

# A MARTINGALE REPRESENTATION FOR MATCHING ESTIMATORS

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

February 2009

## ABSTRACT

Matching estimators are widely used in statistical data analysis. However, the distribution of matching estimators has been derived only for particular cases (Abadie and Imbens, 2006). This article establishes a martingale representation for matching estimators. This representation allows the use of martingale limit theorems to derive the asymptotic distribution of matching estimators. As an illustration of the applicability of the theory, we derive the asymptotic distribution of a matching estimator when matching is carried out without replacement, a result previously unavailable in the literature.

JEL Classification: C13, C14, C21

Keywords: *Matching, Martingales, Treatment Effects, Hot-Deck Imputation*

## I. INTRODUCTION

Matching methods provide simple and intuitive tools for adjusting the distribution of covariates among samples from different populations. Probably because of their transparency and intuitive appeal, matching methods are widely used in evaluation research to estimate treatment effects when all treatment confounders are observed (Rubin, 1973, 1977; Rosenbaum, 2002). In spite of the popularity of matching methods, the asymptotic distribution of matching estimators have been derived only for special cases (Abadie and Imbens, 2006). In the absence of large sample approximation results to the distribution of matching estimators, empirical researchers employing matching methods have often used the bootstrap as a basis for inference. However, recent results have shown that, in general, the bootstrap does not provide valid large sample inference for matching estimators (Abadie and Imbens, 2008).

The main contribution of this article is to establish a martingale representation for matching estimators. This representation allows the use of martingale limit theorems (Hall and Heyde, 1980; Billingsley, 1995; Shorack, 2000) to derive the asymptotic distribution of matching estimators. Because the martingale representation applies to a large class of matching estimators, the applicability of the methods presented in this article is broad. As an illustration of the theory, we apply the martingale methods proposed in this paper to derive the asymptotic distribution of a matching estimator when matching is carried out without replacement, a result previously unavailable in the literature. Despite its simplicity and immediate implications, the martingale representation of matching estimators described in this article seems to have been previously unnoticed in the literature.

## II. MATCHING ESTIMATORS

Empirical researchers often compare the distributions of some variable,  $Y$ , between two groups of units after taking into account the confounding effects of a  $(k \times 1)$  vector of observed covariates,  $X$ . Let  $W$  be a binary variable that indicates membership to a particular population of interest. For example, in discrimination litigation research,  $W$  may

represent membership in a certain demographic group,  $Y$  may represent labor wages, and  $X$  may represent a vector of variables influencing workers productivity, like education or tenure. In evaluation research,  $W$  typically indicates exposure to an active treatment or intervention,  $Y$  is an outcome of interest, and  $X$  is a vector of observed confounders. Let

$$\tau = E[Y|W = 1] - E\left[E[Y|X, W = 0] \middle| W = 1\right].$$

In evaluation research,  $\tau$  is given a causal interpretation as the “average treatment effect on the treated” under unconfoundedness assumptions (Rubin, 1977). Other parameters of interest can be estimated using matching methods. These include the “average treatment effect”, which is of widespread interest in evaluation studies, as well as parameters that focus on features of the distribution of  $Y$  other than the mean (see Imbens, 2004, and Imbens and Wooldridge, 2008, for detailed reviews of the literature). Matching is also used for the analysis of missing data, where it is often referred to as “hot-deck imputation” (Little and Rubin, 2002). For concreteness, and to avoid tedious repetition or unnecessary abstraction, we will discuss matching estimation of  $\tau$  only. However, the techniques proposed in this paper are of immediate application to the estimation of parameters other than  $\tau$  via matching.

Consider random samples of sizes  $N_0$  and  $N_1$  of untreated and treated units, respectively. Pooling together these two samples, we obtain a sample of size  $N = N_0 + N_1$  that contains treated and untreated units. For each unit in the pooled sample we observe the triple  $(Y, X, W)$ . For each treated unit  $i$ , let  $\mathcal{J}_M(i)$  be the indices of  $M$  untreated units with values in the covariates similar to  $X_i$  (where  $M$  is some small positive integer). In other words,  $\mathcal{J}_M(i)$  is a set of  $M$  matches for observation  $i$ . To simplify notation, we will assume that at least one of the variables in the vector  $X$  has a continuous distribution, so perfect matches happen with probability zero. Let  $\|\cdot\|$  be some norm in  $\mathbb{R}^k$  (typically the Euclidean norm). Let  $1_A$  be the indicator function for the event  $A$ . For matching with replacement,

$$\mathcal{J}_M(i) = \left\{ j \in \{1, \dots, N\} \text{ s.t. } W_j = 0, \left( \sum_{k=1}^N (1 - W_k) 1_{\{\|X_i - X_j\| \leq \|X_i - X_k\|\}} \right) \leq M \right\}.$$

For matching without replacement, the elements of  $\{\mathcal{J}_M(i) \text{ s.t. } W_i = 1\}$  are non-overlapping

subsets of  $\{j \in \{1, \dots, N\} \text{ s.t. } W_j = 0\}$  chosen to minimize the sum of the matching discrepancies:

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

In both cases, the matching estimator of  $\tau$  is defined as:

$$\hat{\tau} = \frac{1}{N_1} \sum_{i=1}^N W_i \left( Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \right).$$

Other matching schemes are possible (see, Rosenbaum, 2002; Hansen, 2004; Diamond and Sekhon, 2008) and the results in this article are of broad generality. Notice that in this article we reserve the term “matching” for procedures that use a small number,  $M$ , of matches. Estimators that treat the number of matches as a function of the sample size (with  $M \rightarrow \infty$  as  $N \rightarrow \infty$ ) have been proposed by Heckman, Ichimura, and Todd (1998) and others. These estimators have asymptotically linear representations, so their large sample distributions can be derived using the standard machinery for asymptotically linear estimators.

### III. A MARTINGALE REPRESENTATION FOR MATCHING ESTIMATORS

This section derives a martingale representation for matching estimators. For  $w \in \{0, 1\}$ , let  $\mu_w(x) = E[Y|X = x, W = w]$  and  $\sigma_w^2(x) = \text{var}(Y|X = x, W = w)$ . Assume that these functions are bounded. Abadie and Imbens (2006) derive the following decomposition for matching estimators:

$$\begin{aligned} \hat{\tau} - \tau &= \frac{1}{N_1} \sum_{i=1}^N W_i \left( Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j - \tau \right) \\ &= \frac{1}{N_1} \sum_{i=1}^N W_i \left( (Y_i - \mu_1(X_i)) - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} (Y_j - \mu_0(X_j)) \right) \\ &\quad + \frac{1}{N_1} \sum_{i=1}^N W_i \left( \mu_1(X_i) - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} \mu_0(X_j) - \tau \right) \\ &= \frac{1}{N_1} \sum_{i=1}^N W_i \left( (Y_i - \mu_1(X_i)) - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} (Y_j - \mu_0(X_j)) \right) \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{N_1} \sum_{i=1}^N W_i (\mu_1(X_i) - \mu_0(X_i) - \tau) \\
& + \frac{1}{N_1} \sum_{i=1}^N W_i \left( \mu_0(X_i) - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} \mu_0(X_j) \right).
\end{aligned}$$

That is,  $\hat{\tau} - \tau = D_N + R_N$ , where

$$\begin{aligned}
D_N &= \frac{1}{N_1} \sum_{i=1}^N W_i (\mu_1(X_i) - \mu_0(X_i) - \tau) \\
& + \frac{1}{N_1} \sum_{i=1}^N W_i \left( (Y_i - \mu_1(X_i)) - \frac{1}{M} \sum_{j \in \mathcal{J}(i)} (Y_j - \mu_0(X_j)) \right),
\end{aligned}$$

and

$$R_N = \frac{1}{N_1} \sum_{i=1}^N W_i \left( \mu_0(X_i) - \frac{1}{M} \sum_{j \in \mathcal{J}(i)} \mu_0(X_j) \right).$$

The term  $R_N$  is the conditional bias of matching estimator described in Abadie and Imbens (2004). This term is zero if matches are perfect (that is, if all matching discrepancies,  $X_i - X_j$  for  $j \in \mathcal{J}(i)$ , are zero), or if the regression  $\mu_0$  is a constant function. In general, however, this term is different from zero, as perfect matches happen with probability zero for continuous covariates. The order of magnitude of  $R_N$  depends on the number of continuous covariates, as well as the magnitude of  $N_0$  relative to  $N_1$ . Under appropriate conditions  $\sqrt{N_1} R_N$  converges in probability to zero (see section IV, or Abadie and Imbens, 2004, for the case of matching with replacement).

Next, it will be shown that the term  $D_N$  is a martingale array with respect to certain filtration. First notice that:

$$\begin{aligned}
D_N &= \frac{1}{N_1} \sum_{i=1}^N W_i (\mu_1(X_i) - \mu_0(X_i) - \tau) \\
& + \frac{1}{N_1} \sum_{i=1}^N \left( W_i - (1 - W_i) \frac{K_{N,i}}{M} \right) (Y_i - \mu_{W_i}(X_i)),
\end{aligned}$$

where  $K_{N,i}$  is the number of times that observation  $i$  (with  $W_i = 0$ ) is used as a match.

Therefore, we can write:

$$\sqrt{N_1} D_N = \sum_{k=1}^{2N} \xi_{N,k},$$

where

$$\xi_{N,k} = \begin{cases} \frac{1}{\sqrt{N_1}} W_k (\mu_1(X_k) - \mu_0(X_k) - \tau) & \text{if } 1 \leq k \leq N, \\ \frac{1}{\sqrt{N_1}} \left( W_{k-N} - (1 - W_{k-N}) \frac{K_{N,k-N}}{M} \right) (Y_{k-N} - \mu_{W_{k-N}}(X_{k-N})) & \text{if } N+1 \leq k \leq 2N. \end{cases}$$

Let  $\mathbf{X}_N = \{X_1, \dots, X_N\}$  and  $\mathbf{W}_N = \{W_1, \dots, W_N\}$ . Consider the  $\sigma$ -fields  $\mathcal{F}_{N,k} = \sigma\{\mathbf{W}_N, X_1, \dots, X_k\}$  for  $1 \leq k \leq N$  and  $\mathcal{F}_{N,k} = \sigma\{\mathbf{W}_N, \mathbf{X}_N, Y_1, \dots, Y_{k-N}\}$  for  $N+1 \leq k \leq 2N$ .

Then,

$$\left\{ \sum_{j=1}^i \xi_{N,j}, \mathcal{F}_{N,i}, 1 \leq i \leq 2N \right\}$$

is a martingale for each  $N \geq 1$ . As a result, the asymptotic behavior of  $\sqrt{N_1}D_N$  can be analyzed using martingale methods. Analogous martingale representations hold for alternative matching estimators. Regardless of the choice of matching scheme, a martingale representation holds for  $\sqrt{N_1}D_N$ . The reason is that no matter how matching is implemented, the number of times that unit  $k$  is used as a match,  $K_{N,k}$ , is given conditional on  $\mathbf{X}_N$  and  $\mathbf{W}_N$ , and  $E[Y_k - \mu_{W_k}(X_k) | \mathbf{X}_N, \mathbf{W}_N, Y_1, \dots, Y_{k-1}] = 0$ .

#### IV. APPLICATION: MATCHING WITHOUT REPLACEMENT

In this section, we demonstrate how to apply the martingale representation of matching estimators to derive the asymptotic distribution when matching is done without replacement. To simplify the calculations, we assume that conditional on  $W_i$  the variance of  $Y_i$  does not depend on  $X_i$ . That is,  $\sigma_w^2(x) = \sigma_w^2$  for all  $x$  in the support of  $X$  and  $w \in \{0, 1\}$ . Also, to simplify the exposition, we will concentrate on the case of one-to-one matching ( $M = 1$ ).

The conditional variances of the martingale differences are given by:

$$\begin{aligned} E[\xi_{N,k}^2 | \mathcal{F}_{N,k-1}] &= \frac{1}{N_1} W_k E[(\mu_1(X_k) - \mu_0(X_k) - \tau)^2 | \mathcal{F}_{N,k-1}] \\ &= \frac{1}{N_1} W_k E[(\mu_1(X_k) - \mu_0(X_k) - \tau)^2 | W_k = 1] \end{aligned}$$

for  $1 \leq k \leq N$  and

$$E[\xi_{N,k}^2 | \mathcal{F}_{N,k-1}] = \frac{1}{N_1} E \left[ (W_{k-N} - (1 - W_{k-N}) K_{N,k-N})^2 (Y_{k-N} - \mu_{W_{k-N}}(X_{k-N}))^2 \middle| \mathcal{F}_{N,k-1} \right]$$

$$= \frac{1}{N_1} (W_{k-N}\sigma_1^2 + (1 - W_{k-N})K_{N,k-N}\sigma_0^2)$$

for  $N + 1 \leq k \leq 2N$ . Let

$$V = E[(\mu_1(X_k) - \mu_0(X_k) - \tau)^2 | W_k = 1] + (\sigma_1^2 + \sigma_0^2).$$

Because for all  $1 \leq k \leq N$ ,  $K_{N,k} \leq 1$  with  $\sum_{k=1}^N K_{N,k} = N_1$ ,

$$\sum_{k=1}^{2N} E[\xi_{N,k}^2 | \mathcal{F}_{N,k-1}] = V.$$

To apply a Martingale Central Limit Theorem to  $D_N$ , it is enough to check the Lindeberg condition,

$$\sum_{k=1}^{2N} E[\xi_{N,k}^2 \mathbf{1}_{\{|\xi_{N,k}| \geq \varepsilon\}}] \rightarrow 0 \quad \text{for all } \varepsilon > 0$$

(Billingsley, 1995, see Hall and Heyde, 1980, and Shorack, 2000, for alternative conditions).

Because for all  $\delta > 0$ ,  $|\xi_{N,k}|^2 \mathbf{1}_{\{|\xi_{N,k}| \geq \varepsilon\}} \varepsilon^\delta \leq |\xi_{N,k}|^{2+\delta}$ , we obtain that Lindeberg's condition is implied by Lyapounov's condition:

$$\sum_{k=1}^{2N} E[\xi_{N,k}^{2+\delta}] \rightarrow 0 \quad \text{for some } \delta > 0,$$

which, in turn, can be easily established under usual regularity conditions regarding boundedness of moments. Under these conditions, the Central Limit Theorem for Triangular Martingale Arrays implies:

$$\sqrt{N_1} D_N \xrightarrow{d} N(0, V).$$

The proof concludes by showing that  $\sqrt{N_1} R_N \xrightarrow{p} 0$ . For  $1 \leq i \leq N$  such that  $W_i = 1$ , let  $\|U_{N_0, N_1, i}\|$  be the matching discrepancy for treated unit  $i$  when untreated units are matched without replacement to treated units in such a way that the sum of the matching discrepancies is minimized. If  $\mu_0$  is Lipschitz-continuous, then there exists a constant  $C$  such that

$$\sqrt{N_1} R_N \leq C \frac{1}{\sqrt{N_1}} \sum_{i=1}^N W_i \|U_{N_0, N_1, i}\|.$$

The following proposition (proven in the appendix) provides sufficient conditions under which  $\sqrt{N_1} R_N$  vanishes asymptotically.

PROPOSITION 1: Let  $F_0$  and  $F_1$  be the distributions of  $X$  given  $W = 0$  and  $X$  given  $W = 1$ , respectively. Assume that  $F_0$  and  $F_1$  have a common support that is a Cartesian product of intervals, and that the densities  $f_0(x)$  and  $f_1(x)$  are bounded and bounded away from zero:  $\underline{f} \leq f_0 \leq \bar{f}$  and  $\underline{f} \leq f_1 \leq \bar{f}$ . Assume that there exists  $c > 0$  and  $r > k$  where  $k$  is the number of (continuous) covariates, such that  $N_1^r/N_0 \leq c$ . Then,

$$\frac{1}{\sqrt{N_1}} \sum_{i=1}^N W_i \|U_{N_0, N_1, i}\| \xrightarrow{p} 0.$$

The conditions of Proposition 1 assume that all covariates have continuous distributions. This is done without loss of generality. Discrete covariates with a finite number of support points can be easily dealt with by conditioning on their values, in which case  $k$  is equal to the number of continuous covariates in  $X$ . Under the conditions of Proposition 1, the conditional bias term,  $\sqrt{N_1}R_N$ , is asymptotically negligible, so we obtain:

$$\sqrt{N_1}(\hat{\tau} - \tau) \xrightarrow{d} N(0, V).$$

## V. CONCLUSION

This article establishes a martingale array representation for matching estimators. This representation allows the use of well-known martingale limit theorems to determine the large sample distribution of matching estimators. Because the martingale representation applies to a large class of matching estimators, the applicability of the methods presented in this article is very broad. Specific applications include matching estimators of average treatment effects as well as “hot-deck” imputation methods for missing data.



## APPENDIX

PROOF OF PROPOSITION 1: By changing units of measurement, we can always make the support of the covariates equal to the unit  $k$ -cube. (This only adds a multiplicative constant to our bounds.) Notice that we can always divide a unit  $k$ -cube into  $N_1^k$  identical cubes, for  $N_1 = 1, 2, 3, \dots$

Divide the support of  $F_0$  and  $F_1$  into  $N_1^k$  identical cubes. Let  $Z_{N_0, N_1}$  be the number of such cells that are occupied by more observations from the treated sample than observations from the untreated sample. Let  $M_{N_1}$  be the maximum number of observations from the treated sample in a single cell. Let  $m_{N_0, N_1}$  be the minimum number of untreated observations in a single cell. Notice that for any series,  $f(N_1)$ , such that  $1 \leq f(N_1) < N_1$ , we have:

$$\begin{aligned}
 \Pr(Z_{N_0, N_1} > 0) &\leq \sum_{m=1}^{N_1} \Pr(m_{N_0, N_1} < m) \Pr(M_{N_1} = m) \\
 &\leq \sum_{m=1}^{\lfloor f(N_1) \rfloor} \Pr(m_{N_0, N_1} < m) \Pr(M_{N_1} = m) \\
 &\quad + \sum_{m=\lfloor f(N_1) \rfloor + 1}^{N_1} \Pr(m_{N_0, N_1} < m) \Pr(M_{N_1} = m) \\
 &\leq f(N_1) \Pr(m_{N_0, N_1} < f(N_1)) \\
 &\quad + (N_1 - f(N_1)) \Pr(M_{N_1} > f(N_1)).
 \end{aligned}$$

Let  $D_{N_1, m}$  be the number of cells that contain more than  $m$  treated observations. Let  $0 < \alpha < \min\{r-k, 1\}$ . Consider  $f(N_1) = N_1^\alpha$ . For  $N_1$  large enough,  $\bar{f}/N_1^k < 1$ . Using Markov's Inequality we obtain for  $N_1$  large enough:

$$\begin{aligned}
 \Pr(M_{N_1} > f(N_1)) &= \Pr(D_{N_1, N_1^\alpha} \geq 1) \\
 &\leq E[D_{N_1, N_1^\alpha}] \\
 &\leq N_1^k \Pr(B(N_1, \bar{f}/N_1^k) > N_1^\alpha).
 \end{aligned}$$

Using Bennett's bound for binomial tails (e.g., Shorack and Wellner, 1996, p. 440), we obtain:

$$\begin{aligned}
 \Pr(B(N_1, \bar{f}/N_1^k) > N_1^\alpha) &= \Pr\left(\frac{B(N_1, \bar{f}/N_1^k) - \bar{f}/N_1^{k-1}}{\sqrt{N_1}} > \frac{N_1^\alpha - \bar{f}/N_1^{k-1}}{\sqrt{N_1}}\right) \\
 &\leq \exp\left\{-\frac{\bar{f}/N_1^{k-1}}{1 - \bar{f}/N_1^k} \left[\frac{N_1^{\alpha+k-1}}{\bar{f}} \left(\log\left(\frac{N_1^{\alpha+k-1}}{\bar{f}}\right) - 1\right) + 1\right]\right\} \\
 &= \exp\left\{-\frac{1}{1 - \bar{f}/N_1^k} \left[N_1^\alpha \left(\log\left(\frac{N_1^{\alpha+k-1}}{\bar{f}}\right) - 1\right) + \frac{\bar{f}}{N_1^{k-1}}\right]\right\}.
 \end{aligned}$$

Similarly, let  $C_{N_0, N_1, m}$  be the number of cells with less than  $m$  untreated observations. Then,

$$\begin{aligned}
 \Pr(m_{N_0, N_1} < m) &= \Pr(C_{N_0, N_1, m} \geq 1) \\
 &\leq E[C_{N_0, N_1, m}]
 \end{aligned}$$

$$= \sum_{n=1}^{N_1^k} \Pr(B(N_0, p_n) < m),$$

where  $p_n$  is the probability that an untreated observation falls in cell  $n$ . Then, because for all  $n$ ,  $p_n \geq \underline{f}/N_1^k$ , we obtain:

$$\Pr(m_{N_0, N_1} < m) \leq N_1^k \Pr(B(N_0, \underline{f}/N_1^k) < m).$$

Also, for large enough  $N_1$ , there exists  $\delta$  such that  $(c/\underline{f})/N_1^{r-\alpha-k} < \delta < 1$ . Using Chernoff's bound for the lower tail of a sum of independent Poisson trials (e.g., Motwani and Raghavan, 1995, p. 70), we obtain that for large enough  $N_1$ :

$$\begin{aligned} \Pr(B(N_0, \underline{f}/N_1^k) < N_1^\alpha) &= \Pr\left(B(N_0, \underline{f}/N_1^k) < \underline{f} \frac{N_0}{N_1^k} \frac{N_1^{\alpha+k}}{\underline{f} N_0}\right) \\ &\leq \Pr\left(B(N_0, \underline{f}/N_1^k) < \underline{f} \frac{N_0}{N_1^k} \frac{c/\underline{f}}{N_1^{r-\alpha-k}}\right) \\ &\leq \exp\left(-(\underline{f} N_0/N_1^k)(1 - (c/\underline{f})/N_1^{r-\alpha-k})^2/2\right) \\ &\leq \exp\left(-\underline{f} N_1^{r-k}(1 - \delta)^2/2c\right). \end{aligned}$$

This proves an exponential bound for  $\Pr(Z_{N_0, N_1} > 0)$ .

Rearrange the observations so the first  $N_1$  observations in the sample are the treated observations. For  $1 \leq i \leq N_1$ , let  $\|U_{N_0, N_1, i}\|$  be the matching discrepancy for treated unit  $i$  when untreated units are matched without replacement to treated units in such a way that the sum of the matching discrepancies is minimized. For  $1 \leq i \leq N_1$ , let  $\|V_{N_0, N_1, i}\|$  be the matching discrepancy for treated unit  $i$  when untreated units are matched without replacement to treated units in such a way that the matches are first done within cells and, after all possible within-cell matches are exhausted, untreated units that were not previously used as a match are matched without replacement to previously unmatched treated units in other cells. Notice that:

$$\sum_{i=1}^{N_1} \|U_{N_0, N_1, i}\| \leq \sum_{i=1}^{N_1} \|V_{N_0, N_1, i}\|.$$

Let  $d_{N_1, k}$  be the diameter of the cells. Let  $C_k$  be the diameter of the unit  $k$ -cube. Notice the if the unit  $k$ -cube is divided in  $N_1^k$  identical cells, then  $C_k = N_1 d_{N_1, k}$ . For  $1 \leq n \leq N_1^k$ , let  $A_{N_1, n}$  be the  $n$ -th cell. Then,

$$\begin{aligned} E[\|V_{N_0, N_1, i}\| \mid Z_{N_0, N_1} = 0] &\leq \sum_{n=1}^{N_1^k} d_{N_1, k} \Pr(X_{1, i} \in A_{N_1, n} \mid Z_{N_0, N_1} = 0) \\ &\leq d_{N_1, k} \\ &\leq \frac{C_k}{N_1}. \end{aligned}$$

Now,

$$E\left[\frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} \|U_{N_0, N_1, i}\|\right] \leq E\left[\frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} \|V_{N_0, N_1, i}\|\right]$$

$$\begin{aligned}
&= E \left[ \frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} \|V_{N_0, N_1, i}\| \middle| Z_{N_0, N_1} = 0 \right] \Pr(Z_{N_0, N_1} = 0) \\
&+ E \left[ \frac{1}{\sqrt{N_1}} \sum_{i=1}^{N_1} \|V_{N_0, N_1, i}\| \middle| Z_{N_0, N_1} > 0 \right] \Pr(Z_{N_0, N_1} > 0) \\
&\leq \frac{C_k}{\sqrt{N_1}} + \sqrt{N_1} C_k \Pr(Z_{N_0, N_1} > 0) \longrightarrow 0.
\end{aligned}$$

Markov's Inequality produces the desired result. □

## REFERENCES

- ABADIE, A. and IMBENS, G.W. (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects", *Econometrica*, vol. 74, no. 1, 235-267.
- ABADIE, A. and IMBENS, G.W. (2008), "On the Failure of the Bootstrap for Matching Estimators", *Econometrica*, vol. 76, no. 6, 1537-1558.
- ANGRIST, J.D. and KRUEGER, A.B (1999), "Empirical Strategies in Labor Economics," in A. Ashenfelter and D. Card eds. *Handbook of Labor Economics*, vol. 3. Elsevier Science.
- BILLINGSLEY, P. (1995), *Probability and Measure*, third edition. Wiley, New York.
- DEHEJIA, R. and WAHBA, S. (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94, 1053-1062.
- DIAMOND, A. and SEKHON, J.S. (2008), "Genetic Matching for Estimating Causal Effects: A New Method of Achieving Balance in Observational Studies," UC Berkeley.
- HALL, P. and HEYDE C.C. (1980), *Martingale Limit Theory and its Applications*. Academic Press, New York.
- HANSEN, B.B. (2004), "Full Matching in an Observational Study of Coaching for the SAT", *Journal of the American Statistical Association*, 99, 609-618.
- HECKMAN, J., H. ICHIMURA, and P. TODD (1998), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261-294.
- IMBENS, G.W. (2004), "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review", *Review of Economics and Statistics*, vol. 86, no. 1, 4-29.
- IMBENS, G.W. and WOOLDRIDGE, J. M. (2008) "Recent Developments in the Econometrics of Program Evaluation", NBER Working Paper No. W14251.
- LITTLE, R.J.A. and D.B. RUBIN (2002), *Statistical Analysis with Missing Data*, second edition. Wiley-Interscience, New York.
- MOTWANI, R. and RAGHAVAN, P. (1995), *Randomized Algorithms*. Cambridge University Press, New York.
- ROSENBAUM, P.R. (2002), *Observational Studies*, second edition. Springer, New York.
- RUBIN, D.B. (1973), "Matching to Reduce Bias in Observational Studies", *Biometrics*, 29, 159-183.
- RUBIN, D.B. (1977), "Assignment to Treatment Group on the Basis of a Covariate", *Journal of Educational Statistics*, 2, 1-26.
- SHORACK, G.R. (2000), *Probability for Statisticians*. Springer, New York.
- SHORACK, G.R. and WELLNER, J.A. (1986), *Empirical Processes with Applications to Statistics*. Wiley, New York.