

Revision in progress.
December 17, 2004

Large Sample Properties of Matching Estimators for Average Treatment Effects

Alberto Abadie – Harvard University and NBER

Guido Imbens – UC Berkeley and NBER

First version: July 2001

This version: December 2004

We wish to thank Don Andrews, Joshua Angrist, Gary Chamberlain, Geert Dhaene, Jinyong Hahn, James Heckman, Keisuke Hirano, Hidehiko Ichimura, Whitney Newey, Jack Porter, James Powell, Geert Ridder, Paul Rosenbaum, Ed Vytlačil, a co-editor and two anonymous referees, and seminar participants at various universities for comments, and Don Rubin for many discussions on these topics. Financial support for this research was generously provided through NSF grants SES-0350645 (Abadie), SBR-9818644 and SES-0136789 (Imbens).

LARGE SAMPLE PROPERTIES OF MATCHING ESTIMATORS
FOR AVERAGE TREATMENT EFFECTS

ALBERTO ABADIE AND GUIDO W. IMBENS

ABSTRACT

Matching estimators for average treatment effects are widely used in evaluation research despite the fact that their large sample properties have not been established in many cases. The absence of formal results in this area may be partly due to the fact that standard asymptotic expansions do not apply to simple matching estimators, which are highly nonsmooth functionals of the data. In this article, we develop new methods to analyze the properties of matching estimators and establish a number of new results. First, we show that matching estimators are not $N^{1/2}$ -consistent in general and describe conditions under which matching estimators attain $N^{1/2}$ -consistency. Second, we show that even for cases that matching estimators are $N^{1/2}$ -consistent, simple matching estimators with a fixed number of matches do not attain the semiparametric efficiency bound. Third, we provide new a estimator for the variance that does not require consistent nonparametric estimation of unknown functions.

keywords: *matching estimators, average treatment effects, unconfoundedness, selection-on-observables, potential outcomes*

Alberto Abadie
John F. Kennedy School of Government
Harvard University
79 John F. Kennedy Street
Cambridge, MA 02138
and NBER
alberto_abadie@harvard.edu
<http://www.ksg.harvard.edu/fs/aabadie/>

Guido Imbens
Department of Economics, and Department
of Agricultural and Resource Economics
University of California at Berkeley
661 Evans Hall #3880
Berkeley, CA 94720-3880
and NBER
imbens@econ.berkeley.edu
<http://elsa.berkeley.edu/users/imbens/>

1. INTRODUCTION

Estimation of average treatment effects is an important goal of much evaluation research, both in academic studies, as well as in government sponsored evaluations of social programs. Often, analyses are based on the assumptions that *(i)* assignment to treatment is unconfounded or exogenous, that is, based on observable pretreatment variables only, and *(ii)* there is sufficient overlap in the distributions of the pretreatment variables. Methods for estimating average treatment effects in parametric settings under these assumptions have a long history.¹ Recently, a number of nonparametric implementations of this idea have been proposed. Hahn (1998) calculates the efficiency bound and proposes an asymptotically efficient estimator based on nonparametric series estimation. Heckman, Ichimura, and Todd (1998) and Heckman, Ichimura, Smith, and Todd (1998) focus on the average effect on the treated and consider estimators based on local linear kernel regression methods. Hirano, Imbens, and Ridder (2000) propose an estimator that weights the units by the inverse of their assignment probabilities, and show that nonparametric series estimation of this conditional probability, labeled the propensity score by Rosenbaum and Rubin (1983), leads to an efficient estimator.

Empirical researchers, however, often use simple matching procedures to estimate average treatment effects when assignment for treatment is believed to be unconfounded. Much like nearest neighbor estimators, these procedures match each treated unit to a fixed number of untreated units with similar values for the pretreatment variables. The average effect of the treatment is then estimated by averaging within-match differences in the outcome variable between the treated and the untreated units (see, e.g., Rosenbaum, 1995; Dehejia and Wahba, 1999). Matching estimators have great intuitive appeal, and are widely used in practice. However, their formal large sample properties have not been established. Part of the reason may be that simple matching estimators are highly non-smooth functionals of the distribution of the data, not amenable to standard asymptotic methods for smooth functionals. In this article, we study the large sample properties of matching estimators of average treatment effects and establish a number of new results.

Our results show that some of the formal large sample properties of simple matching estimators are not very attractive. First, we show that when matching is done in multiple dimensions, matching estimators include a conditional bias term whose stochastic order increases with the number of continuous matching variables. We shown that the order of this conditional bias term

¹See for example Cochran and Rubin (1973), Rubin (1977), Barnow, Cain, and Goldberger (1980), Rosenbaum and Rubin (1983), Heckman and Robb (1984), and Rosenbaum (1995).

may be greater than $N^{-1/2}$. As a result, matching estimators are in general not $N^{1/2}$ -consistent. Second, even if the dimension of the covariates is low enough for the conditional bias term to vanish asymptotically, we show that the simple matching estimator with a fixed number of matches does not achieve the semiparametric efficiency bound as calculated by Hahn (1998). However, for the case when only a single continuous covariate is used to match, we show that the efficiency loss can be made arbitrarily close to zero by allowing a sufficiently large number of matches. Despite these poor formal properties, matching estimators do have some attractive features that may account for their popularity. In particular, matching estimators are extremely easy to implement, and they do not require consistent nonparametric estimation of unknown functions.

In this paper we also propose a consistent estimator for the variances of matching estimators that does not require consistent non-parametric estimation of unknown functions.

In the next section we introduce the notation and define the estimators. In Section 3 we discuss the large sample properties of simple matching estimators. In Section 4 we propose an estimator for the large sample variance of matching estimators. Section 5 concludes. The appendix contains proofs.

2. NOTATION AND BASIC IDEAS

2.1. NOTATION

We are interested in estimating the average effect of a binary treatment on some outcome. For unit i , with $i = 1, \dots, N$, let $Y_i(0)$ and $Y_i(1)$ denote the two potential outcomes given the control treatment and given the active treatment, respectively. The variable W_i , for $W_i \in \{0, 1\}$ indicates the treatment received. For unit i , we observe W_i and the outcome for this treatment,

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

as well as a vector of pretreatment variables or covariates, X_i . Following the literature, our main focus is on the population average treatment effect,

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

and the average effect for the treated,

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

See Rubin (1977), Heckman and Robb (1984), and Imbens (2003) for some discussion of these estimands.

We assume that assignment to treatment is unconfounded (Rosenbaum and Rubin, 1983), and that the probability of assignment is bounded away from zero and one.

ASSUMPTION 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.*

ASSUMPTION 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 $\eta > 0$.

The dimension of X , denoted by k , will be seen to play an important role in the properties of matching estimators. We assume that all covariates have continuous distributions.² The combination of the two conditions in Assumption 2 is referred to as strong ignorability (Rosenbaum and Rubin, 1983). These conditions are strong, and in many cases may not be satisfied. Compactness and convexity of the support of the covariates are convenient regularity conditions.

Heckman, Ichimura and Todd (1998) point out that for identification of the average treatment effect, τ , Assumption 2(i) can be weakened to mean independence ($\mathbb{E}[Y(w)|W, X] = \mathbb{E}[Y(w)|X]$ for $w = 0, 1$). For simplicity, we assume full independence, although for most of the results mean-independence is sufficient. When the parameter of interest is the average effect for the treated, τ^t , then the first part of Assumption 2 can be relaxed to require only that $Y(0)$, is independent of W conditional on X . Also, when the parameter of interest is τ^t the second part of Assumption 2 can be relaxed so that the support of X for the treated (\mathbb{X}_1) is a subset of the support of X for the untreated (\mathbb{X}_0).

ASSUMPTION 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 $\eta > 0$.

Under Assumption 2(i), the average treatment effect for the subpopulation with $X = x$ is equal to:

$$\tau(x) = \mathbb{E}[Y(1) - Y(0)|X = x] = \mathbb{E}[Y|W = 1, X = x] - \mathbb{E}[Y|W = 0, X = x]. \quad (1)$$

²Discrete covariates with a finite number of support points can be easily dealt with by analyzing estimation of average treatment effects within subsamples defined by their values. The number of such covariates does not affect the asymptotic properties of the estimators. In small samples, however, matches along discrete covariates may not be exact, so discrete covariates may create the same type of biases as continuous covariates.

Under Assumption 2(ii), the difference on the right hand side of equation (1) is identified for almost all x in \mathbb{X} . Therefore, the average effect of the treatment can be recovered by averaging $\mathbb{E}[Y|W = 1, X = x] - \mathbb{E}[Y|W = 0, X = x]$ over the distribution of X :

$$\tau = \mathbb{E}[\tau(X)] = \mathbb{E}[\mathbb{E}[Y|W = 1, X = x] - \mathbb{E}[Y|W = 0, X = x]].$$

Under Assumption 2'(i), the average treatment effect for the subpopulation with $X = x$ and $W = 1$ is equal to:

$$\tau^t(x) = \mathbb{E}[Y(1) - Y(0)|W = 1, X = x] = \mathbb{E}[Y|W = 1, X = x] - \mathbb{E}[Y|W = 0, X = x]. \quad (2)$$

Under Assumption 2'(ii), the difference on the right hand side of equation (2) is identified for all x in \mathbb{X}_1 . Therefore, the average effect of the treatment on the treated can be recovered by averaging $\mathbb{E}[Y|W = 1, X = x] - \mathbb{E}[Y|W = 0, X = x]$ over the distribution of X conditional on $W = 1$:

$$\tau^t = \mathbb{E}[\tau^t(X)|W = 1] = \mathbb{E}[\mathbb{E}[Y|W = 1, X = x] - \mathbb{E}[Y|W = 0, X = x]|W = 1].$$

Next, we introduce some additional notation. For $x \in \mathbb{X}$ and $w \in \{0, 1\}$, let $\mu(x, w) = \mathbb{E}[Y|X = x, W = w]$, $\mu_w(x) = \mathbb{E}[Y(w)|X = x]$, $\sigma^2(x, w) = \mathbb{V}(Y|X = x, W = w)$, $\sigma_w^2(x) = \mathbb{V}(Y(w)|X = x)$, and $\varepsilon_i = Y_i - \mu_{W_i}(X_i)$. Under Assumption 2, $\mu(x, w) = \mu_w(x)$ and $\sigma^2(x, w) = \sigma_w^2(x)$. Let $f_w(x)$ be the conditional density of X given $W = w$, and let $e(x) = \Pr(W = 1|X = x)$ be the propensity score (Rosenbaum and Rubin, 1983). In part of our analysis, we adopt the following assumption.

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

In some cases, however, treated and untreated are sampled separately and their proportions in the sample may not reflect their proportions in the population. Therefore, we relax Assumption 3 so that conditional on W_i sampling is random. As we will show later, relaxing Assumption 3 is particularly useful when the parameter of interest is the average treatment effect on the treated. The numbers of control and treated units are N_0 and N_1 respectively, with $N = N_0 + N_1$. For reasons that will become clear later, to estimate average effects on the treated we will require that the size of the control group is at least of the same order of magnitude as the size of the treatment group.

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

In this paper we focus on 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.³ Let $j_m(i)$ be the index j that solves $W_j = 1 - W_i$ and

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

where $1\{\cdot\}$ is the indicator function, equal to one if the expression in brackets is true and zero otherwise. In other words, $j_m(i)$ is the index of the unit that is the m -th closest to unit i in terms of the covariate values, among the units with the treatment opposite to that of unit i . In particular, $j_1(i)$, which will be sometimes denoted by $j(i)$, is the nearest match for unit i . For notational simplicity and because we only consider continuous covariates, we ignore the possibility of ties, which happen with probability zero. Let $\mathcal{J}_M(i)$ denote the set of indices for the first M matches for unit i : $\mathcal{J}_M(i) = \{j_1(i), \dots, j_M(i)\}$.⁴ Finally, let $K_M(i)$ denote the number of times unit i is used as a match given that M matches per unit are done:

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

The distribution of $K_M(i)$ will play an important role in the variance of the estimators.

In many analyses of matching methods (e.g., Rosenbaum, 1995), matching is carried out without replacement, so that every unit is used as a match at most once, and $K_M(i) \leq 1$. In this article, however, we focus on matching with replacement, allowing each unit to be used as a match more than once. Matching with replacement produces matches of higher quality than matching without replacement by increasing the set of possible matches.⁵ In addition, matching with replacement has the advantage that it allows us to consider estimators that match all units, treated as well as controls, so that the estimand is identical to the population average treatment effect.

2.2. ESTIMATORS

The unit level treatment effect is $\tau_i = Y_i(1) - Y_i(0)$. For the units in the sample, only one of the potential outcomes, $Y_i(0)$ and $Y_i(1)$, is observed and the other is unobserved or missing. All

³Alternative norms of the form $\|x\|_V = (x'Vx)^{1/2}$ for some positive definite symmetric matrix V are also covered by the results below, because $\|x\|_V = ((Px)'(Px))^{1/2}$ for P such that $P'P = V$.

⁴For this definition to make sense, we assume that $N_0 \geq M$ and $N_1 \geq M$. We maintain this assumption implicit throughout.

⁵As we show below, inexact matches generate bias in matching estimators. Therefore, expanding the set of possible matches will tend to produce smaller biases.

estimators for the average treatment effects we consider impute the expected potential outcomes in some way. The first estimator, the simple matching estimator, uses the following estimates for the expected potential outcomes:

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

leading to the following estimator for the average treatment effect:

$$\hat{\tau}_M^{sm} = \frac{1}{N} \sum_{i=1}^N (\hat{Y}_i(1) - \hat{Y}_i(0)) = \frac{1}{N} \sum_{i=1}^N (2W_i - 1) \left(1 + \frac{K_M(i)}{M} \right) Y_i. \quad (3)$$

The simple matching estimator can easily be modified to estimate the average treatment effect on the treated:

$$\hat{\tau}_M^{sm,t} = \frac{1}{N_1} \sum_{W_i=1} (Y_i - \hat{Y}_i(0)) = \frac{1}{N_1} \sum_{i=1}^N \left(W_i - (1 - W_i) \frac{K_M(i)}{M} \right) Y_i. \quad (4)$$

It is useful to compare matching estimators to covariance-adjustment or regression imputation estimators. Let $\hat{\mu}_w(X_i)$ be a consistent estimator of $\mu_w(X_i)$. Let

$$\bar{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0, \\ \hat{\mu}_0(X_i) & \text{if } W_i = 1, \end{cases} \quad \text{and} \quad \bar{Y}_i(1) = \begin{cases} \hat{\mu}_1(X_i) & \text{if } W_i = 0, \\ Y_i & \text{if } W_i = 1. \end{cases} \quad (5)$$

The regression imputation estimators of τ and τ^t are

$$\hat{\tau}^{reg} = \frac{1}{N} \sum_{i=1}^N (\bar{Y}_i(1) - \bar{Y}_i(0)) \quad \text{and} \quad \hat{\tau}^{reg,t} = \frac{1}{N_1} \sum_{W_i=1} (Y_i - \bar{Y}_i(0)). \quad (6)$$

In our discussion we classify as regression imputation estimators those for which $\hat{\mu}_w(x)$ is a consistent estimator of $\mu_w(x)$. The estimators proposed by Hahn (1998) and some of those proposed by Heckman, Ichimura, and Todd (1997) and Heckman, Ichimura, Smith, and Todd (1998) fall into this category.

If $\mu_w(X_i)$ is estimated using a nearest neighbor estimator with a fixed number of neighbors, then the regression imputation estimator is identical to the matching estimator with the same number of matches. The two estimators differ in the way they change with the sample size. We classify as matching estimators those estimators which use a finite and fixed number of matches. Interpreting matching estimators in this way may provide some intuition for some of the subsequent results. In nonparametric regression methods one typically chooses smoothing parameters to balance bias and variance of the estimated regression function. For example, in

kernel regression a smaller bandwidth leads to lower bias but higher variance. A nearest neighbor estimator with a single neighbor is at the extreme end of this. The bias is minimized within the class of nearest neighbors estimators but the variance no longer vanishes with the sample size. Nevertheless, as we shall show, matching estimators of average treatment effects are consistent under weak regularity conditions. The variance of matching estimators, however, is still relatively high and, as a result, matching with a fixed number of matches does not lead to an efficient estimator.

The first goal of our paper is to derive the properties of the simple matching estimator in large samples, that is, as N increases, for fixed M . The motivation for our fixed- M asymptotics is to provide an approximation to the sampling distribution of matching estimators with a small number of matches, because matching estimators with a small number of matches have been widely used in practice. The properties of interest include bias and variance. Of particular interest is the dependence of these results on the dimension of the covariates. A second goal is to provide methods for conducting inference through estimation of the large sample variance of the matching estimator.

3. SIMPLE MATCHING ESTIMATORS

In this section we investigate the properties of the simple matching estimator, $\hat{\tau}_M^{sm}$, defined in (3). We can write the difference between the matching estimator, $\hat{\tau}_M^{sm}$, and the population average treatment effect τ as

$$\hat{\tau}_M^{sm} - \tau = \left(\overline{\tau(X)} - \tau \right) + E_M^{sm} + B_M^{sm}, \quad (7)$$

where $\overline{\tau(X)}$ is the average conditional treatment effect:

$$\overline{\tau(X)} = \frac{1}{N} \sum_{i=1}^N (\mu_1(X_i) - \mu_0(X_i)), \quad (8)$$

E_M^{sm} is a weighted average of the residuals:

$$E_M^{sm} = \frac{1}{N} \sum_{i=1}^N E_{M,i}^{sm} = \frac{1}{N} \sum_{i=1}^N (2W_i - 1) \left(1 + \frac{K_M(i)}{M} \right) \varepsilon_i, \quad (9)$$

and B_M^{sm} is the conditional bias relative to $\overline{\tau(X)}$:

$$B_M^{sm} = \frac{1}{N} \sum_{i=1}^N B_{M,i}^{sm} = \frac{1}{N} \sum_{i=1}^N (2W_i - 1) \frac{1}{M} \sum_{m=1}^M \left(\mu_{1-W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)}) \right). \quad (10)$$

The first two terms on the right hand side of equation (7), $(\overline{\tau(X)} - \tau)$ and E_M^{sm} , have zero mean. They will be shown to be $N^{1/2}$ -consistent and asymptotically normal. The first term depends only on the covariates, and its variance is $V^{\tau(X)}/N$, where $V^{\tau(X)} = \mathbb{E}[(\tau(X) - \tau)^2]$ is the variance of the conditional average treatment effect $\tau(X)$. Conditional on \mathbf{X} and \mathbf{W} , (the matrix and vector with i -th row equal to X_i' and W_i respectively) the variance of $\widehat{\tau}_M^{sm}$ is equal to the conditional variance of the second term, E_M^{sm} . We will analyze the variances of these two terms in Section 3.2. We will refer to the third term on the right hand side of equation (7), B_M^{sm} , as the conditional bias, and to $\text{Bias}_M^{sm} = \mathbb{E}[B_M^{sm}]$ as the (unconditional) bias. If matching is exact, $X_i = X_{j_m(i)}$ for all i , and the conditional bias is equal to zero. In general it is not and its properties will be analyzed in Section 3.1.

Similarly, we can write the estimator for the average effect for the treated, (4), as

$$\widehat{\tau}_M^{sm,t} - \tau^t = \left(\overline{\tau(X)}^t - \tau^t \right) + E_M^{sm,t} + B_M^{sm,t}, \quad (11)$$

where

$$\overline{\tau(X)}^t = \frac{1}{N_1} \sum_{i=1}^N W_i (\mu(X_i, 1) - \mu_0(X_i)),$$

is the average conditional treatment effect for the treated,

$$E_M^{sm,t} = \frac{1}{N_1} \sum_{i=1}^N E_{M,i}^{sm,t} = \frac{1}{N_1} \sum_{i=1}^N (W_i - (1 - W_i)K_M(i)/M)\varepsilon_i,$$

is the contribution of the residuals, and

$$B_M^{sm,t} = \frac{1}{N_1} \sum_{i=1}^N B_{M,i}^{sm,t} = \frac{1}{N_1} \sum_{i=1}^N W_i \frac{1}{M} \sum_{m=1}^M (\mu_0(X_i) - \mu_0(X_{j_m(i)})),$$

is the conditional bias.

3.1. BIAS

Here we investigate the stochastic order of the conditional bias (10) and its counterpart for the average treatment effect for the treated. The conditional bias consists of terms of the form $\mu_1(X_{j_m(i)}) - \mu_1(X_i)$ or $\mu_0(X_i) - \mu_0(X_{j_m(i)})$. To investigate the nature of these terms expand the difference $\mu_1(X_{j_m(i)}) - \mu_1(X_i)$ around X_i :

$$\begin{aligned} \mu_1(X_{j_m(i)}) - \mu_1(X_i) &= (X_{j_m(i)} - X_i)' \frac{\partial \mu_1}{\partial x}(X_i) \\ &\quad + \frac{1}{2} (X_{j_m(i)} - X_i)' \frac{\partial^2 \mu_1}{\partial x \partial x'}(X_i) (X_{j_m(i)} - X_i) + O(\|X_{j_m(i)} - X_i\|^3). \end{aligned}$$

In order to study the components of the bias it is therefore useful to analyze the distribution of the matching discrepancy $X_{j_m(i)} - X_i$.

First, let us analyze the matching discrepancy at a general level. Fix the covariate value at $X = z$, and suppose we have a random sample X_1, \dots, X_N with density $f(x)$ and distribution function $F(x)$ over the support \mathbb{X} which is bounded. Now, consider the closest match to z in the sample. Let $j_1 = \operatorname{argmin}_{j=1, \dots, N} \|X_j - z\|$, and let $U_1 = X_{j_1} - z$ be the matching discrepancy. We are interested in the distribution of the difference U_1 , which is a $k \times 1$ vector. More generally, we are interested in the distribution of the m -th closest matching discrepancy, $U_m = X_{j_m} - z$, where X_{j_m} is the m -th closest match to z from the random sample of size N . The following lemma describes some key asymptotic properties of the matching discrepancy at interior points of the support of X .

LEMMA 1: (MATCHING DISCREPANCY – ASYMPTOTIC PROPERTIES)

Suppose that $f(z) > 0$ and that f is differentiable in a neighborhood of z . Let $V_m = N^{1/k}U_m$ and f_{V_m} be the density of V_m . Then, as $N \rightarrow \infty$,

$$\lim_{N \rightarrow \infty} f_{V_m}(v) = \frac{f(z)}{(m-1)!} \left(\|v\|^k \frac{f(z)}{k} \frac{2\pi^{k/2}}{\Gamma(k/2)} \right)^{m-1} \exp \left(-\|v\|^k \frac{f(z)}{k} \frac{2\pi^{k/2}}{\Gamma(k/2)} \right),$$

where $\Gamma(y) = \int_0^\infty e^{-t} t^{y-1} dt$ (for $y > 0$) is Euler's Gamma Function, so that $U_m = O_p(N^{-1/k})$.

Moreover, the first three moments of U_m are:

$$\mathbb{E}[U_m] = \Gamma \left(\frac{mk+2}{k} \right) \frac{1}{(m-1)!k} \left(f(z) \frac{\pi^{k/2}}{\Gamma(1+k/2)} \right)^{-2/k} \frac{1}{f(z)} \frac{\partial f}{\partial x}(z) \frac{1}{N^{2/k}} + o \left(\frac{1}{N^{2/k}} \right),$$

$$\mathbb{E}[U_m U_m'] = \Gamma \left(\frac{mk+2}{k} \right) \frac{1}{(m-1)!k} \left(f(z) \frac{\pi^{k/2}}{\Gamma(1+k/2)} \right)^{-2/k} \frac{1}{N^{2/k}} I_k + o \left(\frac{1}{N^{2/k}} \right),$$

and

$$\mathbb{E}[\|U_m\|^3] = O \left(N^{-3/k} \right),$$

where I_k is the identity matrix of size k .

(All proofs are given in the appendix.)

This lemma shows that the order of the matching discrepancy increases with the number of continuous covariates. The lemma also shows that the first term in the stochastic expansion of $N^{1/k}U_m$ has a rotation invariant distribution with respect to the origin. The following lemma shows that for all points in the support, including boundary points, the normalized moments of the matching discrepancies, U_m , are bounded.

LEMMA 2: (MATCHING DISCREPANCY – UNIFORMLY BOUNDED MOMENTS)

If Assumption 1 holds, then all the moments of $N^{1/k}\|U_m\|$ are uniformly bounded in N and $z \in \mathbb{X}$.

These results allow us to bound the stochastic order of the conditional bias.

THEOREM 1: (ORDER OF THE CONDITIONAL BIAS FOR THE AVERAGE TREATMENT EFFECT)

Under assumptions 1, 2 and 3: (i) if $\mu_0(x)$ and $\mu_1(x)$ are Lipschitz on \mathbb{X} , then $B_M^{sm} = O_p(N^{-1/k})$, and (ii) the order of $\text{Bias}_M^{sm} = \mathbb{E}[B_M^{sm}]$ is not in general lower than $N^{-2/k}$.

Consider the implications of this theorem for the asymptotic properties of the simple matching estimator. First notice that, under regularity conditions, $\sqrt{N}(\tau(\bar{X}) - \tau) = O_p(1)$ with a normal limiting distribution, by a standard central limit theorem. Also, it will be shown later that, under regularity conditions $\sqrt{N}E_M^{sm} = O_p(1)$, again with a normal limiting distribution. However, the result of the theorem implies that $\sqrt{N}B_M^{sm}$ is not $O_p(1)$ in general. In particular, if k is large enough the asymptotic distribution of $\sqrt{N}(\hat{\tau}_M^{sm} - \tau)$ is dominated by the bias term and the simple matching estimator is not $N^{1/2}$ -consistent. However, if only one of the covariates is continuously distributed, then $k = 1$ and $B_M^{sm} = O_p(N^{-1})$, so $\sqrt{N}(\hat{\tau}_M^{sm} - \tau)$ will be asymptotically normal.

A similar result holds for the average treatment effect on the treated.

THEOREM 2: (ORDER OF THE CONDITIONAL BIAS FOR THE AVERAGE TREATMENT EFFECT ON THE TREATED)

Under assumptions 1, 2' and 3',

(i) if $\mu_0(x)$ is Lipschitz on \mathbb{X}_0 , then $B_M^{sm,t} = O_p(N_1^{-r/k})$, and

(ii) if \mathbb{X}_1 is a compact subset of the interior of \mathbb{X}_0 , $\mu_0(x)$ has bounded third derivatives in the interior of \mathbb{X}_0 , $f_0(x)$ is differentiable in the interior of \mathbb{X}_0 with bounded derivatives, then

$$\text{Bias}_M^{sm,t} = \mathbb{E}[B_M^{sm,t}] = - \left(\frac{1}{M} \sum_{m=1}^M \Gamma \left(\frac{mk+2}{k} \right) \frac{1}{(m-1)!k} \right) \frac{1}{N_1^{2r/k}} \times \\ \theta^{2/k} \int \left(f_0(x) \frac{\pi^{k/2}}{\Gamma(1+k/2)} \right)^{-2/k} \left\{ \frac{1}{f_0(x)} \frac{\partial f_0}{\partial x'}(x) \frac{\partial \mu_0}{\partial x}(x) + \frac{1}{2} \text{tr} \left(\frac{\partial^2 \mu_0}{\partial x' \partial x}(x) \right) \right\} f_1(x) dx + o \left(\frac{1}{N_1^{2r/k}} \right).$$

This case is particularly relevant because often matching estimators have been used to estimate the average effect for the treated, in settings in which a large number of controls are sampled separately. Generally, in those cases, the conditional bias term has been ignored in the asymptotic approximation to standard errors and confidence intervals. That is justified if N_0 is of sufficiently high order relative to N_1 , or, to be precise, if $r > k/2$. In that case it follows that $B_M^{sm,t} =$

$o_p(N_1^{-1/2})$, and the bias term will get dominated in the large sample distribution by the two other terms, $\overline{\tau(X)}^t - \tau^t$ and $E_M^{sm,t}$, both of which are $O_p(N_1^{-1/2})$.

In part (ii) of Theorem 2, we show that a general expression of the bias $\text{Bias}_M^{sm,t}$ can be calculated if \mathbb{X}_1 is compact and $\mathbb{X}_1 \subset \text{int } \mathbb{X}_0$ (so that the bias is not affected by the geometric characteristics of the boundary of \mathbb{X}_0). Under these conditions, the bias of the matching estimator is at most of order $N_1^{-2/k}$. This bias is further reduced when μ_0 is constant or when μ_0 is linear and f_0 is constant, among other cases. Note, however, that randomizing the treatment, $f_0 = f_1$ does not reduce the order of $\text{Bias}_M^{sm,t}$.

3.2. VARIANCE

In this section we investigate the variance of the simple matching estimator, $\hat{\tau}_M^{sm}$. We focus on the first two terms of the simple matching estimator, (8) and (9), ignoring for the moment the conditional bias term (10). Conditional on \mathbf{X} and \mathbf{W} , the matrix and vector with i -th row equal to X_i' and W_i respectively, the variance of $\hat{\tau}_M^{sm}$ is

$$\mathbb{V}(\hat{\tau}_M^{sm} | \mathbf{X}, \mathbf{W}) = \frac{1}{N^2} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M}\right)^2 \sigma^2(X_i, W_i). \quad (12)$$

For $\hat{\tau}_M^{sm,t}$ we obtain:

$$\mathbb{V}(\hat{\tau}_M^{sm,t} | \mathbf{X}, \mathbf{W}) = \frac{1}{N_1^2} \sum_{i=1}^N (W_i - (1 - W_i)K_M(i)/M)^2 \sigma^2(X_i, W_i). \quad (13)$$

Let $V^E = N\mathbb{V}(\hat{\tau}_M^{sm} | \mathbf{X}, \mathbf{W})$ and $V^{E,t} = N_1\mathbb{V}(\hat{\tau}_M^{sm,t} | \mathbf{X}, \mathbf{W})$ be the corresponding normalized variances. Ignoring the conditional bias term, B_M^{sm} , the conditional expectation of $\hat{\tau}_M^{sm}$ is $\overline{\tau(X)}$. The variance of this conditional mean is therefore $V^{\tau(X)}/N$, where $V^{\tau(X)} = \mathbb{E}[(\tau(X) - \tau)^2]$. Hence the marginal variance of $\hat{\tau}_M^{sm}$, ignoring the conditional bias term, is $\mathbb{V}(\hat{\tau}_M^{sm}) = (\mathbb{E}[V^E] + V^{\tau(X)})/N$. For the estimator for the average effect on the treated the marginal variance is, again ignoring the conditional bias term, $\mathbb{V}(\hat{\tau}_M^{sm,t}) = (\mathbb{E}[V^{E,t}] + V^{\tau(X),t})/N_1$, where $V^{\tau(X),t} = \mathbb{E}[(\tau^t(X) - \tau^t)^2 | W = 1]$.

The following lemma shows that the expectation of the normalized variance is finite. The key is that $K_M(i)$, the number of times that unit i is used as a match, is $O_p(1)$ with finite moments.⁶

LEMMA 3: (i) Suppose assumptions 1-3 hold, then $K_M(i) = O_p(1)$ and its moments are bounded uniformly in N . (ii) If, in addition, $\sigma^2(x, w)$ are Lipschitz in \mathbb{X} for $w = 0, 1$, then $\mathbb{E}[V^E + V^{\tau(X)}] = O(1)$. (iii) Suppose Assumptions 1, 2' and 3', then $(N_0/N_1)\mathbb{E}[K_M(i)^q | W_i = 0]$ is

⁶Notice that, for $1 \leq i \leq N$, $K_M(i)$ are exchangeable random variables, and therefore have identical marginal distributions.

uniformly bounded in N for all $q \geq 1$. (iv) If also $\sigma^2(x, w)$ are Lipschitz in \mathbb{X} for $w = 0, 1$, then $\mathbb{E}[V^{E,t} + V^{\tau(X),t}] = O(1)$.

3.3. CONSISTENCY AND ASYMPTOTIC NORMALITY

In this section we show that the simple matching estimator is consistent for the average treatment effect and that, without the conditional bias term, is $N^{1/2}$ -consistent and asymptotically normal. The next assumption contains a set of weak smoothness restrictions on the conditional distribution of Y given X . Notice that it does not require existence of higher order derivatives.

ASSUMPTION 4: (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.

THEOREM 3: (CONSISTENCY OF THE SIMPLE MATCHING ESTIMATOR)

(i) Suppose assumptions 1-3 and 4(i) hold. Then $\hat{\tau}_M^{sm} - \tau \xrightarrow{p} 0$. (ii) Suppose assumptions 1, 2', 3', and 4(i) hold. Then $\hat{\tau}_M^{sm,t} - \tau^t \xrightarrow{p} 0$.

Notice that the consistency result holds regardless of the dimension of the covariates.

Next, we state the formal result for asymptotic normality. The first result gives an asymptotic normality result for the estimators $\hat{\tau}_M^{sm}$ and $\hat{\tau}_M^{sm,t}$ after subtracting the bias term.

THEOREM 4: (ASYMPTOTIC NORMALITY FOR THE SIMPLE MATCHING ESTIMATOR)

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

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

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

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

Although one generally does not know the conditional bias term, this result is useful for two reasons. First, in some cases the bias term can be ignored because it is of sufficiently low order. Second, as we show in Abadie and Imbens (2003), under some conditions, an estimate of the bias term can be used in the statement of Theorem 4 without changing the result.

In the scalar covariate case, or when only the treated are matched and the size of the control group is of sufficient order of magnitude, there is no need to remove the bias.

COROLLARY 1: (ASYMPTOTIC NORMALITY FOR SIMPLE MATCHING ESTIMATOR – VANISHING BIAS)

(i) Suppose assumptions 1-3 and 4 hold, and $k = 1$. Then

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

(ii) Suppose assumptions 1, 2', 3', and 4 hold, and $r > k/2$. Then

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

3.4. EFFICIENCY

The asymptotic efficiency of the estimators considered here depends on the limit of $\mathbb{E}[V^E]$, which in turn depends on the limiting distribution of $K_M(i)$. It is difficult to work out the limiting distribution of this variable for the general case.⁷ Here we investigate the form of the variance for the special case with a scalar covariate ($k = 1$) and a general M .

THEOREM 5: Suppose $k = 1$. If Assumptions 1 to 4 hold, and f_1 and f_0 are continuous on $\text{int } \mathbb{X}$, then

$$\begin{aligned} N \cdot \mathbb{V}(\hat{\tau}_M^{sm}) &= \mathbb{E} \left[\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1 - e(X)} \right] + V^{\tau(X)} \\ &+ \frac{1}{2M} \mathbb{E} \left[\left(\frac{1}{e(X)} - e(X) \right) \sigma_1^2(X) + \left(\frac{1}{1 - e(X)} - (1 - e(X)) \right) \sigma_0^2(X) \right] + o(1). \end{aligned}$$

Note that with $k = 1$ we can ignore the conditional bias term, B_M^{sm} . The semiparametric efficiency bound for this problem is, as established by Hahn (1998),

$$V^{\text{eff}} = \mathbb{E} \left[\frac{\sigma_1^2(X)}{e(X)} + \frac{\sigma_0^2(X)}{1 - e(X)} \right] + V^{\tau(X)}.$$

The limiting variance of the matching estimator is in general larger. Relative to the efficiency bound it can be written as

$$\lim_{N \rightarrow \infty} \frac{N \cdot \mathbb{V}(\hat{\tau}_M^{sm}) - V^{\text{eff}}}{V^{\text{eff}}} < \frac{1}{2M}.$$

The asymptotic efficiency loss disappears quickly if the number of matches is large enough, and the efficiency loss from using a few matches is very small. For example, the asymptotic variance with a single match is less than 50% higher than the asymptotic variance of the efficient estimator, and with five matches the asymptotic variance is less than 10% higher.

⁷The key is the second moment of the volume of the ‘‘catchment area’’ $A_M(i)$, defined as the subset of \mathbb{X} such that each observation, j , with $W_j = 1 - W_i$ and $X_j \in A_M(i)$ is matched to i . In the single match case with $M = 1$ these objects are studied in stochastic geometry where they are known as Poisson-Voronoi tessellations (Okabe, Boots, Sugihara and Nok Chiu, 2000). The variance of the volume of such objects under uniform $f_0(x)$ and $f_1(x)$, normalized by the mean, has been worked out numerically for the one, two, and three dimensional cases.

4. ESTIMATING THE VARIANCE

Corollary 1 uses the square-roots of $V^E + V^{\tau(X)}$ and $V^{E,t} + V^{\tau(X),t}$ as normalizing factors to attain a limiting normal distribution for matching estimators. In this section, we show how to estimate these terms.

4.1. ESTIMATING THE CONDITIONAL VARIANCE

Estimating the conditional variance, $V^E = \sum_{i=1}^N (1 + K_M(i)/M)^2 \sigma_{W_i}^2(X_i)/N$, is complicated by the fact that it involves the conditional outcome variances, $\sigma_w^2(x)$. We propose an estimator of the conditional variance of the simple matching estimator which does not require consistent nonparametric estimation of $\sigma_w^2(x)$. Our method uses a matching estimator for $\sigma_w^2(x)$ where instead of the original matching of treated to control units, we now match treated units to treated units and control units to control units.

Let $l_j(i)$ be the j -th closest unit to unit i among the units with the same value for the treatment. Then, we estimate the conditional variance as

$$\hat{\sigma}_{W_i}^2(X_i) = \frac{J}{J+1} \left(Y_i - \frac{1}{J} \sum_{j=1}^J Y_{l_j(i)} \right)^2. \quad (14)$$

Notice that if all matches are perfect so $X_{l_j(i)} = X_i$ for all $j = 1, \dots, J$, then $\mathbb{E}[\hat{\sigma}_{W_i}^2(X_i)|X_i = x, W_i = w] = \sigma_w^2(x)$. In practice, if the covariates are continuous, it will not be possible to find perfect matches, so $\hat{\sigma}_{W_i}^2(X_i)$ will be only asymptotically unbiased. Because $\hat{\sigma}_{W_i}^2(X_i)$ is an average of a fixed number of observations, this estimator will not be consistent for $\sigma_{W_i}^2(X_i)$. However, the next theorem shows that averages of the $\hat{\sigma}_{W_i}^2(X_i)$ -s are consistent for V^E and $V^{E,t}$.

THEOREM 6: *Let $\hat{\sigma}_{W_i}^2(X_i)$ be as in equation (14). Let*

$$\hat{V}^E = \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M} \right)^2 \hat{\sigma}_{W_i}^2(X_i), \quad \hat{V}^{E,t} = \frac{1}{N_1} \sum_{i=1}^N \left(W_i - (1 - W_i) \frac{K_M(i)}{M} \right)^2 \hat{\sigma}_{W_i}^2(X_i).$$

If Assumptions 1 to 4 hold, then $|\hat{V}^E - V^E| = o_p(1)$. If Assumptions 1', 3' and 4 hold, then $|\hat{V}^{E,t} - V^{E,t}| = o_p(1)$.

4.2. ESTIMATING THE MARGINAL VARIANCE

Here we develop consistent estimators for $V = V^E + V^{\tau(X)}$ and $V^t = V^{E,t} + V^{\tau(X),t}$. These estimators are based on the same matching approach to estimating the conditional error variance

$\sigma_w^2(x)$ as in the previous subsection. In addition, these estimators exploit the fact that,

$$\mathbb{E}[(\hat{Y}_i(1) - \hat{Y}_i(0) - \tau)^2] \simeq V^{\tau(X)} + \mathbb{E} \left[\varepsilon_i^2 + \frac{1}{M^2} \sum_{m=1}^M \varepsilon_{j_m(i)}^2 \right].$$

The average of the left hand side can be estimated as $\sum_i (\hat{Y}_i(1) - \hat{Y}_i(0) - \hat{\tau}_M^{sm})^2 / N$. The average of the second term on the right hand side can be estimated using the fact that

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[\varepsilon_i^2 + \frac{1}{M^2} \sum_{m=1}^M \varepsilon_{j_m(i)}^2 \mid \mathbf{X}, \mathbf{W} \right] = \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right) \sigma_{W_i}^2(X_i),$$

which can then be combined to estimate $V^{\tau(X)}$. This in turn can be combined with the previously defined estimator for V^E to obtain an estimator of V .

THEOREM 7: *Let $\hat{\sigma}^2(X_i, W_i)$ be as in equation (14). Let*

$$\hat{V} = \frac{1}{N} \sum_{i=1}^N \left(\hat{Y}_i(1) - \hat{Y}_i(0) - \hat{\tau}_M^{sm} \right)^2 + \frac{1}{N} \sum_{i=1}^N \left[\left(\frac{K_M(i)}{M} \right)^2 + \left(\frac{2M-1}{M} \right) \left(\frac{K_M(i)}{M} \right) \right] \hat{\sigma}^2(X_i, W_i)$$

and

$$\hat{V}^t = \frac{1}{N_1} \sum_{W_i=1} \left(Y_i - \hat{Y}_i(0) - \hat{\tau}_M^{sm,t} \right)^2 + \frac{1}{N_1} \sum_{i=1}^N (1 - W_i) \left(\frac{K_M(i)(K_M(i) - 1)}{M^2} \right) \hat{\sigma}^2(X_i, W_i).$$

If Assumptions 1 to 4 hold, then $|\hat{V} - V| = o_p(1)$. If Assumptions 1 2', 3' and 4 hold, then $|\hat{V}^t - V^t| = o_p(1)$.

5. CONCLUSION

In this paper we derive large sample properties of simple matching estimators that are widely used in applied evaluation research. The formal large sample properties turn out to be surprisingly poor. We show that simple matching estimators include a conditional bias term which does not disappear in large samples, under the standard $N^{1/2}$ normalization. Therefore, standard matching estimators are not $N^{1/2}$ -consistent in general. We derive the asymptotic distribution of matching estimators for the cases where the conditional bias can be ignored, and show that matching estimators with a fixed number of matches are not efficient.

APPENDIX

Before proving Lemma 1, we collect some results on integration using polar coordinates that will be useful. See for example Stroock (1999). Let $S_k = \{\omega \in \mathbb{R}^k : \|\omega\| = 1\}$ be the unit k -sphere, and λ_{S_k} be its surface measure. Then, the area and volume of the unit k -sphere are

$$\int_{S_k} \lambda_{S_k}(d\omega) = \frac{2\pi^{k/2}}{\Gamma(k/2)} \quad \text{and} \quad \int_0^1 r^{k-1} \int_{S_k} \lambda_{S_k}(d\omega) dr = \frac{2\pi^{k/2}}{k\Gamma(k/2)} = \frac{\pi^{k/2}}{\Gamma(1+k/2)},$$

respectively. In addition,

$$\int_{S_k} \omega \lambda_{S_k}(d\omega) = 0, \quad \text{and} \quad \int_{S_k} \omega \omega' \lambda_{S_k}(d\omega) = \frac{\int_{S_k} \lambda_{S_k}(d\omega)}{k} I_k = \frac{\pi^{k/2}}{\Gamma(1+k/2)} I_k,$$

where I_k is the k -dimensional identity matrix. For any non-negative measurable function $g(\cdot)$ on \mathbb{R}^k ,

$$\int_{\mathbb{R}^k} g(x) dx = \int_0^\infty r^{k-1} \left(\int_{S_k} g(r\omega) \lambda_{S_k}(d\omega) \right) dr.$$

We will also use the following result on Laplace approximation of integrals.

LEMMA A.1: *Let $a(r)$ and $b(r)$ be two real functions, $a(r)$ is continuous in a neighborhood of zero and $b(r)$ has continuous first derivative in a neighborhood of zero. Let $b(0) = 0$, $b(r) > 0$ for $r > 0$, and that for every $\tilde{r} > 0$ the infimum of $b(r)$ over $r \geq \tilde{r}$ is positive. Suppose that there exist positive real numbers a_0, b_0, α, β such that*

$$\lim_{r \rightarrow 0} a(r)r^{1-\alpha} = a_0, \quad \lim_{r \rightarrow 0} b(r)r^{-\beta} = b_0, \quad \text{and} \quad \lim_{r \rightarrow 0} \frac{db}{dr}(r)r^{1-\beta} = b_0\beta.$$

Suppose also that $\int_0^\infty |a(r)| \exp(-Nb(r)) dr < \infty$ for all sufficiently large N . Then, for $N \rightarrow \infty$

$$\int_0^\infty a(r) \exp(-Nb(r)) dr = \Gamma\left(\frac{\alpha}{\beta}\right) \frac{a_0}{\beta b_0^{\alpha/\beta}} \frac{1}{N^{\alpha/\beta}} + o\left(\frac{1}{N^{\alpha/\beta}}\right).$$

PROOF: It follows from Theorem 7.1 in Olver (1997), page 81.

PROOF OF LEMMA 1: First consider the conditional probability of unit i being the m -th closest match to z , given $X_i = x$:

$$\Pr(j_m = i | X_i = x) = \binom{N-1}{m-1} (\Pr(\|X - z\| > \|x - z\|))^{N-m} (\Pr(\|X - z\| \leq \|x - z\|))^{m-1}.$$

Because the marginal probability of unit i being the m -th closest match to z is $\Pr(j_m = i) = 1/N$, and because the density of X_i is $f(x)$, then the distribution of X_i conditional on it being the m -th closest match is:

$$\begin{aligned} f_{X_i | j_m = i}(x) &= N f(x) \Pr(j_m = i | X_i = x) \\ &= N f(x) \binom{N-1}{m-1} (1 - \Pr(\|X - z\| \leq \|x - z\|))^{N-m} (\Pr(\|X - z\| \leq \|x - z\|))^{m-1}, \end{aligned}$$

and this is also the distribution of X_{j_m} . Now transform to the matching discrepancy $U_m = X_{j_m} - z$ to get

$$f_{U_m}(u) = N \binom{N-1}{m-1} f(z+u) (1 - \Pr(\|X - z\| \leq \|u\|))^{N-m} (\Pr(\|X - z\| \leq \|u\|))^{m-1}. \quad (\text{A.1})$$

Transform to $V_m = N^{1/k}U_m$ with Jacobian N^{-1} to obtain:

$$\begin{aligned} f_{V_m}(v) &= \binom{N-1}{m-1} f\left(z + \frac{v}{N^{1/k}}\right) \left(1 - \Pr\left(\|X - z\| \leq \frac{\|v\|}{N^{1/k}}\right)\right)^{N-m} \left(\Pr\left(\|X - z\| \leq \frac{\|v\|}{N^{1/k}}\right)\right)^{m-1} \\ &= N^{1-m} \binom{N-1}{m-1} f\left(z + \frac{v}{N^{1/k}}\right) \left(1 - \Pr\left(\|X - z\| \leq \frac{\|v\|}{N^{1/k}}\right)\right)^N (1 + o(1)) \left(N \Pr\left(\|X - z\| \leq \frac{\|v\|}{N^{1/k}}\right)\right)^{m-1}. \end{aligned}$$

Note that $\Pr(\|X - z\| \leq \|v\|N^{-1/k})$ is

$$\int_0^{\|v\|/N^{1/k}} r^{k-1} \left(\int_{S_k} f(z + r\omega) \lambda_{S_k}(d\omega) \right) dr,$$

where $S_k = \{\omega \in \mathbb{R}^k : \|\omega\| = 1\}$ is the unit k -sphere, and λ_{S_k} is its surface measure. The derivative w.r.t. N is

$$\left(\frac{-1}{N^2}\right) \frac{\|v\|^k}{k} \int_{S_k} f\left(z + \frac{\|v\|^k}{N^{1/k}} \omega\right) \lambda_{S_k}(d\omega).$$

Therefore,

$$\lim_{N \rightarrow \infty} \frac{\Pr(\|X - z\| \leq \|v\|N^{-1/k})}{1/N} = \frac{\|v\|^k}{k} f(z) \int_{S_k} \lambda_{S_k}(d\omega).$$

In addition, it is easy to check that

$$N^{1-m} \binom{N-1}{m-1} = \frac{1}{(m-1)!} + o(1).$$

Therefore,

$$\lim_{N \rightarrow \infty} f_{V_m}(v) = \frac{f(z)}{(m-1)!} \left(\frac{\|v\|^k}{k} \int_{S_k} \lambda_{S_k}(d\omega) \right)^{m-1} \exp\left(-\frac{\|v\|^k}{k} \int_{S_k} \lambda_{S_k}(d\omega)\right).$$

The previous equation shows that the density of V_m converges pointwise to a non-negative function which is rotation invariant with respect to the origin. As a result, the matching discrepancy U_m is $O_p(N^{-1/k})$ and the limiting distribution of $N^{1/k}U_m$ is rotation invariant with respect to the origin. This finishes the proof of the first result.

Next, given $f_{U_m}(u)$ in (A.1),

$$\mathbb{E}U_m = N \binom{N-1}{m-1} A_m,$$

where

$$A_m = \int_{\mathbb{R}^k} u f(z + u) (1 - \Pr(\|X - z\| \leq \|u\|))^{N-m} (\Pr(\|X - z\| \leq \|u\|))^{m-1} du.$$

Boundedness of \mathbb{X} implies that A_m converges uniformly. (It is easy to relax the bounded support condition here. We maintain it because it is used elsewhere in the article.) Changing variables to polar coordinates gives:

$$A_m = \int_0^\infty r^{k-1} \left(\int_{S_k} r\omega f(z + r\omega) \lambda_{S_k}(d\omega) \right) (1 - \Pr(\|X - z\| \leq r))^{N-m} (\Pr(\|X - z\| \leq r))^{m-1} dr$$

Then rewriting the probability $\Pr(\|X - z\| \leq r)$ as

$$\int_{\mathbb{R}^k} f(x) 1\{\|x - z\| \leq r\} dx = \int_{\mathbb{R}^k} f(z + v) 1\{\|v\| \leq r\} dv = \int_0^r s^{k-1} \left(\int_{S_k} f(z + s\omega) \lambda_{S_k}(d\omega) \right) ds$$

and substituting this into the expression for A_m gives:

$$A_m = \int_0^\infty r^{k-1} \left(\int_{S_k} r\omega f(z+r\omega) \lambda_{S_k}(d\omega) \right) \left(1 - \int_0^r s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds \right)^{N-m} \\ \times \left(\int_0^r s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds \right)^{m-1} dr = \int_0^\infty e^{-Nb(r)} a(r) dr,$$

where

$$b(r) = -\log \left(1 - \int_0^r s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds \right),$$

and

$$a(r) = r^k \left(\int_{S_k} \omega f(z+r\omega) \lambda_{S_k}(d\omega) \right) \frac{\left(\int_0^r s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds \right)^{m-1}}{\left(1 - \int_0^r s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds \right)^m}.$$

That is, $a(r) = q(r)p(r)$, $q(r) = r^k c(r)$, and $p(r) = (g(r))^{m-1}$, where

$$c(r) = \frac{\int_{S_k} \omega f(z+r\omega) \lambda_{S_k}(d\omega)}{1 - \int_0^r s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds}, \quad g(r) = \frac{\int_0^r s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds}{1 - \int_0^r s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds}.$$

First notice that $b(r)$ is continuous in a neighborhood of zero and $b(0) = 0$. By Theorem 6.20 in Rudin (1976), $s^{k-1} \int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega)$ is continuous, and

$$\frac{db}{dr}(r) = \frac{r^{k-1} \left(\int_{S_k} f(z+r\omega) \lambda_{S_k}(d\omega) \right)}{1 - \int_0^r s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds},$$

which is also continuous. Using L'Hospital's rule:

$$\lim_{r \rightarrow 0} b(r)r^{-k} = \lim_{r \rightarrow 0} \frac{1}{kr^{k-1}} \frac{db}{dr}(r) = \frac{1}{k} f(z) \int_{S_k} \lambda_{S_k}(d\omega).$$

Similarly, $c(r)$ is continuous in a neighborhood of zero, $c(0) = 0$, and

$$\lim_{r \rightarrow 0} c(r)r^{-1} = \lim_{r \rightarrow 0} \frac{dc}{dr}(r) = \frac{\partial f}{\partial x}(z) \int_{S_k} \omega \omega' \lambda_{S_k}(d\omega) = \frac{1}{k} \frac{\partial f}{\partial x}(z) \int_{S_k} \lambda_{S_k}(d\omega).$$

Therefore,

$$\lim_{r \rightarrow 0} q(r)r^{-(k+1)} = \lim_{r \rightarrow 0} \frac{dc}{dr}(r) = \frac{1}{k} \frac{\partial f}{\partial x}(z) \int_{S_k} \lambda_{S_k}(d\omega).$$

Similar calculations yield

$$\lim_{r \rightarrow 0} g(r)r^{-k} = \lim_{r \rightarrow 0} \frac{1}{kr^{k-1}} \frac{dg}{dr}(r) = \frac{1}{k} f(z) \int_{S_k} \lambda_{S_k}(d\omega).$$

Therefore

$$\lim_{r \rightarrow 0} p(r)r^{-(m-1)k} = \left(\frac{1}{k} f(z) \int_{S_k} \lambda_{S_k}(d\omega) \right)^{m-1}.$$

Now, it is clear that

$$\begin{aligned} \lim_{r \rightarrow 0} a(r)r^{-(mk+1)} &= \left(\lim_{r \rightarrow 0} p(r)r^{-(m-1)k} \right) \left(\lim_{r \rightarrow 0} q(r)r^{-(k+1)} \right) \\ &= \left(\frac{1}{k} f(z) \int_{S_k} \lambda_{S_k}(d\omega) \right)^{m-1} \frac{1}{k} \frac{\partial f}{\partial x}(z) \int_{S_k} \lambda_{S_k}(d\omega) = \left(\frac{1}{k} f(z) \int_{S_k} \lambda_{S_k}(d\omega) \right)^m \frac{1}{f(z)} \frac{\partial f}{\partial x}(z). \end{aligned}$$

Therefore, the conditions of Lemma A.1 hold for $\alpha = mk + 2$, $\beta = k$

$$a_0 = \left(\frac{1}{k} f(z) \int_{S_k} \lambda_{S_k}(d\omega) \right)^m \frac{1}{f(z)} \frac{\partial f}{\partial x}(z), \quad \text{and} \quad b_0 = \frac{1}{k} f(z) \int_{S_k} \lambda_{S_k}(d\omega).$$

Applying Lemma A.1, we get

$$\begin{aligned} A_m &= \Gamma \left(\frac{mk+2}{k} \right) \frac{a_0}{k b_0^{(mk+2)/k}} \frac{1}{N^{(mk+2)/k}} + o \left(\frac{1}{N^{(mk+2)/k}} \right) \\ &= \Gamma \left(\frac{mk+2}{k} \right) \frac{1}{k} \left(f(z) \frac{\pi^{k/2}}{\Gamma(1 + \frac{k}{2})} \right)^{-2/k} \frac{1}{f(z)} \frac{df}{dx}(z) \frac{1}{N^{(mk+2)/k}} + o \left(\frac{1}{N^{(mk+2)/k}} \right). \end{aligned}$$

Therefore,

$$\mathbb{E}[U_m] = \Gamma \left(\frac{mk+2}{k} \right) \frac{1}{(m-1)!k} \left(f(z) \frac{\pi^{k/2}}{\Gamma(1 + \frac{k}{2})} \right)^{-2/k} \frac{1}{f(z)} \frac{df}{dx}(z) \frac{1}{N^{2/k}} + o \left(\frac{1}{N^{2/k}} \right),$$

which finishes the proof for the second result of the lemma. The results for $\mathbb{E}[U_m U'_m]$ and $\mathbb{E}[\|U_m\|^3]$ follow from similar arguments. \square

PROOF OF LEMMA 2: The proof consists of showing that the density of $V_m = N^{1/k} U_m$, denoted by $f_{V_m}(v)$, is bounded by $\bar{f}_{V_m}(v)$ which does not depend on N or z , followed by a proof that $\int \|v\|^L \bar{f}_{V_m}(v) dv < \infty$ for any $L > 0$. It is enough to show the result for $N > m$ (the bounded support condition implies uniformly bounded moments of V_m over $z \in \mathbb{X}$ for any given N , and in particular for $N = m$.) Recall from the proof of Lemma 1 that

$$f_{V_m}(v) = \binom{N-1}{m-1} f \left(z + \frac{v}{N^{1/k}} \right) \left(1 - \Pr \left(\|X - z\| \leq \frac{\|v\|}{N^{1/k}} \right) \right)^{N-m} \left(\Pr \left(\|X - z\| \leq \frac{\|v\|}{N^{1/k}} \right) \right)^{m-1}.$$

Define $\underline{f} = \inf_{x \in \mathbb{X}} f(x)$ and $\bar{f} = \sup_{x \in \mathbb{X}} f(x)$. By assumption, $\underline{f} > 0$ and \bar{f} is finite. Let \bar{u} be the diameter of \mathbb{X} ($\bar{u} = \sup_{x, y \in \mathbb{X}} \|x - y\|$) which is finite because \mathbb{X} is bounded by assumption. Consider all the balls $B(x, \bar{u})$ with centers $x \in \mathbb{X}$ and radius \bar{u} . Let c be the infimum over $x \in \mathbb{X}$ of the fraction that the intersection with \mathbb{X} represents in volume of the balls. Notice that, because \mathbb{X} has dimension k , then $0 < c < 1$; and that, because \mathbb{X} is convex, this proportion can only increase for a smaller radius. Let $z \in \mathbb{X}$ and $\|v\| \leq N^{1/k} \bar{u}$.

$$\begin{aligned} \Pr \left(\|X - z\| \leq \frac{\|v\|}{N^{1/k}} \right) &= \int_0^{\|v\| N^{-1/k}} r^{k-1} \int_{S_k} f(z + r\omega) \lambda_{S_k}(d\omega) dr \\ &= \int_0^{\|v\| N^{-1/k}} r^{k-1} \int_{S_k} f(z + r\omega) 1\{f(z + r\omega) > 0\} \lambda_{S_k}(d\omega) dr \\ &\geq \underline{f} \int_0^{\|v\| N^{-1/k}} r^{k-1} \int_{S_k} 1\{f(z + r\omega) > 0\} \lambda_{S_k}(d\omega) dr \\ &\geq c \underline{f} \int_0^{\|v\| N^{-1/k}} r^{k-1} \int_{S_k} \lambda_{S_k}(d\omega) dr = c \frac{\|v\|^k}{N} \underline{f} \frac{\pi^{k/2}}{\Gamma(1 + k/2)}. \end{aligned}$$

Similarly,

$$\Pr\left(\|X - z\| \leq \frac{\|v\|}{N^{1/k}}\right) \leq \frac{\|v\|^k}{N} \bar{f} \frac{\pi^{k/2}}{\Gamma(1+k/2)}.$$

Hence, using the fact that for positive a , $\log(a) \leq a - 1$ and thus for all $0 < b < N$ and $N > m$ we have $(1 - b/N)^{(N-m)} \leq \exp(-b(N-m)/N) \leq \exp(-b/(m+1))$. In addition,

$$N^{1-m} \binom{N-1}{m-1} \leq \frac{1}{(m-1)!}.$$

It follows that

$$f_{V_m}(v) \leq \bar{f}_{V_m}(v) = \frac{\bar{f}}{(m-1)!} \exp\left(-c \frac{\|v\|^k}{(m+1)} \bar{f} \frac{2\pi^{k/2}}{\Gamma(k/2)}\right) \left(\|v\|^k \bar{f} \frac{2\pi^{k/2}}{\Gamma(k/2)}\right)^{m-1} = C_1 \|v\|^{k(m-1)} \exp(-C_2 \|v\|^k),$$

with C_1 and C_2 positive. This inequality holds trivially for $\|v\| > N^{1/k} \bar{u}$. This establishes an exponential bound that does not depend on N or z . Hence for all N and z , $\int \|v\|^L \bar{f}_{V_m}(v) dv$ is finite and thus all moments of $N^{1/k} U_m$ are uniformly bounded in N and z . \square

PROOF OF THEOREM 1(i):

Let the unit-level matching discrepancy $U_{m,i} = X_i - X_{j_m(i)}$. Define the unit-level conditional bias from the m -th match as

$$\begin{aligned} B_{m,i} &= W_i (\mu_0(X_i) - \mu_0(X_{j_m(i)})) - (1 - W_i) (\mu_1(X_i) - \mu_1(X_{j_m(i)})) \\ &= W_i (\mu_0(X_i) - \mu_0(X_i + U_{m,i})) - (1 - W_i) (\mu_1(X_i) - \mu_1(X_i + U_{m,i})). \end{aligned}$$

By the Lipschitz assumption on μ_0 and μ_1 , we obtain $|B_{m,i}| \leq C_1 \|U_{m,i}\|$, for some positive constant C_1 . The bias term is

$$B_M^{sm} = \frac{1}{NM} \sum_{i=1}^N \sum_{m=1}^M B_{m,i}.$$

Using the Cauchy-Schwarz Inequality and Lemma 2:

$$\begin{aligned} \mathbb{E}[N^{2/k} (B_M^{sm})^2] &\leq C_1 N^{2/k} \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N \|U_{M,i}\|^2\right] \\ &= C_1 N^{2/k-1} \mathbb{E}\left[\frac{1}{N_0^{2/k}} \sum_{W_i=1} \mathbb{E}\left[N_0^{2/k} \|U_{M,i}\|^2 \mid W_1, \dots, W_N, X_i\right] \right. \\ &\quad \left. + \frac{1}{N_1^{2/k}} \sum_{W_i=0} \mathbb{E}\left[N_1^{2/k} \|U_{M,i}\|^2 \mid W_1, \dots, W_N, X_i\right]\right] \\ &\leq C_2 \mathbb{E}\left[\left(\frac{N}{N_0}\right)^{2/k} \frac{N_1}{N} + \left(\frac{N}{N_1}\right)^{2/k} \frac{N_0}{N}\right], \end{aligned}$$

for some positive constant, C_2 . Using Chernoff's Inequality, it can be seen that any moment of N/N_1 or N/N_0 is uniformly bounded in N (with $N_w \geq M$ for $w = 0, 1$). The result of the theorem follows now from Markov's Inequality. This proves part (i) of the theorem. We defer the proof of Theorem 1(ii) until after the proof of Theorem 2(ii), because the former will follow directly from the latter. \square

PROOF OF THEOREM 2:

The proof of first part of theorem 2 is very similar to the proof of Theorem 1, and therefore is omitted. To prove the second part of Theorem 2, the following auxiliary lemma will be useful.

LEMMA A.2: Let X be distributed with density f on some compact set of dimension k : $\mathbb{X} \subset \mathbb{R}^k$. Let \mathbb{Z} be a compact set of dimension k which is a subset of $\text{int } \mathbb{X}$. Suppose that f is bounded and bounded away from zero on \mathbb{X} , $0 < \underline{f} \leq f(x) \leq \bar{f} < \infty$ for all $x \in \mathbb{X}$. Suppose also that f is differentiable in the interior of \mathbb{X} with bounded derivatives \mathbb{X} , $\sup_{x \in \text{int } \mathbb{X}} \|\partial f(x)/\partial X\| < \infty$. Then $N^{2/k} \|E[U_m]\|$ is bounded by a constant uniformly over $z \in \mathbb{Z}$ and $N > m$.

PROOF OF LEMMA A.2: Fix $z \in \mathbb{Z}$. From the proof of Lemma 1, we know that:

$$\begin{aligned} E[U_m] &= N \binom{N-1}{m-1} \int u f(z+u) \left(1 - \Pr(\|X-z\| \leq \|u\|)\right)^{N-m} \left(\Pr(\|X-z\| \leq \|u\|)\right)^{m-1} du \\ &= N \binom{N-1}{m-1} \int_0^\infty r^{k-1} \left(\int_{S_k} r \omega f(z+r\omega) \lambda_{S_k}(d\omega) \right) \left(1 - \Pr(\|X-z\| \leq r)\right)^{N-m} \left(\Pr(\|X-z\| \leq r)\right)^{m-1} dr. \end{aligned}$$

Let $\bar{r}(z) = \sup\{r > 0 \mid z+r\omega \in \mathbb{X}, \text{ for all } \omega \in S_k\}$. Given the conditions of the lemma, there exists \underline{r} such that $\bar{r}(z) \geq \underline{r} > 0$ for all $z \in \mathbb{Z}$. Let

$$A_m^1 = \int_0^{\bar{r}(z)} \varphi(r) dr, \quad \text{and} \quad A_m^2 = \int_{\bar{r}(z)}^\infty \varphi(r) dr,$$

where

$$\varphi(r) = N \binom{N-1}{m-1} r^{k-1} \left(\int_{S_k} r \omega f(z+r\omega) \lambda_{S_k}(d\omega) \right) \left(1 - \Pr(\|X-z\| \leq r)\right)^{N-m} \left(\Pr(\|X-z\| \leq r)\right)^{m-1}.$$

Then,

$$E[U_m] = A_m^1 + A_m^2.$$

Consider the change of variable $t = rN^{1/k}$, then:

$$\begin{aligned} A_m^1 &= \frac{1}{N^{1/k}} \int_0^{\bar{r}(z)N^{1/k}} \binom{N-1}{m-1} t^k \left(\int_{S_k} \omega f\left(z + \frac{t}{N^{1/k}} \omega\right) \lambda_{S_k}(d\omega) \right) \\ &\quad \times \left(1 - \Pr\left(\|X-z\| \leq \frac{t}{N^{1/k}}\right)\right)^{N-m} \left(\Pr\left(\|X-z\| \leq \frac{t}{N^{1/k}}\right)\right)^{m-1} dt. \end{aligned}$$

Let $\overline{\|\partial f/\partial X\|} = \sup_{x \in \text{int } \mathbb{X}} \|(\partial f/\partial X)(x)\|$. By the Mean Value Theorem, for $\tilde{t} \leq t \leq \bar{r}(z)N^{1/k}$

$$\left| f\left(z + \frac{t}{N^{1/k}} \omega\right) - f(z) \right| = \frac{t}{N^{1/k}} \left\| \omega' \frac{\partial f}{\partial X}(z + \tilde{t}N^{-1/k}\omega) \right\| \leq \frac{t}{N^{1/k}} \overline{\|\partial f/\partial X\|}.$$

As a result, for $t \leq \bar{r}(z)N^{1/k}$, we obtain:

$$\begin{aligned} \left\| \int_{S_k} \omega f\left(z + \frac{t}{N^{1/k}} \omega\right) \lambda_{S_k}(d\omega) \right\| &= \left\| \int_{S_k} \omega \left(f\left(z + \frac{t}{N^{1/k}} \omega\right) - f(z) \right) \lambda_{S_k}(d\omega) \right\| \\ &\leq \frac{t}{N^{1/k}} \overline{\|\partial f/\partial X\|} \int_{S_k} \lambda_{S_k}(d\omega). \end{aligned}$$

Note also that for $t \leq \bar{r}(z)N^{1/k}$ we have:

$$\begin{aligned} \Pr\left(\|X-z\| \leq \frac{t}{N^{1/k}}\right) &= \int_0^{tN^{-1/k}} s^{k-1} \left(\int_{S_k} f(z+s\omega) \lambda_{S_k}(d\omega) \right) ds \\ &\geq \underline{f} \int_0^{tN^{-1/k}} s^{k-1} ds \int_{S_k} \lambda_{S_k}(d\omega) = \frac{1}{N} \frac{t^k}{k} \underline{f} \int_{S_k} \lambda_{S_k}(d\omega). \end{aligned}$$

Similarly:

$$\Pr\left(\|X - z\| \leq \frac{t}{N^{1/k}}\right) \leq \frac{1}{N} \frac{t^k}{k} \bar{f} \int_{S_k} \lambda_{S_k}(d\omega).$$

Therefore

$$\begin{aligned} \|A_m^1\| &\leq \frac{1}{N^{2/k}} \int_0^{\bar{r}(z)N^{1/k}} N^{1-m} \binom{N-1}{m-1} t^{k+1} \left\| \frac{\partial f}{\partial X} \right\| \int_{S_k} \lambda_{S_k}(d\omega) \\ &\quad \times \left(1 - \frac{1}{N} \frac{t^k}{k} \underline{f} \int_{S_k} \lambda_{S_k}(d\omega)\right)^{N-m} \left(\frac{t^k}{k} \bar{f} \int_{S_k} \lambda_{S_k}(d\omega)\right)^{m-1} dt. \end{aligned}$$

It is easy to see that:

$$N^{1-m} \binom{N-1}{m-1} \leq \frac{1}{(m-1)!}.$$

It is also easily seen that for $0 < b < N$ and $N > m$, $(1 - b/N)^{N-m} \leq \exp(-b/(m+1))$. Therefore, for $z \in \mathbb{Z}$ and $N > m$, we obtain:

$$\begin{aligned} \|A_m^1\| &< \frac{1}{N^{2/k}} \int_0^\infty \frac{1}{(m-1)!} t^{k+1} \left\| \frac{\partial f}{\partial X} \right\| \int_{S_k} \lambda_{S_k}(d\omega) \\ &\quad \times \exp\left(-\frac{t^k}{(m+1)k} \underline{f} \int_{S_k} \lambda_{S_k}(d\omega)\right) \left(\frac{t^k}{k} \bar{f} \int_{S_k} \lambda_{S_k}(d\omega)\right)^{m-1} dt \leq C_1 \frac{1}{N^{2/k}}, \end{aligned}$$

for some positive constant, C_1 .

For A_m^2 , notice that, for $N \geq m$:

$$\begin{aligned} \|A_m^2\| &\leq \left(1 - \Pr(\|X - z\| \leq \bar{r}(z))\right)^{N-m} N \binom{N-1}{m-1} \int_{\bar{r}(z)}^\infty r^{k-1} \left\| \int_{S_k} r\omega f(z + r\omega) \lambda_{S_k}(d\omega) \right\| \\ &\quad \times \Pr(\|X - z\| \leq r)^{m-1} dr \\ &< \left(1 - \Pr(\|X - z\| \leq \bar{r}(z))\right)^{N-m} \frac{N^m}{m!} \\ &\quad \times \left(m \int_0^\infty r^{k-1} \left\| \int_{S_k} r\omega f(z + r\omega) \lambda_{S_k}(d\omega) \right\| \Pr(\|X - z\| \leq r)^{m-1} dr\right) \\ &\leq N^m \left(1 - \Pr(\|X - z\| \leq \bar{r}(z))\right)^{N-m} (\bar{u}/m!), \end{aligned}$$

where \bar{u} is the diameter of \mathbb{X} , that is, $\bar{u} = \sup_{x, y \in \mathbb{X}} \|x - y\| < \infty$. The last inequality holds because the last term on the left hand side is equal to the expectation of $\|U_m\|$ when $N = m$ which is bounded by \bar{u} . Consequently, we obtain:

$$N^{2/k} \|A_m^2\| < (\bar{u}/m!) N^{2/k+m} \left(1 - \underline{f} \frac{\bar{r}(z)^k}{k} \int_{S_k} \lambda_{S_k}(d\omega)\right)^{N-m} \quad (\rightarrow 0, \forall z \in \mathbb{Z}).$$

To simplify notation, let $b = \underline{f} \int_{S_k} \lambda_{S_k}(d\omega)/k$. Then:

$$N^{2/k} \|A_m^2\| < (u/m!) N^{2/k+m} (1 - b \bar{r}(z)^k)^{N-m} \leq (u/m!) N^{2/k+m} (1 - b \underline{r}^k)^{N-m},$$

where $0 < b \underline{r}^k < 1$. The right hand side of last equation is maximal at:

$$N = -\frac{(2/k + m)}{\ln(1 - b \underline{r}^k)}.$$

Therefore,

$$\begin{aligned} N^{2/k} \|A_m^2\| &< (u/m!) \frac{(2/k+m)^{2/k+m}}{(-\ln(1-br^k))^{2/k+m}} (1-br^k)^{\left(\frac{-(2/k+m)}{\ln(1-br^k)}-m\right)} \\ &= (2/k+m)^{(2/k+m)} (u/m!) e^{-(2/k+m)} \frac{(1-br^k)^{-m}}{(-\ln(1-br^k))^{2/k+m}} = C_2 < \infty. \end{aligned}$$

Note that this bound does not depend on N or z . As a result, we obtain

$$N^{2/k} \|E[U_m]\| < C_1 + C_2 < \infty,$$

for all $N > m$ and $z \in \mathbb{Z}$. □

The proof of the second part of Theorem 2 is as follows.

$$\begin{aligned} \text{Bias}_M^{sm,t} &= \mathbb{E}[B_M^{sm,t}] = \mathbb{E} \left[\frac{1}{N_1 M} \sum_{i=1}^N \sum_{m=1}^M W_i (\mu_0(X_i) - \mu_0(X_{j_m(i)})) \right] \\ &= \frac{1}{M} \sum_{m=1}^M \mathbb{E} [\mu_0(X_i) - \mu_0(X_{j_m(i)}) | W_i = 1]. \end{aligned}$$

Applying a second order Taylor expansion, we obtain:

$$\mu_0(X_{j_m(i)}) - \mu_0(X_i) = \frac{\partial \mu_0}{\partial x'}(X_i) U_{m,i} + \frac{1}{2} \text{tr} \left(\frac{\partial^2 \mu_0}{\partial x \partial x'}(X_i) U_{m,i} U_{m,i}' \right) + O(\|U_{m,i}\|^3).$$

Therefore, because the trace is a linear operator:

$$\begin{aligned} \mathbb{E} [\mu_0(X_{j_m(i)}) - \mu_0(X_i) | X_i = z, W_i = 1] &= \frac{\partial \mu_0}{\partial x'}(z) \mathbb{E}[U_{m,i} | X_i = z, W_i = 1] \\ &+ \frac{1}{2} \text{tr} \left(\frac{\partial^2 \mu_0}{\partial x \partial x'}(z) \mathbb{E}[U_{m,i} U_{m,i}' | X_i = z, W_i = 1] \right) + O(\mathbb{E}[\|U_{m,i}\|^3 | X_i = z, W_i = 1]). \end{aligned}$$

Lemma 2 implies that the norm of $N_0^{2/k} \mathbb{E}[U_{m,i} U_{m,i}' | X_i = z, W_i = 1]$ and $N_0^{2/k} \mathbb{E}[\|U_{m,i}\|^3 | X_i = z, W_i = 1]$ are uniformly bounded over $z \in \mathbb{X}_1$ and N_0 . Lemma A.2 implies the same result for $N_0^{2/k} \mathbb{E}[U_{m,i} | X_i = z, W_i = 1]$. As a result, $\|N_0^{2/k} \mathbb{E}[\mu_0(X_{j_m(i)}) - \mu_0(X_i) | X_i = z, W_i = 1]\|$ is uniformly bounded over $z \in \mathbb{X}_1$ and N_0 . Applying Lebesgue's Dominated Convergence Theorem along with Lemma 1, we obtain:

$$\begin{aligned} N_0^{2/k} \mathbb{E} [\mu_0(X_{j_m(i)}) - \mu_0(X_i) | W_i = 1] &= \Gamma \left(\frac{mk+2}{k} \right) \frac{1}{(m-1)!k} \times \\ &\int \left(f_0(x) \frac{\pi^{k/2}}{\Gamma(1+k/2)} \right)^{-2/k} \left\{ \frac{1}{f_0(x)} \frac{\partial f_0}{\partial x'}(x) \frac{\partial \mu_0}{\partial x}(x) + \frac{1}{2} \text{tr} \left(\frac{\partial^2 \mu_0}{\partial x' \partial x}(x) \right) \right\} f_1(x) dx + o(1). \end{aligned}$$

Now, the result follows easily from the conditions of the theorem. □

PROOF OF THEOREM 1(ii):

Consider the special case where $\mu_1(x)$ is flat over \mathbb{X} and $\mu_0(x)$ is flat in a neighborhood of the boundary, \mathbb{B} . Then, matching the control units does not create bias. Matching the treated units creates a bias that is similar to the formula in Theorem 2(ii), but with $r = 1$, $\theta = p/(1-p)$, and the integral taken over $\mathbb{X} \cap \mathbb{B}^c$. □

PROOF OF LEMMA 3: Define $\underline{f} = \inf_{x,w} f_w(x)$ and $\bar{f} = \sup_{x,w} f_w(x)$, with $\underline{f} > 0$ and \bar{f} finite. Let $\bar{u} = \sup_{x,y \in \mathbb{X}} \|x-y\|$. Consider all the balls $B(x, u)$ with centers $x \in \mathbb{X}$ and radius u . Let $c(u)$ ($0 < c(u) < 1$)

be the infimum over $x \in \mathbb{X}$ of the proportion that the intersection with \mathbb{X} represents in volume of the balls. Note that, because \mathbb{X} is convex, this proportion nonincreasing in u , so let $c = c(\bar{u})$, and $c(u) \geq c$ for $u \leq \bar{u}$. The proof consists of three parts. First we derive an exponential bound for the probability that the distance to a match, $\|X_{j_m(i)} - X_i\|$ exceeds some value. Second, we use this to obtain an exponential bound on the volume of the catchment area, $A_M(i)$, defined as the subset of \mathbb{X} such that i is matched to each observation, j , with $W_j = 1 - W_i$ and $X_j \in A_M(i)$. That is, if $W_j = 1 - W_i$ and $X_j \in A_M(i)$, then $i \in J_M(j)$.

$$A_M(i) = \left\{ x \mid \sum_{l \mid W_l = W_i} 1 \{ \|X_l - x\| \leq \|X_i - x\| \} \leq M \right\}.$$

Third, we use the exponential bound on the volume of the catchment area to derive an exponential bound on the probability of a large $K_M(i)$, which will be used to bound the moments of $K_M(i)$.

For the first part we bound the probability of the distance to a match. Let $x \in \mathbb{X}$ and $u < N_{1-W_i}^{1/k} \bar{u}$. Then,

$$\begin{aligned} & \Pr \left(\|X_j - X_i\| > u \cdot N_{1-W_i}^{-1/k} \mid W_1, \dots, W_N, W_j = 1 - W_i, X_i = x \right) \\ &= 1 - \int_0^{u N_{1-W_i}^{-1/k}} r^{k-1} \int_{S_k} f_{1-W_i}(x + r\omega) \lambda_{S_k}(d\omega) dr \leq 1 - c \underline{f} \int_0^{u N_{1-W_i}^{-1/k}} r^{k-1} \int_{S_k} \lambda_{S_k}(d\omega) dr \\ &= 1 - c \underline{f} u^k N_{1-W_i}^{-1} \pi^{k/2} / \Gamma(1 + k/2). \end{aligned}$$

Similarly

$$\Pr \left(\|X_j - X_i\| \leq u \cdot N_{1-W_i}^{-1/k} \mid W_1, \dots, W_N, W_j = 1 - W_i, X_i = x \right) \leq \bar{f} u^k N_{1-W_i}^{-1} \pi^{k/2} / \Gamma(1 + k/2).$$

Notice also that

$$\begin{aligned} & \Pr \left(\|X_j - X_i\| > u \cdot N_{1-W_i}^{-1/k} \mid W_1, \dots, W_N, X_i = x, j \in \mathcal{J}_M(i) \right) \\ & \leq \Pr \left(\|X_j - X_i\| > u \cdot N_{1-W_i}^{-1/k} \mid W_1, \dots, W_N, X_i = x, j = j_M(i) \right) \\ & = \sum_{m=0}^{M-1} \binom{N_{1-W_i}}{m} \Pr \left(\|X_j - X_i\| > u \cdot N_{1-W_i}^{-1/k} \mid W_1, \dots, W_N, W_j = 1 - W_i, X_i = x \right)^{N_{1-W_i} - m} \\ & \quad \times \Pr \left(\|X_j - X_i\| \leq u \cdot N_{1-W_i}^{-1/k} \mid W_1, \dots, W_N, W_j = 1 - W_i, X_i = x \right)^m. \end{aligned}$$

In addition,

$$\begin{aligned} & \binom{N_{1-W_i}}{m} \Pr \left(\|X_j - X_i\| \leq u \cdot N_{1-W_i}^{-1/k} \mid W_1, \dots, W_N, W_j = 1 - W_i, X_i = x \right)^m \\ & \leq \frac{1}{m!} \left(u^k \bar{f} \frac{\pi^{k/2}}{\Gamma(1 + k/2)} \right)^m. \end{aligned}$$

Therefore,

$$\begin{aligned} & \Pr \left(\|X_j - X_i\| > u \cdot N_{1-W_i}^{-1/k} \mid W_1, \dots, W_N, X_i = x, j \in \mathcal{J}_M(i) \right) \\ & \leq \sum_{m=0}^{M-1} \frac{1}{m!} \left(u^k \bar{f} \frac{\pi^{k/2}}{\Gamma(1 + k/2)} \right)^m \left(1 - u^k c \underline{f} \frac{\pi^{k/2}}{\Gamma(1 + k/2)} \cdot \frac{1}{N_{1-W_i}} \right)^{N_{1-W_i} - m}. \end{aligned}$$

Then, for some constant $C_1 > 0$,

$$\begin{aligned} & \Pr \left(\|X_j - X_i\| > u \cdot N_{1-W_i}^{-1/k} \mid W_1, \dots, W_N, X_i = x, j \in \mathcal{J}_M(i) \right) \\ & \leq C_1 \max\{1, u^{k(M-1)}\} \sum_{m=0}^{M-1} \left(1 - u^k c_f \frac{\pi^{k/2}}{\Gamma(1+k/2)} \cdot \frac{1}{N_{1-W_i}} \right)^{N_{1-W_i} - m} \\ & \leq C_1 M \max\{1, u^{k(M-1)}\} \exp \left(-\frac{u^k}{(M+1)} c_f \frac{\pi^{k/2}}{\Gamma(1+k/2)} \right). \end{aligned}$$

Notice that this bound also holds for $u \geq N_{1-W_i}^{1/k} \bar{u}$, because in that case the probability that $\|X_{j_m(i)} - X_i\| > u \cdot N_{1-W_i}^{-1/k}$ is zero.

Next, we consider for unit i , the volume $B_M(i)$ of the catchment area $A_M(i)$, defined as:

$$B_M(i) = \int_{A_M(i)} dx.$$

Conditional on W_1, \dots, W_N , $i \in \mathcal{J}_M(j)$, $X_i = x$, and $A_M(i)$, the distribution of X_j is proportional to $f_{1-W_i}(x) \cdot \mathbf{1}\{x \in A_M(i)\}$. Notice that a ball with radius $(b/2)^{1/k} / (\pi^{k/2} / \Gamma(1+k/2))^{1/k}$ has volume $b/2$. Therefore for X_i in $A_M(i)$ and $B_M(i) \geq b$, we obtain

$$\Pr \left(\|X_j - X_i\| > \frac{(b/2)^{1/k}}{(\pi^{k/2} / \Gamma(1+k/2))^{1/k}} \mid W_1, \dots, W_N, X_i = x, A_M(i), i \in \mathcal{J}_M(j) \right) \geq \frac{f}{2\bar{f}}.$$

The last inequality does not depend on $A_m(i)$ (given $B_M(i) \geq b$). Therefore,

$$\Pr \left(\|X_j - X_i\| > \frac{(b/2)^{1/k}}{(\pi^{k/2} / \Gamma(1+k/2))^{1/k}} \mid W_1, \dots, W_N, X_i = x, i \in \mathcal{J}_M(j), B_M(i) \geq b \right) \geq \frac{f}{2\bar{f}}.$$

As a result, if

$$\Pr \left(\|X_j - X_i\| > \frac{(b/2)^{1/k}}{(\pi^{k/2} / \Gamma(1+k/2))^{1/k}} \mid W_1, \dots, W_N, X_i = x, i \in \mathcal{J}_M(j) \right) \leq \delta \frac{f}{2\bar{f}}, \quad (\text{A.2})$$

then it must be the case that $\Pr(B_M(i) \geq b \mid W_1, \dots, W_N, X_i = x, i \in \mathcal{J}_M(j)) \leq \delta$. In fact, the inequality in equation (A.2) has been established above for

$$b = \frac{2u^k}{N_{W_i}} \left(\frac{\pi^{k/2}}{\Gamma(1+k/2)} \right), \text{ and } \delta = \frac{2\bar{f}}{f} C_1 M \max\{1, u^{k(M-1)}\} \exp \left(-\frac{u^k}{(M+1)} c_f \frac{\pi^{k/2}}{\Gamma(1+k/2)} \right).$$

Let $t = 2u^k \pi^{k/2} / \Gamma(1+k/2)$, then

$$\Pr(N_{W_i} B_M(i) \geq t \mid W_1, \dots, W_N, X_i = x, i \in \mathcal{J}_M(j)) \leq C_2 \max\{1, C_3 t^{M-1}\} \exp(-C_4 t),$$

for some positive constants, C_2 , C_3 , and C_4 . This establishes an uniform exponential bound, so all the moments of $N_{W_i} B_M(i)$ exist conditional on $W_1, \dots, W_N, X_i = x, i \in \mathcal{J}_M(j)$ (uniformly in N).

For the third part of the proof, consider the distribution of $K_M(i)$, the number of times unit i is used as a match. Let $P_M(i)$ be the probability that an observation with the opposite treatment is matched to observation i :

$$P_M(i) = \int_{A_M(i)} f_{1-W_i}(x) dx \leq \bar{f} B_M(i).$$

Note that for $n \geq 0$,

$$\begin{aligned} \mathbb{E}[(N_{W_i} P_M(i))^n \mid X_i = x, W_1, \dots, W_N] & \leq \mathbb{E}[(N_{W_i} P_M(i))^n \mid X_i = x, W_1, \dots, W_N, i \in \mathcal{J}_M(j)] \\ & \leq \bar{f}^n \mathbb{E}[(N_{W_i} B_M(i))^n \mid X_i = x, W_1, \dots, W_N, i \in \mathcal{J}_M(j)]. \end{aligned}$$

As a result, $\mathbb{E}[(N_{W_i} P_M(i))^n | X_i = x, W_1, \dots, W_N]$ is uniformly bounded. Conditional on $P_M(i)$, and on $X_i = x, W_1, \dots, W_N$, the distribution of $K_M(i)$ is binomial with parameters N_{1-W_i} and $P_M(i)$. Therefore, conditional on $P_M(i)$, and $X_i = x, W_1, \dots, W_N$, the q -th moment of $K_M(i)$ is

$$\mathbb{E}[K_M^q(i) | P_M(i), X_i = x, W_1, \dots, W_N] = \sum_{n=0}^q \frac{S(q, n) N_{1-W_i}! P_M(i)^n}{(N_{1-W_i} - n)!} \leq \sum_{n=0}^q S(q, n) (N_{1-W_i} P_M(i))^n,$$

where $S(q, n)$ are Stirling numbers of the second kind and $q \geq 1$ (see, e.g., Johnson, Kotz and Kemp, 1992). Then, because $S(q, 0) = 0$ for $q \geq 1$,

$$\mathbb{E}[K_M^q(i) | X_i = x, W_1, \dots, W_N] \leq C \sum_{n=1}^q S(q, n) \cdot \left(\frac{N_{1-W_i}}{N_{W_i}} \right)^n,$$

for some positive constant, C . Using Chernoff's bound for binomial tails, it can be easily seen that $E[(N_{1-W_i}/N_{W_i})^n | X_i = x, W_i] = E[(N_{1-W_i}/N_{W_i})^n | W_i]$ is uniformly bounded in N , for all $n \geq 1$, so the result of the first part of the lemma follows.

Next, consider part (ii) of Lemma 3. Because the variance $\sigma^2(x, w)$ is Lipschitz on a bounded set, it is therefore bounded by some constant, $\bar{\sigma}^2 = \sup_{w,x} \sigma^2(x, w)$. As a result, $\mathbb{E}[(1 + K_M/M)^2 \sigma^2(x, w)]$ is bounded by $\bar{\sigma}^2 \mathbb{E}[(1 + K_M/M)^2]$, which is uniformly bounded in N by the result in the first part of the lemma. Hence $\mathbb{E}[V^E] = O(1)$.

Next, consider part (iii) of Lemma 3. Using the same argument as for $\mathbb{E}[K_M^q(i)]$, we obtain

$$\mathbb{E}[K_M^q(i) | W_i = 0] \leq \sum_{n=1}^q S(q, n) \left(\frac{N_1}{N_0} \right)^n \mathbb{E}[(N_0 P_M(i))^n | W_i = 0].$$

Therefore,

$$\left(\frac{N_0}{N_1} \right) \mathbb{E}[K_M^q(i) | W_i = 0] \leq \sum_{n=1}^q S(q, n) \left(\frac{N_1}{N_0} \right)^{n-1} \mathbb{E}[(N_0 P_M(i))^n | W_i = 0],$$

which is uniformly bounded because $r \geq 1$.

For part (iv) notice that

$$\begin{aligned} \mathbb{E}[V^{E,t}] &= \mathbb{E} \left[\frac{1}{N_1} \sum_{i=1}^N W_i \sigma^2(X_i, W_i) \right] + E \left[\frac{1}{N_1} \sum_{i=1}^N (1 - W_i) \left(\frac{K_M(i)}{M} \right)^2 \sigma^2(X_i, W_i) \right] \\ &= E[\sigma^2(X_i, W_i) | W_i = 1] + \left(\frac{N_0}{N_1} \right) E \left[\left(\frac{K_M(i)}{M} \right)^2 \sigma^2(X_i, W_i) \middle| W_i = 0 \right]. \end{aligned}$$

Therefore, $\mathbb{E}[V^{E,t}]$ is uniformly bounded. \square

PROOF OF THEOREM 3:

We only prove the first part of the theorem. The second part follows the same argument. We can write $\hat{\tau}_M^{sm} - \tau = (\tau(\bar{X}) - \tau) + E_M^{sm} + B_M^{sm}$. We consider each of the three terms separately. First, by assumptions 1 and 4(i), $\mu_w(x)$ is bounded over $x \in \mathbb{X}$ and $w = 0, 1$. Hence $\mu_1(X) - \mu_0(X) - \tau$ has mean zero and finite variance. Therefore, by a standard law of large numbers $\tau(\bar{X}) - \tau \xrightarrow{p} 0$. Second, by Theorem 1, $B_M^{sm} = O_p(N^{-1/k}) = o_p(1)$. Finally, because $\mathbb{E}[\varepsilon_i | \mathbf{X}, \mathbf{W}] = 0$, $\mathbb{E}[\varepsilon_i^2 | \mathbf{X}, \mathbf{W}] \leq \bar{\sigma}^2$ and $\mathbb{E}[\varepsilon_i \varepsilon_j | \mathbf{X}, \mathbf{W}] = 0$ ($i \neq j$), we obtain

$$\mathbb{E} \left[\left(\sqrt{N} E_M^{sm} \right)^2 \right] = \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[\left(1 + \frac{K_M(i)}{M} \right)^2 \varepsilon_i^2 \right] = \mathbb{E} \left[\left(1 + \frac{K_M(i)}{M} \right)^2 \sigma^2(X_i, W_i) \right] = O(1),$$

where the last equality comes from Lemma 3. By Markov's inequality $E_M^{sm} = O_p(N^{-1/2}) = o_p(1)$.

PROOF OF THEOREM 4:

We only prove the first assertion in the theorem as the second follows the same argument. We can write $\sqrt{N}(\hat{\tau}_M^{sm} - B_M^{sm} - \tau) = \sqrt{N}(\tau(\overline{X}) - \tau) + \sqrt{N}E_M^{sm}$. First, consider the contribution of $\sqrt{N}(\tau(\overline{X}) - \tau)$. By a standard central limit theorem

$$\sqrt{N}(\tau(\overline{X}) - \tau) \xrightarrow{d} \mathcal{N}(0, V^{\tau(X)}). \quad (\text{A.3})$$

Second, consider the contribution of $\sqrt{N}E_M^{sm}/\sqrt{V^E} = \sum_{i=1}^N E_{M,i}^{sm}/\sqrt{NV^E}$. Conditional on \mathbf{W} and \mathbf{X} the unit-level terms $E_{M,i}^{sm}$ are independent with zero means and non-identical distributions. The conditional variance of $E_{M,i}^{sm}$ is $(1 + K_M(i)/M)^2 \cdot \sigma^2(X_i, W_i)$. We will use a Lindeberg-Feller central limit theorem for $\sqrt{N}E_M^{sm}/\sqrt{V^E}$. For a given \mathbf{X}, \mathbf{W} , the Lindeberg-Feller condition requires that

$$\frac{1}{N \cdot V^E} \sum_{i=1}^N \mathbb{E} \left[(E_{M,i}^{sm})^2 1\{|E_{M,i}^{sm}| \geq \eta_{LF} \sqrt{N \cdot V^E}\} | \mathbf{X}, \mathbf{W} \right] \rightarrow 0, \quad (\text{A.4})$$

for all $\eta_{LF} > 0$. To prove that (A.4) condition holds, notice that by Hölder's and Markov's inequalities we have

$$\begin{aligned} & \mathbb{E} \left[(E_{M,i}^{sm})^2 1\{|E_{M,i}^{sm}| \geq \eta_{LF} \sqrt{N \cdot V^E}\} | \mathbf{X}, \mathbf{W} \right] \\ & \leq \left(\mathbb{E} [(E_{M,i}^{sm})^4 | \mathbf{X}, \mathbf{W}] \right)^{1/2} \left(\mathbb{E} [1\{|E_{M,i}^{sm}| \geq \eta_{LF} \sqrt{N \cdot V^E}\} | \mathbf{X}, \mathbf{W}] \right)^{1/2} \\ & \leq \left(\mathbb{E} [(E_{M,i}^{sm})^4 | \mathbf{X}, \mathbf{W}] \right)^{1/2} \left(\Pr \left(|E_{M,i}^{sm}| \geq \eta_{LF} \sqrt{NV^E} | \mathbf{X}, \mathbf{W} \right) \right) \\ & \leq \left(\mathbb{E} [(E_{M,i}^{sm})^4 | \mathbf{X}, \mathbf{W}] \right)^{1/2} \frac{\mathbb{E} [(E_{M,i}^{sm})^2 | \mathbf{X}, \mathbf{W}]}{\eta_{LF}^2 \cdot N \cdot V^E}. \end{aligned}$$

Let $\bar{s} = \sup_{w,x} \sigma_w^2(x) < \infty$, $\underline{s} = \inf_{w,x} \sigma^2(x, w) > 0$, and $\bar{K} = \sup_{w,x} \mathbb{E} [\varepsilon_i^4 | X_i = x, W_i = w] < \infty$. Notice that $V^E \geq \underline{s}$. Therefore,

$$\begin{aligned} & \frac{1}{N \cdot V^E} \sum_{i=1}^N \mathbb{E} \left[(E_{M,i}^{sm})^2 1\{|E_{M,i}^{sm}| \geq \eta_{LF} \sqrt{N \cdot V^E}\} | \mathbf{X}, \mathbf{W} \right] \\ & \leq \frac{1}{N \cdot V^E} \sum_{i=1}^N \left((1 + K_M(i)/M)^4 \mathbb{E} [\varepsilon_i^4 | \mathbf{X}, \mathbf{W}] \right)^{1/2} \cdot \frac{(1 + K_M(i)/M)^2 \cdot \sigma^2(X_i, W_i)}{\eta_{LF}^2 \cdot N \cdot V^E} \\ & \leq \frac{\bar{s}\bar{K}^{1/2}}{\eta_{LF}^2 \underline{s}^2} \cdot \frac{1}{N} \left(\frac{1}{N} \sum_{i=1}^N (1 + K_M(i)/M)^4 \right). \end{aligned}$$

Because $\mathbb{E}[(1 + K_M(i)/M)^4]$ is uniformly bounded, by Markov's Inequality, the last term in parentheses is bounded in probability. Hence, the Lindeberg-Feller condition is satisfied for almost all \mathbf{X} and \mathbf{W} . As a result,

$$\frac{N^{1/2} \sum_{i=1}^N E_{M,i}^{sm}}{\left(\sum_{i=1}^N (1 + K_M(i)/M)^2 \sigma^2(X_i, W_i) \right)^{1/2}} = \frac{N^{1/2} \cdot E_M^{sm}}{\sqrt{V^E}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Finally, $\sqrt{N}E_M^{sm}/\sqrt{V^E}$ and $\sqrt{N}(\tau(\overline{X}) - \tau)$ are asymptotically independent (the central limit theorem for $\sqrt{N}E_M^{sm}/\sqrt{V^E}$ holds conditional on \mathbf{X} and \mathbf{W}). Thus, the fact that both converge to standard normal distributions, boundedness of V^E and $V^{\tau(X)}$, and boundedness away from zero of V^E imply that $(V^E + V^{\tau(X)})^{-1/2} N^{1/2} (\hat{\tau}_M^{sm} - B_M^{sm} - \tau)$ converges to a standard normal distribution. \square

PROOF OF THEOREM 5:

The proof of Theorem 5 is long but mechanical, so we omit it here. In this proof, we use Lemma A.1 to characterize the asymptotic behavior of the probability of the events $\{j \in \mathcal{J}_M(1)\}$ and $\{j \in \mathcal{J}_M(1) \cup j \in \mathcal{J}_M(2)\}$ conditional on $X_j = x_j$ and $W_1 = W_2 = 1 - W_j$. These results allow us to establish the asymptotic behavior of $E[K_M(i)|W_i = w, X_i = x]$ and $E[K_M(i)^2|W_i = w, X_i = x]$, which in turn, allows us to establish the limit of $N\mathbb{V}(\widehat{\tau}_M^{sm})$. The derivation of the asymptotic expansion of the conditional probability of $\{j \in \mathcal{J}_M(1) \cup j \in \mathcal{J}_M(2)\}$ depends crucially on $k = 1$. To see why, assume, without loss of generality, that $X_2 \leq X_1$. Then, the restriction that $k = 1$ allows us to study the asymptotic behavior of the conditional probability of $\{j \in \mathcal{J}_M(1) \cup j \in \mathcal{J}_M(2)\}$ under three exhaustive cases: $X_j \leq X_2 \leq X_1$, $X_2 \leq X_j \leq X_1$, and $X_2 \leq X_1 \leq X_j$. The entire proof is available on the web pages of the authors.

Before proving Theorems 6 and 7 we give one preliminary result.

LEMMA A.3: : *Let $q \geq 0$. Under assumptions 1-4:*

$$\frac{1}{N} \sum_{i=1}^N K_M(i)^q \left(\widehat{\sigma}_{W_i}^2(X_i) - \sigma_{W_i}^2(X_i) \right) = o_p(1).$$

Let $q \geq 1$. Under assumptions 1, 2', 3', and 4:

$$\frac{1}{N_1} \sum_{i=1}^N (1 - W_i) K_M(i)^q \left(\widehat{\sigma}_{W_i}^2(X_i) - \sigma_{W_i}^2(X_i) \right) = o_p(1).$$

PROOF: Take N_1 as given. Notice that $N_1^{1/k} \|X_i - X_{l_j(i)}\|$ is identically distributed for all i with $W_i = 1$. By Lemma 2, all the moments of $N_1^{1/k} \|X_i - X_{l_j(i)}\|$ are uniformly bounded in N_1 for all i with $W_i = 1$. Applying Bonferroni's and Markov's inequalities, we obtain

$$N_1^{-\xi} \max_{W_i=1} N_1^{1/k} \|X_i - X_{l_j(i)}\| = o_p(1),$$

as $N_1 \rightarrow \infty$, for all $\xi > 0$. In particular, making $\xi = 1/k$, we obtain

$$\max_{W_i=1} \|X_i - X_{l_j(i)}\| = o_p(1), \tag{A.5}$$

as $N_1 \rightarrow \infty$. Because $N_1 \rightarrow \infty$ almost surely, and because conditional probabilities are bounded, Lebesgue's Dominated Convergence Theorem implies that the result in equation (A.5) holds also as $N \rightarrow \infty$, without conditioning on N_1 . The analogous result holds for $W_i = 0$. Therefore,

$$\max_{i=1 \dots N} \|X_i - X_{l_j(i)}\| = \max_{w=0,1} \left\{ \max_{W_i=w} \|X_i - X_{l_j(i)}\| \right\} = o_p(1). \tag{A.6}$$

In addition, it can be seen that the maximum of number of times that an observation is used as a match within its own treatment group is bounded by $J \cdot \bar{L}(k)$, where $\bar{L}(k) < \infty$ is the "kissing number" in k dimensions. (See Lemma 3.2.1 in Miller et al. 1997. $\bar{L}(k)$ is defined as the maximum number of non-overlapping unit balls in \mathbb{R}^k that can be arranged to overlap with a unit ball.)

Notice that

$$\begin{aligned} \widehat{\sigma}_{W_i}^2(X_i) &= \frac{J}{J+1} \left(Y_i - \frac{1}{J} \sum_{j=1}^J Y_{l_j(i)} \right)^2 \\ &= \frac{J}{J+1} \left(\varepsilon_i - \frac{1}{J} \sum_{j=1}^J \varepsilon_{l_j(i)} + \frac{1}{J} \sum_{j=1}^J \left(\mu_{W_i}(X_i) - \mu_{W_i}(X_{l_j(i)}) \right) \right)^2. \end{aligned}$$

Therefore,

$$\begin{aligned} E[\hat{\sigma}_{W_i}^2(X_i)|\mathbf{X}, \mathbf{W}] &= \frac{J}{J+1} \left[\sigma_{W_i}^2(X_i) + \frac{1}{J^2} \sum_{j=1}^J \sigma_{W_i}^2(X_{l_j(i)}) + \left(\frac{1}{J} \sum_{j=1}^J (\mu_{W_i}(X_i) - \mu_{W_i}(X_{l_j(i)})) \right)^2 \right] \\ &= \frac{J}{J+1} \left[\frac{J+1}{J} \sigma_{W_i}^2(X_i) + \frac{1}{J^2} \sum_{j=1}^J (\sigma_{W_i}^2(X_{l_j(i)}) - \sigma_{W_i}^2(X_i)) + \left(\frac{1}{J} \sum_{j=1}^J (\mu_{W_i}(X_i) - \mu_{W_i}(X_{l_j(i)})) \right)^2 \right]. \end{aligned}$$

Using the Lipschitz conditions on $\mu_w(x)$ and $\sigma_w^2(x)$ (Assumption 4(i)), and the result in equation (A.6), we obtain:

$$\max_{i=1 \dots N} \left| E[\hat{\sigma}_{W_i}^2(X_i)|\mathbf{X}, \mathbf{W}] - \sigma_{W_i}^2(X_i) \right| = o_p(1). \quad (\text{A.7})$$

Therefore,

$$\frac{1}{N} \sum_{i=1}^N K_M(i)^q \left(E[\hat{\sigma}_{W_i}^2(X_i)|\mathbf{X}, \mathbf{W}] - \sigma_{W_i}^2(X_i) \right) = o_p(1). \quad (\text{A.8})$$

To obtain the first result of the lemma, it is left to be proven that

$$\frac{1}{N} \sum_{i=1}^N K_M(i)^q \left(E[\hat{\sigma}_{W_i}^2(X_i)|\mathbf{X}, \mathbf{W}] - \hat{\sigma}_{W_i}^2(X_i) \right) = o_p(1). \quad (\text{A.9})$$

Notice first that:

$$\begin{aligned} &\left(\frac{J+1}{J} \right) \frac{1}{N} \sum_{i=1}^N K_M(i)^q \left(\hat{\sigma}_{W_i}^2(X_i) - E[\hat{\sigma}_{W_i}^2(X_i)|\mathbf{X}, \mathbf{W}] \right) = \frac{1}{N} \sum_{i=1}^N K_M(i)^q \left(\varepsilon_i^2 - \sigma_{W_i}^2(X_i) \right) \\ &+ \frac{1}{N} \sum_{i=1}^N K_M(i)^q \frac{1}{J^2} \sum_{j=1}^J \left(\varepsilon_{l_j(i)}^2 - \sigma_{W_i}^2(X_{l_j(i)}) \right) + \frac{2}{J^2 N} \sum_{i=1}^N K_M(i)^q \sum_{j=1}^J \sum_{h>j} \varepsilon_{l_j(i)} \varepsilon_{l_h(i)} \\ &- \frac{2}{JN} \sum_{i=1}^N K_M(i)^q \varepsilon_i \sum_{j=1}^J \varepsilon_{l_j(i)} + \frac{2}{JN} \sum_{i=1}^N K_M(i)^q \varepsilon_i \sum_{j=1}^J \left(\mu_{W_i}(X_i) - \mu_{W_i}(X_{l_j(i)}) \right) \\ &\quad - \frac{2}{J^2 N} \sum_{i=1}^N K_M(i)^q \sum_{j=1}^J \varepsilon_{l_j(i)} \sum_{j=1}^J \left(\mu_{W_i}(X_i) - \mu_{W_i}(X_{l_j(i)}) \right). \quad (\text{A.10}) \end{aligned}$$

The expectations, conditional on \mathbf{X} and \mathbf{W} , of each term on the right hand side of last equation is equal to zero, so the unconditional expectations are also zero. Applying Bonferroni's Inequality, it is easy to show that for any $\xi > 0$, $E[N^{-\xi} \max_{i=1, \dots, N} K_M(i)^{2q}]$ is uniformly bounded. Using this result, along with finiteness of $\bar{L}(k)$, it can be shown that the variances of all the terms on the right hand side of equation (A.10) are $o(1)$. (This last part of the proof is largely technical and therefore omitted here. It is available on the websites of the authors.)

The proof of the second part of the Lemma follows the same pattern and is omitted here. \square

PROOF OF THEOREM 6: It follows directly from Lemma A.3. \square

PROOF OF THEOREM 7:

For part (i) notice that

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \left(\widehat{Y}_i(1) - \widehat{Y}_i(0) - \widehat{\tau}_M^{sm} \right)^2 &= \frac{1}{N} \sum_{i=1}^N \left(\widehat{Y}_i(1) - \widehat{Y}_i(0) - \tau \right)^2 - (\widehat{\tau}_M^{sm} - \tau)^2 \\ &= \frac{1}{N} \sum_{i=1}^N \left(\widehat{Y}_i(1) - \widehat{Y}_i(0) - \tau \right)^2 + o_p(1). \end{aligned} \quad (\text{A.11})$$

In addition,

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^N \left(\widehat{Y}_i(1) - \widehat{Y}_i(0) - \tau \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^N \left((2W_i - 1) \left(\frac{1}{M} \sum_{m=1}^M \mu_{W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)}) \right) - \tau \right)^2 + \frac{1}{N} \sum_{i=1}^N \left(\varepsilon_i - \frac{1}{M} \sum_{m=1}^M \varepsilon_{j_m(i)} \right)^2 \\ &+ \frac{2}{N} \sum_{i=1}^N \left((2W_i - 1) \left(\frac{1}{M} \sum_{m=1}^M \mu_{W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)}) \right) - \tau \right) (2W_i - 1) \left(\varepsilon_i - \frac{1}{M} \sum_{m=1}^M \varepsilon_{j_m(i)} \right). \end{aligned} \quad (\text{A.12})$$

Because the sample maximum of the norms of the matching discrepancies, $\|X_i - X_{j_m(i)}\|$, is $o_p(1)$, and the regression functions, μ_w , are Lipschitz, we obtain

$$\frac{1}{N} \sum_{i=1}^N \left(\frac{1}{M} \sum_{m=1}^M \mu_{1-W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)}) \right)^2 = o_p(1). \quad (\text{A.13})$$

Consider the first term on the right hand side of equation (A.12):

$$\begin{aligned} &\frac{1}{N} \sum_{i=1}^N \left((2W_i - 1) \left(\frac{1}{M} \sum_{m=1}^M \mu_{W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)}) \right) - \tau \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^N \left((\mu_1(X_i) - \mu_0(X_i) - \tau) + (2W_i - 1) \left(\frac{1}{M} \sum_{m=1}^M \mu_{1-W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)}) \right) \right)^2 \\ &= \frac{1}{N} \sum_{i=1}^N (\mu_1(X_i) - \mu_0(X_i) - \tau)^2 + \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{M} \sum_{m=1}^M \mu_{1-W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)}) \right)^2 \\ &+ \frac{1}{N} \sum_{i=1}^N (\mu_1(X_i) - \mu_0(X_i) - \tau) (2W_i - 1) \left(\frac{1}{M} \sum_{m=1}^M \mu_{1-W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)}) \right) \\ &= \frac{1}{N} \sum_{i=1}^N (\mu_1(X_i) - \mu_0(X_i) - \tau)^2 + o_p(1), \end{aligned} \quad (\text{A.14})$$

by Hölder's Inequality and equation (A.13). Next, consider the second term on the right hand side of equation (A.12):

$$\frac{1}{N} \sum_{i=1}^N \left(\varepsilon_i - \frac{1}{M} \sum_{m=1}^M \varepsilon_{j_m(i)} \right)^2 = \frac{1}{N} \sum_{i=1}^N \left(\varepsilon_i^2 + \frac{1}{M^2} \left(\sum_{m=1}^M \varepsilon_{j_m(i)}^2 + 2 \sum_{m=1}^M \sum_{n>m}^M \varepsilon_{j_m(i)} \varepsilon_{j_n(i)} \right) - \frac{2}{M} \sum_{m=1}^M \varepsilon_i \varepsilon_{j_m(i)} \right).$$

Therefore,

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^N \left(\varepsilon_i - \frac{1}{M} \sum_{m=1}^M \varepsilon_{j_m(i)} \right)^2 - \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right) \sigma^2(X_i, W_i) \\
&= \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right) (\varepsilon_i^2 - \sigma^2(X_i, W_i)) \\
&+ \frac{1}{N} \sum_{i=1}^N \frac{2}{M^2} \left(\sum_{m=1}^M \sum_{n>m} \varepsilon_{j_m(i)} \varepsilon_{j_n(i)} \right) - \frac{1}{N} \sum_{i=1}^N \frac{2}{M} \sum_{m=1}^M \varepsilon_i \varepsilon_{j_m(i)}. \quad (\text{A.15})
\end{aligned}$$

The expectations conditional on \mathbf{X} and \mathbf{W} of each of the three terms on the right hand side of last expression are zero, so the unconditional expectations are also zero. Because the fourth conditional moments of ε_i are uniformly bounded, we obtain:

$$\begin{aligned}
& E \left[\left(\frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right) (\varepsilon_i^2 - \sigma^2(X_i, W_i)) \right)^2 \right] \\
&= \frac{1}{N} E \left[\frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right)^2 (\varepsilon_i^2 - \sigma^2(X_i, W_i))^2 \right] = o(1).
\end{aligned}$$

The variance of the second term divided by $4/M^4$ is

$$\begin{aligned}
& E \left[\left(\frac{1}{N} \sum_{i=1}^N \sum_{m=1}^M \sum_{n>m} \varepsilon_{j_m(i)} \varepsilon_{j_n(i)} \right)^2 \right] = \frac{1}{N} E \left[\frac{1}{N} \sum_{i=1}^N \left(\sum_{m=1}^M \sum_{n>m} \varepsilon_{j_m(i)} \varepsilon_{j_n(i)} \right)^2 \right] \\
&+ \frac{2}{N} E \left[\frac{1}{N} \sum_{i=1}^N \sum_{j>i} \left(\sum_{m=1}^M \sum_{n>m} \varepsilon_{j_m(i)} \varepsilon_{j_n(i)} \right) \left(\sum_{m=1}^M \sum_{n>m} \varepsilon_{j_m(j)} \varepsilon_{j_n(j)} \right) \right] \\
&\leq \frac{1}{N} E \left[\frac{1}{N} \sum_{i=1}^N \frac{(M-1)M}{2} \bar{\sigma}^4 \right] \\
&+ \frac{1}{N} E \left[\frac{1}{N} \sum_{i=1}^N (M-1) K_M(i) (K_M(i) - 1) \bar{\sigma}^4 \right] = o(1).
\end{aligned}$$

(Last inequality holds because there are $K_M(i)(K_M(i)-1)/2$ combinations of two different sets of matches, $J_M(i_1)$ and $J_M(i_2)$ with $1 \leq i_1 < i_2 \leq N$, with observation i in both sets. Moreover, for each of these combinations, there are at most $M-1$ terms of the form $\varepsilon_{j_{m_1}(i_1)} \varepsilon_{j_{n_1}(i_1)} \varepsilon_{j_{m_2}(i_2)} \varepsilon_{j_{n_2}(i_2)}$, containing ε_i^2 and with non-zero expectations.)

The variance of the third term divided by $4/M^2$ is

$$\begin{aligned}
& E \left[\left(\frac{1}{N} \sum_{i=1}^N \sum_{m=1}^M \varepsilon_i \varepsilon_{j_m(i)} \right)^2 \right] = \frac{1}{N} E \left[\frac{1}{N} \sum_{i=1}^N \left(\sum_{m=1}^M \varepsilon_i \varepsilon_{j_m(i)} \right)^2 \right] \\
&+ \frac{2}{N} E \left[\frac{1}{N} \sum_{i=1}^N \left(\sum_{m=1}^M \varepsilon_i \varepsilon_{j_m(i)} \right) \sum_{j>i} \left(\sum_{m=1}^M \varepsilon_j \varepsilon_{j_m(j)} \right) \right] \\
&\leq \frac{1}{N} E \left[\frac{1}{N} \sum_{i=1}^N M \bar{\sigma}^4 \right] \\
&+ \frac{2}{N} E \left[\frac{1}{N} \sum_{i=1}^N M \bar{\sigma}^4 \right] = o(1).
\end{aligned}$$

As a result, we obtain

$$\frac{1}{N} \sum_{i=1}^N \left(\varepsilon_i - \frac{1}{M} \sum_{m=1}^M \varepsilon_{j_m(i)} \right)^2 - \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right) \sigma^2(X_i, W_i) = o_p(1).$$

Finally, consider the last term on the right hand side of equation (A.12). Let

$$\Psi_{M,i} = \left((2W_i - 1) \left(\frac{1}{M} \sum_{m=1}^M \mu_{W_i}(X_i) - \mu_{1-W_i}(X_{j_m(i)}) \right) - \tau \right).$$

Notice that there is a finite bound $\bar{\Psi}$, such that $|\Psi_{M,i}| \leq \bar{\Psi}$ for all i . The conditional expectation of the last term of equation (A.12) is zero, so the unconditional expectation is also zero. The conditional variance of this term (divided by 4) is:

$$\begin{aligned} & E \left[\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \Psi_{M,i} \Psi_{M,j} \left(\varepsilon_i - \frac{1}{M} \sum_{m=1}^M \varepsilon_{j_m(i)} \right) \left(\varepsilon_j - \frac{1}{M} \sum_{m=1}^M \varepsilon_{j_m(j)} \right) \right] \\ & \leq \left| E \left[\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \Psi_{M,i} \Psi_{M,j} \varepsilon_i \varepsilon_j \right] \right| + 2 \left| E \left[\frac{1}{MN^2} \sum_{i=1}^N \sum_{j=1}^N \Psi_{M,i} \Psi_{M,j} \varepsilon_i \sum_{m=1}^M \varepsilon_{j_m(j)} \right] \right| \\ & \quad + \left| E \left[\frac{1}{M^2 N^2} \sum_{i=1}^N \sum_{j=1}^N \Psi_{M,i} \Psi_{M,j} \sum_{m=1}^M \varepsilon_{j_m(i)} \sum_{m=1}^M \varepsilon_{j_m(j)} \right] \right|. \end{aligned}$$

Notice that:

$$\begin{aligned} & \left| E \left[\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \Psi_{M,i} \Psi_{M,j} \varepsilon_i \varepsilon_j \right] \right| = \left| E \left[\frac{1}{N^2} \sum_{i=1}^N \Psi_{M,i}^2 \varepsilon_i^2 \right] \right| \leq \frac{\bar{\Psi}^2}{N} E \left[\frac{1}{N} \sum_{i=1}^N \varepsilon_i^2 \right] = o(1), \\ & \left| E \left[\frac{1}{MN^2} \sum_{i=1}^N \sum_{j=1}^N \Psi_{M,i} \Psi_{M,j} \varepsilon_i \sum_{m=1}^M \varepsilon_{j_m(j)} \right] \right| \leq E \left[\frac{\bar{\Psi}}{MN^2} \sum_{i=1}^N |\Psi_{M,i}| K_M(i) \varepsilon_i^2 \right] \\ & \leq \frac{\bar{\Psi}^2}{MN} E \left[\frac{1}{N} \sum_{i=1}^N K_M(i) \varepsilon_i^2 \right] = o(1), \\ & \left| E \left[\frac{1}{M^2 N^2} \sum_{i=1}^N \sum_{j=1}^N \Psi_{M,i} \Psi_{M,j} \sum_{m=1}^M \varepsilon_{j_m(i)} \sum_{m=1}^M \varepsilon_{j_m(j)} \right] \right| \leq \left| E \left[\frac{1}{M^2 N^2} \sum_{i=1}^N \Psi_{M,i}^2 \left(\sum_{m=1}^M \varepsilon_{j_m(i)} \right)^2 \right] \right| \\ & + 2 \left| E \left[\frac{1}{M^2 N^2} \sum_{i=1}^N \sum_{j>i}^N \Psi_{M,i} \Psi_{M,j} \sum_{m=1}^M \varepsilon_{j_m(i)} \sum_{m=1}^M \varepsilon_{j_m(j)} \right] \right| \leq \frac{\bar{\Psi}^2}{N} E \left[\frac{1}{MN} \sum_{i=1}^N \sigma^2 \right] \\ & \quad + \frac{\bar{\Psi}^2}{N} E \left[\frac{1}{M^2 N} \sum_{i=1}^N K_M(i) (K_M(i) - 1) \varepsilon_i^2 \right] = o(1). \end{aligned}$$

As a result, we obtain:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \left(\hat{Y}_i(1) - \hat{Y}_i(0) - \hat{\tau}_M^{sm} \right)^2 & = \frac{1}{N} \sum_{i=1}^N (\mu_1(X_i) - \mu_0(X_i) - \tau)^2 \\ & + \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right) \sigma^2(X_i, W_i) + o_p(1) \end{aligned}$$

Applying the previous lemma:

$$\left| \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right) \sigma^2(X_i, W_i) - \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right) \hat{\sigma}^2(X_i, W_i) \right| = o_p(1).$$

Therefore,

$$\widehat{V}^{\tau(X)} = \frac{1}{N} \sum_{i=1}^N \left(\widehat{Y}_i(1) - \widehat{Y}_i(0) - \widehat{\tau}_M^{sm} \right)^2 - \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M^2} \right) \hat{\sigma}^2(X_i, W_i) \xrightarrow{p} V^{\tau(X)}.$$

From the lemma, we know that

$$\widehat{V}^E = \frac{1}{N} \sum_{i=1}^N \left(1 + \frac{K_M(i)}{M} \right)^2 \hat{\sigma}^2(X_i, W_i) \xrightarrow{p} V^E.$$

Putting the two pieces together:

$$\begin{aligned} \widehat{V} &= \widehat{V}^E + \widehat{V}^{\tau(X)} \\ &= \frac{1}{N} \sum_{i=1}^N \left(\widehat{Y}_i(1) - \widehat{Y}_i(0) - \widehat{\tau}_M^{sm} \right)^2 \\ &\quad + \frac{1}{N} \sum_{i=1}^N \left[\left(\frac{K_M(i)}{M} \right)^2 + \left(\frac{2M-1}{M} \right) \left(\frac{K_M(i)}{M} \right) \right] \hat{\sigma}^2(X_i, W_i) \xrightarrow{p} V^E + V^{\tau(X)}, \end{aligned}$$

finishing the proof for the first part of the theorem.

The proof for the second part follows the same pattern and is omitted here. It is available from the authors on the web. \square

REFERENCES

- ABADIE, A., G.W. IMBENS (2003), "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects," mimeo.
- BARNOW, B.S., G.G. CAIN AND A.S. GOLDBERGER (1980), "Issues in the Analysis of Selectivity Bias," in *Evaluation Studies*, vol. 5, ed. by E. Stromsdorfer and G. Farkas. San Francisco: Sage.
- COCHRAN, W., (1968) "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies", *Biometrics* 24, 295-314.
- COCHRAN, W., AND D. RUBIN (1973) "Controlling Bias in Observational Studies: A Review" *Sankhya*, 35, 417-46.
- 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 R. ROBB, (1984), "Alternative Methods for Evaluating the Impact of Interventions," in Heckman and Singer (eds.), *Longitudinal Analysis of Labor Market Data*, Cambridge, Cambridge University Press.
- HECKMAN, J., H. ICHIMURA, AND P. TODD, (1998), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies* 65, 261-294.
- 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*.
- IMBENS, G. (2003), "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Survey," forthcoming, *Review of Economics and Statistics*.
- JOHNSON, N., S. KOTZ, AND A. KEMP, (1992), *Univariate Discrete Distributions*, 2nd Edition, Wiley, New York.
- MILLER, G.L., S. TENG, W. THURSTON, AND S.A. VAVASIS, (1997), "Separators for Sphere-Packings and Nearest Neighbor Graphs," *Journal of the ACM*, 44, 1-29.
- OKABE, A., B. BOOTS, K. SUGIHARA, AND S. NOK CHIU, (2000), *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd Edition, Wiley, New York.
- OLVER, F.W.J., (1974), *Asymptotics and Special Functions*. Academic Press, New York.
- ROSENBAUM, P., (1995), *Observational Studies*, Springer Verlag, New York.
- ROSENBAUM, P., AND D. RUBIN, (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, 41-55.
- RUBIN, D., (1973), "Matching to Remove Bias in Observational Studies", *Biometrics*, 29, 159-183.
- RUBIN, D., (1977), "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2(1), 1-26.
- RUDIN, W., (1976), *Principles Mathematical Analysis*, 3rd Edition, McGraw-Hill, New York.
- STROOCK, D.W., (1994), *A Concise Introduction to the Theory of Integration*. Birkhäuser, Boston.