NBER WORKING PAPER SERIES

A MARTINGALE REPRESENTATION FOR MATCHING ESTIMATORS

Alberto Abadie
Guido Imbens

Working Paper 14756
http://www.nber.org/papers/w14756

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2009

We thank Rustam Ibragimov, Don Rubin, and seminar participants at Harvard/MIT, Brown, Georgetown, UPenn, Montreal, and the 2009 Conference on Causal Inference in Statistics and the Quantitative Sciences at the Banff International Research Station for useful comments and suggestions. We are also grateful to Greg Weyland, from the US Census Bureau, for patiently explaining to us the intricacies of the hot deck imputation algorithms employed in the Current Population Survey. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

A Martingale Representation for Matching Estimators
Alberto Abadie and Guido Imbens
NBER Working Paper No. 14756
February 2009, Revised April 2010
JEL No. C13,C14,C21

ABSTRACT

Matching estimators (Rubin, 1973a, 1977; Rosenbaum, 2002) are widely used in statistical data analysis. However, the large sample 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 large sample 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. In addition, we apply the techniques proposed in this article to derive a correction to the standard error of a sample mean when missing data are imputed using the "hot deck", a matching imputation method widely used in the Current Population Survey (CPS) and other large surveys in the social sciences. We demonstrate the empirical relevance of our methods using two Monte Carlo designs based on actual data sets. In these realistic Monte Carlo exercises the large sample distribution of matching estimators derived in this article provides an accurate approximation to the small sample behavior of these estimators. In addition, our simulations show that standard errors that do not take into account hot deck imputation of missing data may be severely downward biased, while standard errors that incorporate the correction proposed in this article for hot deck imputation perform extremely well. This result demonstrates the practical relevance of the standard error correction for the hot deck proposed in this article.

Alberto Abadie
John F. Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
and NBER
alberto_abadie@harvard.edu

Guido Imbens
Department of Economics
Littauer Center
Harvard University
1805 Cambridge Street
Cambridge, MA 02138
and NBER
imbens@fas.harvard.edu

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, 1977; Dehejia and Wahba, 1999; Rosenbaum, 2002, Hansen, 2004). Matching is also used for the analysis of missing data, where it is often referred to as "hot deck imputation" (Little and Rubin, 2002). As a notorious example, missing weekly earnings are currently imputed using hot deck methods for more than 30 percent of the records with weekly earnings data in the monthly U.S. Current Population Survey (CPS) files (Bollinger and Hirsch, 2009).

In spite of the pervasiveness of matching methods, the asymptotic distribution of matching estimators has 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 sometimes 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). Similarly, the properties of statistics based on data imputed using sequential hot deck methods, like those employed in the CPS and other large surveys, are not well-understood, and empirical researchers using these surveys typically ignore missing data imputation issues when they construct standard errors. Andridge and Little (2010) provide a recent survey on hot deck imputation methods.

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 very broad. 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. The use of martingale methods is attractive because the limit behavior of martingale sequences has been extensively studied in the statistics literature (see, for example, Hall and Heyde, 1980).

As an illustration of the usefulness 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. In addition, we apply the techniques proposed in this article to derive a correction to the standard error of a sample mean when missing data are imputed using the hot deck.

Finally, we demonstrate the empirical relevance of our methods using two Monte Carlo designs based on actual data sets. In these realistic Monte Carlo exercises the large sample distribution of matching estimators derived in this article provides an accurate approximation to the small sample behavior of these estimators. In addition, our simulations show that standard errors that do not take into account hot deck imputation of missing data may be severely downward biased while standard errors that incorporate the correction proposed in this article for hot deck imputation perform extremely well. This result demonstrates the practical relevance of the standard error correction for the hot deck proposed in this article.

The rest of the article is organized as follows. Section II describes matching estimators. Section III presents the main result of the article, which establishes a martingale representation for matching estimators. In section IV, we apply martingale techniques to analyze the large sample properties of a matching estimator when matching is carried out without replacement. In section V, we apply martingale techniques to study hot deck imputation. Section VI describes of the Monte Carlo simulation exercises and reports the results. Section VII concludes.

II. Matching Estimators

Let W be a binary variable that indicates membership to a particular population of interest. Empirical researchers often compare the distributions of some variable, Y, between units with $W = 1$ and units with $W = 0$ after adjusting for the differences in a $(k \times 1)$ vector of observed covariates, X. 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 describing job and/or worker characteristics. 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. As in that literature, we will say that units with $W = 1$ are "treated" and units with $W = 0$ are "untreated". Let

$$\tau = E[Y|W = 1] - E\Big[E[Y|X, W = 0]\Big|W = 1\Big]. \tag{1}$$

In evaluation research, τ is given a causal interpretation as the "average treatment effect on the treated" under unconfoundedness assumptions (Rubin, 1977). Applied researchers often use matching methods to estimate τ. Other parameters of interest that can be estimated by matching methods include: (i) the "average treatment effect", which is of widespread interest in evaluation studies, (ii) parameters that focus on features of the distribution of Y other than the mean, (iii) parameters estimated by hot deck imputation methods in the presence of missing data. Rosenbaum (2002), Imbens (2004), and Rubin (2006) provide detailed surveys of the literature. For concreteness, and to avoid tedious repetition or unnecessary abstraction, in this section we discuss matching estimation of τ only. However, the techniques proposed in this paper are of immediate application to the estimation of parameters other than τ via matching (see, for example, section V).

Also, to avoid notational clutter, we consider only estimators with a fixed number of matches, M, per unit. However, as it will be explained later, our techniques can be immediately applied to estimators for which the number of matches may differ across units (see, e.g., Hansen, 2004). Consider two random samples of sizes N_0 and N_1 of untreated and treated units, respectively. Pooling these two samples, we obtain a sample of size $N = N_0 + N_1$ containing both 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 $\mathcal{J}_M(i)$ consists of the indices of the M untreated observations with the closest value covariate values to X_i:

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

For matching without replacement, the elements of $\{\mathcal{J}_M(i) \text{ s.t. } W_i = 1\}$ are non-overlapping subsets of $\{j \in \{1,\ldots,N\} \text{ s.t. } W_j = 0\}$, 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 τ is defined as:

$$\widehat{\tau} = \frac{1}{N_1} \sum_{i=1}^{N} W_i \Big(Y_i - \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j \Big). \tag{2}$$

Many other matching schemes are possible (see, e.g., Gu and Rosenbaum, 1993; Rosenbaum, 2002; Hansen, 2004; Diamond and Sekhon, 2008; Iacus, King, and Porro, 2009), 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 $M \to \infty$ as $N \to \infty$) have been proposed by Heckman, Ichimura, and Todd (1998) and others. Under certain conditions, these estimators have asymptotically linear representations, so their large sample distributions can be derived using the standard machinery for asymptotically linear estimators. In contrast, despite the pervasiveness of matching estimators that use a small number of matches (e.g., hot deck imputation in the CPS), the previous literature does not provide a general framework for establishing their large sample properties.

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. Given equation (2), we can write $\widehat{\tau} - \tau = D_N + R_N$, where

$$D_N = \frac{1}{N_1}\sum_{i=1}^{N} W_i\big(\mu_1(X_i) - \mu_0(X_i) - \tau\big) + \frac{1}{N_1}\sum_{i=1}^{N} W_i\Big(\big(Y_i - \mu_1(X_i)\big) - \frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}\big(Y_j - \mu_0(X_j)\big)\Big),$$

and

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

The term R_N is the conditional bias of matching estimator described in Abadie and Imbens (2006). This term is zero if all matches are perfect (that is, if all matching discrepancies, $X_i - X_j$ for $j \in \mathcal{J}_M(i)$, are zero), or if the regression μ_0 is a constant function. In most cases of interest, 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 for the case of matching without replacement, or Abadie and Imbens, 2006, for the case of matching with replacement).

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

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

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

$$K_{N,i} = \sum_{j=1}^{N} 1_{\{i\in\mathcal{J}_M(j)\}}.$$

Therefore, we can write:

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

where

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

Let $\mathbf{X}_N = \{X_1, \ldots, X_N\}$ and $\mathbf{W}_N = \{W_1, \ldots, W_N\}$. Consider the σ-fields $\mathcal{F}_{N,k} = \sigma\{\mathbf{W}_N, X_1, \ldots, X_k\}$ for $1 \leq k \leq N$ and $\mathcal{F}_{N,k} = \sigma\{\mathbf{W}_N, \mathbf{X}_N, Y_1, \ldots, Y_{k-N}\}$ for $N+1 \leq k \leq 2N$. Then, and this is the key insight in this article,

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

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, if matches depend only on the covariates X, a martingale representation holds for $\sqrt{N_1} D_N$. The reason is that no matter how matching is implemented, (i) the number of times that unit k is used as a match, $K_{N,k}$, is a deterministic function of $\mathbf{X}_N$ and $\mathbf{W}_N$, and (ii) $E[Y_k - \mu_{W_k}(X_k) \,|\, \mathbf{X}_N, \mathbf{W}_N, Y_1, \ldots, Y_{k-1}] = 0$.

So far, we have considered the case where $K_{N,i}$ is fixed given $\mathbf{X}_N$ and $\mathbf{W}_N$, for all $1 \leq i \leq N$. This assumption does not hold for certain matching schemes that break matching ties using randomization. Notice, however, that any sequence of randomized tie-breaks can be included in the set of variables that span $\mathcal{F}_{N,k}$ for $N+1 \leq k \leq 2N$ to preserve the martingale representation of D_N.

IV. Application: Matching without Replacement

In this section, we illustrate the usefulness of the martingale representation of matching estimators by deriving the asymptotic distribution of a matching estimator when matching is done without replacement, so $K_{N,i} \in \{0, 1\}$ for every unit i with $W_i = 0$.

For $1 \leq k \leq N$, the conditional variances of the martingale differences are given by:

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

$$= \frac{1}{N_1} W_k E[(\mu_1(X_k) - \mu_0(X_k) - \tau)^2 | W_k = 1].$$

For $N+1 \le k \le 2N$, the conditional variances of the martingale differences are given by:

$$\begin{aligned} E[\xi_{N,k}^2 | \mathcal{F}_{N,k-1}] &= \frac{1}{N_1} E\left[\left(W_{k-N} - (1 - W_{k-N}) \frac{K_{N,k-N}}{M} \right)^2 \left(Y_{k-N} - \mu_{W_{k-N}}(X_{k-N}) \right)^2 \middle| \mathcal{F}_{N,k-1} \right] \\ &= \frac{1}{N_1} \left(W_{k-N} \sigma_1^2(X_{k-N}) + (1 - W_{k-N}) \frac{K_{N,k-N}}{M^2} \sigma_0^2(X_{k-N}) \right) \\ &= \frac{1}{N_1} W_{k-N} \left(\sigma_1^2(X_{k-N}) + \frac{\sigma_0^2(X_{k-N})}{M} \right) + r_{N,k-N}, \end{aligned}$$

where

$$r_{N,k-N} = \frac{1}{N_1} \left((1 - W_{k-N}) \frac{K_{N,k-N}}{M^2} \sigma_0^2(X_{k-N}) - W_{k-N} \frac{\sigma_0^2(X_{k-N})}{M} \right).$$

Assume that the conditional variance function $\sigma_0^2(x)$ is Lipschitz-continuous, with Lipschitz constant equal to c_1. For $1 \le i \le N$ such that $W_i = 1$, let $\|U_{N_0,N_1,i}^{(M,m)}\|$ be the m-th 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. That is, if unit i is a treated observation, and unit j is the m-th match for unit i, then $\|U_{N_0,N_1,i}^{(M,m)}\| = \|X_i - X_j\|$. Lipschitz-continuity of $\sigma_0^2(x)$ implies:

$$\left| \sum_{k=N+1}^{2N} r_{N,k-N} \right| \le \frac{c_1}{M^2} \frac{1}{N_1} \sum_{i=1}^{N} \sum_{m=1}^{M} W_i \|U_{N_0,N_1,i}^{(M,m)}\|.$$

Because the average matching discrepancy converges to zero in probability (see Proposition 1 in the appendix for a stronger result), the Weak Law of Large Numbers implies

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

where

$$\sigma^2 = E[(\mu_1(X) - \mu_0(X) - \tau)^2 | W = 1] + E\left[\sigma_1^2(X) + \frac{\sigma_0^2(X)}{M} \middle| W = 1 \right]. \tag{3}$$

In view of this result, to apply a Martingale Central Limit Theorem to D_N, it is sufficient to check the Lindeberg condition,

$$\sum_{k=1}^{2N} E[\xi_{N,k}^2 1_{\{|\xi_{N,k}| \ge \varepsilon\}}] \to 0 \qquad \text{for all } \varepsilon > 0$$

(Billingsley, 1995, see Hall and Heyde, 1980, and Shorack, 2000, for alternative conditions). Because for all $\delta > 0$, $|\xi_{N,k}|^2 1_{\{|\xi_{N,k}| \geq \varepsilon\}} \varepsilon^\delta \leq |\xi_{N,k}|^{2+\delta}$, it follows that Lindeberg's condition is implied by Lyapounov's condition:

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

For the matching estimators considered in this section, this can be easily established under usual regularity conditions regarding boundedness of moments. Under these conditions, the Central Limit Theorem for Triangular Martingale Arrays implies:

$$\sqrt{N_1} D_N \xrightarrow{d} N(0, \sigma^2).$$

The proof concludes by showing that $\sqrt{N_1} R_N \xrightarrow{p} 0$. If μ_0 is Lipschitz-continuous, then there exists a constant c_2 such that

$$\sqrt{N_1} R_N \leq c_2 \frac{1}{\sqrt{N_1}} \frac{1}{M} \sum_{i=1}^{N} \sum_{m=1}^{M} W_i \| U_{N_0,N_1,i}^{(M,m)} \|.$$

Proposition 1 in the appendix shows that under some conditions, and if 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,

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

so $\sqrt{N_1} R_N$ vanishes asymptotically.

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. The proof of Proposition 1 indicates that the support conditions in this proposition can also be relaxed. However, the requirement that the size of the untreated group is of larger order of magnitude than the size of the treated group is crucial to the result in the proposition. To see that $r = 1$ is not sufficient (even in the one-dimensional case), consider the case with $M = 1$ and $N_0 = N_1$. Then, because matching is done without replacement and all treated units are matched, the matching estimator is equal to the difference in sample means of Y between treated and nontreated, regardless of the total sample size N.

Proposition 1 provides conditions under which matching discrepancies are negligible in large samples. In practical terms, Proposition 1 demonstrates the benefits of having a large "donor pool" of control units for matching estimators. Notice however that, for particular applications, researchers can assess the quality of the matches directly from the data. When matching discrepancies are large the resulting bias can be eliminated or reduced using the bias correction techniques in Rubin (1973b), Quade (1982), and Abadie and Imbens (2009). These authors propose a bias-corrected matching estimator that adjusts each matched pair for its contribution to the conditional bias term:

$$\widehat{\tau}_{\text{bc}} = \frac{1}{N_1}\sum_{i=1}^{N} W_i\Big((Y_i - \widehat{\mu}_0(X_i)) - \frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}(Y_j - \widehat{\mu}_0(X_{j(i)}))\Big), \tag{4}$$

where $\widehat{\mu}_0(\cdot)$ is an estimator of $\mu_0(\cdot)$. Abadie and Imbens (2009) show that under certain conditions this bias-correction technique eliminates the asymptotic bias of matching estimators without affecting the asymptotic variance.

Under the conditions of Proposition 1, the conditional bias term, $\sqrt{N_1}R_N$, is asymptotically negligible, so we obtain:

$$\sqrt{N_1}\big(\widehat{\tau} - \tau\big) \xrightarrow{d} N(0, \sigma^2),$$

where σ^2 is given in equation (3). Straightforward calculations show that the variance estimator

$$\widehat{\sigma}^2 = \frac{1}{N_1 - 1}\sum_{i=1}^{N} W_i\Big(Y_i - \frac{1}{M}\sum_{j\in\mathcal{J}_M(i)} Y_j - \widehat{\tau}\Big)^2 \tag{5}$$

is consistent for σ^2.

Despite the simplicity of this result, to our knowledge the validity of $\widehat{\sigma}^2/N_1$ as an estimator of the variance of $\widehat{\tau}$ when matching is done without replacement has not been established previously. Conversely, it is known that $\widehat{\sigma}^2/N_1$ is not a valid estimator of the variance of $\widehat{\tau}$ when matching is done with replacement (Abadie and Imbens, 2006).

V. Application: Hot Deck Imputation

In this section, we consider a "cell hot deck" imputation scheme where incomplete records of Y are imputed using complete observations within the same "cell" of the covariates, X. That is, the support of the covariates is partitioned into T cells, $\mathcal{C}_1, \ldots \mathcal{C}_T$, and each incomplete record of Y is filled using a complete record from the same cell. Other hot deck imputation procedures are possible (see, for example, Little and Rubin, 2002). However, the cell hot deck methods is probably the most widely used in practice, as it is the one used by the US Census Bureau to impute missing data in the Current Population Survey (CPS), the decennial census, the Survey of Income and Program Participation (SIPP), and other large surveys. Derivations similar to the ones presented in this section can be applied to alternative hot deck imputation schemes.

Cell hot deck imputation methods like the one employed in the CPS can be justified by a "Missing and Coarsening at Random" assumption. Let W be an indicator for complete record, that is $W = 1$ indicates that Y is observed. A missing and coarsening at random assumption states that Y is independent of (X, W) conditional on $X \in \mathcal{C}_t$, for $1 \leq t \leq T$. Missing and coarsening at random may be a strong assumption in many contexts where data are imputed using the cell hot deck. However, without this assumption, or a similar one, the cell hot deck will produce inconsistent estimators in general. Therefore, in our analysis we assume missing and coarsening at random. Let $\mu = E[Y]$, $\mu(x) = E[Y|X = x]$, $\mu_t = E[Y|X \in \mathcal{C}_t]$ and $\sigma_t^2 = \text{var}(Y|X \in \mathcal{C}_t)$. Let $j(i)$ be the index of the observation used to impute Y for observation i (if $W_i = 1$, then $j(i) = i$). Let

$$
\begin{aligned}
\bar{Y} &= \frac{1}{N} \sum_{i=1}^{N} Y_{j(i)} \\
&= \frac{1}{N} \sum_{i=1}^{N} W_i (1 + K_{N,i}) Y_i, \qquad (6)
\end{aligned}
$$

where now $K_{N,i}$ is the number of times that observation i is used to impute an incomplete record. The variables $K_{N,i}$ depend on how imputations are chosen from the complete records within a cell. One possibility is the *random cell hot deck*, which imputes missing records using a record chosen at random among the complete observation in the same cell.

The CPS and other large surveys use a more complicated procedure called the *sequential cell hot deck*. The sequential cell hot deck imputes missing records using the last complete record in the same cell. That is, unlike the random cell hot deck, the sequential cell hot deck uses information about the order of the observations in the sample.

Notice that

$$\begin{aligned}\bar{Y} - \mu = {} & \frac{1}{N}\sum_{i=1}^{N}(\mu(X_i) - \mu) \\ & + \frac{1}{N}\sum_{i=1}^{N} W_i(1 + K_{N,i})(Y_i - \mu(X_i)) \\ & + \frac{1}{N}\sum_{i=1}^{N}(\mu(X_{j(i)}) - \mu(X_i)).\end{aligned}$$

By the Missing and Coarsening at Random assumption, $\mu(X_{j(i)}) - \mu(X_i) = 0$ for all i. Assume that the second moment of $K_{N,i}$ exists, and that for each cell, t, we have:

$$\left| \frac{1}{N_t}\sum_{i=1}^{N} 1_{\{X_i \in \mathcal{C}_t\}} W_i (1 + K_{N,i})^2 - E\left[\frac{1}{N_t}\sum_{i=1}^{N} 1_{\{X_i \in \mathcal{C}_t\}} W_i (1 + K_{N,i})^2\right]\right| \xrightarrow{p} 0, \qquad (7)$$

which can be usually established using negative association properties of $\{K_{N,i}$ s.t. $W_i = 1, X_i \in \mathcal{C}_t\}$ (Joag-Dev and Proschan, 1983). We can write:

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{N}} = \sum_{k=1}^{2N} \xi_{N,k},$$

where

$$\sigma^2 = E\left[\sum_{t=1}^{T}\left(\frac{N_t}{N}\right)(\mu_t - \mu)^2\right] + E\left[\sum_{t=1}^{T}\left(\frac{N_t}{N}\right)\sigma_t^2 \frac{1}{N_t}\sum_{i=1}^{N} 1_{\{X_i \in \mathcal{C}_t\}} W_i (1 + K_{N,i})^2\right],$$

and

$$\xi_{N,k} = \begin{cases} \dfrac{1}{\sigma\sqrt{N}}\left(\mu(X_k) - \mu\right) & \text{if } 1 \le k \le N, \\ \dfrac{1}{\sigma\sqrt{N}} W_{k-N}(1 + K_{N,k-N})\big(Y_{k-N} - \mu(X_{k-N})\big) & \text{if } N+1 \le k \le 2N. \end{cases}$$

Let $\mathbf{X}_N = \{X_1, \ldots, X_N\}$, $\mathbf{W}_N = \{W_1, \ldots, W_N\}$ and $\mathbf{J}_N = \{j(1), \ldots, j(N)\}$. Consider the σ-fields $\mathcal{F}_{N,k} = \sigma\{\mathbf{W}_N, X_1, \ldots, X_k\}$ for $1 \le k \le N$ and $\mathcal{F}_{N,k} = \sigma\{\mathbf{W}_N, \mathbf{X}_N,$

$\mathbf{J}_N, Y_1, \ldots, Y_{k-N}\}$ 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 $N \geq 1$. Equation (7) along with the Central Limit Theorem for martingale arrays (e.g., Theorem 3.2 in Hall and Heyde, 1980) imply:

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{N}} \xrightarrow{d} N(0,1).$$

Consider now the usual variance estimator that ignores missing data imputation:

$$\widehat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_{j(i)} - \bar{Y})^2. \tag{8}$$

Notice that

$$\left| \widehat{\sigma}^2 - \sum_{t=1}^{T} \left(\frac{N_t}{N} \right) (\mu_t - \mu)^2 - \sum_{t=1}^{T} \left(\frac{N_t}{N} \right) \sigma_t^2 \right| \xrightarrow{p} 0.$$

In addition, because $\sum_{i=1}^{N} 1_{\{X_i \in \mathcal{C}_t\}} W_i (1 + K_{N,i}) = N_t$, then

$$\frac{1}{N_t} \sum_{i=1}^{N} 1_{\{X_i \in \mathcal{C}_t\}} W_i (1 + K_{N,i})^2 = 1 + \frac{1}{N_t} \sum_{i=1}^{N} 1_{\{X_i \in \mathcal{C}_t\}} W_i (K_{N,i}^2 + K_{N,i}).$$

This suggests using the following estimator of the variance of the re-scaled estimator:

$$\begin{aligned} \widehat{\sigma}^2_{\text{adj}} &= \widehat{\sigma}^2 + \frac{1}{N} \sum_{t=1}^{T} \left(\sum_{i=1}^{N} 1_{\{X_i \in \mathcal{C}_t\}} W_i (K_{N,i}^2 + K_{N,i}) \right) \widehat{\sigma}_t^2 \\ &= \widehat{\sigma}^2 + \sum_{t=1}^{T} \left(\frac{N_t}{N} \right) \left(\frac{1}{N_t} \sum_{i=1}^{N} 1_{\{X_i \in \mathcal{C}_t\}} W_i (K_{N,i}^2 + K_{N,i}) \right) \widehat{\sigma}_t^2. \end{aligned} \tag{9}$$

where $\widehat{\sigma}_t^2$ is the sample variance of Y calculated from the complete observations in cell $\mathcal{C}_t$. Notice that this formula applies no matter how imputation is done within the cells (for example, randomized or based on the order of the observations in the sample) as long as equation (7) holds.

VI. Monte Carlo Analysis

This section reports the results of two Monte Carlo simulations based on actual data. Section VI.A uses the Boston HMDA data set, a data set collected by the Federal Reserve

Bank of Boston to investigate racial discrimination in mortgage credit markets, to assess the quality of the large sample approximation to the distribution of matching estimators derived in section IV. Section VI.B uses CPS data to investigate the performance of the standard error correction for missing data imputation derived in section V.

A. Matching without Replacement in the Boston HMDA Dataset

In order to detect potential discriminatory practices of mortgage credit lenders against minority applicants, the U.S. Home Mortgage Disclosure Act (HMDA) of 1975 requires lenders to routinely disclose information on mortgage applications, including the race and ethnicity of the applicants. The information collected under the HMDA does not include, however, data on the credit histories of the applicants, and other loan and applicant characteristics that are considered to be important factors in determining the approval or denial of mortgage loans. The absence of such information has generated some skepticism about whether the HMDA data can effectively be used to detect discrimination in the mortgage credit market. To overcome this criticism, the Federal Reserve Bank of Boston collected an additional set of 38 variables included in mortgage applications for a sample of applications in the Boston metropolitan area in 1990. The Boston HMDA data set includes all mortgage applications by black and Hispanic applicants in the Boston metropolitan area in 1990, as well as a random sample of mortgage applications by white applicants in the same year and geographical area. Regression analysis of the Boston HMDA data indicated that minority applicants were more likely to be denied mortgage than white applicants with the same characteristics (Munnell et al., 1996).

In this section, we use the Boston HMDA data set to evaluate the empirical performance of the large sample approximation to the distribution of matching estimators derived in section IV. The HMDA data provides a relevant context for this evaluation because the Federal Reserve System employs matching in the HMDA data as an screening device for fair lending regulation compliance (Avery, Beeson, and Calem, 1997, Avery, Canner, and Cook, 2005). We restrict our sample to single-family residences and male applicants who are white non-Hispanic or black non-Hispanic, not self-employed, who were approved for private mortgage insurance, and who do not have a public record of default or bankruptcy at the time of the application. This leaves us with a sample of 148 black applicants and 1336 white applicants, for a total of 1484 applicants.

In the context of this application, the outcome variable, Y, is an indicator variable that takes value one if the mortgage application was denied, and zero if the mortgage application was approved, W is a binary indicator that takes value one for black applicants, and X is a vector of six applicant and loan characteristics used in Munnell et al. (1996): housing expense to income ratio, total debt payments to income ratio, consumer credit history, mortgage credit history, regional unemployment rate in the applicant's industry, and loan amount to appraised value ratio (see Munnell et al., 1996, for a precise definition of these variables).

To run our simulations for samples sizes of N_1 black observations and N_0 white observations we proceed in five steps. First, for the entire sample, we estimate a logistic model of the mortgage denial indicator on the black indicator and the covariates in X. Second, we draw (with replacement) N_1 observations from the empirical distribution of X for black applicants and N_0 observations from the empirical distribution of X for white applicants. Third, for each individual in the simulated sample, we generate the mortgage denial indicator, Y, using the logistic model estimated in the first step. Fourth, for the simulated sample, we compute $\widehat{\tau}$, the matching estimator in equation (2), matching without replacement, the bias-corrected version of this estimator, $\widehat{\tau}_{\rm bc}$, in equation (4), and the variance estimator, $\widehat{\sigma}^2$, in equation (5). All covariates are normalized to have unit variance prior to matching, and a logistic model is employed to calculate the bias correction. Finally, we repeat steps two to four for a total number of 10000 simulations. That is, in this simulation we sample from a population distribution of the covariates that is equal to the distribution of the covariates in the HMDA sample of 1484 applicants. The distribution of Y conditional W and X in our simulation is given by a logistic model with parameters equal to those estimated in the HMDA sample of 1484 applicants. In this Monte Carlo design, the parameter τ in equation (1) is equal to 0.099, which represents the difference in the probability of denial between black applicants and white applicants of the same characteristics in our simulation.

Table I reports the results of the simulation, for different sample sizes, N_1 and N_0. Column (1) reports the bias of $\widehat{\tau}$ relative to τ. As suggested by the results in section IV, our simulation results indicate that for a fixed N_1 the bias of $\widehat{\tau}$ decreases when N_0 increases. For small samples, however, the bias of $\widehat{\tau}$ may be substantial, reflecting the high dimensionality of the vector of matching variables. The bias-corrected estimator in column (2) generates much smaller biases. Columns (3) and (4) report the variance of $\widehat{\tau}$ across simulations and the average, also across simulations, of the variance estimator of $\widehat{\tau}$ in equation (5). Even in fairly small samples ($N_1 = 25$ and $N_0 = 250$), $\widehat{\sigma}^2/N_1$ provides a very precise approximation to the variance of $\widehat{\tau}$. Finally, columns (5) and (6) report coverage rates of nominal 95% confidence intervals constructed with $(\widehat{\tau}, \widehat{\sigma}^2)$ and $(\widehat{\tau}_{\mathrm{bc}}, \widehat{\sigma}^2)$, respectively. The results indicate that, in this simulation, the Normal approximation to the distribution of matching estimators derived in section IV is very accurate, especially when the bias of the matching estimator is corrected using the bias correction techniques in Rubin (1973b), Quade (1982), and Abadie and Imbens (2009).

B. Hot Deck Imputation in the Current Population Survey

Hot deck methods have long been used to impute missing data in large surveys (see, for example, Andridge and Little, 2010). However, the sampling properties of complex hot deck imputation methods, like the sequential hot deck used by the Census Bureau in the CPS, are largely unknown. This void in the literature has become an object of serious concern in recent years, because the proportion of observations in the CPS with imputed values of weekly earnings has increased steadily: from around 16 percent in 1979, when weekly earnings were included in the monthly survey questionnaire, to more than 30 percent in recent years (Hirsch and Schumacher, 2004; Bollinger and Hirsch, 2009).

In this section we investigate the performance of the approximation to the distribution of a sample mean proposed in section V, when data are imputed using a sequential hot deck like in the CPS. In order to make our exercise as realistic as possible we base our Monte Carlo design on actual CPS data.

Hot deck imputation in the CPS Outgoing Rotation Groups is done through a series of steps, each one imputing a specific survey item. Here, we focus on imputation of missing earnings, because earnings are affected by imputation rates that are much higher than for other survey items. As for other missing survey items, imputation of weekly earnings for non-hourly workers is implemented through a cell hot deck procedure. Observations are assigned to cells defined by age, race, gender, education, occupation, hours worked, and receipt of overtime wages, tips, or commissions, for a total of 11,520 cells (see Bollinger and Hirsch, 2006, for details). Then each missing record is imputed using the value of weekly earnings of last complete record in the same cell.

The imputation of weekly earnings in the CPS Outgoing Rotation Groups cannot be perfectly reproduced with the CPS public use data files. The main reason is that the race variable used by the imputation algorithm is different from the one included in the public use data release. Nevertheless, the Monte Carlo exercise carried out in this section is designed to reproduce as closely as possible the imputation algorithm used by the Census Bureau for weekly earnings. In our simulation we use data from the CPS monthly file of August 2009. In order to simplify the analysis, we first restrict our sample to male individuals working for a pay, who are white, aged 25 to 64, have a high school diploma or equivalent, hold one job only, have a tertiary occupation, do not receive overtime wages, tips, or commissions, and work 40 hours/week. In addition, we discard four observations with zero recorded weekly earnings. This leaves us with 856 observations in 30 of the 11,520 original hot deck cells. The 30 hot deck cells are defined by three categories of age, two of education, and five of occupation. The average number of observations per cell is 28.53, the minimum is 2, and the maximum is 149. In this sample the percentage of observations with missing weekly earnings is 32.83, and each cell has at least two complete observations.

For a fixed number of observations, N, the simulation proceeds as follows. First, for each cell t we simulate two observations of log weekly earnings, $Y^*_{t,1}$ and $Y^*_{t,2}$, from a normal distribution with the same mean and variance as in the distribution of log weekly earnings for complete the CPS observations in the same cell. In our simulation, $Y^*_{t,1}$ and $Y^*_{t,2}$ represent the last two complete observations in cell t in previous CPS waves. Second, we sample N observation from the multinomial distribution of cell frequencies in the CPS sample. For each of these N observations, we simulate log weekly earnings using a normal distribution with the same mean and variance as log weekly earnings for complete CPS observations in the same cell. Then, for each observation we mark weekly earnings as unrecorded with probability equal to the proportion of missing weekly earnings in the same cell of the CPS sample. Third, in our simulated sample of N observations, we impute missing log weekly earnings using the last complete observation in the cell (which may possibly be $Y^*_{t,2}$). This creates a partially imputed sample with N values of log weekly earnings. Four, we calculate the sample average, $\bar{Y}$ in equation (6), as well as the usual and adjusted variance estimators: $\widehat{\sigma}^2$ and $\widehat{\sigma}^2_{\text{adj}}$ in equations (8) and (9), respectively. To compute the intra-cell variances, $\widehat{\sigma}^2_t$ of equation (9), we use all the complete simulated observations in the cell plus $Y^*_{t,1}$ and $Y^*_{t,2}$. Simulating two complete observations per cell, $Y^*_{t,1}$ and $Y^*_{t,2}$, that correspond to the last two complete observations in the cell in previous CPS waves allows us to compute $\widehat{\sigma}^2_t$ even for cells with no other complete observations in the simulation. Finally, we repeat steps one to four for a total number of 50000 simulations.

The results are reported on Table II for sample sizes 50, 100, 200, and 856, the actual number of observations in the CPS sample. The average of our adjusted variance estimator across simulations, in column (2), closely approximates the variance of $\bar{Y}$, in column (1), even for fairly small sample sizes. In contrast, columns (3) and (4) show that the usual variance estimator is severely downward biased, and that the bias of this estimator (as a percentage of the true variance) increases with the sample size. For 856 observations, that is the actual size of the CPS data sample used in the simulation, the usual variance estimator is only 58 percent of the true variance of $\bar{Y}$. Large sample sizes make possible that some observations are repeatedly used for imputation, increasing the difference between the adjusted and unadjusted variances in equation (9). This happens when missing observations arrive consecutively to a cell, without the observation used for imputation being "refreshed" by another complete observation. Columns (5) and (6) report coverage rates of nominal 95%

confidence intervals constructed with $\widehat{\sigma}^2_{\mathrm{adj}}$ and $\widehat{\sigma}^2$, respectively. The results show coverage rates close to nominal coverage in column (5), when the adjusted variance estimator is used to construct confidence interval. In contrast, confidence intervals calculated with the usual variance estimator suffer from severe under-coverage, as reported in column (6).

VII. 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. Two realistic simulations demonstrate the empirical relevance of the results of this article.

Appendix

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,*

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

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 $N_1 = 1, 2, 3, \ldots$.

Divide the support of F_0 and F_1 into N_1^k identical cubes. Let Z_{M,N_0,N_1} be the number of such cells where the number of untreated observation is less than M times the number of observations from the treated 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:

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

Let $D_{N_1,n}$ be the number of cells where the number of treated observations is larger than n. 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 Bonferroni Inequality we obtain for N_1 large enough:

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

where $B(N,p)$ denotes a Binomial random variable with parameters (N,p). Using Bennett's bound for binomial tails (e.g., Shorack and Wellner, 1996, p. 440), we obtain:

$$\Pr\left(B(N_1, \bar{f}/N_1^k) > N_1^{\alpha}\right) = \Pr\left(\frac{B(N_1, \bar{f}/N_1^k) - \bar{f}/N_1^{k-1}}{\sqrt{N_1}} > \frac{N_1^{\alpha} - \bar{f}/N_1^{k-1}}{\sqrt{N_1}} \right)$$

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

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

$$\begin{aligned} \Pr(m_{N_0,N_1} < m) &= \Pr(C_{N_0,N_1,m} \geq 1) \\ &\leq \sum_{n=1}^{N_1^k} \Pr\big(B(N_0, p_n) < m\big), \end{aligned}$$

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:

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

Also, for large enough N_1, there exists δ such that $(Mc/\underline{f})/N_1^{r-\alpha-k} < \delta < 1$. 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:

$$\begin{aligned} \Pr\big(B(N_0, \underline{f}/N_1^k) < MN_1^{\alpha}\big) &= \Pr\left(B(N_0, \underline{f}/N_1^k) < \underline{f}\frac{N_0}{N_1^k}\frac{MN_1^{\alpha+k}}{\underline{f}N_0}\right) \\ &\leq \Pr\left(B(N_0, \underline{f}/N_1^k) < \underline{f}\frac{N_0}{N_1^k}\frac{Mc/\underline{f}}{N_1^{r-\alpha-k}}\right) \\ &\leq \exp\left(-(\underline{f}N_0/N_1^k)\big(1-(Mc/\underline{f})/N_1^{r-\alpha-k}\big)^2/2\right) \\ &\leq \exp\left(-\underline{f}N_1^{r-k}(1-\delta)^2/2c\right). \end{aligned}$$

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

Rearrange the observations so the first N_1 observations in the sample are the treated observations. For $1 \leq i \leq N_1$, let $\|U_{N_0,N_1,i}^{(M,m)}\|$ be the m-th 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 $\|V_{N_0,N_1,i}^{(M,m)}\|$ be the m-th 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:

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

Let $d_{N_1,k}$ be the diameter of the cells. Let C_k be the diameter of the unit k-cube. Notice that 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,

$$
\begin{aligned}
E\left[\|V^{(M,m)}_{N_0,N_1,i}\| \,\middle|\, Z_{M,N_0,N_1} = 0\right] &\le \sum_{n=1}^{N_1^k} d_{N_1,k} \Pr(X_{1,i} \in A_{N_1,n} | Z_{N_0,N_1} = 0) \\
&\le d_{N_1,k} \\
&= \frac{C_k}{N_1}.
\end{aligned}
$$

Now,

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

Markov's Inequality produces the desired result. $\square$

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.

Abadie, A. and Imbens, G.W. (2009), "Bias Corrected Matching Estimators for Average Treatment Effects," *Journal of Business and Economic Statistics* (forthcoming).

Andridge, R.R. and Little, R.J.A. (2010), "A Review of Hot Deck Imputation for Survey Non-response," *International Statistical Review* (forthcoming).

Avery, R.B., Beeson, P.E., and Calem, P.S. (1997), "Using HMDA Data as a Regulatory Screen for Fair Lending Compliance," *Journal of Financial Services Research*, vol. 11, 9-42.

Avery, R.B, Canner, G.B., and Cook, R.E. (2005), "New Information Reported Under HMDA and Its Application in Fair Lending Enforcement," *Federal Reserve Bulletin*, vol. 91, 344-394.

Billingsley, P. (1995), *Probability and Measure*, third edition. Wiley, New York.

Bollinger, C.R. and Hirsch, B.T. (2009), "Wage Gap Estimation with Proxies and Nonresponse," mimeo.

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.

Gu, X.S. and Rosenbaum, P.R. (1993), "Comparison of Multivariate Matching Methods: Structures, Distances and Algorithms," *Journal of Computational and Graphical Statistics*, 2, 405-420.

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., Ichimura, H., and Todd, P. (1998), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, vol. 65, 261-294.

Hirsch, B.T. and Schumacher, E.J. (2004), "Match Bias in Wage Gap Estimates Due to Earnings Imputation," *Journal of Labor Economics*, vol. 22, no. 3, 689-722.

Iacus, S.M., King, G., and Porro, G. (2009),"Causal Inference Without Balance Checking: Coarsened Exact Matching," mimeo.

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.

Little, R.J.A. and Rubin, D.B. (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.

Munnell, A.H., Tootell, G.M.B., Browne, L.E. and McEneaney, J. (1996), “Mortgage Lending in Boston: Interpreting HMDA Data,” *American Economic Review*, vol. 86, no. 1, 25-53.

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

Rosenbaum, P.R. (2002), *Observational Studies*, second edition. Springer, New York.

Rubin, D.B. (1973a), “Matching to Reduce Bias in Observational Studies,” *Biometrics*, 29, 159-183.

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

Rubin, D.B. (1977), “Assignment to Treatment Group on the Basis of a Covariate”, *Journal of Educational Statistics*, 2, 1-26.

Rubin, D.B. (2006), *Matched Sampling for Causal Effects.* Cambridge University Press, New York.

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.

Table I – Boston HMDA Data, Simulation Results
Black-White Difference in Mortgage Denial Probability for Matched Pairs
(Number of simulations = 10000)

Sample sizes		Bias		Variance		Coverage of 95% C.I.	
		(1)	(2)	(3)	(4)	(5)	(6)
		$\lvert E[\widehat{\tau}] - \tau\rvert$	$\lvert E[\widehat{\tau}_{\mathrm{bc}}] - \tau\rvert$	$\mathrm{var}(\widehat{\tau})$	$E[\widehat{\sigma}^2/N_1]$	$\widehat{\tau} \pm 1.96\,\widehat{\sigma}/\sqrt{N_1}$	$\widehat{\tau}_{\mathrm{bc}} \pm 1.96\,\widehat{\sigma}/\sqrt{N_1}$
$N_1 = 25$	$N_0 = 250$	0.0143	0.0012	0.0091	0.0091	0.9225	0.9348
	$N_0 = 500$	0.0106	0.0001	0.0092	0.0091	0.9244	0.9394
	$N_0 = 1000$	0.0077	0.0002	0.0090	0.0091	0.9263	0.9430
$N_1 = 50$	$N_0 = 500$	0.0106	0.0011	0.0045	0.0045	0.9427	0.9458
	$N_0 = 1000$	0.0073	0.0009	0.0044	0.0046	0.9427	0.9456
$N_1 = 100$	$N_0 = 1000$	0.0090	0.0001	0.0023	0.0023	0.9436	0.9468

Table II – Current Population Survey Data, Simulation Results
Average Log Weekly Earnings
(Number of simulations = 50000)

Sample size	Variance			Ratio	Coverage of 95% C.I.	
	(1)	(2)	(3)	(4)	(5)	(6)
N	$\text{var}(\bar{Y})$	$E[\widehat{\sigma}^2_{\text{adj}}/N]$	$E[\widehat{\sigma}^2/N]$	(3)/(1)	$\bar{Y} \pm 1.96\, \widehat{\sigma}_{\text{adj}}/\sqrt{N}$	$\bar{Y} \pm 1.96\, \widehat{\sigma}/\sqrt{N}$
50	0.0072	0.0071	0.0052	0.7262	0.9436	0.8973
100	0.0039	0.0039	0.0026	0.6701	0.9476	0.8888
200	0.0021	0.0021	0.0013	0.6342	0.9492	0.8799
856	0.0005	0.0005	0.0003	0.5834	0.9482	0.8661