge, currently there is no valid bootstrap procedure to approximate the distributions of $\sqrt{N}(\tilde{\tau} - \tau)$ and $\sqrt{N_1}(\tilde{\tau}^t - \tau^t)$ under the fixed-M asymptotics. Indeed, Abadie and Imbens (2008, p. 1546) conjectured that a wild bootstrap method may be employed to estimate consistently the distributions of $\sqrt{N}(\tilde{\tau} - \tau)$ and $\sqrt{N_1}(\tilde{\tau}^t - \tau^t)$. In this paper, we confirm their conjecture by developing a new wild bootstrap method that is asymptotically valid under the fixed-M asymptotics.

3. Wild Bootstrap

In this section, we present asymptotically valid bootstrap procedures to estimate the distributions of $\sqrt{N}(\tilde{\tau}-\tau)$ and $\sqrt{N_1}(\tilde{\tau}^t-\tau^t)$ based on the bias corrected estimators $\tilde{\tau}=\hat{\tau}-\hat{B}_N$ and $\tilde{\tau}^t=\hat{\tau}^t-\hat{B}_N^t$ for the treatment effects.

We first describe the bootstrap procedure for $\tilde{\tau}$. By the definitions of $\hat{\tau}$ and $\hat{B}_N$ and the fact that $\hat{\mu}(1-W_i,X_j)=\hat{\mu}(W_j,X_j)$ for $j\in\mathcal{J}(i)$, the estimator $\tilde{\tau}$ is written as a linear form

$$\begin{aligned}\tilde{\tau} &= \frac{1}{N}\sum_{i=1}^{N}(2W_i-1)\left(\{Y_i-\hat{\mu}(1-W_i,X_i)\}+\frac{1}{M}\sum_{j\in\mathcal{J}(i)}\{Y_j-\hat{\mu}(1-W_i,X_j)\}\right)\\ &= \tilde{\tau}+\frac{1}{N}\sum_{i=1}^{N}(2W_i-1)\left(\hat{e}_i+\frac{1}{M}\sum_{j\in\mathcal{J}(i)}\hat{e}_j\right)+\hat{\xi}_i, \qquad (6)\end{aligned}$$

where $\hat{e}_i=Y_i-\hat{\mu}(W_i,X_i)$ and $\hat{\xi}_i=(2W_i-1)\{\hat{\mu}(W_i,X_i)-\hat{\mu}(1-W_i,X_i)\}-\tilde{\tau}$. The expression in (6) is insightful since $\hat{e}_i$ and $\hat{\xi}_i$ can be interpreted as residual components. Indeed, if we consider the population counterparts $e_i=Y_i-\mu(W_i,X_i)$ and $\xi_i=(2W_i-1)\{\mu(W_i,X_i)-\mu(1-W_i,X_i)\}-\tau$ of $\hat{e}_i$ and $\hat{\xi}_i$, respectively, then the variance components σ_{1N}^2 and σ_2^2 appeared in (2) can be written as

$$\begin{aligned}\sigma_{1N}^2 &= Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(2W_i-1)\left(e_i+\frac{1}{M}\sum_{j\in\mathcal{J}(i)}e_j\right)\middle|\,\mathbf{W},\mathbf{X}\right),\\ \sigma_2^2 &= Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i\right),\end{aligned}$$

respectively, where $\mathbf{W}=(W_1,\ldots,W_N)$ and $\mathbf{X}=(X_1,\ldots,X_N)$. Therefore, it is natural to resample the residuals $\hat{e}_i$ and $\hat{\xi}_i$ by the wild bootstrap (Wu, 1986) to recover the asymptotic variance σ_N^2 of the bias corrected estimator $\tilde{\tau}$. Let $\mathbf{Z}=(Z_1,\ldots,Z_n)$, where $Z_i=(Y_i,W_i,X_i')'$. Precisely speaking, we construct the bootstrap counterpart of $\tilde{\tau}$ as

$$\tilde{\tau}^*=\tilde{\tau}+\frac{1}{N}\sum_{i=1}^{N}(2W_i-1)\left(\hat{e}_i\epsilon_i^*+\frac{1}{M}\sum_{j\in\mathcal{J}(i)}\hat{e}_j\epsilon_j^*\right)+\hat{\xi}_i\epsilon_i^*, \qquad (7)$$

where $\{\epsilon_i^*\}_{i=1}^N$ is an i.i.d. sequence satisfying

$$E[\epsilon_i^*|\mathbf{Z}]=0,\qquad E[\epsilon_i^{*2}|\mathbf{Z}]=1,\qquad E[\epsilon_i^{*4}|\mathbf{Z}]<+\infty.$$

For example, as in Mammen (1993), we can draw ϵ_i^* from the two-point distribution,

$$\epsilon_i^*=\begin{cases}-(\sqrt{5}-1)/2 & \text{with probability} \quad (\sqrt{5}+1)/2\sqrt{5}\\ (\sqrt{5}+1)/2 & \text{with probability} \quad (\sqrt{5}-1)/2\sqrt{5}\end{cases}. \qquad (8)$$

In the simulation study below, we employ this distribution. For more discussions on the choice for ϵ_i^*, see Liu (1988) and Davidson and Flachaire (2008).

In the same manner, we can construct the bootstrap counterpart of $\tilde{\tau}^t$ as

$$\tilde{\tau}^{t*} = \tilde{\tau}^t + \frac{1}{N_1} \sum_{W_i=1} \left(\hat{e}_i \epsilon_i^* + \frac{1}{M} \sum_{j \in \mathcal{J}(i)} \hat{e}_j \epsilon_j^* \right) + \hat{\xi}_i \epsilon_i^* \tag{9}$$

where ϵ_i^* satisfies the same conditions above.

Let $\lambda = (\lambda_1, \ldots, \lambda_k)'$ be a k-dimensional vector of non-negative integers and $\partial^\lambda a(x) = \partial^{\sum_{l=1}^k \lambda_l} a(x)/\partial x_1^{\lambda_1} \cdots \partial x_k^{\lambda_k}$. Define $|a(\cdot)|_d = \max_{\sum_{l=1}^k \lambda_l \leq d} \sup_{x \in \mathbb{X}} |\partial^\lambda a(x)|$. Our main theorem, validity of the wild bootstrap procedures, is presented as follows.

Theorem 1. *Suppose that for $w \in \{0,1\}$, the derivatives $\partial^\lambda \mu(w,x)$ exist for all λ with $\sum_{l=1}^k \lambda_l = k$ and satisfy $\sup_{x \in \mathbb{X}} |\partial^\lambda \mu(w,x)| \leq C$ for some $C > 0$, and $|\hat{\mu}(w,\cdot) - \mu(w,\cdot)|_{k-1} = o_p(N^{-1/2+1/k})$.*

(i): *Under Assumption M,*

$$\sup_t |\Pr\{\sqrt{N}(\tilde{\tau}^* - \tilde{\tau}) \leq t | \mathbf{Z}\} - \Pr\{\sqrt{N}(\tilde{\tau} - \tau) \leq t\}| \xrightarrow{p} 0.$$

(ii): *Under Assumption Mt,*

$$\sup_t |\Pr\{\sqrt{N_1}(\tilde{\tau}^{t*} - \tilde{\tau}^t) \leq t | \mathbf{Z}\} - \Pr\{\sqrt{N_1}(\tilde{\tau}^t - \tau^t) \leq t\}| \xrightarrow{p} 0.$$

This theorem says that the distributions of the wild bootstrap counterparts $\sqrt{N}(\tilde{\tau}^* - \tilde{\tau})$ and $\sqrt{N_1}(\tilde{\tau}^{t*} - \tilde{\tau}^t)$ consistently estimate those of the target objects $\sqrt{N}(\tilde{\tau} - \tau)$ and $\sqrt{N_1}(\tilde{\tau}^t - \tau^t)$ under the Kolmogorov distance, respectively. For example, let q_α^* be the $(1-\alpha)$-th quantile of $|\tilde{\tau}^* - \tilde{\tau}|$, which can be estimated by simulating ϵ_i^*. Then based on Theorem 1 (i), the $100(1-\alpha)\%$ symmetric bootstrap confidence interval of τ is obtained as $[\tilde{\tau} - q_\alpha^*, \tilde{\tau} + q_\alpha^*]$. In the simulation below, we use this confidence interval.

The condition on $\hat{\mu}(w,\cdot)$ is required to guarantee sufficiently fast convergence rates on the bias estimators $\hat{B}_N$ and $\hat{B}_N^t$ in (5), i.e., $\sqrt{N}(\hat{B}_N - B_N) \xrightarrow{p} 0$ and $\sqrt{N_1}(\hat{B}_N^t - B_N^t) \xrightarrow{p} 0$. For example, $\hat{\mu}(w,\cdot)$ can be a series estimator with a suitable choice of the series length (see, Lemma A.1 of Abadie and Imbens, 2011). Other candidates of $\hat{\mu}(w,\cdot)$ are the kernel estimator and nearest neighborhood estimator with adequate trimming (Stone, 1977). For any choice of $\hat{\mu}(w,\cdot)$, we need to choose a tuning constant that varies with N to guarantee the fast convergence of $\hat{B}_N$ and $\hat{B}_N^t$.

4. Simulation

In this section, we present some simulation evidence to illustrate our theoretical result. Based on Frölich (2004) and Busso, DiNardo and McCrary (2014), we consider the following data generating process for $\{Y_i, W_i, X_i\}_{i=1}^N$,

$$\begin{aligned} Y_i(1) &= \tau + m(X_i) + \epsilon_i, \\ Y_i(0) &= m(X_i) + \epsilon_i, \\ W_i &= \mathbb{I}\{X_i \leq \nu_i\}, \\ X_i &= \alpha + \beta Z_i, \end{aligned}$$

where ν_i and Z_i are uniformly distributed on $[0,1]$. The average treatment effect is set as $\tau = 0$. For the error term ϵ_i, we consider two cases, $\epsilon_i \sim N(0, 0.2^2)$ and centered lognormal with the standard deviation 0.2. We consider 48 data generating processes (four designs of (α, β) and six curves of $m(\cdot)$) summarized in Table 1.

Although $k = 1$ and thus $\hat{\tau}$ is $\sqrt{N}$-consistent without bias correction, we consider the bias corrected estimator $\tilde{\tau} = \hat{\tau} - \hat{B}_N$, where $\hat{B}_N$ is estimated by the OLS of the linear regression $\hat{\mu}_i = \hat{\gamma}_0 + \hat{\gamma}_1 X_i$. The perturbation ϵ_i^* for the wild bootstrap resample is drawn from the two-point distribution in (8). For each simulation, we report 90% and 95% coverage rates of the wild bootstrap and asymptotic t confidence intervals.[1] The standard error of the asymptotic t is computed by the method in Abadie and Imbens (2006). We also report average lengths (over Monte Carlo replications) of the confidence intervals. The sample size is set as $N = 100$. Simulation results based on 10,000 replications are displayed in Tables 2 and 3 for the normal and lognormal ϵ cases, respectively.

Our findings are summarized as follows. First, across all cases the coverages of our wild bootstrap confidence interval are reasonably close to the nominal ones. This result shows that the wild bootstrap approximation works well even for the moderate sample size $N = 100$. Second, compared to the asymptotic t confidence interval, the wild bootstrap shows better coverage properties. In particular, for the case of Design 1 with Curve 6, the coverage rates of the 90% wild bootstrap confidence intervals are 0.8948 and 0.9011 for the normal and lognormal cases, respectively. However the asymptotic t confidence intervals show under-coverage, 0.7852 and 0.7461, respectively. Finally, the average lengths of the wild bootstrap and asymptotic t confidence intervals are comparable except for some cases where the asymptotic t shows under-coverage.

To assess the power properties, we plot the power curves of the wild bootstrap and asymptotic t tests for the null hypothesis $H_0 : \tau = 0$ with the 5% significance level for the case of Design 1 with Curve 1 in Figures 1 (normal case) and 2 (lognormal case). For the normal case, the power curves of two tests are very similar. For the lognormal case, the wild bootstrap test is slightly more powerful than the asymptotic t test.

Overall our simulation results are encouraging: the wild bootstrap method developed in this paper is favorably comparable to the asymptotic t method in Abadie and Imbens (2006, 2011).

5. Empirical Application

In this section, we apply the wild bootstrap method to the National Supported Work (NSW) data. The NSW demonstration was a program that aimed to provide temporary job experiments for individuals with longstanding employment problems. The candidates were randomly assigned to the treatment group (received all benefit of the NSW program) and control group (did not receive any benefit). Due to its random assignment property, researchers are able to estimate

[1] In a preliminary simulation study, we also computed the naive bootstrap confidence intervals (i.e., resample $\{Y_i, W_i, X_i\}_{i=1}^N$ with uniform weights), which is inconsistent under the fixed-M asymptotics. For all cases, the 90% and 95% coverage rates are 1 and its confidence intervals are longer than the asymptotic t and wild bootstrap ones. For example, for Design 1 and Curve 1, the average length of the naive bootstrap confidence intervals is 0.2250 (the asymptotic t is 0.1737 and wild bootstrap is 0.1701).

consistently the treatment effect τ^t for the treated population. Together with non-experimental data from outside, we can evaluate the bias of non-experimental estimators. The data was first analyzed by Lalonde (1986) and subsequently by Heckman and Hotz (1989), Dehejia and Wahba (1999, 2002) (hereafter, DW), Smith and Todd (2005), among others. Here we use the sample uploaded by DW. The first experimental sample is the one used by Lalonde (1986). It includes a male sample with ex-addict, ex-offender, and/or high school dropout having complete ex-ante and ex-post earning data. The second experimental data is the one used by DW. This is a subset of the first sample satisfying either randomly assigned in January through April of 1976 or randomly assigned in October of 1976 through August of 1977 with zero earning in 13-24 months before the random assignment. Two non-experimental data were created by Lalonde (1986) and then used by DW. The first sample is drawn from the Current Population Survey (CPS) with matched Social Security earnings data. It includes all male samples in the CPS less than 55 years old. The second sample is drawn from the Panel Study Income Dynamics (PSID). It contains all male household heads who were less than 55 years old and did not classify themselves as retired in 1975. To summarize the data, Table 4 reports the mean, standard error, and normalized distance between the experimental data and non-experimental control groups. Two NSW data has very similar characteristics except the real earnings in 1975 (RE 75). This is because the DW data contains more portion of zero earning data. We can also see the clear difference between the experimental and non-experimental samples. Specifically, age, marital status, ethnicity, and earnings are largely different. Following Abadie and Imbens (2011), we report the normalized distance between the experimental data and non-experimental control groups in the last six rows. The normalized distance is defined as

$$\frac{\bar{X}_1 - \bar{X}_0}{\sqrt{(S_0^2 + S_1^2)/2}},$$

where $\bar{X}_w = \sum_{i:W_i=w} X_i/N_w$ and $S_w = \sum_{i:W_i=w}(X_w - \bar{X}_w)^2/(N_w - 1)$.[2] As expected, the experimental and non-experimental control groups have larger difference (in absolute value), which implies difficulty of finding a good match.

For each treatment control pair and the number of matches M, we compute three different standard errors to assess the difference of three inference methods: (i) the asymptotic standard error by Abadie and Imbens (2006), (ii) wild bootstrap proposed in this paper, and (iii) naive bootstrap which is inconsistent under the fixed-M asymptotics.[3]

Table 5 summarizes the results for these matching estimates and standard errors. The first row presents the mean differences of treatment/control group. The unbiased estimates of the average treatment effect on treated are 886.30 and 1794.34 for the Lalonde and DW samples, respectively. The simple covariate matching works well for the DW sample, but not for the Lalonde sample. For the standard errors, we find that the asymptotic and wild bootstrap methods provide similar values in general. On the other hand, the naive bootstrap standard error tends to take somewhat

[2] Note that the normalized distance is different from the t-statistic $(\bar{X}_1 - \bar{X}_0)/\sqrt{S_0^2/N_0 + S_1^2/N_1}$ for null hypothesis $E[X_1] = E[X_0]$.

[3] The matching is based on the Mahalanobis distance. We use the simple linear regression to compute $\hat{\mu}(w, x)$ for the bias correction and wild bootstrap.

different values compared to the others. This result suggests that statistical inference based on the naive bootstrap can be problematic and it is reasonable to employ the asymptotic normal or wild bootstrap inference, which are consistent under the fixed-M asymptotics.

6. Conclusion

This paper proposes a wild bootstrap inference method for the matching estimator of treatment effects. In contrast to the naive bootstrap, our wild bootstrap method is asymptotically valid under the asymptotics with a fixed number of matches. Our method is applicable for both the average treatment effect and the counterpart for the treated population. Simulation results indicate that the wild bootstrap method works well in finite samples and favorably comparable with the asymptotic normal approximation. Although it is beyond the scope of this paper, it would be interesting to extend the higher order analysis for the wild bootstrap (Kline and Santos, 2012) to the fixed M-asymptotics and to investigate higher order properties of our wild bootstrap method. Also an extension of the wild bootstrap method to the propensity score matching is currently under investigation by the authors.

Appendix A. Mathematical Appendix

A.1. **Proof of Theorem 1 (i).** We first show $\sqrt{N}(\hat{B}_N - B_N) \xrightarrow{p} 0$. Since $|2W_i - 1| = 1$, the argument in Abadie and Imbens (2011, p. 9) based on Taylor expansions implies

$$
\begin{aligned}
|\hat{B}_N - B_N| &\leq \max_{i=1,\dots,N, j\in\mathcal{J}_M(i)} \sum_{w=0,1} |\{\hat{\mu}(w, X_i) - \hat{\mu}(w, X_j)\} - \{\mu(w, X_i) - \mu(w, X_j)\}| \\
&\leq \sum_{w=0,1} |\hat{\mu}(w, \cdot) - \mu(w, \cdot)|_{k-1} \sum_{d=1}^{k-1} \frac{1}{d!} \sum_{\lambda\in\Lambda_d} \max_{i=1,\dots,N, j\in\mathcal{J}_M(i)} |X_i - X_j|^d + o_p(N^{-1/2}),
\end{aligned}
$$

where Λ_d is the set of k-vectors λ of non-negative integers with $\sum_{l=1}^k \lambda_l = d$. By Abadie and Imbens (2006, Lemma 1), $\max_{i=1,\dots,N, j\in\mathcal{J}_M(i)} |X_i - X_j|^d = O_p(N^{-d/k})$. Thus, the assumption $|\hat{\mu}(w, \cdot) - \mu(w, \cdot)|_{k-1} = o_p(N^{-1/2+1/k})$ implies $\sqrt{N}(\hat{B}_N - B_N) \xrightarrow{p} 0$.

We now consider the bootstrap counterpart $\tilde{\tau}^*$. Let us introduce some notation

$$
\begin{aligned}
K_M(i) &= \sum_{l=1}^{N} \mathbb{I}\{i \in \mathcal{J}_M(l)\}, \qquad e_i = Y_i - \mu(W_i, X_i), \\
\xi_i &= (2W_i - 1)\{\mu(W_i, X_i) - \mu(1 - W_i, X_i)\} - \tau, \\
\tau_i &= \mu(1, X_i) - \mu(0, X_i).
\end{aligned}
$$

Then $\tilde{\tau}^*$ in (7) is rewritten and decomposed as

$$
\begin{aligned}
\tilde{\tau}^* &= \tilde{\tau} + \frac{1}{N}\sum_{i=1}^{N}(2W_i - 1)\left(\hat{e}_i\epsilon_i^* + \frac{1}{M}\sum_{j\in\mathcal{J}(i)} \hat{e}_j\epsilon_j^*\right) + \hat{\xi}_i\epsilon_i^*, \\
&= \tilde{\tau} + \frac{1}{N}\sum_{i=1}^{N}\left[(2W_i - 1)\{1 + M^{-1}K_M(i)\}\hat{e}_i + \hat{\xi}_i\right]\epsilon_i^* \\
&= \tilde{\tau} + T_N^* + R_{1N}^* + R_{2N}^*,
\end{aligned}
$$

where

$$
\begin{aligned}
T_N^* &= \frac{1}{N}\sum_{i=1}^{N}\left[(2W_i - 1)\{1 + M^{-1}K_M(i)\}e_i + \xi_i\right]\epsilon_i^*, \\
R_{1N}^* &= \frac{1}{N}\sum_{i=1}^{N}(2W_i - 1)\{1 + M^{-1}K_M(i)\}\{\mu(W_i, X_i) - \hat{\mu}(W_i, X_i)\}\epsilon_i^*, \\
R_{2N}^* &= \frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_i - \xi_i.
\end{aligned}
$$

Thus, it is enough to show that

$$
\sup_t |\Pr\{\sqrt{N}T_N^* \leq t|\mathbf{Z}\} - \Pr\{\sqrt{N}(\tilde{\tau} - \tau) \leq t\}| \xrightarrow{p} 0, \tag{10}
$$

$$
\sqrt{N}(R_{1N}^* + R_{2N}^*) \xrightarrow{p} 0. \tag{11}
$$

First, we show (10). From Polya's theorem and (1) shown by Abadie and Imbens (2006, Theorem 4), it is enough for (10) to verify

$$Var(\sqrt{N}T_N^*|\mathbf{Z}) - \sigma_N^2 \xrightarrow{p} 0, \tag{12}$$

$$\Pr\{\sqrt{N}T_N^*/\sigma_N \le t|\mathbf{Z}\} - \Phi(t) \xrightarrow{p} 0 \quad \text{for all } t \in \mathbb{R}, \tag{13}$$

where $\Phi(t)$ is the standard normal distribution function.

To show (12), denote $Var(\sqrt{N}T_N^*|\mathbf{Z}) = \hat{\sigma}_{1N}^2 + \hat{\sigma}_2^2 + 2C_N$ where

$$\begin{aligned}
\hat{\sigma}_{1N}^2 &= \frac{1}{N}\sum_{i=1}^N \{1 + M^{-1}K_M(i)\}^2 e_i^2, \\
\hat{\sigma}_{2N}^2 &= \frac{1}{N}\sum_{i=1}^N (\tau_i - \tau)^2, \\
C_N &= \frac{1}{N}\sum_{i=1}^N (2W_i - 1)\{1 + M^{-1}K_M(i)\}e_i(\tau_i - \tau).
\end{aligned}$$

By the standard law of large number, $\hat{\sigma}_{2N}^2 \xrightarrow{p} \sigma_2^2$. For $\hat{\sigma}_{1N}^2$ note that

$$\begin{aligned}
E[(\hat{\sigma}_{1N}^2 - \sigma_N^2)^2] &= \frac{1}{N}E\left[\{1 + M^{-1}K_M(i)\}^4 E[\{e_i^2 - \sigma^2(W_i, X_i)\}^2|W_i, X_i]\right] \\
&\le \frac{1}{N}E[\{1 + M^{-1}K_M(i)\}^4] \sup_{w\in\{0,1\}, x\in\mathbb{X}} E[e_i^4|W_i = w, X_i = x] \\
&\to 0,
\end{aligned}$$

where the convergence follows from Assumption M (iv) and boundedness of $E[K_M(i)^q]$ for all $q > 0$ uniformly over N (Lemma 3 of Abadie and Imbens, 2006). Thus, the Markov inequality implies $|\hat{\sigma}_{1N}^2 - \sigma_{1N}^2| \xrightarrow{p} 0$. Similarly, by Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
E[C_N^2] &= \frac{1}{N}E\left[\{1 + M^{-1}K_M(i)\}^2 e_i^2(\tau_i - \tau)^2\right] \\
&\le \frac{1}{N}E[\{1 + M^{-1}K_M(i)\}^4]^{1/2}\left(E[(\tau_i - \tau)^4] \sup_{w\in\{0,1\}, x\in\mathbb{X}} E[e_i^4|W_i = w, X_i = x]\right)^{1/2} \\
&\to 0
\end{aligned}$$

and thus, $C_N \xrightarrow{p} 0$. Therefore, (12) is verified.

Now denote $\eta_i = (2W_i - 1)\{1 + M^{-1}K_M(i)\}e_i + \xi_i$. so that $T_N^* = N^{-1}\sum_{i=1}^N \eta_i\epsilon_i^*$. Since η_i is a deterministic function of $\mathbf{Z}$, T_N^* is a sum of independent variables given $\mathbf{Z}$. Thus, for (13), it is enough to verify the Lindeberg condition

$$\frac{1}{N}\sum_{i=1}^N E[(\eta_i\epsilon_i^*\sigma_N^{-1})^2\mathbb{I}\{|\eta_i\epsilon_i^*\sigma_1^{-1}| > \delta\sqrt{N}\}|\mathbf{Z}] \xrightarrow{p} 0, \tag{14}$$

for each $\delta > 0$. By the Cauchy-Schwarz and Markov inequalities, we obtain

$$\begin{aligned}
E[(\eta_i\epsilon_i^*\sigma_N^{-1})^2\mathbb{I}\{|\eta_i\epsilon_i^*\sigma_N^{-1}| > \delta\sqrt{N}\}|\mathbf{Z}] &\le E[(\eta_i\epsilon_i^*\sigma_N^{-1})^4|\mathbf{Z}]/(\delta^2 N) \\
&= \eta_i^4 E[\epsilon_i^{*4}|\mathbf{Z}]/(\sigma_N^4\delta^2 N).
\end{aligned}$$

Since (a) $E[\epsilon_i^{*4}|\mathbf{Z}]$ is bounded (assumption on ϵ_i^{*4}), (b) σ_N^4 is uniformly bounded from below ($\sigma_N^4 \geq \inf_{w\in\{0,1\},x\in\mathbb{X}} \sigma^4(w,x)$), and (c) $\frac{1}{N}\sum_{i=1}^N \eta_i^4 = O_p(1)$ (Lemma 1 below), the Lindeberg condition in (14) follows and (13) is verified.

Next, we show (11). By $|2W_i - 1| = 1$ and the Cauchy-Schwarz inequality, the term R_{1N}^* satisfies

$$E[NR_{1N}^{*2}] \leq \sqrt{E[\{1+M^{-1}K_M(i)\}^4\epsilon_i^{*4}]E[\{\mu(W_i,X_i)-\hat{\mu}(W_i,X_i)\}^4]} \to 0,$$

where the convergence follows from $E[\{\mu(W_i,X_i)-\hat{\mu}(W_i,X_i)\}^4] \to 0$ (by the assumption on $\hat{\mu}(w,\cdot)$) and boundedness of $E[K_M(i)\}^4]$ and $E[\epsilon_i^{*4}|\mathbf{Z}]$. Thus the Markov inequality implies $R_{1N}^* = o_p(N^{-1/2})$. For R_{2N}^*, decompose

$$\begin{aligned} R_{2N}^* &= \frac{1}{N}\sum_{i=1}^N (2W_i-1)\{\mu(W_i,X_i)-\hat{\mu}(W_i,X_i)\}\epsilon_i^* \\ &\quad -\frac{1}{N}\sum_{i=1}^N (2W_i-1)\{\mu(1-W_i,X_i)-\hat{\mu}(1-W_i,X_i)\}\epsilon_i^*. \end{aligned}$$

By applying a similar argument to R_{1N}^*, we obtain $R_{2N}^* = o_p(N^{-1/2})$. Therefore, the conclusion follows.

Lemma 1. *Under Assumption M, it holds* $\frac{1}{N}\sum_{i=1}^N \eta_i^4 = O_p(1)$.

Proof: Decompose

$$\begin{aligned} \eta_i^4 &= \{1+M^{-1}K_M(i)\}^4 e_i^4 + 4\{1+M^{-1}K_M(i)\}^3 e_i^3\xi_i + 6\{1+M^{-1}K_M(i)\}^2 e_i^2\xi_i^2 \\ &\quad +4\{1+M^{-1}K_M(i)\}e_i\xi_i^3 + \xi_i^4. \end{aligned}$$

Since $E[Y^4|W=w,X=x]$ is uniformly bounded over w and x, $\max_{q\in\{1,\ldots,4\}} E[|e_i|^q|\mathbf{Z}] < C$ for some $C > 0$. By Abadie and Imbens (2006, Lemma 3), $E[K_M(i)^q]$ is also uniformly bounded for any q. By Lipschitz continuity of μ and compact support of X, it holds $|\xi_i| < C'$ for some $C' > 0$. Thus, we have

$$\begin{aligned} E\left[\left|\frac{1}{N}\sum_{i=1}^N \eta_i^4\right|\right] &\leq \frac{1}{N}\sum_{i=1}^N E[\{1+M^{-1}K_M(i)\}^4 e_i^4] + \frac{4}{N}\sum_{i=1}^N E[\{1+M^{-1}K_M(i)\}^3|e_i|^3|\xi_i|] \\ &\quad +\frac{6}{N}\sum_{i=1}^N E[\{1+M^{-1}K_M(i)\}^2 e_i^2\xi_i^2] + \frac{4}{N}\sum_{i=1}^N E[\{1+M^{-1}K_M(i)\}|e_i||\xi_i|^3] + \frac{1}{N}\sum_{i=1}^N E[\xi_i^4] \\ &\leq E[\{1+M^{-1}K_M(i)\}^4]C + 4E[\{1+M^{-1}K_M(i)\}^3]CC' + 6E[\{1+M^{-1}K_M(i)\}^2]CC'^2 \\ &\quad +4E[\{1+M^{-1}K_M(i)\}]CC'^3 + C'^4 \\ &= C'' < \infty. \end{aligned}$$

By Markov inequality,

$$\Pr\left\{\left|\frac{1}{N}\sum_{i=1}^N \eta_i^4\right| > c\right\} < \frac{1}{c}E\left[\left|\frac{1}{N}\sum_{i=1}^N \eta_i^4\right|\right] < \frac{C''}{c},$$

for any $c > 0$, and the conclusion follows.

A.2. **Proof of Theorem 1 (ii).** The proof is similar to that of Part (i). By the same argument, it holds $\sqrt{N_1}(\hat{B}_N^t - B_N^t) \xrightarrow{p} 0$. Note that (9) can be rewritten as

$$\begin{aligned}
\tilde{\tau}^{t*} &= \tilde{\tau}^t + \frac{1}{N_1}\sum_{W_i=1}\left(\hat{e}_i\epsilon_i^* + \frac{1}{M}\sum_{j\in\mathcal{J}(i)}\hat{e}_j\epsilon_j^*\right) + \hat{\xi}_i\epsilon_i^*, \\
&= \tilde{\tau}^t + \frac{1}{N_1}\sum_{i=1}^{N}\left\{W_i(\hat{e}_i + \hat{\xi}_i) + (1-W_i)M^{-1}K_M(i)\hat{e}_i\right\}\epsilon_i^* \\
&= \tilde{\tau}^t + T_N^{t*} + R_{1N}^{t*} + R_{2N}^{t*},
\end{aligned}$$

where

$$\begin{aligned}
T_N^{t*} &= \frac{1}{N_1}\sum_{i=1}^{N}\left\{W_i(e_i + \xi_i) + (1-W_i)M^{-1}K_M(i)e_i\right\}\epsilon_i^*, \\
R_{1N}^{t*} &= \frac{1}{N_1}\sum_{i=1}^{N}\big(W_i\{\mu(W_i,X_i) - \hat{\mu}(W_i,X_i)\} \\
&\qquad + (1-W_i)\{1 + M^{-1}K_M(i)\}\{\mu(W_i,X_i) - \hat{\mu}(W_i,X_i)\}\big)\,\epsilon_i^*, \\
R_{2N}^{t*} &= \frac{1}{N_1}\sum_{i;W_i=1}\hat{\xi}_i - \xi_i.
\end{aligned}$$

Thus, the following conditions are sufficient for the result.

$$\sup_t |\Pr\{\sqrt{N_1}T_N^{t*} \le t|\mathbf{Z}\} - \Pr\{\sqrt{N_1}(\tilde{\tau}^t - \tau^t) \le t\}| \xrightarrow{p} 0, \tag{15}$$

$$\sqrt{N_1}(R_{1N}^{t*} + R_{2N}^{t*}) \xrightarrow{p} 0. \tag{16}$$

First we show (15). Since Abadie and Imbens (2006) have shown the asymptotic normality of $\sqrt{N_1}(\tilde{\tau}^t - \tau^t)/\sigma_N^t$, it is enough to show the following condition

$$Var(\sqrt{N_1}T_N^{t*}|\mathbf{Z}) - (\sigma_N^t)^2 \xrightarrow{p} 0, \tag{17}$$

$$|\Pr\{\sqrt{N_1}T_N^{t*} \le t|\mathbf{Z}\} - \Phi(t)| \xrightarrow{p} 0, \quad \text{for all } t \in \mathbb{R} \tag{18}$$

We first show (17). Note that

$$\begin{aligned}
Var(\sqrt{N_1}T_N^{t*}|\mathbf{Z}) &= (\hat{\sigma}_{1N}^t)^2 + (\hat{\sigma}_{2N}^t)^2 + 2C_N^t, \\
(\hat{\sigma}_{1N}^t)^2 &= \frac{1}{N_1}\sum_{i=1}^{N}\left\{W_i + (1-W_i)M^{-1}K_M(i)\right\}^2 e_i^2, \\
(\hat{\sigma}_{2N}^t)^2 &= \frac{1}{N_1}\sum_{i:W_i=1}(\tau_i - \tau)^2, \\
C_N^t &= \frac{1}{N_1}\sum_{i=1}^{N}\left\{W_i + (1-W_i)M^{-1}K_M(i)\right\}e_i(\tau_i - \tau).
\end{aligned}$$

By the law of large numbers, $(\hat{\sigma}_{2N}^t)^2 \xrightarrow{p} (\sigma_2^t)^2$. For $(\hat{\sigma}_{1N}^t)^2$, note that

$$\begin{aligned}
& E[\{(\hat{\sigma}_{1N}^t)^2 - (\sigma_{1N}^t)^2\}^2] \\
= & E\left[\frac{1}{N_1^2}\sum_{i=1}^N W_i\{e_i^2 - \sigma^2(W_i, X_i)\}^2\right] + E\left[\frac{1}{N_1^2}\sum_{i=1}^N (1-W_i)\left(M^{-1}K_M(i)\right)^4 \{e_i^2 - \sigma^2(W_i, X_i)\}^2\right] \\
\leq & \frac{1}{N_1}\sup_{w\in\{0,1\},x\in\mathbb{X}} E[e_i^4|W_i = w, X_i = x] \\
& + \frac{1}{N_1}\frac{N_0}{N_1}E[\{M^{-1}K_M(i)\}^4|W_i = 0]\sup_{w\in\{0,1\},x\in\mathbb{X}} E[e_i^4|W_i = w, X_i = x] \\
\to & 0,
\end{aligned}$$

where the last line follows from the uniform boundedness of $\frac{N_0}{N_1}E[\{M^{-1}K_M(i)\}^4|W_i = 0]$ (Lemma 3 (iii) Abadie and Imbens 2006). By Markov inequality, it holds $|(\hat{\sigma}_{1N}^t)^2 - (\sigma_{1N}^t)^2| \xrightarrow{p} 0$. Finally, by Cauchy-Schwarz inequality,

$$\begin{aligned}
E[C_N^{t2}] = & E\left[\frac{1}{N_1^2}\sum_{i=1}^N W_i e_i^2(\tau_i - \tau)^2\right] + E\left[\frac{1}{N_1^2}\sum_{i=1}^N (1-W_i)\left(M^{-1}K_M(i)\right)^2 e_i^2(\tau_i - \tau)^2\right] \\
\leq & \frac{1}{N_1}E[(\tau_i - \tau)^2|W_i = 1]\sup_{w\in\{0,1\},x\in\mathbb{X}} E[e_i^4|W_i = w, X_i = x] \\
& + \frac{1}{N_1}\frac{N_0}{N_1}E[\{M^{-1}K_M(i)\}^4|W_i = 0]^{1/2}\left(E[(\tau_i - \tau)^4|W_i = 0]\sup_{w\in\{0,1\},x\in\mathbb{X}} E[e_i^4|W_i = w, X_i = x]\right)^{1/2} \\
\to & 0.
\end{aligned}$$

holds. By Markov inequality, it holds $E[C_N^t] \xrightarrow{p} 0$ and thus (17) follows.

Let $\eta_i^t = W_i(e_i + \xi_i) + (1 - W_i)M^{-1}K_M(i)e_i$ so that $T_N^{t*} = N_1^{-1}\sum_{i=1}^N \eta_i^t\epsilon_i^*$. Conditional on the sample $\mathbf{Z}$, T_N^{t*} is a sum of independent random variable. Thus, the Lindeberg condition

$$\frac{1}{N_1}\sum_{i=1}^N E[(\eta_i^t\epsilon_i^*/\sigma_N^t)^2\mathbb{I}\{|\eta_i^t\epsilon_i^*/\sigma_N^t| > \delta\sqrt{N}\}|\mathbf{Z}] \xrightarrow{p} 0, \tag{19}$$

is sufficient for (18). Note that

$$\begin{aligned}
E[(\eta_i^t\epsilon_i^*/\sigma_N^t)^2\mathbb{I}\{|\eta_i^t\epsilon_i^*/\sigma_N^t| > \delta\sqrt{N}\}|\mathbf{Z}] &\leq E[(\eta_i^t\epsilon_i^*/\sigma_N^t)^4|\mathbf{Z}]/(\delta^2 N) \\
&= \eta_i^t E[(\epsilon_i^*/\sigma_{1N}^t)^4|\mathbf{Z}]/(\delta^2 N).
\end{aligned}$$

Since (a) $E[\epsilon_i^{*4}|\mathbf{Z}]$ is bounded (assumption on ϵ_i^{*4}), (b) $(\sigma_N^t)^4$ is uniformly bounded from below ($(\sigma_N^t)^4 \geq \inf_{w\in\{0,1\},x\in\mathbb{X}}\sigma^4(w,x)$), and (c) $\frac{1}{N_1}\sum_{i=1}^N(\eta_i^t)^4 = O_p(1)$ (similar argument to Lemma 1), the Lindeberg condition follows.

The proof of (16) is similar to that of Part (i) and thus is omitted.

Appendix B. Tables and Figures

Table 1. Simulation designs

Design	α	β	Control-treated ratio
1	0.15	0.7	1:1
2	0.3	0.4	1:1
3	0.5	0.4	7:3
4	0.6	0.2	7:3

Curves	$m(X)$
1	$0.15 + 0.7X$
2	$0.1 + X/2 + \exp(-200(X-0.7)^2)/2$
3	$0.8 - 2(X-0.9)^2 - 5(X-0.7)^3 - 10(X-0.6)^{10}$
4	$0.2 + \sqrt{1-X} - 0.6(0.9-X)^2$
5	$0.2 + \sqrt{1-X} - 0.6(0.9-X)^2 - 0.1X\cos(30X)$
6	$0.4 + 0.25\sin(8X-5) + 0.4\exp(-16(4X-2.5)^2)$

Table 2. Coverage rates for normal ϵ case

		Wild bootstrap			Asymptotic t		
Design	Curve	90% CI	95% CI	Average CI length	90% CI	95% CI	Average CI length
1	1	0.8798	0.9349	0.1701	0.8843	0.9379	0.1737
	2	0.8864	0.9393	0.2007	0.8366	0.9035	0.1803
	3	0.8831	0.9350	0.1745	0.8764	0.9310	0.1733
	4	0.8800	0.9342	0.1729	0.8792	0.9323	0.1735
	5	0.8790	0.9346	0.1759	0.8753	0.9316	0.1750
	6	0.8948	0.9457	0.2296	0.7852	0.8588	0.1789
2	1	0.8878	0.9409	0.1618	0.8897	0.9440	0.1642
	2	0.9034	0.9496	0.1785	0.8776	0.9333	0.1664
	3	0.8891	0.9395	0.1622	0.8906	0.9402	0.1641
	4	0.8861	0.9383	0.1620	0.8879	0.9399	0.1639
	5	0.8915	0.9428	0.1645	0.8886	0.9422	0.1641
	6	0.9285	0.9685	0.1943	0.8744	0.9298	0.1655
3	1	0.8746	0.9265	0.1775	0.8777	0.9307	0.1818
	2	0.8556	0.9164	0.2432	0.7352	0.8102	0.1885
	3	0.8669	0.9249	0.1783	0.8684	0.9271	0.1814
	4	0.8686	0.9219	0.1785	0.8741	0.9251	0.1827
	5	0.8715	0.9275	0.1842	0.8679	0.9236	0.1844
	6	0.8957	0.9465	0.2113	0.8388	0.9018	0.1845
4	1	0.8804	0.9328	0.1730	0.8838	0.9363	0.1765
	2	0.8991	0.9480	0.2108	0.8362	0.8970	0.1793
	3	0.8806	0.9346	0.1730	0.8863	0.9384	0.1766
	4	0.8818	0.9380	0.1730	0.8876	0.9404	0.1766
	5	0.8755	0.9287	0.1765	0.8731	0.9291	0.1777
	6	0.8910	0.9402	0.1766	0.8890	0.9386	0.1770

Table 3. Coverage rates for lognormal ϵ case

		Wild bootstrap			Asymptotic t		
Design	Curve	90% CI	95% CI	Average CI length	90% CI	95% CI	Average CI length
1	1	0.8993	0.9542	0.1563	0.8877	0.9480	0.1615
	2	0.9007	0.9523	0.1909	0.8236	0.8927	0.1689
	3	0.8992	0.9500	0.1609	0.8730	0.9349	0.1605
	4	0.9050	0.9591	0.1577	0.8877	0.9450	0.1593
	5	0.9058	0.9571	0.1640	0.8842	0.9419	0.1636
	6	0.9011	0.9530	0.2231	0.7461	0.8270	0.1678
2	1	0.9079	0.9548	0.1493	0.8950	0.9480	0.1531
	2	0.9331	0.9693	0.1698	0.8860	0.9414	0.1568
	3	0.9140	0.9594	0.1505	0.8978	0.9524	0.1539
	4	0.9118	0.9603	0.1498	0.8976	0.9525	0.1531
	5	0.9191	0.9635	0.1538	0.8971	0.9523	0.1545
	6	0.9455	0.9816	0.1870	0.8658	0.9269	0.1557
3	1	0.8725	0.9294	0.1614	0.8640	0.9264	0.1673
	2	0.8764	0.9291	0.2340	0.7254	0.8052	0.1732
	3	0.8924	0.9407	0.1622	0.8774	0.9340	0.1667
	4	0.8832	0.9365	0.1620	0.8727	0.9323	0.1679
	5	0.8970	0.9446	0.1689	0.8750	0.9339	0.1700
	6	0.9371	0.9730	0.2002	0.8526	0.9195	0.1692
4	1	0.8907	0.9395	0.1576	0.8791	0.9335	0.1628
	2	0.9343	0.9737	0.1998	0.8507	0.9158	0.1649
	3	0.8903	0.9400	0.1586	0.8811	0.9346	0.1636
	4	0.8835	0.9394	0.1577	0.8758	0.9342	0.1627
	5	0.9026	0.9511	0.1608	0.8848	0.9392	0.1631
	6	0.8866	0.9391	0.1627	0.8670	0.9225	0.1638

TABLE 4. Summary statistics.

	No. of obs	Age	Education	Black	Hispanic	No degree	Married	RE74	RE 75
NSW/Lalonde									
Treated	297	24.63	10.38	.80	.09	.73	.17		3,066
		(6.69)	(1.82)	(.40)	(.29)	(.37)	(.44)		(4,874)
Control	425	24.45	10.19	.84	.11	.81	.16		3,026
		(6.59)	(1.62)	(0.40)	(0.32)	(0.36)	(0.39)		(5,201)
NSW/DW									
Treated	185	24.45	10.19	.84	.06	.71	.19	2,095	1,532
		(7.16)	(2.01)	(.36)	(.24)	(.39)	(.46)	(4,886)	(3219)
Control	260	25.05	10.09	.83	.11	.83	.15	2,107	1,267
		(7.06)	(1.61)	(.38)	(.31)	(.36)	(.37)	(5,687)	(3,102)
PSID	2,490	34.85	12.11	.25	.03	.31	.87	19,429	19,063
		(10.44)	(3.08)	(.43)	(.18)	(.46)	(.34)	(13,406)	(13,596)
CPS	15,992	33.22	12.02	.07	.07	.29	.71	14,016	13,650
		(11.05)	(2.87)	(.26)	(.25)	(.45)	(.46)	(9,570)	(9,270)
Normalized dif									
NSW Lalonde treatment/control		0.03	0.11	0.003	-0.06	-0.20	0.03		0.008
NSW Lalond/PSID		-1.17	-0.69	1.32	-1.95	0.26	0.94		-1.57
NSW Lalond/CPS		-0.94	-0.69	2.16	-1.31	0.08	0.97		-1.43
NSW DW treatment/control		0.11	0.14	0.04	-0.17	-0.30	0.09	-0.002	0.08
NSW DW/PSID		-1.01	-0.68	1.48	0.13	0.88	-1.84	-1.72	-1.77
NSW DW/CPS		-0.80	-0.68	2.43	-0.05	0.90	-1.23	-1.57	-1.75

* Earnings are expressed in 1982 dollars.

Table 5. Simple covariate matching

	NSW Lalonde sample			NSW DW sample		
	NSW	CPS	PSID	NSW	CPS	PSID
Mean difference	886.30			1794.34		
	(488.20)			(671.00)		
Covariate matching						
$K = 1$	361.80	19.80	-638.54	2046.21	2186.92	742.54
Asymptotic	(619.24)	(687.29)	(1139.77)	(739.40)	(924.67)	(1262.48)
Wild boot	(590.08)	(655.89)	(1018.55)	(762.77)	(891.21)	(1294.87)
Naive boot	(613.46)	(674.26)	(977.71)	(805.07)	(849.30)	(1055.76)
$K = 4$	881.31	-340.29	-1122.87	1908.52	1688.37	1460.62
Asymptotic	(511.36)	(560.47)	(1008.49)	(678.94)	(735.24)	(902.77)
Wild boot	(500.76)	(546.06)	(984.54)	(704.27)	(734.14)	(831.73)
Naive boot	(534.03)	(605.76)	(925.64)	(699.86)	(763.08)	(898.42)
$K = 16$	952.86	-427.62	-928.56	1959.27	1430.83	1687.08
Asymptotic	(487.19)	(508.27)	(785.43)	(664.68)	(654.88)	(846.48)
Wild boot	(502.91)	(530.95)	(777.75)	(668.78)	(657.65)	(800.84)
Naive boot	(522.84)	(535.95)	(774.82)	(712.53)	(663.70)	(790.79)

FIGURE 1. Power functions for Design 1 with Curve 1 and normal ϵ

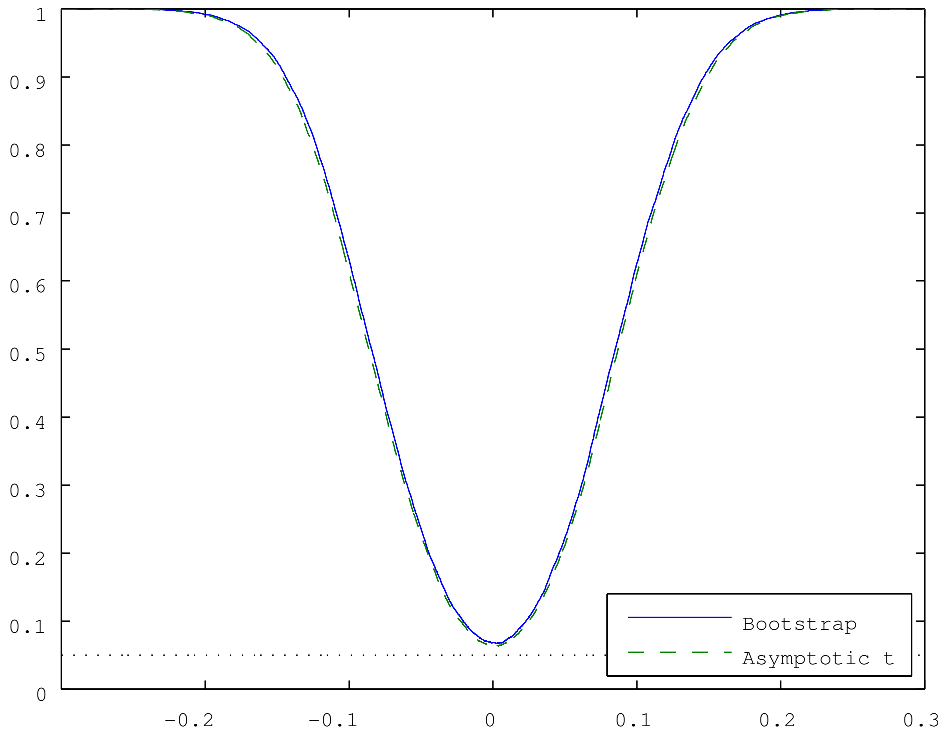
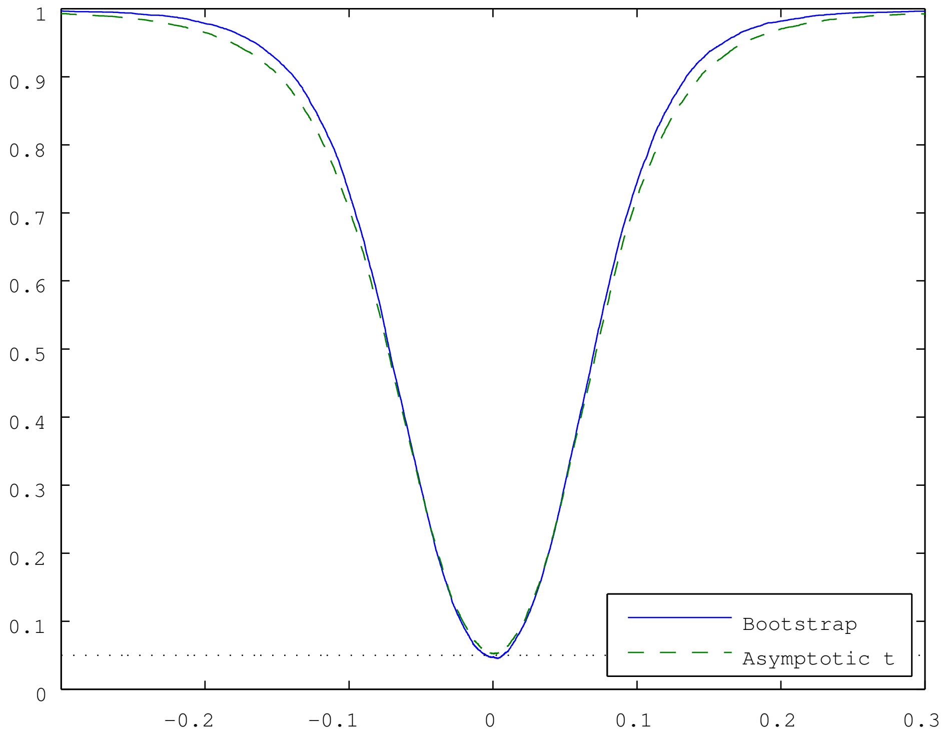


FIGURE 2. Power functions for Design 1 with Curve 1 and lognormal ϵ



References

[1] Abadie, A. and G. W. Imbens (2006) Large sample properties of matching estimators for average treatment effects, *Econometrica*, 74, 235-267.

[2] Abadie, A. and G. W. Imbens (2008) On the failure of the bootstrap for matching estimators, *Econometrica*, 76, 1537-1557.

[3] Abadie, A. and G. W. Imbens (2011) Bias-corrected matching estimators for average treatment effects, *Journal of Business & Economic Statistics*, 29, 1-11.

[4] Abadie, A. and G. W. Imbens (2012) A Martingale representation for matching estimators, *Journal of the American Statistical Association*, 107, 833-843.

[5] Busso, M., DiNardo, J. and J. McCrary (2014) New evidence on the finite sample properties of propensity score matching and reweighting estimators, *Review of Economics and Statistics*, 96, 885-897.

[6] Davidson, R. and E. Flachaire (2008) The wild bootstrap, tamed at last, *Journal of Econometrics*, 146, 162-169.

[7] 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.

[8] Dehejia, R. and S. Wahba (2002) Propensity score-matching method for nonexperimental causal studies, *Review of Economics and Statistics*, 84, 151-161.

[9] Frölich, M. (2004) Finite-sample properties of propensity-score matching and weighting estimators, *Review of Economics and Statistics*, 86, 77-90.

[10] Heckman, J. and J. Hotz (1989) Choosing among alternative nonexperimental methods for estimating the impact of social programs: the case of manpower training, *Journal of the American Statistical Association*, 84, 862-874.

[11] Heckman, J., Ichimura, H. and P. Todd (1998) Matching as an econometric evaluation estimator, *Review of Economic Studies*, 65, 261-294.

[12] Kline, P. and A. Santos (2012) Higher order properties of the wild bootstrap under misspecification, *Journal of Econometrics*, 171, 54-70.

[13] Lalonde, R. J. (1986) Evaluation the econometric evaluations of training programs with experimental data, *American Economic Review*, 76, 604-620.

[14] Hirano, K., Imbens, G. and G. Ridder (2003) Efficient estimation of average treatment effects using the estimated propensity score, *Econometrica*, 71, 1161-1189.

[15] Liu, R. Y. (1988) Bootstrap procedures under some non-I.I.D. models, *Annals of Statistics*, 16, 1696-1708.

[16] Mammen, E. (1993) Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics*, 21, 255-285.

[17] Politis, N. and J. Romano (1994) Large sample confidence regions based on subsamples under minimal assumptions, *Annals of Statistics*, 22, 2031-2050.

[18] Smith, J. and P. Todd (2005) Does matching address LaLonde's critique of nonexperimental estimators?, *Journal of Econometrics*, 125, 305-353.

[19] Stone, C. J. (1977) Consistent nonparametric regression, *Annals of Statistics*, 5, 595-620.

[20] Wu, C. F. J. (1986) Jackknife, bootstrap, and other resampling methods in regression analysis, *Annals of Statistics*, 14, 1261-1295.

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

E-mail address: `t.otsu@lse.ac.uk`

Department of Economics, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, WI 53706, USA.

E-mail address: `yrai@wisc.edu`