UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

WORKING PAPERS SES

A Wild Bootstrap Algorithm for Propensity Score Matching Estimators

Hugo Bodory,
Lorenzo Camponovo,
Martin Huber,
and Michael Lechner

N. 470
VII.2016

Faculté des sciences économiques et sociales
Wirtschafts- und sozialwissenschaftliche Fakultät

A Wild Bootstrap Algorithm for Propensity Score Matching Estimators

Hugo Bodory*, Lorenzo Camponovo*, Martin Huber+, Michael Lechner*,**

* University of St. Gallen, Dept. of Economics

+ University of Fribourg, Dept. of Economics

Abstract: We introduce a wild bootstrap algorithm for the approximation of the sampling distribution of pair or one-to-many propensity score matching estimators. Unlike the conventional iid bootstrap, the proposed wild bootstrap approach does not construct bootstrap samples by randomly resampling from the observations with uniform weights. Instead, it fixes the covariates and constructs the bootstrap approximation by perturbing the martingale representation for matching estimators. We also conduct a simulation study in which the suggested wild bootstrap performs well even when the sample size is relatively small. Finally, we provide an empirical illustration by analyzing an information intervention in rural development programs.

Keywords: Inference, Propensity Score Matching Estimators, Wild Bootstrap.
JEL classification: C14, C15, C21.

We are grateful to Alberto Abadie for his advice concerning the program code of his variance estimator. Addresses for correspondence: Hugo Bodory, University of St. Gallen, Varnbüelstrasse 14, CH-9000 St. Gallen, hugo.bodory@unisg.ch; Lorenzo Camponovo, University of St. Gallen, Bodanstrasse 6, CH-9000 St. Gallen, lorenzo.camponovo@unisg.ch; Martin Huber, University of Fribourg, Bd. de Pérolles 90, CH-1700 Fribourg, martin.huber@unifr.ch; Michael Lechner, University of St. Gallen, Varnbüelstrasse 14, CH-9000 St. Gallen, michael.lechner@unisg.ch. **: Michael Lechner is also affiliated with CEPR, London, CESIfo, Munich, IAB, Nuremberg, and IZA, Bonn.

1 Introduction

Propensity score matching estimators are widely used in empirical economics for the evaluation and estimation of average effects of binary treatments; see, e.g., Rosenbaum and Rubin (1983, 1985), Heckman, Ichimura, and Todd (1998), and Dehejia and Wahba (1999), among others. By adjusting only for the propensity score (i.e., the conditional probability of assignment to a treatment given a vector of covariates), propensity score matching estimators have the great advantage of reducing the dimensionality of matching to a single dimension. Because of this desirable feature, they may be preferred to direct matching estimators with a matching process based on the full (large) set of covariates in many settings.

Abadie and Imbens (2016) derived the asymptotic properties of propensity score matching estimators, in particular root-N consistency and their normal limit distribution with zero asymptotic bias. Despite these desirable asymptotic features, Abadie and Imbens (2008) showed that the conventional iid bootstrap does not consistently estimate the distribution of pair or one-to-many matching estimators (on the propensity score). The source of this inconsistency is related to the incapability of the iid bootstrap of correctly reproducing the distribution of the number of times each unit is used as a match.

To overcome this problem, recently Otsu and Rai (2015) introduced and proved the consistency of a wild bootstrap procedure for matching estimators in the spirit of Wu (1986), Liu (1988) and Mammen (1993). However, their approach does not apply to matching estimators with estimated propensity score. The main contribution of this paper is to extend the definition of the wild bootstrap algorithm also in this latter setting, which is typically the standard case in applied work. In line with Otsu and Rai (2015), the definition of our wild bootstrap procedure relies on the martingale representation for matching estimators suggested in Abadie and Imbens (2012). Unlike the conventional iid bootstrap, the wild bootstrap does not construct bootstrap samples by randomly resampling from the observations with uniform weights. Instead, it fixes the covariates and constructs the bootstrap approximation by perturbing the martingale representation for matching estimators.

We investigate the finite sample behavior of the proposed wild bootstrap algorithm in simulation settings previously considered in the supplementary materials in Abadie and Imbens (2016). Inference based on our bootstrap algorithm seems to outperform inference based on the limit distributions derived in Abadie and Imbens (2016), in particular when the sample size is relatively small. The empirical coverage implied by the wild bootstrap tends to be closer to the nominal coverage rate than that based on the asymptotic approximation. Similar empirical findings are also confirmed in Bodory, Camponovo, Huber and Lechner (2016).

Finally, we provide an empirical illustration using survey data from rural (FYR) Macedonia that were previously analyzed in Huber, Kotevska, Martinovska-Stojcheska, and Solovyeva (2016) to evaluate the effects of an information brochure about public rural development programs. The sample size is $N = 256$. Even though the p-values implied by the wild bootstrap and the limit distribution differ somewhat, both approaches suggest that the information intervention reinforced the sentiment that such programs increase the administrative burden of the targeted households. However, it is interesting to highlight that in line with the Monte Carlo results, confidence intervals constructed using the limit distribution tend to be shorter than those constructed using the wild bootstrap approach. This result may indicate an undercoverage of asymptotic confidence intervals when the sample size is relatively small.

The remainder of the paper is organized as follows. In Section 2, we present the model and notation. In Section 3, we introduce the wild bootstrap algorithm. In Section 4, we study the finite sample properties of our approach in Monte Carlo simulations. In Section 5, we present the real data application. Section 6 concludes.

2 Model and Notation

We consider a similar setting and notation as introduced in Abadie and Imbens (2006, 2016). For each of units $i = 1, \ldots, N$ (with N denoting the sample size), let $Y_i(1)$ and $Y_i(0)$ denote the two potential outcomes when receiving a (binary) treatment or not, respectively. The variable $W_i \in \{0, 1\}$ indicates the treatment status. For each unit i, we observe the outcome Y_i under treatment W_i only,

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

and a vector of pretreatment covariates denoted by X_i. The parameters of interest are the population average treatment effect (ATE), denoted by τ and defined as

$$\tau = E[Y_i(1) - Y_i(0)],$$

and the population average treatment effect on the treated (ATET), denoted by τ_t and defined as

$$\tau_t = E[Y_i(1) - Y_i(0)|W_i = 1].$$

We consider propensity score matching estimators for τ and τ_t, where the propensity score is defined as $p(X_i) = P(W_i = 1|X_i)$. Following Rosenbaum and Rubin (1983), we consider a generalized linear specification for the propensity score, $p(X_i) = F(X_i'\theta)$, where F is a known

function (usually specified as Logit or Probit), while θ is an unknown parameter. Let $\hat{\theta}_N$ be the maximum likelihood estimator of θ defined as

$$\hat{\theta}_N = \arg\max_{\theta} \sum_{i=1}^{N} W_i \ln F(X_i'\theta) + (1 - W_i) \ln(1 - F(X_i'\theta)). \tag{1}$$

Furthermore, let M be the number of matches per unit (e.g. $M = 1$ for pair matching), and let $\mathcal{J}_M(i, \hat{\theta}_N)$ be the set of matches for unit i, defined as

$$\mathcal{J}_M(i, \hat{\theta}_N) = \left\{ j = 1, \ldots, N : W_j = 1 - W_i, \left(\sum_{k: W_k = 1 - W_i} \mathbb{I}_{\{|F(X_i'\hat{\theta}_N) - F(X_k'\hat{\theta}_N)| \leq |F(X_i'\hat{\theta}_N) - F(X_j'\hat{\theta}_N)|\}} \right) \leq M \right\},$$

where $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function. Then, the propensity score matching estimators for τ and τ_t are defined as

$$\hat{\tau}_N = \frac{1}{N} \sum_{i=1}^{N} (2W_i - 1) \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i, \hat{\theta}_N)} Y_j \right), \tag{2}$$

$$\hat{\tau}_{t,N} = \frac{1}{N_1} \sum_{i=1}^{N} W_i \left(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i, \hat{\theta}_N)} Y_j \right), \tag{3}$$

respectively, where $N_1 = \sum_{i=1}^{N} W_i$ is the number of treated units in the sample.

Under some regularity conditions, Abadie and Imbens (2016) showed root-N consistency and asymptotic normality of the proposed matching estimators. Furthermore, they also proposed estimators of the asymptotic variances. Therefore, these results allow constructing confidence intervals or testing hypotheses on the parameters of interest based on asymptotic approximations. As an alternative for conducting inference, the next section introduces a wild bootstrap algorithm. As highlighted in the Monte Carlo analysis, the wild bootstrap approach seems to outperform approximations based on the limit distribution, in particular when the sample size is relatively small.

3 Wild Bootstrap Algorithm

For the setting introduced in Section 2, the conventional iid bootstrap constructs random samples $(Z_1^*, \ldots, Z_N^*)$ by resampling from the observations $(Z_1, \ldots, Z_N)$ with uniform weights $1/N$, where $Z_i = (Y_i, W_i, X_i')$. Unfortunately and as demonstrated in Abadie and Imbens (2008), this approach does not provide a valid method for approximating the distribution of (propensity score) matching estimators. In particular, the source of this inconsistency is related to the incapability of the iid bootstrap of correctly reproducing the distribution of the number of times each unit is used as a match.

To overcome this problem, recently Otsu and Rai (2015) introduced and proved the consistency of a wild bootstrap procedure for matching estimators. However, their approach does not apply to matching estimators with estimated propensity score. In this section, we fill this gap by extending the definition of the wild bootstrap algorithm for the settings presented in Section 2. More precisely, in Sections 3.1 and 3.2 we present wild bootstrap methods for ATE and ATET, respectively. Finally, in Section 3.3 we discuss some theoretical and computational aspects of our procedures.

3.1 Wild Bootstrap for ATE

The definition of the wild bootstrap approach relies on the martingale representation for matching estimators suggested in Abadie and Imbens (2012, 2016). In particular, note that as shown in Abadie and Imbens (2016), we can decompose the matching estimator as $\sqrt{N}(\hat{\tau}_N - \tau) = R_{1N} + R_{2N} + o_p(1)$, where

$$\begin{aligned} R_{1N} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mu(1, F(X_i'\hat{\theta}_N)) - \mu(0, F(X_i'\hat{\theta}_N)) - \tau), \\ R_{2N} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (2W_i - 1) \left(1 + \frac{K_M(i)}{M}\right) \left(Y_i - \mu(W_i, F(X_i'\hat{\theta}_N))\right), \end{aligned}$$

$\mu(w,p) = E[Y|W = w, p(X) = p]$, and $K_M(i)$ is the number of times that observations i is used as a match

$$K_M(i) = \sum_{j=1}^{N} \mathbb{I}_{\{i \in \mathcal{J}_M(j, \hat{\theta}_N)\}}.$$

In line with the wild bootstrap procedure proposed in Otsu and Rai (2015), we can use this representation to reproduce the sampling distribution of $\sqrt{N}(\hat{\tau}_N - \tau)$. However, the definition of wild bootstrap procedures for matching estimators in the setting presented in Section 2 requires some care. Indeed, in this case the wild bootstrap algorithm needs also to capture the variability implied by the estimation of the propensity score.

To overcome this problem, we propose following approach. We generate random bootstrap treatments $(W_1^*, \ldots, W_N^*)$ according to

$$W_i^* = \begin{cases} 0, & \text{with prob. } 1 - F(X_i'\hat{\theta}_N), \\ 1, & \text{with prob. } F(X_i'\hat{\theta}_N). \end{cases} \tag{4}$$

Let $\hat{\theta}_N^*$ be the bootstrap maximum likelihood estimator of θ defined as the solution of (1) by

replacing $(W_1, \ldots, W_N)$ with the bootstrap treatments $(W_1^*, \ldots, W_N^*)$. Then, we compute

$$\begin{aligned}
\mathcal{J}_M^*(i, \hat{\theta}_N^*) &= \left\{ j = 1, \ldots, n : W_j^* = 1 - W_i^*, \left(\sum_{k: W_k^* = 1 - W_i^*} \mathbb{I}_{\{|F(X_i'\hat{\theta}_N^*) - F(X_k'\hat{\theta}_N^*)| \leq |F(X_i'\hat{\theta}_N^*) - F(X_j'\hat{\theta}_N^*)|\}} \right) \leq M \right\}, \\
K_M^*(i) &= \sum_{j=1}^{N} \mathbb{I}_{\{i \in \mathcal{J}_M^*(j, \hat{\theta}_N^*)\}}.
\end{aligned}$$

$\mathcal{J}_M^*(i, \hat{\theta}_N^*)$ is the set of matches for unit i among the units with a different value of treatment, while $K_M^*(i)$ denotes the number of times unit i is used as a match given that M matches per unit are used, based on the bootstrap treatments. By adopting this approach we are able to reproduce the variability implied in the estimation of the parameter θ for the computation of the estimated propensity score. Furthermore, let $\hat{\mu}(0, p)$ and $\hat{\mu}(1, p)$ be some kernel estimators of $\mu(0, p)$ and $\mu(1, p)$, respectively, and let $\hat{\epsilon}_i^* = Y_i - \hat{\mu}(W_i, F(X_i'\hat{\theta}_N^*))$. We introduce the error terms $\hat{\epsilon}_{i,0}^*$ and $\hat{\epsilon}_{i,1}^*$ defined as

$$\hat{\epsilon}_{i,0}^* = \begin{cases} \hat{\epsilon}_i^*, & \text{if } W_i = 0, \\ \hat{\epsilon}_{\mathcal{J}_1(i, \hat{\theta}_N)}^*, & \text{if } W_i = 1, \end{cases} \tag{5}$$

and

$$\hat{\epsilon}_{i,1}^* = \begin{cases} \hat{\epsilon}_{\mathcal{J}_1(i, \hat{\theta}_N)}^*, & \text{if } W_i = 0, \\ \hat{\epsilon}_i^*, & \text{if } W_i = 1. \end{cases} \tag{6}$$

In Equations (5) and (6), we construct bootstrap error terms for different treatments by matching on the fitted residuals $(\hat{\epsilon}_1^*, \ldots, \hat{\epsilon}_N^*)$. Finally, we approximate the term $R_{1N} + R_{2N}$ with the bootstrap decomposition $R_{1N}^* + R_{2N}^*$, where

$$R_{1N}^* = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\hat{\mu}(1, F(X_i'\hat{\theta}_N^*)) - \hat{\mu}(0, F(X_i'\hat{\theta}_N^*)) - \hat{\tau}_N) u_i, \tag{7}$$

$$R_{2N}^* = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (2W_i^* - 1) \left(1 + \frac{K_M^*(i)}{M} \right) \hat{\epsilon}_{i, W_i^*}^* u_i, \tag{8}$$

$(u_1, \ldots, u_n)$ are iid random variables with

$$E[u_i | \mathcal{Z}] = 0, \qquad E[u_i^2 | \mathcal{Z}] = 1, \qquad E[u_i^4 | \mathcal{Z}] < \infty, \tag{9}$$

and $\mathcal{Z} = (Z_1, \ldots, Z_n)$. Next, we formally introduce the wild bootstrap algorithm.

Algorithm 1. Wild Bootstrap Algorithm for ATE.

(ATE-1) Generate the bootstrap treatments $(W_1^*, \ldots, W_N^*)$ according to (4).

(ATE-2) Let $\hat{\theta}_N^*$ be the bootstrap maximum likelihood estimator of θ defined as the solution of (1) by replacing $(W_1, \ldots, W_N)$ with the bootstrap treatments $(W_1^*, \ldots, W_N^*)$. Then, compute $\mathcal{J}_M^*(i, \hat{\theta}_N^*)$ and $K_M^*(i) = \sum_{j=1}^N \mathbb{I}_{\{i \in \mathcal{J}_M^*(j, \hat{\theta}_N^*)\}}$.

(ATE-3) Let $\hat{\mu}(0,p)$ and $\hat{\mu}(1,p)$ be some kernel estimators of $\mu(0,p)$ and $\mu(1,p)$, respectively, and let $\hat{\epsilon}_i^* = Y_i - \hat{\mu}(W_i, F(X_i'\hat{\theta}_N^*))$. For $i = 1, \ldots, N$, compute the error terms $\hat{\epsilon}_{i,0}^*$ and $\hat{\epsilon}_{i,1}^*$ defined in (5) and (6), respectively.

(ATE-4) Generate a sequence of iid random variables $(u_1, \ldots, u_n)$ satisfying (9). Furthermore, compute $R_{1N}^* + R_{2N}^*$ defined in (7) and (8).

(ATE-5) Repeat steps (ATE-1)-(ATE-4) many times. The empirical distribution of $R_{1N}^* + R_{2N}^*$ approximates the sampling distribution of $\sqrt{N}(\hat{\tau}_N - \tau)$.

In Step (ATE-1), we generate a sequence of random bootstrap treatments $(W_1^*, \ldots, W_N^*)$ based on the estimated propensity score. Using this sequence, in Step (ATE-2) we compute $\hat{\theta}_N^*$, $\mathcal{J}_M^*(i, \hat{\theta}_N^*)$, and $K_M^*(i)$. In Step (ATE-3) we introduce the error terms $\hat{\epsilon}_{i,0}^*$ and $\hat{\epsilon}_{i,1}^*$ for the different treatment states. Finally, in Step (ATE-4) we compute the wild bootstrap statistics $R_{1N}^* + R_{2N}^*$. The empirical distribution of $R_{1N}^* + R_{2N}^*$ obtained by repeating Steps (ATE-1)-(ATE-4) many times approximates the sampling distribution of $\sqrt{N}(\hat{\tau}_N - \tau)$.

3.2 Wild Bootstrap for ATET

In this section, we extend the definition of the wild bootstrap for the ATET. To this end, note that as shown in Abadie and Imbens (2016), we can write the matching estimator as $\sqrt{N}(\hat{\tau}_{t,N} - \tau_t) = S_{1N} + S_{2N} + o_p(1)$, where

$$\begin{aligned} S_{1N} &= \frac{\sqrt{N}}{N_1} \sum_{i=1}^N W_i(\mu(1, F(X_i'\hat{\theta}_N)) - \mu(0, F(X_i'\hat{\theta}_N)) - \tau_t), \\ S_{2N} &= \frac{\sqrt{N}}{N_1} \sum_{i=1}^N \left(W_i - (1 - W_i)\frac{K_M(i)}{M} \right) \left(Y_i - \mu(W_i, F(X_i'\hat{\theta}_N)) \right). \end{aligned}$$

In line with the wild bootstrap procedure proposed in Otsu and Rai (2015), we can use this representation to reproduce the sampling distribution of $\sqrt{N}(\hat{\tau}_{t,N} - \tau_t)$. However, also in this case the definition of the wild bootstrap algorithm has to capture the variability implied by the estimation of the propensity score. To this end, consider the bootstrap treatments $(W_1^*, \ldots, W_N^*)$ generated according to (4), and let $\hat{\theta}_N^*$ be the bootstrap maximum likelihood estimator of θ defined as the solution of (1) by replacing $(W_1, \ldots, W_N)$ with the bootstrap treatments $(W_1^*, \ldots, W_N^*)$. Furthermore, let $K_M^*(i) = \sum_{j=1}^N \mathbb{I}_{\{i \in \mathcal{J}_M^*(j, \hat{\theta}_N^*)\}}$, and compute the error

terms $\hat{\epsilon}^*_{i,0}$ and $\hat{\epsilon}^*_{i,1}$ defined in (5) and (6), respectively. Then, we approximate the term $S_{1N}+S_{2N}$ with the bootstrap decomposition $S^*_{1N}+S^*_{2N}$, where

$$S^*_{1N} = \frac{\sqrt{N}}{N^*_1}\sum_{i=1}^{N} W^*_i(\mu(1,F(X_i'\hat{\theta}^*_N))-\mu(0,F(X_i'\hat{\theta}^*_N))-\tau_t)u_i, \tag{10}$$

$$S^*_{2N} = \frac{\sqrt{N}}{N^*_1}\sum_{i=1}^{N}\left(W^*_i-(1-W^*_i)\frac{K^*_M(i)}{M}\right)\hat{\epsilon}^*_{i,W^*_i}u_i. \tag{11}$$

$N^*_1=\sum_{i=1}^N W^*_i$, and $(u_1,\ldots,u_n)$ are iid random variables satisfying (9). Next, we formally introduce the wild bootstrap algorithm.

Algorithm 2. Wild Bootstrap Algorithm for ATET.

(ATET-1) Generate the bootstrap treatments $(W^*_1,\ldots,W^*_N)$ according to (4).

(ATET-2) Let $\hat{\theta}^*_N$ be the bootstrap maximum likelihood estimator of θ defined as the solution of (1) by replacing $(W_1,\ldots,W_N)$ with the bootstrap treatments $(W^*_1,\ldots,W^*_N)$. Then, compute $\mathcal{J}^*_M(i,\hat{\theta}^*_N)$ and $K^*_M(i)=\sum_{j=1}^N \mathbb{I}_{\{i\in\mathcal{J}^*_M(j,\hat{\theta}^*_N)\}}$.

(ATET-3) Let $\hat{\mu}(0,p)$ and $\hat{\mu}(1,p)$ be some kernel estimators of $\mu(0,p)$ and $\mu(1,p)$, respectively, and let $\hat{\epsilon}^*_i=Y_i-\hat{\mu}(W_i,F(X_i'\hat{\theta}^*_N))$. For $i=1,\ldots,N$, compute the error terms $\hat{\epsilon}^*_{i,0}$ and $\hat{\epsilon}^*_{i,1}$ defined in (5) and (6), respectively.

(ATET-4) Generate a sequence of iid random variables $(u_1,\ldots,u_n)$ satisfying (9). Furthermore, compute $S^*_{1N}+S^*_{2N}$ defined in (10) and (11).

(ATET-5) Repeat steps (ATET-1)-(ATET-4) many times. The empirical distribution of $S^*_{1N}+S^*_{2N}$ approximates the sampling distribution of $\sqrt{N}(\hat{\tau}_{t,N}-\tau_t)$.

In Step (ATET-1), we generate a sequence of random bootstrap treatments $(W^*_1,\ldots,W^*_N)$ based on the estimated propensity score. Using this sequence, in Step (ATET-2) we compute $\hat{\theta}^*_N$, $\mathcal{J}^*_M(i,\hat{\theta}^*_N)$, and $K^*_M(i)$. In Step (ATET-3) we introduce the error terms $\hat{\epsilon}^*_{i,0}$ and $\hat{\epsilon}^*_{i,1}$ for the different treatment states. Finally, in Step (ATET-4) we compute the wild bootstrap statistics $S^*_{1N}+S^*_{2N}$. The empirical distribution of $S^*_{1N}+S^*_{2N}$ obtained by repeating Steps (ATET-1)-(ATET-4) many times approximates the sampling distribution of $\sqrt{N}(\hat{\tau}_{t,N}-\tau_t)$.

3.3 Theoretical and Computational Aspects

We conclude this section by discussing some theoretical and computational aspects of our procedures. First, suppose that the true propensity score $p(X_i)=P(W_i=1|X_i)$ is known and

does not depend on some unknown parameters. Then, in the definition of our wild bootstrap approach it is not necessary to generate bootstrap treatments $(W_1^*, \ldots, W_N^*)$ and re-estimate the parameter θ. More precisely, in Algorithms 1 and 2, for $i = 1, \ldots, N$ we can simply replace W_i^* and $F(X_i'\hat{\theta}_N^*)$ with W_i and $p(X_i)$, respectively. Note that in this case our approach is asymptotically equivalent to the wild bootstrap procedure proposed in Otsu and Rai (2015). On the other hand, when the propensity score is unknown and has to be estimated, the wild bootstrap algorithm has to be properly modify. Indeed, as shown in Abadie and Imbens (2016), matching estimators based on known or estimated propensity score have different limit distributions. Therefore, the wild bootstrap algorithm has to be adapted in order to capture the variability implied by the estimation of the propensity score.

The implementation of our wild bootstrap approach requires the computation of the maximum likelihood estimator $\hat{\theta}_N^*$ in each bootstrap sample. The computational costs of our procedures can be substantially reduced by adopting the fast resampling approach introduced in Davidson and MacKinnon (1999) and Andrews (2002). More precisely, instead of solving equation (1) for each bootstrap sample, we can replace $\hat{\theta}_N^*$ with the first-step Newton-Raphson estimator $\hat{\theta}_{N,1}^*$ defined as

$$\hat{\theta}_{N,1}^* = \hat{\theta}_N - \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'} \rho(W_i^*, X_i, \hat{\theta}_N) \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial\theta} \rho(W_i^*, X_i, \hat{\theta}_N) \right),$$

where $\rho(W_i^*, X_i, \theta) = W_i^* \ln F(X_i'\theta) + (1 - W_i^*) \ln(1 - F(X_i'\theta))$.

Finally, the implementation of the wild bootstrap algorithms also requires the selection of kernel estimators $\hat{\mu}(0, p)$ and $\hat{\mu}(1, p)$ for $\mu(0, p)$ and $\mu(1, p)$, respectively, with appropriate convergence rates; see, e.g., Abadie and Imbens (2011) and Otsu and Rai (2015). In the Monte Carlo analysis presented in the next section, we consider the nonparametric estimators suggested in Adadie and Imbens (2011).

4 Monte Carlo Simulations

In this section, we investigate the finite sample properties of the wild bootstrap procedures by Monte Carlo simulations. We consider a setting also analyzed in the supplementary material in Abadie and Imbens (2016). For $i = 1, \ldots, N$, let $X_i = (X_{i1}, X_{i2})'$ be a vector of covariates, where X_{i1} and X_{i2} are uniformly distributed random variables with support $[-0.5, 0.5]$, and independent of each other. The two potential outcomes are generated by $Y_i(0) = 3X_{i1} - 3X_{i2} + U_{i0}$, and $Y_i(1) = 5 + 5X_{i1} + X_{i2} + U_{i1}$, where U_{i0} and U_{i1} are independent standard normal random

Table 1: Empirical Coverage and Length of Confidence Intervals for the ATE

$N = 100$	asymptotic theory	wild bootstrap
	0.929 (1.199)	0.941 (1.254)
$N = 200$	asymptotic theory	wild bootstrap
	0.936 (0.867)	0.943 (0.932)
$N = 400$	asymptotic theory	wild bootstrap
	0.941 (0.619)	0.944 (0.650)

Note: We report the empirical coverage and the length of confidence intervals (in brackets) based on asymptotic theory (second column), and the wild bootstrap algorithm (third column). Pair matching is on the estimated propensity score. The sample sizes are $N = 100, 200, 400$, and the nominal coverage probability is 0.95.

variables. Moreover, we assume a logistic propensity score

$$P(W_i = 1|X_i) = F(X_i'\theta) = \frac{\exp(\theta_1 X_{i1} + \theta_2 X_{i2})}{1 + \exp(\theta_1 X_{i1} + \theta_2 X_{i2})},$$

where $\theta = (\theta_1, \theta_2)'$, with $\theta_1 = 1$ and $\theta_2 = 2$. We generate 5000 Monte Carlo samples of sizes $N = 100, 200$, and 400 according to these parameter selections. Using the wild bootstrap approach defined in Algorithms 1 and 2, we construct 0.95-confidence intervals for ATE and ATET. For simplicity we use $M = 1$ matches, and $B = 299$ bootstrap replications. For comparison, we also construct confidence intervals based on the limit distributions derived in Theorems 1 and 2 in Abadie and Imbens (2016), with asymptotic variances estimated as outlined in Section 4 in Abadie and Imbens (2016). Tables 1 and 2 report the empirical findings for ATE and ATET, respectively.

The results in Tables 1 and 2 are qualitatively very similar and highlight the accuracy of the wild bootstrap approximation, whose empirical coverage is always closer to the nominal coverage rate than that based on the asymptotic approximation. For instance, in Table 2 for $N = 100$, the empirical coverage rate using the wild bootstrap is 0.94. In the same setting, the rate of the asymptotic approximation is instead 0.921. Even though the accuracy of the approximations based on the limit distribution increases in the sample size N (as expected), it is in all scenarios dominated by the wild bootstrap.

Table 2: Empirical Coverage and Length of Confidence Intervals for the ATET

$N = 100$	asymptotic theory	wild bootstrap
	0.921 (1.469)	0.940 (1.513)
$N = 200$	asymptotic theory	wild bootstrap
	0.935 (1.081)	0.944 (1.114)
$N = 400$	asymptotic theory	wild bootstrap
	0.938 (0.774)	0.946 (0.782)

Note: We report the empirical coverage and the length of confidence intervals (in brackets) based on asymptotic theory (second column), and the wild bootstrap algorithm (third column). Pair matching is on the estimated propensity score. The sample sizes are $N = 100, 200, 400$, and the nominal coverage probability is 0.95.

5 Empirical Application

We apply our method to survey data from rural (FYR) Macedonia (spring 2015) that were considered by Huber, Kotevska, Martinovska-Stojcheska, and Solovyeva (2016) to investigate the impact of an information brochure about public rural development programs on the farmers' opinion on/intention to make use of the programs. The information intervention had initially been planned as experiment, however, randomization was not properly conducted. Therefore, Huber, Kotevska, Martinovska-Stojcheska, and Solovyeva (2016) analyze the effectiveness of the brochure based on OLS and reweighting and control for a range of observed covariates, for instance farmer's age, gender, education, household size, farm size, share of income from farming, farm profitability, and others.

We reconsider their main evaluation sample for propensity score matching and our wild bootstrap procedure, with the exception that we drop one observation with missing values in the outcomes, which ultimately entails 256 observations. Table 3 provides the means and standard deviations by treatment states of the control variables (among them several dummies for missing covariates) and two outcome variables. The latter consist of the (self-assessed) intention of the interviewee to make use of public rural development programs in the next 3-5 years and a judgement on whether such programs increase the administrative burden for households (which might be a reason for non-application).

Table 3: Descriptive Statistics Across Treatment States

Variables	Treated		Non-treated		St.diff.	Probit (%)	
	mean	std.	mean	std.	(%)	m.eff.	s.e.
Outcomes							
Intention to use programs in next 3-5 years+	2.49	0.78	2.57	0.76	-7		
Programs increase administrative burden for households++	3.29	0.48	3.10	0.52	27		
Covariates							
Age	43.90	7.32	45.76	7.48	-18	-0.74	0.55
Male (binary)	0.78	-	0.72	-	9	4.68	6.74
Education: primary (binary)	0.04	-	0.14	-	-25	-10.37	28.15
Education: high school (binary)	0.73	-	0.68	-	8	14.76	26.97
Education: college/university (binary)	0.13	-	0.13	-	0	9.35	25.24
Frequency of cooperation*	3.77	1.52	3.58	1.53	9	-2.02	2.30
Household size	4.17	1.00	4.04	1.37	8	1.86	2.69
Household head's occupation: agriculture (binary)	0.50	-	0.54	-	-6	-0.05	7.08
Years in farming	21.78	7.85	22.48	9.45	-6	0.19	0.53
Share of agricultural production sold on a market	86.44	17.68	87.87	15.61	-6	-0.09	0.29
Share of income from farming	50.32	23.44	53.63	22.59	-10	-0.03	0.15
Profitable farm**	3.63	0.51	3.42	0.64	25	14.34	5.26
Subsidy dependence***	2.02	0.78	2.17	0.84	-13	-2.22	3.96
Capacity: Farmed area (ha)	1.60	1.08	1.71	1.13	-7	-3.33	3.11
Capacity: total livestock (number of heads)	1.07	2.75	1.20	2.80	-3	-0.90	1.10
Education missing (binary)	0.10	-	0.05	-	14	19.14	23.41
Frequency of cooperation missing (binary)	0	-	0.01	-	-10	-60.31	4.89
Household head's occupation missing (binary)	0.01	-	0.02	-	-4	-10.53	23.75
Share of agricultural production sold missing (binary)	0.03	-	0	-	16	17.30	17.23
Share of income from farming missing (binary)	0.01	-	0	-	11	28.00	9.95
Subsidy dependence missing (binary)	0.01	-	0	-	8	22.24	8.94
Linear index of propensity score model: $X\hat{\theta}$	0.39	0.35	0.10	0.45	51		
Propensity score: $\Phi(X\hat{\theta})$	0.64	0.12	0.54	0.16	51		
Number of observations, Pseudo-R2 (%)	156		100			11.49	

Note: +scale: 1=very low,..., 5= very strong. ++scale: 1=strongly disagree,..., 5=strongly agree. *scale: 1=never, ..., 5=always. **scale: 1=very unprofitable,..., 5=very profitable. ***scale: 1=not dependent, 2=slightly dependent, 3=very dependent.

St.diff. (standardized difference) is defined as the difference of means normalized by the square root of the sum of estimated variances of the particular variables in both subsamples. Mean, std., s.e. stand for mean, standard deviation, and standard error, respectively. Standard deviations for binary variables (not informative) are not reported. M.eff.: Marginal effects evaluated at the mean in the probit model for treatment selection based on discrete changes for binary variables and derivatives otherwise. $\hat{\theta}$ denotes the estimated probit coefficients and $\Phi(X\hat{\theta})$ is the c.d.f. of the standard normal distribution evaluated at $X\hat{\theta}$. Pseudo-R^2 is the so-called Efron's R^2 $\left\{1 - \frac{\sum_{i=1}^{n}[D_i - \Phi(X_i\tilde{\theta})]^2}{\sum_{i=1}^{n}[D_i - n^{-1}\sum_{i=1}^{n} D_i]^2}\right\}$.

Table 4 reports the ATET estimates on the outcomes using pair matching on the propensity score, which is estimated by a probit model. It further displays standard errors based on the asymptotic variance estimator of Abadie and Imbens (2016), and the wild bootstrap presented in Algorithm 2 with $B = 4999$ bootstrap replications. For simplicity, we report bootstrap standard errors instead of constructing confidence intervals using the bootstrap empirical distribution. However, in both cases the empirical findings are qualitatively very similar.

We find no statistically significant effect at the 10% level on the intention to make use of the programs, but a positive impact on the view that such programs increase the administrative burden of the targeted households. The latter effect is statistically significant at the 10% level when using the wild bootstrap-based standard errors, and at the 5% when using the analytical standard errors based on the limit distribution. It is interesting to highlight that in line with the Monte Carlo results, confidence intervals constructed using the limit distribution tend to be shorter that those constructed using the wild bootstrap approach. This result may indicate an undercoverage of asymptotic confidence intervals when the sample size is relatively small.

Table 4: ATET Estimates and Standard Errors

Outcomes	ATET	Standard errors	
		asymptotic	wild boot
Intention to use programs in next 3-5 years	-0.18	0.12	0.17
Programs increase administrative burden for households	0.14	0.05	0.08

Note: 'asymptotic' denotes the asymptotic standard errors based on Theorem 2 of Abadie and Imbens (2016). 'wild boot' refers to the standard errors computed using the wild bootstrap approach presented in Algorithm 2.

6 Conclusion

In this paper, we propose a novel wild bootstrap algorithm for approximating the sampling distribution of propensity score matching estimators. Unlike the conventional iid bootstrap, our approach does not resample from observations with uniform weights, but fixes the covariates and constructs the bootstrap approximation by perturbing the martingale representation for matching estimators. We investigate the finite sample performance of our method in a simulation study. Inference based on the wild bootstrap seems to outperform inference based on the limit distribution in particular when the sample size is relatively small. As an empirical illustration, we provide an application to an information campaign on public rural development policies.

References

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

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

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

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

[5] Abadie, A., and Imbens, G.W. (2016). Matching on the estimated propensity score, *Econometrica*, **84**, 781-807.

[6] Andrews, D.W.K., (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators, *Econometrica*, **70**, 119-162.

[7] Bodory, H., Camponovo, L., Huber, M., and Lechner, M. (2016). The finite sample performance of inference methods for propensity score matching and weighting estimators, working paper.

[8] Davidson, R., and MacKinnon, J.G. (1999). Bootstrap testing in nonlinear models, *International Economic Review*, **40**, 487-508.

[9] Dehejia, R. H., and Wahba, S. (1999). Causal effects in non-experimental studies: reevaluating the evaluation of training programmes, *Journal of American Statistical Association*, **94**, 1053-1062.

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

[11] Huber, M., Kotevska, A., Martinovska-Stojcheska, A., and Solovyeva, A. (2016). Evaluating an information campaign about rural development policies in (FYR) Macedonia, working paper.

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

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

[14] Otsu, T., and Rai, Y. (2015). Bootstrap inference of matching estimators for average treatment effects, working paper.

[15] Rosenbaum, P. R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, **70**, 41-55.

[16] Rosenbaum, P. R., and Rubin, D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *The American Statistician*, **39**, 33-38.

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

Working Paper 470 july 2016

Authors

Hugo BODORY
University of St. Gallen, Department of Economics, Swiss Institute for Empirical Economic Research, Varnbüelstrasse 14, 9000 St. Gallen, Switzerland. Phone: +41 71 224 2767; Email: hugo.bodory@unisg.ch; Website: www.sew.unisg.ch/de/ueber_uns/

Lorenzo CAMPONOVO
University of St. Gallen, Department of Economics, Maths and Statistics, Varnbüelstrasse 14, 9000 St. Gallen, Switzerland. Phone: +41 71 224 2432; Email: lorenzo.camponovo@unisg.ch; Website: www.mathstat.unisg.ch/en/people

Martin HUBER
University of Fribourg, Faculty of Economics and Social Sciences, Chair of Applied Econometrics - Evaluation of Public Policies, Bd. de Pérolles 90, 1700 Fribourg, Switzerland. Phone: +41 26 300 8274; Email: martin.huber@unifr.ch; Website: http://www.unifr.ch/appecon/en/team/martin-huber

Michael LECHNER
University of St. Gallen, Department of Economics, Swiss Institute for Empirical Economic Research, Varnbüelstrasse 14, 9000 St. Gallen, Switzerland. Phone: +41 71 224 2320; Email: michael.lechner@unisg.ch; Website: www.michael-lechner.eu

Abstract

We introduce a wild bootstrap algorithm for the approximation of the sampling distribution of pair or one-to-many propensity score matching estimators. Unlike the conventional iid bootstrap, the proposed wild bootstrap approach does not construct bootstrap samples by randomly resampling from the observations with uniform weights. Instead, it fixes the covariates and constructs the bootstrap approximation by perturbing the martingale representation for matching estimators. We also conduct a simulation study in which the suggested wild bootstrap performs well even when the sample size is relatively small. Finally, we provide an empirical illustration by analyzing an information intervention in rural development programs.

Citation proposal

Hugo Bodory, Lorenzo Camponovo, Martin Huber, Michael Lechner. 2016. «A Wild Bootstrap Algorithm for Propensity Score Matching Estimators». Working Papers SES 470, Faculty of Economics and Social Sciences, University of Fribourg (Switzerland)

Jel Classification

C14, C15, C21

Keywords

Inference, Propensity Score Matching Estimators, Wild Bootstrap

Working Papers SES collection

Last published

464 Grossmann V., Strulik H.: Optimal Social Insurance and Health Inequality; 2015
465 Baguet M., Dumas C.: Birth weight and long-term outcomes in a developing country; 2015
466 Bodory H., Camponovo L., Huber M., Lechner M.: The finite sample performance of inference methods for propensity score matching and weighting estimators; 2016
467 Denisova-Schmidt E., Huber M., Leontyeva E.: On the Development of Students' Attitudes towards Corruption and Cheating in Russian Universities; 2016
468 Imhof D., Karagök Y., Rutz S.: Screening for Bid-rigging – Does it Work?; 2016
469 Huber M., Kotevska A., Martinovska Stojcheska A., Solovyeva A.: Evaluating an information campaign about rural development policies in (FYR) Macedonia; 2016

Catalogue and download links

http://www.unifr.ch/ses/wp
http://doc.rero.ch/collection/WORKING_PAPERS_SES

Publisher

Université de Fribourg, Suisse, Faculté des sciences économiques et sociales
Universität Freiburg, Schweiz, Wirtschafts- und sozialwissenschaftliche Fakultät
University of Fribourg, Switzerland, Faculty of Economics and Social Sciences

Bd de Pérolles 90, CH-1700 Fribourg
Tél.: +41 (0) 26 300 82 00
decanat-ses@unifr.ch www.unifr.ch/ses