Biometrika (2024), **111**, 4, *pp.* 1421–1428
https://doi.org/10.1093/biomet/asae026
Advance Access publication 14 June 2024

On propensity score matching with a diverging number of matches

By YIHUI HE

School of Mathematical Sciences, Peking University, Beijing, BJ 100080, China
2000010770@stu.pku.edu.cn

AND FANG HAN

Department of Statistics, University of Washington,
B-317 Padelford Hall, Seattle, Washington 98195, U.S.A.
fanghan@uw.edu

SUMMARY

This paper re-examines the work of Abadie & Imbens (2016) on propensity score matching for average treatment effect estimation. We explore the asymptotic behaviour of these estimators when the number of nearest neighbours, M, grows with the sample size. It is shown, while not surprising, but technically nontrivial, that the modified estimators can improve upon the original fixed M-estimators in terms of efficiency. Additionally, we demonstrate the potential to attain the semiparametric efficiency lower bound when the propensity score admits some special structures, echoing the insight of Hahn (1998).

Some key words: Diverging-M asymptotics; Le Cam's discretization device; Le Cam's third lemma; Semiparametric efficiency.

1. INTRODUCTION

1.1. *Main result*

Consider a quadruple $\{X, W, Y(0), Y(1)\}$, where $W \in \{0, 1\}$ denotes the treatment status, $X \in \mathbb{R}^k$ represents the pretreatment variables, and $Y(0)$ and $Y(1)$ signify the potential outcomes (Neyman, 1923; Rubin, 1974) under treatment and control. This paper's primary focus is on inferring the average treatment effect, mathematically defined as $\tau = E\{Y(1) - Y(0)\}$, based on N independent observations of $\{X, W, Y(W)\}$.

A significant focus of this paper centres on a nearest-neighbour matching estimator (Abadie & Imbens, 2006, 2011, 2012, 2016; Lin et al., 2023); this estimator matches the subjects under study with those in the opposite treatment group who possess similar propensity scores, with these scores being estimated using the same dataset.

In detail, let $p(x) = \mathrm{pr}(W = 1 \mid X = x)$ be the propensity score that was introduced by Rosenbaum & Rubin (1983). We assume that $p(x)$ can be reliably quantified using a preset family of functions $\{p(x; \theta), \theta \in \Theta \subset \mathbb{R}^k\}$. The following theorem, presented informally here and to be rigorously detailed in § 3, constitutes our central result.

THEOREM 1 (MAIN THEOREM, INFORMAL). *Consider the estimator $\hat{\tau}_N(\bar{\theta}_N)$, which relies on propensity score–based nearest-neighbour matching using the estimated propensity scores $p(X_i; \bar{\theta}_N)$. Here, $\bar{\theta}_N$ denotes an asymptotically discrete (van der Vaart, 1998; Le Cam & Yang, 2000) maximum likelihood estimator of truth θ^*, computed from the same dataset. Then, under certain regularity conditions,*

© The Author(s) 2024. Published by Oxford University Press on behalf of the Biometrika Trust.
All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/biomet/article/111/4/1421/7693698 by Ann Nez user on 23 July 2025

the following results hold for $\hat{\tau}_N(\bar{\theta}_N)$ as the number of nearest neighbours, denoted M, approaches infinity:

(i) $N^{1/2}\{\hat{\tau}_N(\bar{\theta}_N) - \tau\}$ *is approximately asymptotically normal (Theorem 2);*

(ii) *the asymptotic variance of $\hat{\tau}_N(\bar{\theta}_N)$ is strictly smaller than those of a fixed M, and it is possible to attain the semiparametric efficiency lower bound for estimating τ (Hahn, 1998) (an implication of Theorem 2);*

(iii) *there exists a consistent estimator of the asymptotic variance (Theorem 3).*

1.2. *Related literature*

Our findings are anchored in the seminal contributions of Abadie and Imbens, particularly their pioneering work on propensity score nearest-neighbour matching with a fixed value of M as presented in Abadie & Imbens (2016), among their other influential works (Abadie & Imbens, 2006, 2011, 2012). What sets this paper apart from Abadie & Imbens (2016) is our re-evaluation of the conditions for M, which we force to grow to infinity as $N \to \infty$. In this context, our results also align with the recent studies by Lin et al. (2023) and Lin & Han (2023b), who explored diverging-M matching using the original values of the X_i.

Methodologically, the estimator $\hat{\tau}_N(\bar{\theta}_N)$ is part of a broader family of propensity score-based estimators for the average treatment effect. This family has been extensively explored in the literature, including influential works by Rosenbaum (1987), Robins et al. (1994), Rubin & Thomas (1996), Heckman et al. (1997), Hahn (1998), Scharfstein et al. (1999), Hirano et al. (2003), Frölich (2004, 2005), Huber et al. (2013), Chernozhukov et al. (2018), Su et al. (2023), among many others. Comprehensive reviews can be found in Imbens (2004, 2015) and Stuart (2010).

Secondly, the estimator $\hat{\tau}_N(\bar{\theta}_N)$ falls within the category of *substituting estimators*, where a portion of the parameters is initially estimated. A broader discussion of this class of estimators can be found in Pierce (1982), Randles (1982), Pollard (1989), Andrews (1994), Newey & McFadden (1994), Andreou & Werker (2012), along with various works on causal inference (Robins et al., 1992; Henmi & Eguchi, 2004; Hitomi et al., 2008; Lok, 2022) and some interesting recent developments about optimal transport-based statistical inference (Hallin et al., 2022, 2023).

Thirdly, the estimator $\hat{\tau}_N(\bar{\theta}_N)$ is part of the graph-based statistics family, which aims to estimate a functional of the probability measure using random graphs constructed from an empirical realization of the underlying probability distribution. In this context, the fixed-M asymptotics, as explored by Abadie and Imbens, relates interestingly to recent research on Sourav Chatterjee's rank correlation based on nearest-neighbour graphs with a fixed M (Azadkia & Chatterjee, 2021; Chatterjee, 2021); in particular, they both exhibit asymptotic normality (Abadie & Imbens, 2006; Lin & Han, 2022) and bootstrap inconsistency (Abadie & Imbens, 2008; Lin & Han, 2024). Both of them also benefit from using a diverging M for enhancing efficiency (Lin et al., 2023; Lin & Han, 2023a).

Theoretical underpinnings of our study differ from previous works like Lin et al. (2023) and Lin & Han (2023b). Our main theorem is essentially an adaptation of the existing analysis of Abadie & Imbens (2016), while addressing the limit of M going to infinity throughout our proof. This adaptation is facilitated by the relative simplicity of handling estimated propensity scores, which are essentially random scalars. Like Abadie & Imbens (2016), we employ Le Cam's third lemma and Le Cam's discretization technique, detailed in van der Vaart (1998, Ch. 5.7), for avoiding establishing uniform convergence results. For a justification of the use of an asymptotically discrete estimator, see Le Cam & Yang (2000, Ch. 6.3).

1.3. *Notation and set-up*

This paper closely follows the notation system of Abadie & Imbens (2016). Throughout, it is assumed that there exists a known form of generalized linear specification for the propensity score

Downloaded from https://academic.oup.com/biomet/article/111/4/1421/7693698 by Ann Nez user on 23 July 2025

$p(x)$, written as

$$p(x) = F(x^{\mathrm{T}}\theta^*) = p(x;\theta^*), \tag{1}$$

where F is known a priori while θ^* is unknown to us. Potential model misspecification of $p(\cdot)$, while not elaborated upon in this context, can still produce consistent average treatment effect estimators provided it results in a valid balancing score (Rosenbaum & Rubin, 1983; Abadie & Imbens, 2016).

For any $\theta \in \Theta$, we write pr_θ to represent the joint distribution of $\{X, W, Y(0), Y(1)\}$, with $\mathrm{pr}(W = 1 \mid X = x)$ now taking the value $p(x;\theta) = F(x^{\mathrm{T}}\theta)$ while keeping the distribution of X and the conditional distribution of $\{Y(0), Y(1)\} \mid X, W$ unchanged. Let E_θ and var_θ respectively denote the expectation and variance under pr_θ. In particular, the data-generating distribution is understood to be $\mathrm{pr} = \mathrm{pr}_{\theta^*}$. We similarly shorthand E_{θ^*} and var_{θ^*} as E and var. Let $Y = Y(W)$.

Adopting Le Cam's view on limits of experiments (Le Cam, 1972), we consider such $\theta = \theta_N$ that are allowed to change with the sample size. Accordingly, it is necessary to emphasize that, throughout this paper, we implicitly consider a triangular array setting, where sampling from a sequence of probability measures pr_{θ_N} is allowed. In this paper, the interest is in the sequence

$$\theta_N = \theta^* + h/N^{1/2}$$

with h a conformable vector of constants. Let

$$Z_{N,i} = (X_{N,i}, W_{N,i}, Y_{N,i})$$

represent the observed data, with the subscript N often suppressed in the sequel.

Following Abadie & Imbens (2016), we write $\mu(w,x) = E(Y \mid W = w,\ X = x)$ and $\sigma^2(w,x) = \mathrm{var}(Y \mid W = w,\ X = x)$ to denote the conditional mean and variance of Y given $W = w$ and $X = x$. Additionally, let $\bar\mu(w,p) = E\{Y \mid W = w,\ p(X) = p\}$ and $\bar\sigma^2(w,p) = \mathrm{var}\{Y \mid W = w,\ p(X) = p\}$ represent the conditional mean and variance of Y given $W = w$ and $p(X) = p$. Similarly, define $\bar\mu_\theta(w,p) = E_\theta\{Y \mid W = w,\ p(X;\theta) = p\}$ and $\bar\sigma^2_\theta(w,p) = \mathrm{var}_\theta\{Y \mid W = w,\ p(X;\theta) = p\}$.

The primary focus of the main text is on average treatment effect estimation. Discussions pertaining to a parallel problem of estimating the average treatment effect on the treated, along with some numeric results and all proofs, are relegated to the Supplementary Material.

2. Propensity score matching

In order to describe Abadie and Imbens's propensity score matching estimator, we first introduce some statistics about the M-nearest-neighbour matching based on the values of a general propensity score estimate $p(X_i;\theta)$.

Let $\mathcal{J}_M(i,\theta)$ represent the index set of the M matches of unit i, measured based on the values of $p(X_i;\theta)$, among the units whose $W = 1 - W_i$. In other words, define

$$\mathcal{J}_M(i,\theta) = \left\{ j\colon\ W_j = 1 - W_i,\ \sum_{k\colon W_k = 1 - W_i} \mathbb{1}\{|p(X_i;\theta) - p(X_k;\theta)| \leqslant |p(X_i;\theta) - p(X_j;\theta)|\} \leqslant M \right\}$$

with $\mathbb{1}(\cdot)$ representing the indicator function. Furthermore, introduce the set of natural numbers

$$K_{M,\theta}(\cdot)\colon \{1,\ldots,N\} \to \mathbb{N}$$

to be the number of matched times of unit i, i.e.,

$$K_{M,\theta}(i) = \sum_{j\colon W_j = 1 - W_i} \mathbb{1}\left[\sum_{k\colon W_k = W_i} \mathbb{1}\{|p(X_k;\theta) - p(X_j;\theta)| \leqslant |p(X_i;\theta) - p(X_j;\theta)|\} \leqslant M \right].$$

Downloaded from https://academic.oup.com/biomet/article/111/4/1421/7693698 by Ann Nez user on 23 July 2025

The propensity score matching estimator of τ, based on the $p(X_i;\theta)$ with a generic θ, is then defined as

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

To complete the estimation process, one needs an estimate of θ^* to be substituted into $\hat{\tau}_N(\cdot)$. In light of (1) and following Abadie & Imbens (2016), we estimate θ^* by maximizing the loglikelihood function,

$$L(\theta \mid Z_1, \ldots, Z_N) = \sum_{i=1}^{N}[W_i \log\{F(X_i^{\mathrm{T}}\theta)\} + (1 - W_i)\log\{1 - F(X_i^{\mathrm{T}}\theta)\}],$$

yielding the maximum likelihood estimator $\hat{\theta}_N$. The final estimator of τ is then defined to be $\hat{\tau}_N(\hat{\theta}_N)$.

3. THEORY

We first outline the main assumptions needed to establish Theorem 1.

Assumption 1. Suppose that

(i) $\{Y(0), Y(1)\}$ is independent of W conditional on X almost surely;
(ii) $p(X)$ is supported over $[\underline{p}, \bar{p}]$ with $0 < \underline{p} < \bar{p} < 1$, and has a Lebesgue density that is continuous over $[\underline{p}, \bar{p}]$.

Assumption 2. Suppose that $\{Z_i\}_{i=1}^N$ are N independent draws from $\mathrm{pr} = \mathrm{pr}_{\theta^*}$.

Assumption 3. Suppose that

(i) $\theta^* \in \mathrm{int}(\Theta)$ with a compact Θ, X has a bounded support and $E[XX^{\mathrm{T}}]$ is nonsingular;
(ii) $F\colon \mathbb{R} \to (0, 1)$ is continuously differentiable with a strictly positive and continuous derivative function f;
(iii) there exists a component of X that is continuously distributed, has nonzero coefficient in θ^* and admits a continuous density function conditional on the rest of X;
(iv) there exist some constants $\varepsilon > 0$ and $C_{\bar{\mu}} < \infty$ such that, for all θ with the Euclidean distance $\|\theta - \theta^*\| \leqslant \varepsilon$ and $w \in \{0, 1\}$,
   (a) $|\bar{\mu}_\theta(w, p_1) - \bar{\mu}_\theta(w, p_2)| \leqslant C_{\bar{\mu}}|p_1 - p_2|$ holds for any p_1 and p_2;
   (b) $\bar{\sigma}^2_\theta(w, p)$ are equicontinuous in p;
   (c) $E_\theta\{Y^4 \mid W = w,\, p(X;\theta) = p\}$ are uniformly bounded.

Assumption 4. For any $\mathbb{R}^{k+2} \to \mathbb{R}$ bounded and measurable function $r(y, w, x)$ that is continuous in x, and for any sequence $\tilde{\theta}_N \to \theta^*$, it is assumed that $E_{\tilde{\theta}_N}\{r(Y, W, X) \mid W, p(X;\tilde{\theta}_N)\}$ converges to $E\{r(Y, W, X) \mid W, p(X)\}$ almost surely under $\mathrm{pr}_{\tilde{\theta}_N}$.

Assumptions 1–4 are modified and reordered versions of Assumptions 1–5 of Abadie & Imbens (2016). In particular, Assumption 1 is identical to Assumption 1 combined with Assumption 2(i) of Abadie & Imbens (2016). Assumption 2 is Assumption 3 of Abadie & Imbens (2016); both allow for a triangular array setting intrinsically. Assumption 3 is a modified version of Assumption 4 of Abadie & Imbens (2016). Compared to Abadie & Imbens (2016), we additionally require a global Lipschitz constant in Assumption 3(iv). Also, we require f to be continuous and $\bar{\sigma}^2_\theta$ to be equicontinuous.

Downloaded from https://academic.oup.com/biomet/article/111/4/1421/7693698 by Ann Nez user on 23 July 2025

Besides, we require the uniform boundedness of the conditional fourth moment of Y as in Abadie & Imbens (2006), in order to prove the convergence of the variance estimator. Assumptions 2(ii) and (iii) of Abadie & Imbens (2016) are the corollary of Assumption 3(iv). Lastly, Assumption 4 is exactly Assumption 5 of Abadie & Imbens (2016).

Next, we formally introduce Le Cam's discretization trick. For any positive constant d, following Abadie & Imbens (2016), we transform $\hat{\theta}_N$ to an asymptotically discrete estimator, $\bar{\theta}_N$, by setting $\bar{\theta}_{N,j} = (d/N^{1/2})[N^{1/2}\hat{\theta}_{N,j}/d]$, with $[\cdot]$ outputting the input's nearest integer.

The following is the main theorem of this paper.

THEOREM 2. *Suppose that Assumptions 1–4 hold. Assume further that $M = O(N^v)$ for some $v < 1/2$ and $M \to \infty$ as $N \to \infty$. Then, the true distribution satisfies*

$$\lim_{d\downarrow 0}\lim_{N\to\infty} \mathrm{P}[N^{1/2}(\sigma^2 - c^{\mathrm{T}} I_{\theta^*}^{-1} c)^{-1/2}\{\hat{\tau}_N(\bar{\theta}_N) - \tau\} \leqslant z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \mathrm{d}x,$$

where

$$I_\theta = E\left[\frac{f(X^{\mathrm{T}}\theta)^2}{p(X;\theta)\{1-p(X;\theta)\}} XX^{\mathrm{T}}\right]$$

is the Fisher information at θ,

$$c = E\left(\left[\frac{\mathrm{cov}\{X, \mu(1,X) \mid p(X)\}}{p(X)} + \frac{\mathrm{cov}\{X, \mu(0,X) \mid p(X)\}}{1-p(X)}\right] f(X^{\mathrm{T}}\theta^*)\right)$$

signifies the cross term, and

$$\sigma^2 = E([\bar{\mu}\{1,p(X)\} - \bar{\mu}\{0,p(X)\} - \tau]^2) + E\left[\frac{\bar{\sigma}^2\{1,p(X)\}}{p(X)}\right] + E\left[\frac{\bar{\sigma}^2\{0,p(X)\}}{1-p(X)}\right]$$

is the asymptotic variance of $\hat{\tau}_N(\theta^*)$.

It is straightforward to check that the asymptotic variance of $\hat{\tau}_N(\bar{\theta}_N)$, of the form $\sigma^2 - c^{\mathrm{T}} I_{\theta^*}^{-1} c$, is strictly smaller than those matching estimators with a fixed M; cf. Abadie & Imbens (2016, Proposition 1 and Theorem 1). In addition, the asymptotic variance is no greater than that of $\hat{\tau}_N(\theta^*)$, the matching estimator using the oracle θ^*, a well-known phenomenon. Furthermore, when

$$\bar{\mu}\{w, p(x)\} = \mu(w,x) \quad \text{and} \quad \bar{\sigma}^2\{w,p(x)\} = \sigma^2(w,x) \quad \text{for } w = 0, 1, \tag{2}$$

the asymptotic variance of $\hat{\tau}_N(\bar{\theta}_N)$ attains the following semiparametric efficiency lower bound for estimating τ (Hahn, 1998):

$$\sigma^{2,\mathrm{eff}} = E[\{\mu(1,X) - \mu(0,X) - \tau\}^2] + E\left\{\frac{\sigma^2(1,X)}{p(X)}\right\} + E\left\{\frac{\sigma^2(0,X)}{1-p(X)}\right\}.$$

Equation (2) would be satisfied if both $\mu(w,x)$ and $\sigma^2(w,x)$ are functions of $p(x)$. Under Assumption 3(ii), condition (2) is equivalent to assuming that $\mu(w,x)$ and $\sigma^2(w,x)$ exhibit a single-index form as functions of $x^{\mathrm{T}}\theta^*$ for each w; cf. Hahn (1998) and Angrist & Pischke (2009, Ch. 3.3.2).

Downloaded from https://academic.oup.com/biomet/article/111/4/1421/7693698 by Ann Nez user on 23 July 2025

We next consider estimating the asymptotic variance. Inspired by Abadie & Imbens (2016), we can estimate $\sigma^2 - c^{\mathrm{T}} I_{\theta^*} c$ using $\hat{\sigma}^2 - \hat{c}^{\mathrm{T}} \hat{I}_{\theta^*} \hat{c}$, where

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left[(2W_i - 1)\left\{Y_i - \frac{1}{M}\sum_{j\in\mathcal{J}_M(i,\hat{\theta}_N)} Y_j\right\} - \hat{\tau}_N(\hat{\theta}_N)\right]^2$$
$$+ \frac{1}{N}\sum_{i=1}^{N}\left[\left\{\frac{K_{M,\hat{\theta}_N}(i)}{M}\right\}^2 + \frac{2M-1}{M}\left\{\frac{K_{M,\hat{\theta}_N}(i)}{M}\right\}\right]\tilde{\sigma}^2\{W_i, p(X_i)\},$$

$$\hat{c} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\hat{\mathrm{C}}\{X_i, \mu(1, X_i) \mid p(X_i)\}}{p(X_i;\hat{\theta}_N)} + \frac{\hat{\mathrm{C}}\{X_i, \mu(0, X_i) \mid p(X_i)\}}{1 - p(X_i;\hat{\theta}_N)}\right] f(X_i^{\mathrm{T}}\hat{\theta}_N)$$

$$\text{and} \quad \hat{I}_{\theta^*} = \frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i^{\mathrm{T}}\hat{\theta}_N)^2}{p(X_i;\hat{\theta}_N)\{1 - p(X_i;\hat{\theta}_N)\}} X_i X_i^{\mathrm{T}},$$

with $\tilde{\sigma}^2\{W_i, p(X_i)\}$ and $\hat{\mathrm{C}}\{X_i, \mu(w, X_i) \mid p(X_i)\}$ defined in the same way as in Abadie & Imbens (2016):

$$\tilde{\sigma}^2\{W_i, p(X_i)\} = \frac{1}{Q-1}\sum_{j\in\mathcal{H}_Q(i,\hat{\theta}_N)}\left\{Y_j - \frac{1}{Q}\sum_{k\in\mathcal{H}_Q(i,\hat{\theta}_N)} Y_k\right\}^2,$$

$$\hat{\mathrm{C}}\{X_i, \mu(W_i, X_i) \mid p(X_i)\} = \frac{1}{L-1}\sum_{j\in\mathcal{H}_L(i,\hat{\theta}_N)}\left\{X_j - \frac{1}{L}\sum_{k\in\mathcal{H}_L(i,\hat{\theta}_N)} X_k\right\}\left\{Y_j - \frac{1}{L}\sum_{k\in\mathcal{H}_L(i,\hat{\theta}_N)} Y_k\right\},$$

$$\text{and} \quad \hat{\mathrm{C}}\{X_i, \mu(1 - W_i, X_i) \mid p(X_i)\} = \frac{1}{L-1}\sum_{j\in\mathcal{J}_L(i,\hat{\theta}_N)}\left\{X_j - \frac{1}{L}\sum_{k\in\mathcal{J}_L(i,\hat{\theta}_N)} X_k\right\}\left\{Y_j - \frac{1}{L}\sum_{k\in\mathcal{J}_L(i,\hat{\theta}_N)} Y_k\right\}.$$

Here, for a generic $\theta \in \Theta$, $\mathcal{H}_M(i, \theta)$ represents the set of M matches of unit i, based on the propensity score matching with $p(X_i;\theta)$, among the units whose $W = W_i$. In other words,

$$\mathcal{H}_M(i,\theta) = \left\{j :\ W_j = W_i,\ \sum_{k:\ W_k = W_i} \mathbb{1}\{|p(X_i;\theta) - p(X_k;\theta)| \leqslant |p(X_i;\theta) - p(X_j;\theta)|\} \leqslant M\right\}.$$

The variance estimator under examination is exactly that presented in Abadie & Imbens (2016) in practical applications. However, the following theorem distinguishes itself by investigating distinct asymptotic behaviours for parameter M, forcing it to diverge.

THEOREM 3. *Suppose that Assumptions 1–4 hold. Assume further that there exists a constant $\nu < 1/2$ such that, as $N \to \infty$, $M = O(N^\nu)$ for $M \to \infty$, $Q = O(N^\nu)$ for $Q \to \infty$ and $L \geqslant 2$ is a fixed finite positive integer. Then, under* pr, *$(\hat{\sigma}^2, \hat{c}, \hat{I}_{\theta^*})$ is a consistent estimator of $(\sigma^2, c, I_{\theta^*})$.*

Remark 1. Theorem 3 differs from that presented in Abadie & Imbens (2016) by incorporating asymptotically diverging values for both M and Q. The introduction of a diverging M is a necessity, as our analysis relies on M approaching infinity. The introduction of a diverging Q, on the other hand, is convenient for managing the $I_{6,2}$ term in the proof of Theorem 3. In fact, even for the study of finite-M matching (Abadie & Imbens, 2016), one could still employ a diverging Q to consistently estimate the variance.

SUPPLEMENTARY MATERIAL

The Supplementary Material contains further simulations and proofs of Theorems 2 and 3.

Downloaded from https://academic.oup.com/biomet/article/111/4/1421/7693698 by Ann Nez user on 23 July 2025

References

Abadie, A. & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica* **74**, 235–67.

Abadie, A. & Imbens, G. W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica* **76**, 1537–57.

Abadie, A. & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. *J. Bus. Econ. Statist.* **29**, 1–11.

Abadie, A. & Imbens, G. W. (2012). A martingale representation for matching estimators. *J. Am. Statist. Assoc.* **107**, 833–43.

Abadie, A. & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica* **84**, 781–807.

Andreou, E. & Werker, B. J. (2012). An alternative asymptotic analysis of residual-based statistics. *Rev. Econ. Statist.* **94**, 88–99.

Andrews, D. W. (1994). Empirical process methods in econometrics. In *Handbook of Econometrics*, vol. 4, pp. 2247–94. Amsterdam: Elsevier.

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Azadkia, M. & Chatterjee, S. (2021). A simple measure of conditional dependence. *Ann. Statist.* **49**, 3070–102.

Chatterjee, S. (2021). A new coefficient of correlation. *J. Am. Statist. Assoc.* **116**, 2009–22.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Economet. J.* **21**, C1–68.

Frölich, M. (2004). Finite-sample properties of propensity-score matching and weighting estimators. *Rev. Econ. Statist.* **86**, 77–90.

Frölich, M. (2005). Matching estimators and optimal bandwidth choice. *Statist. Comp.* **15**, 197–215.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66**, 315–31.

Hallin, M., La Vecchia, D. & Liu, H. (2022). Center-outward R-estimation for semiparametric VARMA models. *J. Am. Statist. Assoc.* **117**, 925–38.

Hallin, M., La Vecchia, D. & Liu, H. (2023). Rank-based testing for semiparametric VAR models: a measure transportation approach. *Bernoulli* **29**, 229–73.

Heckman, J. J., Ichimura, H. & Todd, P. E. (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. *Rev. Econ. Statist.* **64**, 605–54.

Henmi, M. & Eguchi, S. (2004). A paradox concerning nuisance parameters and projected estimating functions. *Biometrika* **91**, 929–41.

Hirano, K., Imbens, G. W. & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71**, 1161–89.

Hitomi, K., Nishiyama, Y. & Okui, R. (2008). A puzzling phenomenon in semiparametric estimation problems with infinite-dimensional nuisance parameters. *Economet. Theory* **24**, 1717–28.

Huber, M., Lechner, M. & Wunsch, C. (2013). The performance of estimators based on the propensity score. *J. Economet.* **175**, 1–21.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. *Rev. Econ. Statist.* **86**, 4–29.

Imbens, G. W. (2015). Matching methods in practice: three examples. *J. Hum. Resour.* **50**, 373–419.

Le Cam, L. (1972). Limits of experiments. In *Proc. 6th Berkeley Symp. Math. Statist. Prob.*, vol. 1, Ed. L. M. Le Cam, J. Neyman and E. L. Scott, pp. 245–61. Berkeley, CA: University of California Press.

Le Cam, L. & Yang, G. L. (2000). *Asymptotics in Statistics: Some Basic Concepts*, 2nd ed. New York: Springer.

Lin, Z., Ding, P. & Han, F. (2023). Estimation based on nearest neighbor matching: from density ratio to average treatment effect. *Econometrica* **91**, 2187–217.

Lin, Z. & Han, F. (2022). Limit theorems of Chatterjee's rank correlation. *arXiv:* 2204.08031v3.

Lin, Z. & Han, F. (2023a). On boosting the power of Chatterjee's rank correlation. *Biometrika* **110**, 283–99.

Lin, Z. & Han, F. (2023b). On regression-adjusted imputation estimators of the average treatment effect. *arXiv:* 2212.05424v2.

Lin, Z. & Han, F. (2024). On the failure of the bootstrap for Chatterjee's rank correlation. *Biometrika*, doi: 10.1093/biomet/asae004.

Lok, J. J. (2022). How estimating nuisance parameters can reduce the variance (with consistent variance estimation). *arXiv:* 2109.02690v3.

Newey, W. K. & McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, vol. 4, pp. 2111–245. Amsterdam: Elsevier.

Neyman, J. (1923). Sur les applications de la théorie des probabilités aux experiences agricoles: essai des principes. *Roczniki Nauk Rolniczych* **10**, 1–51.

Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10**, 475–8.

Downloaded from https://academic.oup.com/biomet/article/111/4/1421/7693698 by Ann Nez user on 23 July 2025

Pollard, D. (1989). Asymptotics via empirical processes. *Statist. Sci.* **4**, 341–54.
Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10**, 462–74.
Robins, J. M., Mark, S. D. & Newey, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* **48**, 479–95.
Robins, J. M., Rotnitzky, A. & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Am. Statist. Assoc.* **89**, 846–66.
Rosenbaum, P. R. (1987). Model-based direct adjustment. *J. Am. Statist. Assoc.* **82**, 387–94.
Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66**, 688–701.
Rubin, D. B. & Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics* **52**, 249–64.
Scharfstein, D. O., Rotnitzky, A. & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Am. Statist. Assoc.* **94**, 1096–120.
Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Statist. Sci.* **25**, 1–21.
Su, F., Mou, W., Ding, P. & Wainwright, M. (2023). When is the estimated propensity score better? High-dimensional analysis and bias correction. *arXiv:* 2303.17102v1.
van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

[*Received on* 25 *December* 2023. *Editorial decision on* 10 *May* 2024]

Downloaded from https://academic.oup.com/biomet/article/111/4/1421/7693698 by Ann Nez user on 23 July 2025