*Journal of Econometrics and Statistics*
Vol. 3, Issue 2, 2023, pp. 129-139
https://doi.org/10.47509/JES.2023.v03i02.02

# In-Person vs Online Teaching: Empirical Analysis based on Bootstrap for Matching Estimators

Lorenzo Camponovo, Emanuele Delucchi, Matteo Garzoni,
Slobodan Krstic, Stefano Scaravaggi, Oliver Villa *[a]

[a] *University of Applied Sciences and Arts of Southern Switzerland*

In this note, we statistically evaluate the performance of students participating in different in-person and online teaching formats. To this end, we introduce and prove the validity of a novel nonparametric bootstrap procedure based on matching estimators. Our approach supports various didactic evaluation methods as presented in the pedagogical literature. The results of the empirical analysis tend to show better performance of students attending lectures in the traditional in-person format. In general, students slightly prefer the in-person format, but are also in favour of a combination of in-person and online lectures. Based on these results, it seems appropriate to propose courses that combine both in-person and online teaching formats. However, in order to avoid student performance deficits, the planning and didactic approach of online lectures has to be adapted accordingly and requires further investigation.

**keywords:** Bootstrap, Matching Estimators, Teaching-Learning Settings.

*Corresponding authors: Lorenzo Camponovo, University of Applied Sciences and Arts of Southern Switzerland, CH-6900 Lugano, lorenzo.camponovo@supsi.ch.

**Article History**
Received : 13 August 2023; Revised : 05 September 2023; Accepted : 16 September 2023; Published : 30 December 2023

**To cite this paper**

Lorenzo Camponovo, Emanuele Delucchi, Matteo Garzoni, Slobodan Krstic, Stefano Scaravaggi, Oliver Villa (2023). In-Person vs Online Teaching: Empirical Analysis based on Bootstrap for Matching Estimators. *Journal of Econometrics and Statistics.* 3(2), 129-139.

## 1 Introduction

The COVID-19 pandemic and its aftermath have dramatically increased the awareness among the teaching community of hitherto little known teaching formats and sheer availability. This has naturally led to a thorough reflection on established didactical standards and to a lively debate in the teaching community. Thus, the question of a sound empirical method for a rigorous evaluation of didactical settings acquires particular urgence.

Our project aims at developing such a method, by proposing and testing a new statistical framework that is especially tailored for (and adaptable to) some requirements that have been established in the pedagogical literature regarding empirical evaluation of teaching-learning situations. For instance, Klieme and Rakoczy (2008), and references therein, ask for a multi-dimensional and subject-specific approach that goes beyond the mere test-score-averaging. This is needed in order to achieve a testing method that takes into account all facets of competency-based teaching and learning as well as the many aspects that constitute a correspondingly aligned lecture, e.g., in the sense of Biggs and Tang (2011).

To this end, we replace naive and standard test-score-averaging methods, with a novel bootstrap inference procedure based on treatment effect and matching estimators. A large body of studies in empirical economics, political sciences, sociology, epidemiology, and other field is devoted to the evaluation of the effect of some binary treatment under a *selection-on-observables* or *conditional independence* assumption; see, e.g., Imbens (2004) and Imbens and Wooldridge (2009). Researchers applying treatment effect estimators typically aim to assess the average causal effect of the intervention on some outcome variable, by controlling for differences in observed characteristics across treated and non-treated subsamples. In our empirical study, the intervention is the attendance of the lecture in the online teaching format, while the outcome variable is the performance of the students, taking into account also personal feelings and preferences.

In their seminal work, Abadie and Imbens (2006) derived the asymptotic properties of matching estimators, in particular $\sqrt{N}$-consistency and their normal limit distribution, where N denotes the sample size. 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. To overcome this problem, recently Otsu and Rai (2017) introduced and proved the consistency of a wild bootstrap procedure. Bodory, Camponovo, Huber and Lechner (2020), by analyzing the finite sample performance of inference methods for matching estimators, show desirable properties of (wild) bootstrap methods. Following this literature, Adusumilli (2018) extend the wild bootstrap approach to propensity score matching estimators with estimated propensity score, while Bodory, Camponovo, Huber and Lechner (2022) define a nonparametric bootstrap method when the propensity score is known. In this paper, in Section 2 we provide a first contribution to this literature by defining and proving the validity of a novel nonparametric bootstrap for matching estimators.

The mathematical features of our inference approach allow it to be both ductile enough as to model a complex, multi-dimensional setting as required in Klieme and Rakoczy (2008), as well as robust enough in order to control for unwanted effects and produce statistically significant outcomes even with relatively limited sample size. We demonstrated this by carrying out a simple empirical experiment on the performances of students that attend lectures with different in-person and online formats, by controlling for their average grades in *Calculus 1*, *Algebra 1*, *Numerics* and teaching format preferences.

The empirical findings of our analysis tend to show better performances of students that attend lectures in the classic in-person format. In general, students slightly prefer in-person lectures, but are also in favour of a combination between in-person and online lectures. Based on these results, it seems appropriate to propose courses that combine both in-person and online teaching formats. However, in order to avoid under-performances of students, the scheduling and didactic approach of online lectures has to be properly adapted and requires further investigation.

For the sake of brevity, in this paper we present only the real data application. However, we point out that several Monte Carlo results, available from the authors on request, confirm the reliability and the accuracy of our nonparametric bootstrap approach.

## 2 Econometric Approach

In Section 2.1 we introduce the model and the notation that we adopt for developing our empirical analysis. In Section 2.2 we define and prove the validity of a novel nonparametric bootstrap procedure for inference.

### 2.1 Model and Notation

We consider similar settings 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 (in our empirical study, $W_i = 0$ means that student $i$ attend the lecture in-person, while $W_i = 1$ means that the student $i$ attend the lecture online). 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 parameter of interest is the population average treatment effect (ATE), denoted by $\tau$ and defined as

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

Let $\mathbb{I}_A$ be the indicator function for an event A and $|x|$ be the Euclidean norm. To estimate the unknown parameter of interest $\tau$, we consider the matching estimator $\hat{\tau}_N$ defined as

$$\hat{\tau}_N = \frac{1}{N} \sum_{i=1}^{N} \{\hat{Y}_i(1) - \hat{Y}_i(0)\}, \tag{2}$$

where $\hat{Y}_i(0)$ and $\hat{Y}_i(1)$ are defined as,

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

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

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

Under some regularity conditions (see Theorem 1 below), Abadie and Imbens (2006) proved $\sqrt{N}$-consistency and asymptotic normality of the matching estimator defined in (2). Therefore, these results allow us to make inference on the parameter of interest defined in (1). Nevertheless, in small samples the asymptotic approximation may work poorly; see, e.g., Otsu and Rai (2017). To overcome this problem, in the next section we develop a nonparametric bootstrap approach in the spirit of Bodory, Camponovo, Huber and Lechner (2022).

### 2.2 Nonparametric Bootstrap

For the setting introduced in Section 2.1, 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 matching estimators. To overcome this problem, Otsu and Rai (2017) introduced and proved the validity of a wild bootstrap approach. In this study, we define instead a nonparametric bootstrap method in the spirit of Bodory, Camponovo, Huber, Lechner (2022). To this end, first we introduce following assumptions.

$\{Y_i, W_i, X_i\}$ is an iid sample of $(Y, W, X)$.

**Assumption 1.** *(i) $W$ is independent of $(Y(0), Y(1))$ conditional on $X = x$, for almost every $x$. (ii) There exists $c > 0$ such that $P(W = 1|X = x) \in (c, 1-c)$ for almost every $x$.*

**Assumption 2.** *(i) $X$ is continuously distributed with compact and convex support $\mathbb{X} \subset \mathbb{R}^k$. The density of $X$ is bounded and bounded away from zero on $\mathbb{X}$. (ii) $\mu(w,x) = E[Y|W = w, X = x]$ and $\sigma^2(w,x) = Var[Y|W = w, X = x]$ are Lipschitz-continuous on $\mathbb{X}$, (iii) $E[|Y|^4|W = w, X = x]$ is uniformly bounded on $\mathbb{X}$.*

The same comments to Abadie and Imbens (2006) apply. Assumption 1 contains standard unconfoundedness and overlap conditions to identify $\tau$. Assumption 2 lists boundedness and smoothness conditions for the conditional mean and variance functions.

Under Assumptions 2.2-2, Abadie and Imbens (2006) proved that

$$\sqrt{N}(\hat{\tau}_N - B_N - \tau) \to_d N(0, \sigma^2),$$

where the bias term $B_N$ is defined as

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

and the asymptotic variance term $\sigma^2$ is the sum of the variance of the conditional mean and the marginale variance; see, e.g, Abadie and Imbens (2006) for the exact definition of $\sigma^2$. The bias term $B_N$ can be easily estimated using nonparametric estimators $\hat{\mu}(w,x)$ of $\mu(w,x)$; see Assumption 3 below for the exact properties of $\hat{\mu}(w,x)$. Next, we show how to define a nonparametric bootstrap approach for the approximation of the sampling distribution of $\sqrt{N}(\hat{\tau}_N - B_N - \tau)$.

The definition of our nonparametric bootstrap approach relies on the martingale representation for matching estimators introduced in Abadie and Imbens (2012, 2016). In particular, note that as shown in Abadie and Imbens (2012), we can decompose the matching estimator as,

$$\hat{\tau}_N - B_N - \tau = R_{1N} + R_{2N} + o_p(1),$$

where

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

and $K_M(i)$ is the number of times that observations $i$ is used as a match,

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

Let $\hat{\mu}(w,p)$ be a nonparametric estimator of $\mu(w,p)$ that satisfies following condition.

**Assumption 3.** $|\hat{\mu}(w,\cdot)-\mu(w,\cdot)|=o_p(N^{1/2})$.

For instance, series estimators with a suitable choice of the series length satisfies this condition; see, e.g., Abadie and Imbens (2011). Other candidates are the kernel estimators and nearest neighborhood estimators with adequate trimming; see, e.g., Stone (1977).

Furthermore, we define

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

and, for $=1,\ldots,N$, let $\hat{r}_i=\hat{r}_{1i}+\hat{r}_{2i}$, where,

$$\begin{aligned} \hat{r}_{1i} &= \hat{\mu}(1,X_i)-\hat{\mu}(0,X_i)-\hat{\tau}_N-\hat{R}_{1N}, \\ \hat{r}_{2i} &= (2W_i-1)\left(1+\frac{K_M(i)}{M}\right)(Y_i-\hat{\mu}(W_i,X_i))-\hat{R}_{2N}. \end{aligned}$$

Finally, we construct the nonparametric bootstrap sample $(\hat{r}_1^*,\ldots,\hat{r}_N^*)$, by randomly resampling with replacement from the original sample $(\hat{r}_1,\ldots,\hat{r}_N)$ with uniform weights $1/N$, and compute,

$$\hat{R}_N^*=\frac{1}{N}\sum_{i=1}^{N}\hat{r}_i^*.$$

The validity of the nonparametric bootstrap approach is established in the next theorem.

**Theorem 1.** *Under Assumptions 2.2-3, the conditional law of* $\sqrt{N}\hat{R}_N^*$ *converges weakly to the normal distribution with mean* 0 *and variance* $\sigma^2$.

The result in Theorem 1 allows to construct confidence intervals for the parameter $\tau$ using the empirical bootstrap distribution of $\sqrt{N}\hat{R}_N^*$. Let $\sigma_R^*$ be the standard deviation of the bootstrap distribution $\sqrt{N}\hat{R}_N^*$, and $z_\alpha$ be the $N(0,1)$ $\alpha$-quantile. Finally, let $\hat{B}_N$ be the bias-term estimate. Then, the $(1-\alpha)$-bootstrap confidence interval for $\tau$ is defined as,

$$\left[\hat{\tau}_N-\hat{B}_n+z_{\alpha/2}\frac{\sigma_R^*}{\sqrt{N}};\hat{\tau}_N-\hat{B}_N+z_{(1-\alpha)/2}\frac{\sigma_R^*}{\sqrt{N}}\right].$$

## 3 In-person vs Online Teaching Formats

In Section 3.1 we explain the structure of our project, while in Section 3.2 we present the empirical results.

### 3.1 Structure of the Project

In the Spring term 2022, we proposed to first year bachelor students of engineering at DTI-SUPSI to attend a special lecture on *Big-o and little-o notations*. In particular, we have provided the lecture in two different formats: in-person, i.e., with teacher and students physically present in the same room; and online with streaming live connection. Before the project, all students contacted have attended at least one term of *Calculus 1*, *Algebra 1*, *Numerics*, and were familiar with the notions of *Sequences* and *Limits*.

For both in-person and online lectures, the teacher adopted the same class-notes. In the first 60 minutes of the lecture, the teacher introduced theory through slides and oral presentation, combined with hand-written comments and exercises. In the final 30 minutes the students had to solve an online test with 20 true/false questions and exercises about *Big-o and little-o notations*.

Besides the online test, the students also had to fill out a form indicating their: (i) average grade in *Calculus 1*, $[0-1]$; (ii) average grade in *Algebra 1*, $[0-1]$; (iii) average grade in *Numerics*, $[0-1]$; and (iv) teaching format preferences between in-person and online from 0 to 1, where 0 means completely in favour of in-person lectures, while 1 means completely in favour of online lectures. More information on the project and the empirical results are available at: https://www.icorsi.ch/course/view.php?id=13882.

### 3.2 Econometric Approach and Empirical Results

The objective of our project is to statistically evaluate which teaching format provides the best results in term of proportion of corrected answers in the online test. To this end, for each student $i$, we observe the outcome of the proportion of the correct answer in the test $Y_i$, a 4-dimensional vector of covariates $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$, and whether the student attended the lecture in-person ($W_i = 0$) or online ($W_i = 1$). More precisely, for each student $i$ we observe,

$$\begin{aligned}
Y_i &= \text{proportion of correct answers in the test: } [0-1] \\
X_{i1} &= \text{average grade in } Calculus \text{ 1: } [0-1] \\
X_{i2} &= \text{average grade in } Algebra \text{ 1: } [0-1] \\
X_{i3} &= \text{average grade in } Numerics\text{: } [0-1] \\
X_{i4} &= \text{teaching format preferences: } [0-1] \\
W_i &= \text{lecture attendance: } \{0,1\}
\end{aligned}$$

In our study, $N = 100$ students participated to the project: 52 students attended the lecture in-person, while 48 online. Instead of applying naive and standard test-score-averaging methods, we

| | $CI$ | Asymptotic | Nonparametric Bootstrap |
|---|---|---|---|
| $\hat{\tau}_N = -0.104$ | 90% | $[-0.022; -0.186]$ | $[-0.009; -0.199]$ |
| | 95% | $[-0.006; -0.202]$ | $[0.009; -0.217]$ |

Table 1: In the first column, we estimate $\tau$ with the matching estimator and obtain $\hat{\tau}_N = -0.104$. Furthermore, we compute 90% and 95% confidence intervals using (i) the asymptotic distribution derived in Abadie and Imbens (2006) in column 3, and (ii) our nonparametric bootstrap in column 4.

consider the bootstrap inference procedure proposed in the last section that controls for different covariates in the subsamples. Our parameter of interest is the population average treatment effect defined in Equation (1) above. We estimate $\tau$ with the matching estimator defined in Equation (2) above with $M = 1$. Furthermore, we compute 90% and 95% confidence intervals using (i) the asymptotic distribution derived in Abadie and Imbens (2006), and (ii) our nonparametric bootstrap. In Table 1, we present the numerical results.

The in-person teaching format seems to outperform the online teaching format. Indeed, in our sample we obtain $\hat{\tau}_N = -0.104$. Empirically, this means that by controlling average grades in *Calculus 1*, *Algebra 1*, *Numerics* and teaching format preferences, single students tend to underperform from $-10.4\%$ by attending lectures in online format instead of the classic offline format. Furthermore, it is interesting to note that 90% confidence intervals based on all procedures adopted do not include 0. Empirically, this means that with a significance level of $\alpha = 0.10$, we can conclude that the online teaching format implies under-performance of students. On the other hand, it is also interesting to note that this conclusion is no longer true at the significance level $\alpha = 0.05$, when using our nonparametric bootstrap approach. Indeed, 95% confidence intervals based on the bootstrap method include 0. Therefore, at the significance level $\alpha = 0.05$, based on the bootstrap procedure we cannot conclude that in-person teaching format outperforms online teaching format.

To complete our empirical analysis, in Table 2 we present also the teaching format preferences of the students that attended our project. In particular, we report the mean of the preferences $\hat{X}_{4N} = \frac{1}{N}\sum_{i=1}^{N} X_{i4}$, and 90% and 95% confidence intervals using (i) the standard asymptotic theory and (ii) the conventional iid bootstrap. In general the students seem slightly to prefer the in-person teaching format, but are also in favour of a combination between in-person and online lectures. Indeed, in our sample we obtain $\hat{X}_{4N} = 0.212$. Empirically, this means that students would like to have 21.2% of lectures in

| | $CI$ | Asymptotic | iid Bootstrap |
|---|---|---|---|
| $\hat{X}_{4N} = 0.212$ | 90% | $[0.139; 0.285]$ | $[0.133; 0.291]$ |
| | 95% | $[0.125; 0.299]$ | $[0.119; 0.305]$ |

Table 2: In the first column, we report the mean of the preferences $\hat{X}_{4N} = \frac{1}{N}\sum_{i=1}^{N} X_{i4}$. Furthermore, we compute 90% and 95% confidence intervals using (i) the standard asymptotic distribution in column 3, and (ii) the iid bootstrap in column 4.

online format. Confidence intervals based on both standard asymptotic theory and iid bootstrap are very similar. Based on these results, it seems appropriate to propose corses that combine in-person and online lectures. However, in order to avoid under-performances of students, the scheduling and didactic approach of online lectures has to be properly adapted and requires further investigation.

## 4 Conclusions

We propose a novel inference approach for evaluation of didactical settings. Instead of naive and standard test-score-averaging methods, we introduce and prove the validity of a nonparametric bootstrap procedure based on treatment effect and matching estimators. The mathematical features of our approach allow it to be both ductile enough for the modelling of a complex, multi-dimensional setting as well as robust enough in order to control for unwanted effects and produce statistically significant outcomes even with relatively limited sample size.

We demonstrated the accuracy of our approach by carrying out a simple empirical experiment on the performances of students that attend lectures with different in-person and online formats, by controlling for their average grades in *Calculus 1*, *Algebra 1*, *Numerics* and teaching format preferences. In our analysis, empirical findings tend to show better performances of students that attend lectures in the classic in-person format. In general, students slightly prefer in-person lectures, but are also in favour of a combination between in-person and online lectures. Based on these results, it seems appropriate to propose courses that combine both in-person and online teaching formats. However, in order to avoid under-performances of students, the scheduling and didactic approach of online lectures has to be properly adapted and requires further investigation.

## Appendix: Proof

*Proof of Theorem 1:*

Consider the terms,

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

and, for $=1,\ldots,N$, let $\hat{r}_i=\hat{r}_{1i}+\hat{r}_{2i}$, where,

$$
\begin{aligned}
\hat{r}_{1i} &= \hat{\mu}(1,X_i)-\hat{\mu}(0,X_i)-\hat{\tau}_N-\hat{R}_{1N},\\
\hat{r}_{2i} &= (2W_i-1)\left(1+\frac{K_M(i)}{M}\right)(Y_i-\hat{\mu}(W_i,X_i))-\hat{R}_{2N}.
\end{aligned}
$$

Next, consider the bootstrap sample $(\hat{r}_1^*,\ldots,\hat{r}_N^*)$, by randomly resampling with replacement from the original sample $(\hat{r}_1,\ldots,\hat{r}_N)$ with uniform weights $1/N$. Then, by construction,

$$E^*[\hat{r}_i^*]=E[\hat{r}_i^*|(Z_1,\ldots,Z_N)]=0,$$

for $i=1,\ldots,n$, where $Z_i=(Y_i,W_i,X_i')$. Furthermore, for $i=1,\ldots,n$,

$$E^*\left[\left(\frac{\hat{r}_i^*}{\sqrt{N}}\right)^2\right]=E\left[\left(\frac{\hat{r}_i^*}{\sqrt{N}}\right)^2\middle|(Z_1,\ldots,Z_N)\right]=\frac{1}{N}\sum_{j=1}^{N}\hat{r}_j^2.$$

Note that by Assumptions 2.2-2 and 3,

$$
\begin{aligned}
\hat{r}_{1i} &= \mu(1,X_i)-\mu(0,X_i)-\tau+o_p(1),\\
\hat{r}_{2i} &= (2W_i-1)\left(1+\frac{K_M(i)}{M}\right)(Y_i-\mu(W_i,X_i))+o_p(1).
\end{aligned}
$$

Finally, it turns out that as in Abadie and Imbens (2006), $\frac{1}{N}\sum_{j=1}^{N}\hat{r}_j^2$ converges in probability to $\sigma^2$, and consequently the conditional law of $\sqrt{N}\hat{R}_N^*$ converges weakly to the normal distribution with mean 0 and variance $\sigma^2$.

## 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] Adusumilli, K. (2018). Bootstrap inference for propensity score matching, Working Paper.

[7] Biggs, J. B., Tang, C. (2011). Teaching for Quality Learning at University. *McGraw Hill Education and Open University Press.*

[8] Bodory, H., Camponovo, L., Huber, M., and Lechner, M. (2020). The finite sample performance of inference methods for propensity score matching and weighting estimators, *Journal of Business and Ecanomics Statistics*, **38**, 183-200.

[9] Bodory, H., Camponovo, L., Huber, M., and Lechner, M. (2022). Nonparametric Bootstrap for Propensity Score Matching Estimators, Working Paper.

[10] Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review, *The Review of Economics and Statistics*, **86**, 4-29.

[11] Imbens, G. W., and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation, *Journal of Economic Literature*, **47**, 5-86.

[12] Klieme, E. and Rakoczy K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. *Zeitschrift für Pädagogie*, **54** (2008), 222-237.

[13] Otsu, T., and Rai, Y. (2017). Bootstrap inference of matching estimators for average treatment effects, *Journal of the American Statistical Association*, **112**, 1720-1732.

[14] 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.

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