Bias Corrected Matching Estimators for Average Treatment Effects

Alberto Abadie and Guido W. Imbens *
Harvard University and NBER

August 2007

Abstract

In Abadie and Imbens (2006), we have shown that simple nearest-neighbor matching estimators include a conditional bias term that converges at a rate that may be slower than $N^{1/2}$. As a result, matching estimators are not $N^{1/2}$-consistent, in general. In this article, we propose a bias correction that renders matching estimators $N^{1/2}$-consistent and asymptotically normal. To demonstrate the methods proposed in this article, we apply them to the National Supported Work (NSW) data, originally employed in Lalonde (1986). We also carry out a small simulation study based on the NSW example where a simple implementation of the bias-corrected matching estimator performs well compared to both simple matching estimators and to regression estimators in terms of bias, root-mean-squared-error and coverage rates. Software to compute the estimators proposed in this paper is available in the authors' web pages.

JEL classification: C13, C14, H31
Keywords: Treatment effects; Selection on observables.

---

* *E-mail addresses:* alberto_abadie@harvard.edu (A. Abadie) and imbens@fas.harvard.edu (G. Imbens).
Financial support for this research was generously provided through NSF grants SES-0350645 (Abadie), SBR-0452590 (Imbens). A previous version of this paper circulated under the title "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects". Statistical software for implementing the methods studied here is available at: `http://www.economics.harvard.edu/faculty/imbens/software.html`.

1. Introduction

The purpose of this article is to investigate the properties of estimators that combine matching with the bias correction proposed in Rubin (1973) and Quade (1982) and a non-parametric extensions of this bias correction. We show that a nonparametric implementation of the bias correction removes the conditional bias of matching asymptotically, rendering them $N^{1/2}$-consistent, without affecting the asymptotic variance.

We apply simple matching estimators (without bias-correction) and the bias-corrected matching estimators investigated in this article to the National Supported Work Demonstration data, analyzed originally by Lalonde (1986) and subsequently by Heckman and Hotz (1989), Dehejia and Wahba (1999), Smith and Todd (2001, 2005), and Imbens (2003). For this data set, we show that traditional matching estimators without bias correction are very sensitive to the choice for the number of matches, whereas a simple implementation of the bias correction using linear least squares is much more robust. Moreover, in a small simulation study designed to mimic the data from the NSW application, we find that the simple linear least squares based implementation of the bias-corrected matching estimator performs well compared to both matching estimators without and to regression estimators, in terms of both bias and root-mean-squared-error.

Bias-corrected matching estimators combine some of the advantages and disadvantages of both matching and regression estimators. Compared to matching estimators without bias correction they have the advantage of being $N^{1/2}$-consistent and asymptotically normal irrespective of the number of covariates. However, bias-corrected matching estimators may be more difficult to implement than matching estimators without bias correction, especially if the bias correction is calculated using nonparametric smoothing techniques and, therefore, involves the choice of smoothing parameters as functions of the sample size. Compared to estimators based on regression adjustment without matching (e.g., Heckman, Ichimura, and Todd, 1997; Hahn, 1998; Heckman, Ichimura, Smith, and Todd, 1998) or weighting estimators (Hirano, Imbens, and Ridder, 2003; Abadie, 2005) bias-corrected matching estimators have the advantage of an additional layer of robustness, because matching ensures consistency for any given value of the smoothing parameters, without accurate approximations to either the regression function or the propensity score. However, in contrast to some regression adjustment estimators, bias-corrected matching estimator have the disadvantage of not being fully efficient (Abadie and Imbens, 2006).

2. Matching Estimators

2.1. Setting and Notation

Matching estimators are often used in evaluation research to estimate treatment effects in the absence of experimental data. As in Rubin (1974), we use the classical framework of potential outcomes to causality. Let W_i be a binary variable that indicates exposure of individual i to treatment, so $W_i = 1$ if individual i was exposed to treatment, and $W_i = 0$ otherwise. The variables $Y_i(0)$ and $Y_i(1)$ represent potential outcomes with and without treatment, respectively. Depending on the value of W_i one of the potential outcomes is realized:

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

The usual goal of evaluation research is to estimate the average treatment effect,

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

or the average effect of the treatment on the treated:

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

In general, a simple comparison of average outcomes between treated and non-treated does not identify the average effect of the treatment. The reason is that this comparison may be contaminated by the effect of other variables that correlate with the treatment, W_i, and the potential outcomes, $Y_i(1)$ and $Y_i(0)$. The presence of these confounders may create a correlation between W_i and Y_i even if the treatment has no effect on the outcome. Randomization of the treatment eliminates the correlation between any potential confounder and W_i. However, if exposure to treatment is not randomized, the parameters τ and τ^t are not identified in general. The following set of assumption allow identification and estimation of τ and τ^t in the presence of observed confounders.

A.1: *Let X be a random vector of dimension k of continuous covariates distributed on $\mathbb{R}^k$ with compact and convex support $\mathbb{X}$, with (a version of the) density bounded, and bounded away from zero on its support.*

A.2: *For almost every* $x \in \mathbb{X}$*,*

(i) (unconfoundedness) W *is independent of* $(Y(0), Y(1))$ *conditional on* $X = x$*;*

(ii) (overlap) $\eta < \Pr(W = 1|X = x) < 1 - \eta$*, for some* $\eta > 0$*.*

A.3: $\{(Y_i, W_i, X_i)\}_{i=1}^N$ *are independent draws from the distribution of* (Y, W, X)*.*

A.4: *Let* $\mu(x, w) = \mathbb{E}[Y_i|X_i = x, W_i = w]$ *and* $\sigma^2(x, w) = \mathbb{E}[(Y_i - \mu(x, w))^2|X_i = x, W_i = w]$*. Then, (i)* $\mu(x, w)$ *and* $\sigma^2(x, w)$ *are Lipschitz in* $\mathbb{X}$ *for* $w = 0, 1$*, (ii) the fourth moments of the conditional distribution of* Y *given* $W = w$ *and* $X = x$ *exist and are uniformly bounded, and (iii)* $\sigma^2(x, w)$ *is bounded away from zero.*

Assumption A.1 requires that all variables in X have a continuous distribution. Notice, however, that discrete covariates with a finite number of support points can be easily accommodated in our analysis by conditioning on their values. Assumption A.2(i) states that, conditional on X_i, the treatment, W_i, is "as good as randomized," that is, it is independent of the potential outcomes, $Y_i(1)$ and $Y_i(0)$. That would be the case, in particular, if all potential confounders are included in X. Therefore, conditional on $X_i = x$, a simple comparison of average outcomes between treated and non-treated is equal to the average effect of the treatment given $X_i = x$. Assumption A.2(ii) is the usual support condition invoked for matching estimators. Assumption A.3 refers to the sampling process. Finally, Assumption A.4 collects regularity conditions that will be used later. Abadie and Imbens (2006) discuss assumptions A.1-A.4 in greater detail. Identification conditions for matching estimators are also discussed in Dehejia and Wahba (1999), Lechner (2002), and Imbens (2004), among others.

Let $N_1 = \sum_{i=1}^N W_i$ and $N_0 = N - N_1$. Estimation of τ^t requires weaker assumptions than estimation of τ. In particular, when the parameter of interest is τ^t, Assumptions A.2 and A.3 can be weakened as follows.

A.2′: *For almost every* $x \in \mathbb{X}$*,*

(i) W *is independent of* $Y(0)$ *conditional on* $X = x$*;*

(ii) $\Pr(W = 1|X = x) < 1 - \eta$*, for some* $\eta > 0$*.*

A.3′: *Conditional on* $W_i = w$ *the sample consists of independent draws from* $Y, X|W = w$*, for* $w = 0, 1$*. For some* $r \geq 1$*,* $N_1^r/N_0 \to \theta$*, with* $0 < \theta < \infty$*.*

As in Abadie and Imbens (2006) we consider matching "with replacement", allowing each unit to be used as a match more than once. For $x \in \mathbb{X}$, let $\|x\| = (x'x)^{1/2}$ be the standard Euclidean vector norm. Let $j_m(i)$ be the index of the m-th match to unit i. That is, among the units in the opposite treatment group to unit i, unit $j_m(i)$ is the m-th closest unit to unit i in terms of covariate values. Unit j in the m-th best match to unit i, or $j_m(i) = j$, if $W_j = 1 - W_i$ and

$$\sum_{l:W_l=1-W_i} 1\Big\{\|X_l - X_i\| \leq \|X_j - X_i\|\Big\} = m,$$

where $1\{\cdot\}$ is the indicator function, equal to one if the expression in brackets is true and zero otherwise. For notational simplicity we ignore matching ties, which happen with probability zero. Let $\mathcal{J}_M(i) = \{j_1(i), \ldots, j_M(i)\}$ denote the set of indices for the first M matches for unit i, for M such that $M \leq N_0$ and $M \leq N_1$. Finally, let $K_M(i)$ denote the number of times unit i is used as a match if M matches are done per unit:

$$K_M(i) = \sum_{l=1}^{N} 1\{i \in \mathcal{J}_M(l)\}.$$

2.2. Estimators

For $i = 1, \ldots, N$, let

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

Using this notation, we can write the matching estimators of τ and τ^t, that uses M matches per units with replacement, as

$$\hat{\tau}_M^m = \frac{1}{N} \sum_{i=1}^{N} \Big(\hat{Y}_i(1) - \hat{Y}_i(0)\Big) \qquad \text{and} \qquad \hat{\tau}_M^{m,t} = \frac{1}{N_1} \sum_{W_i=1} \Big(Y_i - \hat{Y}_i(0)\Big). \tag{1}$$

Notice also that,

$$\hat{\tau}_M^m = \frac{1}{N} \sum_{i=1}^{N} (2W_i - 1) \left(1 + \frac{K_M(i)}{M}\right) Y_i \quad \text{and} \quad \hat{\tau}_M^{m,t} = \frac{1}{N_1} \sum_{W_i=1} \left(W_i - (1 - W_i) \frac{K_M(i)}{M}\right) Y_i. \tag{2}$$

In empirical applications, matching estimators are implemented with small M and large N. Therefore, to approximate the distribution of matching estimator in that setting will consider their asymptotics as N increases while M is held fixed.

Let $\mu_w(x) = \mathbb{E}[Y(w)|X = x]$, and let $\hat{\mu}_w(X_i)$ be a consistent estimator of $\mu_w(X_i)$.[1] A regression imputation estimator uses $\hat{\mu}_0(X_i)$ and $\hat{\mu}_1(X_i)$ to impute missing values of $Y_i(0)$ and $Y_i(1)$, respectively. That is, for

$$\bar{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0, \\ \hat{\mu}_0(X_i) & \text{if } W_i = 1 \end{cases} \qquad \text{and} \qquad \bar{Y}_i(1) = \begin{cases} \hat{\mu}_1(X_i) & \text{if } W_i = 0, \\ Y_i & \text{if } W_i = 1, \end{cases}$$

the regression imputation estimators of τ and τ^t are

$$\hat{\tau}^{reg} = \frac{1}{N}\sum_{i=1}^{N}\left(\bar{Y}_i(1) - \bar{Y}_i(0)\right) \qquad \text{and} \qquad \hat{\tau}^{reg,t} = \frac{1}{N_1}\sum_{W_i=1}\left(Y_i - \bar{Y}_i(0)\right).$$

As in Abadie and Imbens (2006) we classify as regression imputation estimators those for which $\hat{\mu}_w(x)$ is a consistent estimator of $\mu_w(x)$.[2] The matching estimators in (1) are similar to the regression imputation estimators, as they can be interpreted as imputing $Y_i(0)$ and $Y_i(1)$ with a nearest-neighbor estimate of $\mu_0(X_i)$ and $\mu_1(X_i)$, respectively. However, because M is held fixed, $\hat{Y}_i(0)$ and $\hat{Y}_i(1)$ do not estimate $\mu_0(X_i)$ and $\mu_1(X_i)$ consistently.

Finally, we consider a bias-corrected matching estimator where the difference within the matches is regression-adjusted for the difference in covariate values:

$$\tilde{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0, \\ \dfrac{1}{M}\displaystyle\sum_{j\in\mathcal{J}_M(i)}\left(Y_j + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_j)\right) & \text{if } W_i = 1, \end{cases}$$

and

$$\tilde{Y}_i(1) = \begin{cases} \dfrac{1}{M}\displaystyle\sum_{j\in\mathcal{J}_M(i)}\left(Y_j + \hat{\mu}_1(X_i) - \hat{\mu}_1(X_j)\right) & \text{if } W_i = 0, \\ Y_i & \text{if } W_i = 1, \end{cases}$$

with corresponding estimators

$$\hat{\tau}^{bcm}_M = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{Y}_i(1) - \tilde{Y}_i(0)\right) \qquad \text{and} \qquad \hat{\tau}^{bcm,t}_M = \frac{1}{N_1}\sum_{W_i=1}\left(Y_i - \tilde{Y}_i(0)\right). \tag{3}$$

[1]Notice that Assumption 2(i) implies $\mu_w(x) = \mu(x, w)$ for $w = 0, 1$ and for almost every $x \in \mathbb{X}$. Assumption 2′(i) implies that $\mu_0(x) = \mu(x, 0)$ for almost every $x \in \mathbb{X}$.

[2]Heckman, Ichimura, and Todd (1997), Hahn (1998), and Heckman, Ichimura, Smith, and Todd (1998), Chen, Hong, and Tarozzi (2004), and Imbens, Newey, and Ridder (2005) propose regression imputation estimators of τ and τ^t.

Rubin (1979) and Quade (1982) discuss such estimators in the context of matching without replacement and with linear covariance adjustment.

2.3. Large Sample Properties of Matching Estimators

For completeness, in this subsection, we collect some results on the large sample properties of matching estimators derived in Abadie and Imbens (2006), which motivate the use of bias-corrected matching estimators.

First, we introduce some additional notation. Let $\mathbf{X}$ be the $N \times k$ matrix with i-th row equal to X_i'. Similarly, let $\mathbf{W}$ be $N \times 1$ vector with i-th element equal to W_i. Let $\tau(x) = E[Y(1) - Y(0)|X = x] = \mu_1(x) - \mu_0(x)$ be the average effect of the treatment conditional on $X = x$. Similarly, let $\tau(x)^t = E[Y(1) - Y(0)|X = x, W = 1]$ be the average effect of the treatment conditional on $X = x$ and $W = 1$. Notice that Assumption $2'$ implies $\tau(x)^t = \mu(x, 1) - \mu_0(x)$. Therefore, under Assumption $2'$, the sample average treatment effect and the sample average treatment effect for the treated conditional on $\mathbf{X}$ and $\mathbf{W}$ are:

$$\overline{\tau(X)} = \frac{1}{N}\sum_{i=1}^{N}\big(\mu_1(X_i) - \mu_0(X_i)\big) \qquad \text{and} \qquad \overline{\tau(X)}^{\,t} = \frac{1}{N_1}\sum_{i=1}^{N} W_i\left(\mu(X_i, 1) - \mu_0(X_i)\right),$$

respectively. For $i = 1, \ldots, N$, let

$$B^m_{M,i} = (2W_i - 1)\frac{1}{M}\sum_{m=1}^{M}\big(\mu_{1-W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)})\big)$$

and

$$B^{m,t}_{M,i} = W_i\frac{1}{M}\sum_{m=1}^{M}\big(\mu_0(X_i) - \mu_0(X_{j_m(i)})\big).$$

Finally, let

$$B^m_M = \frac{1}{N}\sum_{i=1}^{N} B^m_{M,i} = \frac{1}{N}\sum_{i=1}^{N}(2W_i - 1)\frac{1}{M}\sum_{m=1}^{M}\Big(\mu_{1-W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)})\Big),$$

and

$$B^{m,t}_M = \frac{1}{N_1}\sum_{i=1}^{N} B^{m,t}_{M,i} = \frac{1}{N_1}\sum_{i=1}^{N} W_i\frac{1}{M}\sum_{m=1}^{M}(\mu_0(X_i) - \mu_0(X_{j_m(i)})).$$

Notice that B_M^m is the conditional bias of $\hat{\tau}_M^m$ with respect to $\overline{\tau(X)}$ given $\mathbf{X}$ and $\mathbf{W}$. Similarly, $B_M^{m,t}$ is the conditional bias of $\hat{\tau}^t$ with respect to $\overline{\tau(X)}^{\,t}$ given $\mathbf{X}$ and $\mathbf{W}$. These conditional bias terms would be equal to zero if matching was exact, that is, if $X_i = X_{j_m(i)}$ for all matched units, i, and $m = 1, \dots, M$. However, because X is continuous, exact matches happen with probability zero and the conditional bias terms are not equal to zero.

Let $\sigma^2(x, w)$ be the variance of Y conditional on $X = x$ and $W = w$. Notice that $K_M(i)$ is given conditional on $\mathbf{X}$ and $\mathbf{W}$. Therefore, equation (2) implies that the variances of $\hat{\tau}_M^m$ and $\hat{\tau}_M^{m,t}$ conditional on $\mathbf{X}$ and $\mathbf{W}$ are

$$\mathbb{V}(\hat{\tau}_M^m|\mathbf{X}, \mathbf{W}) = \frac{1}{N^2} \sum_{i=1}^{N} \left(1 + \frac{K_M(i)}{M}\right)^2 \sigma^2(X_i, W_i)$$

and

$$\mathbb{V}(\hat{\tau}_M^{m,t}|\mathbf{X}, \mathbf{W}) = \frac{1}{N_1^2} \sum_{i=1}^{N} \left(W_i - (1 - W_i)\frac{K_M(i)}{M}\right)^2 \sigma^2(X_i, W_i),$$

respectively. Let $V^E = N \cdot \mathbb{V}(\hat{\tau}_M^m|\mathbf{X}, \mathbf{W})$ and $V^{E,t} = N_1 \cdot \mathbb{V}(\hat{\tau}_M^{m,t}|\mathbf{X}, \mathbf{W})$ be the corresponding normalized variances. In addition, let $V^{\tau(X)} = \mathbb{E}[(\tau(X) - \tau)^2]$ and $V^{\tau(X),t} = \mathbb{E}[(\tau^t(X) - \tau^t)^2 | W = 1]$. In Abadie and Imbens (2006) we prove the following result.

Theorem 1: (Asymptotic Normality for the Simple Matching Estimator)

(i) *Suppose assumptions 1-3 and 4 hold. Then*

$$\left(V^E + V^{\tau(X)}\right)^{-1/2} \sqrt{N}(\hat{\tau}_M^{sm} - B_M^m - \tau) \xrightarrow{d} \mathcal{N}(0, 1).$$

(ii) *Suppose assumptions 1, 2′, 3′, and 4 hold. Then*

$$\left(V^{E,t} + V^{\tau(X),t}\right)^{-1/2} \sqrt{N_1}(\hat{\tau}_M^{sm,t} - B_M^{m,t} - \tau^t) \xrightarrow{d} \mathcal{N}(0, 1).$$

In Abadie and Imbens (2006) we propose consistent estimators of the variance terms in (i) under assumptions 1-3 and 4, and of the variance terms in (ii) under assumptions 1, 2′, 3′, and 4.

The result of Theorem 1 shows that, after subtracting the conditional bias terms B_M^m and $B_M^{m,t}$, simple matching estimators are $N^{1/2}$-consistent and asymptotically normal. Moreover, we show in Abadie and Imbens (2006) that the same result holds without subtracting the conditional bias terms, if matching is done for only one covariate (e.g., matching on the propensity score). However, when the dimension of X is large, the conditional bias terms converge to zero slowly, which makes matching estimators not be $N^{1/2}$-consistent in general.

3. Bias Corrected Matching

In this section we analyze the properties of the bias corrected matching estimators, defined in equation (3). In order to establish the asymptotic behavior of the bias-corrected estimator, we consider a nonparametric series estimator for the two regression functions, $\mu_0(x)$ and $\mu_1(x)$, with $K(N)$ terms in the series, where $K(N)$ increases with N. An important disadvantage of this estimator is that it will rely on selecting smoothing parameters as functions of the sample size, something that the simple matching estimator allows one to avoid. The advantage of the bias-corrected matching estimator is that it is root-N consistent for any dimension of the covariates, k. In both these properties the bias-corrected matching estimator is similar to the regression imputation estimator. However, it has the same large sample variance as the simple matching estimator and, therefore, it is in general not as efficient as the regression imputation estimator in large samples. Compared to the regression imputation estimator the bias-corrected matching estimator is more robust in the sense that the latter is consistent for a fixed value of the smoothing parameters. Because choosing smoothing parameters as functions of the sample size is precisely what matching estimators allow one to avoid, in the empirical analysis and simulations of sections 4 and 5 we will investigate the performance of a simple implementation of the bias correction by linear least squares.

Let $\lambda = (\lambda_1, ..., \lambda_k)$ be a multi-index of dimension k, that is, a k-dimensional vector of non-negative integers, with $|\lambda| = \sum_{i=1}^k \lambda_i$, and let $x^\lambda = x_1^{\lambda_1} \dots x_k^{\lambda_k}$. The λ-th partial derivative of a function $g(x)$ is given by $\partial^{|\lambda|} g(x)/\partial x_1^{\lambda_1} ... \partial x_k^{\lambda_k}$. Consider a series $\{\lambda(r)\}_{r=1}^\infty$ containing all distinct such vectors such that $|\lambda(r)|$ is nondecreasing. Let $p_r(x) = x^{\lambda(r)}$, where $p^K(x) = (p_1(x), ..., p_K(x))'$. Following Newey (1995), the nonparametric series estimator of the regression function $\mu_w(x)$ is given by:

$$\hat{\mu}_w(x) = p^{K(N)}(x)' \left(\sum_{i:W_i=w} p^{K(N)}(X_i) p^{K(N)}(X_i)' \right)^- \sum_{i:W_i=w} p^{K(N)}(X_i) Y_i,$$

where $(\cdot)^-$ denotes a generalized inverse. Given the estimated regression function, let $\hat{B}_M^{sm}$ be the estimated bias term:

$$\hat{B}_M^m = \frac{1}{N} \sum_{i=1}^N \left\{ W_i \left(\frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} \Big(\hat{\mu}_0(X_i) - \hat{\mu}_0(X_j) \Big) \right) - (1 - W_i) \left(\frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} \Big(\hat{\mu}_1(X_i) - \hat{\mu}_1(X_j) \Big) \right) \right\},$$

and

$$\hat{B}_M^{m,t} = \frac{1}{N_1} \sum_{i=1}^{N} W_i \left(\frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} \left(\hat{\mu}_0(X_i) - \hat{\mu}_0(X_j) \right) \right),$$

for the average effect for the treated, so that $\hat{\tau}_M^{bcm} = \hat{\tau}_M^m - \hat{B}_M^m$, and $\hat{\tau}_M^{bcm,t} = \hat{\tau}_M^{m,t} - \hat{B}_M^{m,t}$. The following theorem shows that the bias correction removes the bias without affecting the asymptotic variance.

THEOREM 2: (BIAS CORRECTED MATCHING ESTIMATOR)
Suppose that Assumptions 1 to 4 hold. Assume also that (i) the support of X, $\mathbb{X} \subset \mathbb{R}^k$, is a Cartesian product of compact intervals; (ii) $K(N) = O(N^\nu)$, with $\nu > 0$, $\nu < 2/(4k+3)$, and $\nu < 2/(4k^2 - k)$; (iii) there is a constant C such that for each multi-index λ the λ-th partial derivative of $\mu_w(x)$ exists for $w = 0, 1$ and its norm is bounded by $C^{|\lambda|}$. Then,

$$\sqrt{N} \left(B_M^m - \hat{B}_M^m \right) \xrightarrow{p} 0 \quad \textit{and} \quad \left(V^E + V^{\tau(X)} \right)^{1/2} \sqrt{N} (\hat{\tau}_M^{bcm} - \tau) \xrightarrow{d} \mathcal{N}(0, 1).$$

If assumptions 1, 2', 3' and 4 hold in addition to (i)-(iii), we obtain:

$$\sqrt{N_1} \left(B_M^{m,t} - \hat{B}_M^{m,t} \right) \xrightarrow{p} 0 \quad \textit{and} \quad \left(V^{E,t} + V^{\tau(X),t} \right)^{1/2} \sqrt{N_1} (\hat{\tau}_M^{bcm,t} - \tau^t) \xrightarrow{d} \mathcal{N}(0, 1).$$

Thus, the bias corrected matching estimator has the same normalized variance as the simple matching estimator.[3]

4. An Application to the Evaluation of a Labor Market Program

In this section we apply the estimators studied in this article to data from the National Supported Work Demonstration (NSW), an evaluation of a subsidized work program first analyzed by Lalonde (1986) and subsequently by Heckman and Hotz (1989), Dehejia and Wahba (1999) and Smith and Todd (2001). The specific sample we use here is the one employed by Dehejia and Wahba (1999). The data set contains an experimental sample from a randomized evaluation of the NSW program, and also a nonexperimental sample from the Panel Study of Income Dynamics

[3] It is easy to check that the same result holds if $\mu_w(x)$ has a finite series representation which is estimated by least squares. This feature will be used later to construct standard errors for the bias corrected estimator when the bias correction is implemented using a simple linear regression.

(PSID). Using the experimental data we obtain an unbiased estimate of the average effect of the program. We then compute non-experimental matching estimators using the experimental participants and the nonexperimental comparison group from the PSID, and compare them to the experimental estimate. In line with previous studies using these data, we focus on the average effect for the treated and therefore only match the treated units.

Table 1 presents summary statistics for the three groups used in our analysis. The first two columns present the summary statistics for the experimental treatment group. The second pair of columns presents summary statistics for the experimental controls. The third pair of columns presents summary statistics for the non-experimental comparison group constructed from the PSID. The last two columns present t-statistics for the hypotheses that the differences in population averages for the experimental treated and controls, and for the experimental treated and the PSID comparison group, respectively, are zero. Panel A contains the results for pretreatment variables and Panel B for outcomes. Notice the large differences in background characteristics between the program participants and the PSID sample. This is what makes drawing causal inferences from comparisons between the PSID sample and the treatment group a tenuous task. From Panel B, we can obtain an unbiased estimate of the effect of the NSW program on earnings in 1978 by comparing the averages for the experimental treated and controls, $6.35 - 4.55 = 1.80$, with a standard error of 0.67 (earnings are measured in thousands of dollars). Using a normal approximation to the limiting distribution of the effect of the program on earnings in 1978, we obtain a 95% confidence interval, which is [0.49, 3.10].

Table 2 presents estimates of the causal effect of the NSW program on earnings using various matching and regression adjustment estimators. Panel A reports estimates for the experimental data (treated and controls). Panel B reports estimates based on the experimental treated and the PSID comparison group. The first set of rows in each case reports matching estimates for M equal to 1, 4, 16, 64 and 2490 (the size of the PSID comparison group).[4]

[4]The matching estimates include simple matching with no bias adjustment, and bias-adjusted matching. All matching estimators use the Euclidean norm to measure the distance between different values for the covariates, after normalizing the covariates to have zero mean and unit variance. The bias adjustment uses linear regression on the nine pretreatment covariates in Table 1, panel A, but not higher order terms or interactions. The bias correction is estimated using only the matched units in the comparison group. The confidence intervals are based on variance estimator proposed in Abadie and Imbens (2006). The last three rows of each panel report estimates based on differences in means, linear regression including terms for all covariates, and linear regression including

The experimental estimates range from 1.16 (bias corrected matching with one match) to 2.27 (quadratic regression). The non-experimental estimates have a much wider range, from -15.20 (simple difference) to 3.26 (quadratic regression). For the non-experimental sample, using a single match, there is little difference between the simple matching estimator and its bias-corrected version, 2.07 and 2.42 respectively. However, simple matching, without bias-correction, produces radically different estimates when the number of matches changes, a troubling result for the empirical implementation of these estimators. With $M \geq 16$ the simple matching estimator produces results outside the experimental 95% confidence interval. In contrast, the bias-corrected matching estimator shows a much more robust behavior when the number of matches changes: only with $M = 2490$ (that is, when all units in the comparison group are matched to each treated) the bias-corrected estimate deteriorates to 0.84, still inside the experimental 95% confidence interval.

To see how well the simple matching estimator performs in terms of balancing the covariates, Table 3 reports average differences within the matched pairs. First, all the covariates are normalized to have zero mean and unit variance. The first two columns report the averages of the normalized covariates for the PSID comparison group and the experimental treated. Before matching, the averages for some of the variables are more than one standard deviation apart, e.g., the earnings and employment variables. The next pair of columns reports the within-matched-pairs average difference and the standard deviation of this within-pair difference. For all the indicator variables the matching is exact. The other, more continuously distributed variables are not matched exactly, but the quality of the matches appears very high: the average difference within the pairs is very small compared to the average difference between treated and comparison units before the matching, and it is also small compared to the standard deviations of these differences. If we increase the number of matches the quality of the matches goes down, with even the indicator variables no longer matched exactly, but in most cases the average difference is still very small until we get to 16 or more matches. As expected, match quality deteriorates when the number of matches increases. This explains why, as shown in Table 2, the bias-correction matters more for larger M. The last row reports matching differences for logistic estimates of

also quadratic terms and a full set of interactions, respectively.

the propensity score. Although the matching is not directly on the propensity score, with single matches the average difference in the propensity score is only 0.21, whereas without matching the difference between treated and comparison units is 8.16, almost 40 times higher.

5. A Monte Carlo Study

In this section, we discuss some simulations designed to assess the performance of the various matching estimators. To mimic as closely as possible the behavior of matching estimators in real applications, we simulated data sets that aim to resemble the NSW data set analyzed in the previous section.

In the simulation we have nine regressors, designed to match the following variables in the NSW data set: age, education, black, hispanic, married, earnings1324, unemployed1324, earnings1975, unemployed1975. For each simulated data set we sampled with replacement 185 observations from the empirical covariate distribution of the experimental treated, and 2490 observations from the empirical covariate distribution of the PSID comparison group. This gives us the joint distribution of covariates and treatment indicators. For the conditional distribution of the outcome given covariates, we estimated a two-part model on the PSID comparison group, where the probability of zero earnings is a logistic function of the covariates with a full set of quadratic terms and interactions.[5] Conditional on log earnings being positive, the log of earnings is modeled as a linear function of the covariates with again a full set of quadratic terms and interactions. We fix the treatment effect at 2.0 for all units.

We performed 10,000 replications. For each estimator we report the mean and median bias, the root-mean-squared-error (rmse), the median-absolute-error (mae), the standard deviation, the average estimated standard error, and the coverage rates for nominal 95% and 90% confidence intervals based on the matching estimator for the variance. We implemented an extremely simple version of the bias adjustment, using only linear terms in the covariates. The results are reported in Table 4.

In terms of rmse and mae, the bias-adjusted matching estimator is best with 4 or 16 matches.

[5] With the nine variables, this gives 54 covariates in the set with all linear terms, quadratic terms and interactions, not including the intercept. Out of these 54 we drop 8 because they are perfectly collinear by definition (e.g., the interaction of earnings 1975 and unemployed 1975).

The simple matching estimator does not perform as well, in terms of bias or rmse. The pure regression adjustment estimators perform poorly. They have high rmse and substantial bias. Coverage rates of confidence intervals centered on the bias-corrected matching estimator are also much more robust than those centered on the simple matching estimator. Confidence intervals for the regression estimators have considerably lower than nominal coverage rates.

6. Conclusion

We propose a nonparametric bias-adjustment that renders matching estimators $N^{1/2}$-consistent. In simulations based on a realistic setting for nonexperimental program evaluations, a simple implementation of this estimator, where the bias-adjustment is based on linear regression, performs well compared to both matching estimators without bias-adjustment and regression-based estimators in terms of bias and mean-squared error. It also has good coverage rates for 90% and 95% confidence intervals, suggesting it may be a useful estimator in practice.

Appendix

Before proving Theorem 2 we state two auxiliary lemmas. Let λ be a multi-index of dimension k, that is, an k-dimensional vector of non-negative integers, with $|\lambda| = \sum_{i=1}^{k} \lambda_i$, and let Λ_l be the set of λ such that $|\lambda| = l$. Furthermore, let $x^\lambda = x_1^{\lambda_1} \dots x_k^{\lambda_k}$, and let $\partial^\lambda g(x) = \partial^{|\lambda|} g(x)/\partial x_1^{\lambda_1} \dots \partial x_k^{\lambda_k}$. For $d \geq 0$, define $|g|_d = \max_{|\lambda \leq d|} \sup_x |\partial^\lambda g(x)|$.

Lemma A.1: (Uniform Convergence of Series Estimators of Regression Functions, Newey 1995)
Suppose the conditions in Theorem 2 hold. Then for any $\xi > 0$ and non-negative integer d,

$$|\hat{\mu}_w - \mu_w|_d = O_p\left(K^{1+2d}\left((K/N)^{1/2} + K^{-\xi}\right)\right),$$

for $w = 0, 1$.

Proof: Assumptions 3.1, 4.1, 4.2 and 4.3 in Newey (1995) are satisfied for $\mu_w(x)$ and $N_w \to \infty$, implying that Newey's Theorems 4.2 and 4.4 apply. The result of the lemma holds because $N/N_w = O_p(1)$ for $w = 0, 1$. □

Lemma A.2: (Unit-level Bias Correction)
Suppose the conditions in Theorem 2 hold. Then

$$\max_{i=1,\dots N} |\hat{\mu}_w(X_i) - \hat{\mu}_w(X_{j_m(i)}) - \left(\mu_w(X_i) - \mu_w(X_{j_m(i)})\right)| = o_p(N^{-1/2}),$$

for $w = 0, 1$.

Proof: Let $U_{m,i} = X_{j_m(i)} - X_i$. Use a Taylor series expansion around X_i to write

$$\left|\mu_w(X_{j_m(i)}) - \mu_w(X_i) - \sum_{1 \leq l \leq k-1} \frac{1}{l!} \sum_{\lambda \in \Lambda_l} \partial^\lambda \mu_w(X_i) U_{m,i}^\lambda\right| \leq \frac{C^k}{k!} \sum_{\lambda \in \Lambda_k} |U_{m,i}^\lambda| \leq \frac{C^k}{k!} \sum_{\lambda \in \Lambda_k} \|U_{m,i}\|^k.$$

Because all moments of $N_{1-W_i}^{1/k} \|U_{m,i}\|$ and N/N_{1-W_i} are uniformly bounded, applying Bonferroni's and Markov's inequalities, we obtain that for any $\varepsilon > 0$:

$$\max_{i=1,\dots,N} \left|\mu_w(X_{j_m(i)}) - \mu_w(X_i) - \sum_{1 \leq l \leq k-1} \frac{1}{l!} \sum_{\lambda \in \Lambda_l} \partial^\lambda \mu_w(X_i) U_{m,i}^\lambda\right| = o_p(N^{-1+\varepsilon}).$$

Because we can choose $\varepsilon \leq 1/2$, it follows that the left hand side of last equation is $o_p(N^{-1/2})$. Similarly, for any $\varepsilon > 0$:

$$\left|\hat{\mu}_w(X_{j_m(i)}) - \hat{\mu}_w(X_i) - \sum_{1 \leq l \leq k-1} \frac{1}{l!} \sum_{\lambda \in \Lambda_l} \partial^\lambda \hat{\mu}_w(X_i) U_{m,i}^\lambda\right| \leq \frac{1}{k!} \sum_{\lambda \in \Lambda_k} |\hat{\mu}_w - \mu_w|_k \|U_{m,i}\|^k + \frac{C^k}{k!} \sum_{\lambda \in \Lambda_k} \|U_{m,i}\|^k.$$

Therefore, for arbitrary $\xi > 0$ and $\varepsilon > 0$:

$$\max_{i=1,\ldots,N} \left| \hat{\mu}_w(X_{j_m(i)}) - \hat{\mu}_w(X_i) - \sum_{1\le l\le k-1} \frac{1}{l!} \sum_{\lambda\in\Lambda_l} \partial^\lambda \hat{\mu}_w(X_i) U^\lambda_{m,i} \right|$$
$$= O_p\left(K^{1+2k}\left((K/N)^{1/2} + K^{-\xi}\right)\right) o_p\left(N^{-1+\varepsilon}\right) + o_p(N^{-1+\varepsilon}).$$

Because $\nu < 2/(4k+3)$, we can choose ξ and ε so that the left hand side of last equation becomes $o_p(N^{-1/2})$. Therefore,

$$\max_{i=1\ldots,N} |\hat{\mu}_w(X_{j_m(i)}) - \hat{\mu}_w(X_i) - \left(\mu_w(X_{j_m(i)}) - \mu_w(X_i)\right)|$$
$$\le \max_{i=1\ldots,N} \sum_{1\le l\le k-1} \frac{1}{l!} \sum_{\lambda\in\Lambda_l} \left|\partial^\lambda \hat{\mu}_w(X_i) - \partial^\lambda \mu_w(X_i)\right| \cdot \left|U^\lambda_{m,i}\right| + o_p(N^{-1/2})$$
$$\le |\hat{\mu}_w - \mu_w|_{k-1} \sum_{1\le l\le k-1} \frac{1}{l!} \sum_{\lambda\in\Lambda_l} \max_{i=1\ldots,N} \|U_{m,i}\|^{|\lambda|} + o_p(N^{-1/2})$$
$$= O_p\left(K^{2k-1}\left((K/N)^{1/2} + K^{-\xi}\right)\right) o_p\left(N^{-1/k+\varepsilon}\right) + o_p(N^{-1/2}),$$

for arbitrary $\xi > 0$ and $\varepsilon > 0$. Now, it can be easily seen that $\nu < 2/(4k^2 - k)$ guarantees that the result of Lemma A.2 holds. □

PROOF OF THEOREM 2:
We focus on the result for the average treatment effect. The second part of the theorem for the average effect for the treated follows the same pattern. The difference $|\hat{B}^m_M - B^m_M|$ can be written as

$$\left|\hat{B}^m_M - B^m_M\right| \le \frac{1}{N}\sum_{i=1}^{N} \frac{1}{M}\sum_{i=1}^{M} \left|\hat{\mu}_{1-W_i}(X_i) - \hat{\mu}_{1-W_i}(X_{j_m(i)}) - \left(\mu_{1-W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)})\right)\right|$$
$$\le \max_{i=1,\ldots,N} \sum_{w=0,1} \left|\hat{\mu}_w(X_i) - \hat{\mu}_w(X_{j_m(i)} - \left(\mu_w(X_i) - \mu_w(X_{j_m(i)})\right)\right| = o_p(N^{-1/2}),$$

by Lemma A.2. □

References

Abadie, A., (2005) "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, 72, 1-19.

Abadie, A. and G. Imbens, (2006) "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74(1), 235-267.

Chen, X., H. Hong, and A. Tarozzi (2004) "Semiparametric Efficiency in GMM Models of Nonclassical Measurement Error, Missing Data and Treatment Effects," unpublished manuscript.

Dehejia, R., and S. Wahba, (1999) "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94, 1053-1062.

Hahn, J., (1998) "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66 (2), 315-331.

Heckman, J., and J. Hotz, (1989) "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," (with discussion), *Journal of the American Statistical Association*, 84, 862-874.

Heckman, J., H. Ichimura, and P. Todd, (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, 64, 605-654.

Heckman, J., H. Ichimura, J. Smith, and P. Todd, (1998) "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017-1098.

Hirano, K., G. Imbens, and G. Ridder, (2003) "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161-1189.

Imbens, G., (2003) "Sensitivity to Exogeneity Assumptions in Program Evaluation," *American Economic Review Papers and Proceedings*, 93(2), 126-132.

Imbens, G., (2004) "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4-30.

Imbens, G., W. Newey and G. Ridder (2005) "Mean-squared-error Calculations for Average Treatment Effects," unpublished manuscript.

Lalonde, R.J., (1986) "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76, 604-620.

Lechner, M., (2002) "Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programmes by Matching Methods," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 165, 59-82.

Newey, W., (1995) “Convergence Rates for Series Estimators,” in G.S. Maddala, P.C.B. Phillips and T.N. Srinavasan eds. Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao. Cambridge: Blackwell.

Quade, D., (1982) “Nonparametric Analysis of Covariance by Matching”, *Biometrics*, 38, 597-611.

Rubin, D., (1973b) “The Use of Matched Sampling and Regression Adjustments to Remove Bias in Observational Studies”, *Biometrics*, 29, 185-203.

Rubin, D., (1974) “Estimating causal effects of treatments in randomized and nonrandomized studies”, *Journal of Educational Psychology*, 66, 688-701.

Rubin, D., (1979) “Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies”, *Journal of the American Statistical Association*, 74, 318-328.

Smith, J. and P. Todd, (2001) “Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods,” *American Economic Review*, Papers and Proceedings, 91, 112-118.

Smith, J. and P. Todd, (2005) “Does Matching Address LaLonde’s Critique of Nonexperimental Estimators,” *Journal of Econometrics*, 125, 305-353.

Table 1
Summary Statistics

	Experimental Data				PSID		T-statistic	
	Treated (185 obs.)		Controls (260 obs.)		(2490 obs.)		Treat/	Treat/
	mean	(s.d.)	mean	(s.d.)	mean	(s.d.)	Contr	PSID
Panel A: Pretreatment Variables								
Age	25.8	(7.16)	25.05	(7.06)	34.85	(10.44)	[1.1]	[-16.0]
Education	10.4	(2.01)	10.09	(1.61)	12.12	(3.08)	[1.4]	[-11.1]
Black	0.84	(0.36)	0.83	(0.38)	0.25	(0.43)	[0.5]	[21.0]
Hispanic	0.06	(0.24)	0.11	(0.31)	0.03	(0.18)	[-1.9]	[1.5]
Married	0.19	(0.39)	0.15	(0.36)	0.87	(0.34)	[1.0]	[-22.8]
Earnings 13-24	2.10	(4.89)	2.11	(5.69)	19.43	(13.41)	[-0.0]	[-38.6]
Unemployed 13-24	0.71	(0.46)	0.75	(0.43)	0.09	(0.28)	[-1.0]	[18.3]
Earnings '75	1.53	(3.22)	1.27	(3.10)	19.06	(13.60)	[0.9]	[-48.6]
Unemployed '75	0.60	(0.49)	0.68	(0.47)	0.10	(0.30)	[-1.8]	[13.8]
Panel B: Outcomes								
Earnings '78	6.35	(7.87)	4.55	(5.48)	21.55	(15.56)	[2.7]	[-23.1]
Unemployed '78	0.24	(0.43)	0.35	(0.48)	0.11	(0.32)	[-2.7]	[4.0]

Note: Earnings data are in thousands of 1978 dollars. Earnings 13-24 and Unemployed 13-24 refers to earnings and unemployment during the period 13 to 24 months prior to randomization.

Table 2
Experimental and Non-experimental Estimates for the NSW Data

	$M=1$		$M=4$		$M=16$		$M=64$		$M=2490$	
	est	(s.e.)	est	(s.e.)	est	(s.e.)	est	(s.e.)	est	(s.e.)
Panel A: Experimental Estimates										
simple matching	1.22	(0.84)	1.99	(0.74)	1.75	(0.74)	2.20	(0.70)	1.80	(0.67)
bias-adjusted matching	1.16	(0.84)	1.84	(0.74)	1.54	(0.75)	1.74	(0.71)	1.72	(0.68)
Regression Estimates										
mean difference	1.79	(0.67)								
linear	1.72	(0.65)								
quadratic	2.27	(0.73)								
Panel B: Non-experimental Estimates										
simple matching	2.07	(1.13)	1.62	(0.91)	0.47	(0.85)	-0.11	(0.75)	-15.20	(0.61)
bias-adjusted matching	2.42	(1.13)	2.51	(0.90)	2.48	(0.83)	2.26	(0.71)	0.84	(0.63)
Regression Estimates										
mean difference	-15.20	(0.66)								
linear	0.84	(0.86)								
quadratic	3.26	(0.98)								

Note: The outcome is earnings in 1978 in thousands of dollars.

Table 3
Mean Covariate Differences in Matched Groups

	Average		$M=1$		$M=4$		$M=16$		$M=64$		$M=2490$	
	PSID	Treated	mean	(s.d.)	mean	(s.d.)	mean	(s.d.)	mean	(s.d.)	mean	(s.d.)
Age	0.06	-0.80	-0.02	(0.65)	-0.06	(0.60)	-0.30	(0.41)	-0.57	(0.57)	-0.86	(0.68)
Education	0.04	-0.54	-0.10	(0.44)	-0.20	(0.48)	-0.25	(0.39)	-0.24	(0.42)	-0.58	(0.66)
Black	-0.09	1.21	-0.00	(0.00)	0.09	(0.32)	0.35	(0.47)	0.70	(0.66)	1.30	(0.80)
Hispanic	-0.01	0.14	-0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.01	(0.03)	0.15	(1.30)
Married	0.12	-1.64	0.00	(0.00)	-0.06	(0.30)	-0.33	(0.46)	-0.90	(0.85)	-1.76	(1.02)
Earnings 13-24	0.09	-1.18	-0.01	(0.10)	-0.01	(0.12)	-0.05	(0.17)	-0.15	(0.30)	-1.26	(0.36)
Unemployed 13-24	-0.13	1.72	0.00	(0.00)	0.02	(0.17)	0.24	(0.40)	0.41	(0.72)	1.85	(1.36)
Earnings '75	0.09	-1.18	-0.04	(0.17)	-0.07	(0.15)	-0.11	(0.19)	-0.19	(0.26)	-1.26	(0.23)
Unemployed '75	-0.10	1.36	0.00	(0.00)	0.00	(0.05)	0.03	(0.28)	0.10	(0.41)	1.46	(1.44)
Log Odds												
Prop Score	-7.08	1.08	0.21	(0.99)	0.56	(1.13)	1.70	(1.14)	3.20	(1.49)	8.16	(2.13)

Note: In this table all covariates have been normalized to have mean zero and unit variance. The first two columns present the averages for the experimental treated and the PSID comparison units. The remaining pairs of columns present the average difference within the matched pairs and the standard deviation of this difference for matching based on 1, 4, 16, 64 and 2490 matches. For the last variable the logarithm of the odds ratio of the propensity score is used. This log odds ratio has mean -6.52 and standard deviation 3.30 in the sample.

Table 4
Simulation Results (10,000 replications)

M	Estimator	mean bias	median bias	rmse	mae	s.d.	mean s.e.	coverage rate (nom. 95%)	coverage rate (nom. 90%)
1	simple matching	-0.49	-0.44	0.87	0.53	0.73	0.88	0.94	0.90
	linear bias-adjusted	0.05	0.08	0.73	0.47	0.73	0.89	0.96	0.94
4	simple matching	-0.83	-0.82	1.03	0.84	0.59	0.63	0.75	0.62
	linear bias-adjusted	0.05	0.07	0.61	0.39	0.60	0.64	0.95	0.91
16	simple matching	-1.81	-1.74	1.88	1.79	0.57	0.53	0.08	0.04
	linear bias-adjusted	0.19	0.18	0.63	0.41	0.60	0.53	0.90	0.84
64	simple matching	-3.17	-3.24	3.33	3.25	0.61	0.52	0.00	0.00
	linear bias-adjusted	0.17	0.17	0.67	0.44	0.65	0.52	0.87	0.80
	mean difference	-19.06	-19.09	19.06	19.09	0.61	1.63	0.00	0.00
	linear regression	-2.04	-2.06	2.26	2.06	1.00	0.98	0.44	0.33
	quadratic regression	2.72	2.65	3.01	2.65	1.35	1.24	0.40	0.27