Matching on the Estimated Propensity Score

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

October 2009

Abstract

Propensity score matching estimators (Rosenbaum and Rubin, 1983) are widely used in evaluation research to estimate average treatment effects. In this article, we derive the large sample distribution of propensity score matching estimators. Our derivations take into account that the propensity score is itself estimated in a first step, previous to matching. We prove that first step estimation of the propensity score affects the large sample distribution of propensity score matching estimators. Moreover, we derive an adjustment to the large sample variance of propensity score matching estimators that corrects for first step estimation of the propensity score. In spite of the great popularity of propensity score matching estimators, these result were previously unavailable in the literature.

Alberto Abadie, John F. Kennedy School of Government, 79 John F. Kennedy Street, Cambridge MA 02138, USA. E-mail: alberto_abadie@harvard.edu, web: http://ksghome.harvard.edu/ aabadie/. Guido W. Imbens, Department of Economics, 1830 Cambridge Street, Cambridge, MA 02138. E-mail: imbens@fas.harvard.edu, web: http://www.economics.harvard.edu/faculty/imbens/imbens.html.

We are grateful to Ben Hansen, Paul Rosenbaum, James Robins, Donald Rubin, and participants in seminars at Harvard University, the University of Pennsylvania, and the Banff Center, for comments and discussions. Software implementing these methods is available on our websites.

I. Introduction

Propensity score matching estimators (Rosenbaum and Rubin, 1983) are widely used to estimate treatment effects when all treatment confounders are measured. Rosenbaum and Rubin (1983) define the propensity score as the conditional probability of assignment to a treatment given a vector of covariates including the values of all treatment confounders. Their key insight is that adjusting for the propensity score is enough to remove the bias created by all treatment confounders. Relative to matching directly on the covariates, propensity score matching has the advantage of reducing the dimensionality of matching to a single dimension. This greatly facilitates the matching process, because units with dissimilar covariate values may nevertheless have similar values in their propensity scores.

Propensity score values are rarely observed in practice. Usually the propensity score has to be estimated prior to matching. In spite of the great popularity that propensity score matching methods have gained since they were proposed by Rosenbaum and Rubin in 1983, their large sample distribution has not yet been derived for this case. A possible reason for this void in the literature is that matching estimators are highly non-smooth functionals of the distribution of the matching variables, which makes it difficult to establish an asymptotic approximation to the distribution of matching estimators when a matching variable is estimated in a first step. Moreover, it has been shown that the bootstrap is not in general valid for matching estimators (Abadie and Imbens, 2008).

In this article, we derive the large sample distribution of propensity score matching estimators. Our derivations take into account that the propensity score is itself estimated in a first step. We prove that first step estimation of the propensity score affects the large sample distribution of propensity score matching estimators. Moreover, we derive an adjustment to the large sample variance of propensity score matching estimators that corrects for first step estimation of the propensity score. Finally, we use a small simulation exercise to illustrate the implications of our theoretical results.

To preview our results, let $F(x'\theta)$ be a parametric model for the propensity score, with unknown parameters θ, and let $\hat{\theta}_N$ be the maximum likelihood estimator for θ. We show that, under regularity conditions, the estimator $\hat{\tau}_N$, for the average treatment effect $\tau = E[Y(1) - Y(0)]$, based on matching on the estimated propensity score $F(X_i'\hat{\theta}_N)$, satisfies

$$\sqrt{N}(\hat{\tau}_N - \tau) \xrightarrow{d} N(0, \sigma^2 - c' I_\theta^{-1} c).$$

In this expression, σ^2 is the variance of the matching estimator based on matching on the true propensity score $F(X_i'\theta)$ (which follows from results in Abadie and Imbens, 2006), I_θ is the Fisher information matrix for the parametric model for the propensity score, and c is a vector that depends on the covariance between the covariates and the outcome, conditional on the propensity score and the treatment. Thus, matching on the estimated propensity score has a smaller asymptotic variance than matching on the true propensity score. This is in line with results in Rubin and Thomas (1992ab) who argue that, in settings with normally distributed covariates, matching on the estimated rather than the true propensity score improves the properties of matching estimators. Hirano, Imbens and Ridder (2003) obtain a similar result for weighting estimators.

The rest of the article is organized as follows. Section II provides an introduction to propensity score matching. Section III is the main section of the article. In this section we derive the large sample properties of an estimator that match on estimated propensity scores. Section IV proposes an estimator for the adjusted standard errors derived in section III. In section V we report the results of a small simulation exercise. Section VI concludes.

II. Matching on the Estimated Propensity Score

In evaluation research the focus of the analysis is typically the effect of a binary treatment, represented in this paper by the indicator variable W, on some outcome variable, Y. More specifically, $W = 1$ indicates exposure to treatment, while $W = 0$ indicates lack of exposure to treatment. Following Rubin (1974), we define treatment effects in terms of potential outcomes. We define $Y(1)$ as the potential outcome under exposure to treatment, and $Y(0)$ as the potential outcome under no exposure to treatment. Our goal is to estimate the average treatment effect,

$$\tau = E\Big[Y(1) - Y(0)\Big],$$

where the expectation is taken over the population of interest, based on a random sample from this population. Estimation of treatment effects is complicated by the fact that for each unit in the population, the observed outcome reflects only one of the potential outcomes:

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

Let X be a vector of covariates that includes treatment confounders, that is, variables that affect the probability of treatment exposure and the potential outcomes. The propensity score is $p(X) = \Pr(W = 1|X)$. The following assumption is often referred to as "strong ignorability" (Rosenbaum and Rubin (1983)).

Assumption 1: (i) $Y(1), Y(0) \perp\!\!\!\perp W|X$ almost surely; (ii) $0 < p(X) < 1$ almost surely.

Assumption 1(i) will hold if all treatment confounders are included in X; so, after controlling for X, treatment exposure is independent of the potential outcomes. Assumption 1(ii) states that for almost all values of X the population includes treated and untreated units.

Let $\mu(w, x) = E[Y|W = w, X = x]$ and $\bar{\mu}(w, p) = E[Y|W = w, p(X) = p]$ be the regression of the outcome on the treatment indicator and the covariates, and on the treatment indicator and the propensity score respectively. Rosenbaum and Rubin (1983) prove that, under Assumption 1,

$$\tau = E\Big[\bar{\mu}(1, p(X)) - \bar{\mu}(0, p(X))\Big].$$

In other words, adjusting for the propensity score is enough to eliminate the bias created by all treatment confounders.

This result by Rosenbaum and Rubin (1983) motivates the use of propensity score matching estimators. Following Rosenbaum and Rubin (1983) and the vast majority of the empirical literature, consider a generalized linear specification for the propensity score $p(X) = F(x'\theta)$. In empirical research the link function F is usually specified as logit or probit. Assume for the moment that the parameters of the propensity score, θ, are known. For each observation, i, let $\mathcal{J}_M(i, \theta)$ be a set of M observations in the treatment group opposite to i and with propensity score values similar to $F(X_i'\theta)$. A propensity score matching estimator can be defined as:

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

In this article we will consider matching with replacement, so each unit in the sample can be used as a match multiple times. In the absence of matching ties, the sets $\mathcal{J}_M(i,\theta)$ can be defined as:

$$\mathcal{J}_M(i,\theta) = \left\{ j : W_j = 1 - W_i,\ \left(\sum_{k=1}^{N} 1_{\{W_k = 1-W_i\}}\, 1_{\{|F(X_i'\theta)-F(X_k'\theta)|\le|F(X_i'\theta)-F(X_j'\theta)|\}}\right) \le M \right\}.$$

Let $K_{N,i}(\theta)$ be the number of times that observation i is used as a match (when matching on $F(X'\theta)$):

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

The estimator $\widehat{\tau}_N(\theta)$ can be represented as:

$$\widehat{\tau}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N}(2W_i - 1)\left(1 + \frac{K_{N,i}(\theta)}{M}\right) Y_i.$$

In practice, propensity scores are not directly observed and estimators that match on the true propensity score are therefore unfeasible. For some random sample $\{Y_i, W_i, X_i\}_{i=1}^N$, let $\widehat{\theta}_N$ be an estimator of θ. A matching estimator of τ that matches on estimated propensity scores is given by:

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

We assume, in concordance with the literature, that $\widehat{\theta}_N$ is the Maximum Likelihood estimator of θ. In the next section, we derive the large sample distribution of $\widehat{\tau}_N(\widehat{\theta}_N)$.

III. Large Sample Distribution

We begin by introducing a decomposition of $\widehat{\tau}_N(\theta)$ that will be used later in this section. Define

$$T_N(\theta) = \sqrt{N}\big(\widehat{\tau}_N(\theta) - \tau\big)$$

$$= \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}(2W_i - 1)\left(1 + \frac{K_{N,i}(\theta)}{M}\right)Y_i - \tau\right).$$

Notice that $T_N(\theta) = D_N(\theta) + R_N(\theta)$, where

$$D_N(\theta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big(\bar{\mu}(1, F(X_i'\theta)) - \bar{\mu}(0, F(X_i'\theta)) - \tau\Big)$$
$$= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(2W_i - 1)\left(1 + \frac{K_{N,i}(\theta)}{M}\right)\Big(Y_i - \bar{\mu}(W_i, F(X_i'\theta))\Big),$$

and

$$R_N(\theta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(2W_i - 1)\Big(\bar{\mu}(1 - W_i, F(X_i'\theta)) - \frac{1}{M}\sum_{j\in\mathcal{J}_M(i)}\bar{\mu}(1 - W_i, F(X_i'\theta))\Big).$$

Let P^θ be the distribution of $Z = \{Y, W, X\}$, induced by the propensity score, $F(x'\theta)$, the marginal distribution of X, and the conditional distribution of Y given X and W. To simplify the exposition, we will implicitly assume that Y and X are bounded, so all moments exist for these two variables. Consider $Z_{N,i} = \{Y_{N,i}, W_{N,i}, X_{N,i}\}$ with distribution given by the local "shift" P^{θ_N} with $\theta_N = \theta + h/\sqrt{N}$, where h is a conformable vector of constants.

Assumption 2: *(i) For some $\varepsilon > 0$, all x in the support of X, and all $\theta^* \in \mathbb{R}^k$ such that $\|\theta - \theta^*\| \leq \varepsilon$, the distribution of $F(X'\theta^*)$ is continuous with support equal to an interval bounded away from zero and one. (ii) For all $\theta^* \in \mathbb{R}^k$ such that $\|\theta - \theta^*\| \leq \varepsilon$, all F in the support of $F(X'\theta^*)$, and all $w = 0, 1$, the regression function $E[Y_{N,i}|W_{N,i} = w, F(X_{N,i}'\theta^*) = F]$ is Lipschitz-continuous in F.*

Proposition 1: *If Assumption 2 holds, $R_N(\theta_N) \xrightarrow{p} 0$ under P^{θ_N}.*

(All proof are provided in the appendix.)

Let

$$\Lambda_N(\theta|\theta_N) = \sum_{i=1}^{N}\log\frac{dP^\theta}{dP^{\theta_N}}(Z_{N,i}),$$

and

$$\Delta_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_{N,i} \frac{W_{N,i} - F(X'_{N,i}\theta)}{F(X'_{N,i}\theta)(1 - F(X'_{N,i}\theta))} f(X'_{N,i}\theta).$$

Let $\widehat{\theta}_N$ be the Maximum Likelihood estimator of θ, and let

$$I_\theta = E\left[\frac{f(X'\theta)^2}{F(X'\theta)(1 - F(X'\theta))} XX'\right],$$

be the Fisher Information Matrix for θ.

Assumption 3: *Under P^{θ_N}:*

$$\Lambda_N(\theta|\theta_N) = -h'\Delta_N(\theta_N) - \frac{1}{2}h'I_\theta h + o_p(1), \tag{1}$$

$$\Delta_N(\theta_N) \xrightarrow{d} N(0, I_\theta), \tag{2}$$

and

$$\sqrt{N}(\widehat{\theta}_N - \theta_N) = I_\theta^{-1}\Delta_N(\theta_N) + o_p(1). \tag{3}$$

For regular parametric models, equation (1) can be established using Proposition 2.1.2 in Bickel et al. (1998). Also for regular parametric models, equation (2) is derived in the proof of Proposition 2.1.2 in Bickel et al. (1998). Equation (3) can be established using the same set of results plus classical conditions for asymptotic linearity of maximum likelihood estimators (see, e.g., van der Vaart (1998) Theorem 5.39; Lehmann and Romano (2005) Theorem 12.4.1).

The following assumption collects some technical regularity conditions that will be used later in this section.

Assumption 4: *(i) The function F has a continuous derivative. (ii) There is some $\varepsilon > 0$, such that for all θ^* such that $\|\theta^* - \theta\| \leq \varepsilon$ the density of $F(X'\theta^*)$ is bounded and bounded away from zero. (iii) For all bounded functions $h(Y, W, X)$, $E_{\theta_N}[h(Y, W, X)|F(X'\theta_N), W]$ converges to $E[h(Y, W, X)|F(X'\theta), W]$ (where E_{θ_N} denotes an expectation with respect to P^{θ_N}).*

Assumption 4(i) is satisfied in the most usual binary choice models employed for the estimation of the propensity score (Probit, Logit). We adopt Assumption 4(ii) for technical reasons, because it simplifies matters considerably in the proof of our main theorem. This assumption typically implies some trimming on the population of interest to discard low-density values of the propensity score. (To avoid cluttering, we leave such trimming implicit in our notation.) Primitive conditions for assumption 4(iii) can be established using the results in Ganssler and Pfanzagl (1971).

Our derivation of the limit distribution of $\sqrt{N}(\widehat{\tau}_N - \tau)$ is based on the techniques developed in Andreou and Werker (2005) to analyze to limit distribution of residual-based statistics. We proceed in four steps. First, we derive the joint limit distribution of $(T_N(\theta_N), \Delta_N(\theta_N))$ under P^{θ_N}. The following result is useful in that respect.

PROPOSITION 2: *Suppose that Assumption 3 holds. Then, under P^{θ_N}:*

$$\left(\begin{array}{c} D_N(\theta_N) \\ \Delta_N(\theta_N) \end{array} \right) \xrightarrow{d} N\left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} \sigma^2 & c' \\ c & I_\theta \end{array} \right) \right),$$

where σ^2 is the asymptotic variance of $T_N(\theta)$ and

$$c = E\left[\text{cov}(X, \mu(W,X)|F(X'\theta), W) f(X'\theta) \left(\frac{W}{F(X'\theta)^2} + \frac{1-W}{(1-F(X'\theta))^2} \right) \right].$$

Notice that propositions 1 and 2 imply:

$$\left(\begin{array}{c} T_N(\theta_N) \\ \Delta_N(\theta_N) \end{array} \right) \xrightarrow{d} N\left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} \sigma^2 & c' \\ c & I_\theta \end{array} \right) \right), \tag{4}$$

under P^{θ_N}.

Second, we use equation (4) ,along with Assumption 3, to obtain the joint limit distribution of $(T_N(\theta_N), \sqrt{N}(\widehat{\theta}_N - \theta_N), \Lambda_N(\theta|\theta_N))$ under P^{θ_N}:

$$\left(\begin{array}{c} T_N(\theta_N) \\ \sqrt{N}(\widehat{\theta}_N - \theta_N) \\ \Lambda_N(\theta|\theta_N) \end{array} \right) \xrightarrow{d} N\left(\left(\begin{array}{c} 0 \\ 0 \\ -h'I_\theta h/2 \end{array} \right), \left(\begin{array}{ccc} \sigma^2 & c'I_\theta^{-1} & -c'h \\ I_\theta^{-1}c & I_\theta^{-1} & -h \\ -h'c & -h' & h'I_\theta h \end{array} \right) \right).$$

Third, applying Le Cam's third lemma, we obtain

$$\left(\begin{array}{c} T_N(\theta_N) \\ \sqrt{N}(\widehat{\theta}_N - \theta_N) \end{array} \right) \xrightarrow{d} N\left(\left(\begin{array}{c} -c'h \\ -h \end{array} \right), \left(\begin{array}{cc} \sigma^2 & c'I_\theta^{-1} \\ I_\theta^{-1}c & I_\theta^{-1} \end{array} \right) \right),$$

or equivalently:

$$\begin{pmatrix} T_N(\theta + h/\sqrt{N}) \\ \sqrt{N}(\widehat{\theta}_N - \theta) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} -c'h \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & c' I_\theta^{-1} \\ I_\theta^{-1} c & I_\theta^{-1} \end{pmatrix} \right),$$

under P^θ, for any $h \in \mathbb{R}^k$. Finally, we calculate the limit distribution of $T_N(\widehat{\theta}_N) = \sqrt{N}(\widehat{\tau}_N - \tau)$ as the limit distribution of $T_N(\theta + h/\sqrt{N})$ conditional on $\widehat{\theta}_N = \theta + h/\sqrt{N}$ (i.e. $\sqrt{N}(\widehat{\theta}_N - \theta) = h$), integrated over the distribution of $\sqrt{N}(\widehat{\theta}_N - \theta)$.

Theorem 1: *Under P^θ*

$$\sqrt{N}(\widehat{\tau}_N - \tau) \xrightarrow{d} N(0, \sigma^2 - c' I_\theta^{-1} c).$$

The asymptotic variance of $\widehat{\tau}_N$ is adjusted by $-c' I_\theta^{-1} c$ to account for first-step estimation of the propensity score. In this case, the adjustment reduces the asymptotic variance. This need not be the case for matching estimators of other treatment parameters, such as the average treatment effect on the treated.

Formally, the proof of Theorem 1 requires a discretization of the first step estimator $\widehat{\theta}_N$ (see Andreou and Werker, 2005, for details). This discretization can be arbitrarily fine and the result of Theorem 1 arises in the limit, as we make the discretization increasingly finer.

IV. Estimation of the Asymptotic Variance

Let $\mathcal{H}_J(i, \theta)$ be the set of the J units with $W = W_i$ and closest values of $F(X'\theta)$ to $F(X_i'\theta)$, and let $\bar{Y}_i^{(J,\theta)}$ be the average of Y for for the units in $\{i \cup \mathcal{H}_J(i, \theta)\}$. Consider the following estimator of $\widehat{\text{var}}(Y_i | F(X_i'\theta), W_i)$:

$$\widehat{\sigma}^2_{N,i}(\theta) = \frac{1}{J} \sum_{j \in \{i \cup \mathcal{H}_J(i,\theta)\}} (Y_j - \bar{Y}_i^{(J,\theta)})(Y_j - \bar{Y}_i^{(J,\theta)}).$$

Abadie and Imbens (2006) propose the following estimator for $\sigma^2(\theta)$:

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

Let $\bar{X}_i^{(J,\theta)}$ be the averages of X for for the units in $\{i \cup \mathcal{H}_J(i,\theta)\}$. Notice that $Y = \mu(W,X) + \varepsilon$, where $E[\varepsilon \,|\, X, W] = 0$. As a result:

$$\text{cov}(X, \mu(W,X)|F(X'\theta), W) = \text{cov}(X, Y|F(X'\theta), W).$$

Consider the following estimator of $\text{cov}(X, Y|F(X'\theta), W)$:

$$\widehat{\text{cov}}(X_i, Y_i|F(X_i'\theta), W_i) = \frac{1}{J} \sum_{j \in \{i \cup \mathcal{H}_J(i,\widehat{\theta}_N)\}} (X_j - \bar{X}_i^{(J,\widehat{\theta}_N)})(Y_j - \bar{Y}_i^{(J,\widehat{\theta}_N)}).$$

Our estimator of c is:

$$\widehat{c} = \frac{1}{N} \sum_{i=1}^{N} \widehat{\text{cov}}(X_i, Y_i|F(X_i'\theta), W_i) f(X_i'\widehat{\theta}_N) \left(\frac{W_i}{F(X_i'\widehat{\theta}_N)^2} + \frac{(1-W_i)}{(1-F(X_i'\widehat{\theta}_N))^2} \right).$$

Finally, let

$$\widehat{I}_{\theta,N} = \frac{1}{N} \sum_{i=1}^{N} \frac{f(X_i'\widehat{\theta}_N)^2}{F(X_i'\widehat{\theta}_N)(1-F(X_i'\widehat{\theta}_N))} X_i X_i'.$$

Because I_θ is non-singular, the inverse of $\widehat{I}_{\theta,N}$ exists with probability approaching one. Our estimator of the large sample variance of the propensity score matching estimator, adjusted for first step estimation of the propensity score, is:

$$\widehat{\sigma}^2_{\text{adj},N}(\widehat{\theta}_N) = \widehat{\sigma}^2_N(\widehat{\theta}_N) - \widehat{c}'\widehat{I}_{\theta,N}^{-1}\widehat{c}.$$

Consistency of this estimator can be shown using the results in Abadie and Imbens (2006) and the contiguity arguments employed in section III.

V. A Small Simulation Exercise

In this section, we run a small Monte Carlo exercise to investigate the sampling distribution of propensity score matching estimators and of the approximation to that distribution that we propose in the article.

We use a simple Monte Carlo design. The outcome variable is generated by $Y = 5W + 4(X_1 + X_2) + U$, where X_1 and X_2 are independent and uniform on $[0,1]$ and U is a standard Normal variable independent of (W, X_1, X_2). The treatment variable, W, is related to (X_1, X_2) through the propensity score, which is logistic

$$\Pr(W = 1|X_1 = x_1, X_2 = x_2) = \frac{\exp(1 + x_1 - x_2)}{1 + \exp(1 + x_1 - x_2)}.$$

Table I reports the results of our Monte Carlo simulation for $M = 1$ and $N = 5000$. As in our theoretical results, the variance of $\widehat{\tau}_N(\theta)$, the estimator that matches on the true propensity score, is larger than the variance of $\widehat{\tau}_N(\widehat{\theta}_N)$, the estimator that matches on the estimated propensity score. The estimator of the variance of $\widehat{\tau}_N(\theta)$ proposed in Abadie and Imbens (2006), $\widehat{\sigma}^2(\theta)$, is centered at the variance of $\widehat{\tau}_N(\theta)$. $\widehat{\sigma}^2_N(\widehat{\theta}_N)$ is the estimator of the variance that treats the first step estimate of the propensity score $\widehat{\theta}_N$ as if it was the true propensity score, and $\widehat{\sigma}^2_{\text{adj},N}(\widehat{\theta}_N)$ is the adjusted estimator of the variance that takes into account that the propensity score is itself estimated in a first step. Finally, the table reports also confidence interval constructed with adjusted and unadjusted standard errors. In concordance with out theoretical results, the simulation shows that $\widehat{\sigma}^2_N(\widehat{\theta}_N)$ is biased and too large on average. As a result, confidence intervals constructed with $\widehat{\sigma}^2_N(\widehat{\theta}_N)$ have larger than nominal coverage rates. In contrast, $\widehat{\sigma}^2_{\text{adj},N}(\widehat{\theta}_N)$ is unbiased and produce confidence intervals that have coverage rates close to nominal rates.

VI. Conclusions

In this article, we propose a method to correct to the asymptotic variance of propensity score matching estimators when the propensity scores are estimated in a first step. Our results allow valid large sample inference for propensity score matching estimators.

Appendix

Proof of Proposition ??: See Abadie and Imbens (2006). □

For the proof of Proposition 1 we will need some preliminary lemmas.

Lemma A.1: *Consider two independent samples of sizes n_0 and n_1 from continuous distributions F_0 and F_1 with common support: $V_{0,1},\dots,V_{0,n_0} \sim$ i.i.d. F_0 and $V_{1,1},\dots,V_{1,n_1} \sim$ i.i.d. F_1. Let $N = n_0 + n_1$. Assume that the support of F_0 and F_1 is an interval inside $[0,1]$. Let f_0 and f_1 be the densities of F_0 and F_1, respectively. Suppose that for any v in the supports of F_0 and F_1, $f_1(v)/f_0(v) \le \bar{r}$. For $1 \le i \le n_1$ and $1 \le m \le M \le n_0$, let $|U_{n_0,n_1,i}|_{(m)}$ be the m-th order statistic of $\{|V_{1,i} - V_{0,1}|,\dots,|V_{1,i} - V_{0,n_0}|\}$. Then, for $n_0 \ge 3$:*

$$E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{n_1}\frac{1}{M}\sum_{m=1}^{M}|U_{n_0,n_1,i}|_{(m)}\right] \le \bar{r}\,\frac{n_1}{N^{1/2}\lfloor n_0^{3/4}\rfloor} + M\frac{n_1}{N^{1/2}}\,n_0^{M-1/4}\exp(-n_0^{1/4}).$$

Proof: Consider N balls assigned at random among n bins of equal probability. It is known that the mean of the number of bins with exactly m balls is equal to

$$n\begin{pmatrix} N \\ m \end{pmatrix}\left(\frac{1}{n}\right)^m\left(1-\frac{1}{n}\right)^{N-m}$$

(see Johnson and Kotz, 1977, p. 114). Because $f_1(v)/f_0(v) \le \bar{r}$, for any measurable set A:

$$\Pr(V_{1,i} \in A) = \int_A f_1(v)\,dv = \int_A\left(\frac{f_1(v)}{f_0(v)}\right)f_0(v)\,dv \le \bar{r}\,\Pr(V_{0,i} \in A).$$

Divide the support of F_0 and F_1 in $\lfloor n_0^{3/4}\rfloor$ cells of equal probability $1/\lfloor n_0^{3/4}\rfloor$ under F_0. Let Z_{M,n_0} be the number of such cells are not occupied by at least M observations from the second sample: $V_{0,1},\dots,V_{0,n_0}$. For $i = 1,\dots,N$. Let $\mu_{M,n_0} = E[Z_{M,n_0}]$. Notice that $n_0 \ge 3$ implies $\lfloor n_0^{3/4}\rfloor \ge 2$. Then,

$$\begin{aligned}
\mu_{M,n_0} &= \sum_{m=0}^{M-1}\lfloor n_0^{3/4}\rfloor\begin{pmatrix} n_0 \\ m \end{pmatrix}\left(\frac{1}{\lfloor n_0^{3/4}\rfloor}\right)^m\left(1-\frac{1}{\lfloor n_0^{3/4}\rfloor}\right)^{n_0-m} \\
&\le \sum_{m=0}^{M-1}\lfloor n_0^{3/4}\rfloor\frac{n_0^m}{m!}\left(\frac{1}{\lfloor n_0^{3/4}\rfloor}\right)^m\left(1-\frac{1}{\lfloor n_0^{3/4}\rfloor}\right)^{n_0-m} \\
&\le Mn_0^{M-1/4}\left(1-\frac{1}{n_0^{3/4}}\right)^{n_0}.
\end{aligned}$$

Using Markov's inequality,

$$\Pr(Z_{M,n_0} > 0) = \Pr(Z_{M,n_0} \ge 1) \le \mu_{M,n_0} \le Mn_0^{M-1/4}\left(1-\frac{1}{n_0^{3/4}}\right)^{n_0}.$$

Notice that for any positive a, we have that $a - 1 \geq \log(a)$. Therefore, for any $b < N$, we have that $\log(1 - b/N) \leq -b/N$ and $(1 - b/N)^N \leq \exp(-b)$. As a result, we obtain:

$$\left(1 - \frac{1}{n_0^{3/4}}\right)^{n_0} = \left(1 - \frac{n_0^{1/4}}{n_0}\right)^{n_0} \leq \exp(-n_0^{1/4}).$$

Putting together the last two displayed equations, we obtain the following exponential bound for $\Pr(Z_{M,n_0} > 0)$:

$$\Pr(Z_{M,n_0} > 0) \leq M n_0^{M-1/4} \exp(-n_0^{1/4}).$$

Notice that $|U_{n_0,n_1,i}|_{(m)} \leq 1$. For $0 \leq n \leq \lfloor n_0^{3/4} \rfloor$, let $c_{n_0,n} = F^{-1}(n/\lfloor n_0^{3/4} \rfloor)$, then

$$\begin{aligned}
E\left[\sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \,\middle|\, Z_{M,n_0} = 0\right] &\leq \sum_{n=1}^{\lfloor n_0^{3/4} \rfloor} M\big(c_{n_0,n} - c_{n_0,n-1}\big) \Pr\left(c_{n_0,n-1} \leq V_{1,i} \leq c_{n_0,n} \middle| Z_{M,n_0} = 0\right) \\
&= \frac{M\bar{r}}{\lfloor n_0^{3/4} \rfloor} \sum_{n=1}^{\lfloor n_0^{3/4} \rfloor} \big(c_{n_0,n} - c_{n_0,n-1}\big) \\
&\leq \frac{M\bar{r}}{\lfloor n_0^{3/4} \rfloor}.
\end{aligned}$$

Now,

$$\begin{aligned}
E\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} \frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)}\right] &= E\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} \frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \middle| Z_{M,n_0} = 0\right] \Pr(Z_{M,n_0} = 0) \\
&+ E\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} \frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \middle| Z_{M,n_0} > 0\right] \Pr(Z_{M,n_0} > 0) \\
&\leq E\left[\frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} \frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \middle| Z_{M,n_0} = 0\right] \\
&+ \frac{n_1}{N^{1/2}} \Pr(Z_{M,n_0} > 0) \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^{n_1} E\left[\frac{1}{M} \sum_{m=1}^{M} |U_{n_0,n_1,i}|_{(m)} \middle| Z_{M,n_0} = 0\right] \\
&+ \frac{n_1}{N^{1/2}} \Pr(Z_{M,n_0} > 0) \\
&\leq \bar{r}\, \frac{n_1}{N^{1/2} \lfloor n_0^{3/4} \rfloor} + M \frac{n_1}{N^{1/2}}\, n_0^{M-1/4} \exp(-n_0^{1/4}).
\end{aligned}$$

□

Lemma A.2: *(Inverse Moments of the Doubly Truncated Binomial Distribution) Let N_0 be a Binomial variable with parameters $(N, (1-p))$ that is left-truncated for values smaller than M and right-truncated for values greater than $N - M$, where $M < N/2$. Then, for any $r > 0$, there exist a constant C_r, such that*

$$E\left[\left(\frac{N}{N_0}\right)^r\right] \leq C_r,$$

for all $N > 2M$*.*

PROOF: Let $N_1 = N - N_0$. For all $\bar{q} > 0$,

$$\begin{aligned} E\left[\left(\frac{N}{N_0}\right)^r\right] &= E\left[\left(\frac{N}{N_0}\right)^r 1\left\{\frac{N}{N_0} > \bar{q}\right\}\right] + E\left[\left(\frac{N}{N_0}\right)^r 1\left\{\frac{N}{N_0} \leq \bar{q}\right\}\right] \\ &\leq \left(\frac{N}{M}\right)^r \Pr\left(\frac{N}{N_0} > \bar{q}\right) + \bar{q}^r \\ &= \left(\frac{N}{M}\right)^r \Pr\left(N_1 > \left(1 - \frac{1}{\bar{q}}\right) N\right) + \bar{q}^r \end{aligned}$$

Notice that:

$$\Pr\left(N_1 > \left(1 - \frac{1}{\bar{q}}\right) N\right) = \frac{\sum\limits_{x > (1-1/\bar{q})N, x \geq M}^{x \leq N-M} \binom{N}{x} p^x (1-p)^{N-x}}{\sum\limits_{x \geq M}^{x \leq N-M} \binom{N}{x} p^x (1-p)^{N-x}}$$

For $N > 2M$ the denominator can be bounded away from zero. Therefore, for some positive constant C, and $\bar{q} > 1/(1-p)$,

$$\begin{aligned} \Pr\left(N_1 > \left(1 - \frac{1}{\bar{q}}\right) N\right) &\leq C \sum_{x > (1-1/\bar{q})N, x \geq M}^{x \leq N-M} \binom{N}{x} p^x (1-p)^{N-x} \\ &\leq C \sum_{x > (1-1/\bar{q})N} \binom{N}{x} p^x (1-p)^{N-x} \\ &\leq C \exp\left\{-2(1 - 1/\bar{q} - p)^2 N\right\}, \end{aligned}$$

by Hoeffding's Inequality (e.g. van der Vaart and Wellner (1996), p. 459). Therefore $E[(N/N_0)^r]$ is uniformly bounded for $N > 2M$. □

LEMMA A.3: *Suppose that the propensity score,* $\Pr(W = 1|X)$*, is continuously distributed and that there exist* $c_L > 0$ *and* $c_U < 1$ *such that* $c_L \leq \Pr(W = 1|X = x) \leq c_U$ *for all* $x \in \mathcal{X}$*. Let* f_1 *be the distribution of the propensity score conditional on* $W = 1$*, and let* f_0 *be the distribution of the propensity score conditional on* $W = 0$*. Then, the ratio* $f_1(p)/f_0(p)$ *is bounded and bounded away from zero.*

PROOF: Use Bayes' Theorem to show that $f_1(p)/f_0(p) = (p/(1-p))(\Pr(W = 1)/\Pr(W = 0))$. □

PROOF OF PROPOSITION 1: Let $f_1^{\theta_N}$ be the distribution of the propensity score conditional on $W = 1$, and let $f_0^{\theta_N}$ be the distribution of the propensity score conditional on $W = 0$. By lemma A.3 the ratio $f_1^{\theta_N}(p)/f_0^{\theta_N}(p)$ is uniformly bounded by some constant $\bar{r}$. Consider N_0 and N_1 as in Lemma A.2. Let

$$\psi^{(1)}_{M,N_0,N_1} = \bar{r}\,\frac{N_1}{N^{1/2}\lfloor N_0^{3/4}\rfloor} + M\frac{N_1}{N^{1/2}}\,N_0^{M-1/4}\exp(-N_0^{1/4}).$$

Then, $\psi^{(1)}_{M,N_0,N_1} \xrightarrow{p} 0$. Rearrange the observations in the sample so that the first N_1 observations have $W = 1$ and the remaining $N_0 = N - N_1$ observations have $W = 0$. For $1 \leq i \leq N_1$ and $1 \leq m \leq M$, let $|U_{N_0,N_1,i}|_{(m)}$ be the m-th order statistic of $\{|F(X'_{N,i}\theta_N) - F(X'_{N,N_1+1}\theta_N)|, \ldots, |F(X'_{N,i}\theta_N) - F(X'_{N,N}\theta_N)|\}$. For $N_1 + 1 \leq i \leq N$ and $1 \leq m \leq M$, let $|U_{N_0,N_1,i}|_{(m)}$ be the m-th order statistic of $\{|F(X'_{N,i}\theta_N) - F(X'_{N,1}\theta_N)|, \ldots, |F(X'_{N,i}\theta_N) - F(X'_{N,N_1}\theta_N)|\}$. Lemma A.1 implies that for large enough N:

$$E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N_1}\frac{1}{M}\sum_{m=1}^{M}|U_{N_0,N_1,i}|_{(m)}\right] \leq E\left[\psi^{(1)}_{M,N_0,N_1}\right]. \tag{A.1}$$

Therefore, to prove that the left-hand-side of equation (A.1) converges to zero, it is enough to show that $\psi^{(1)}_{M,N_0,N_1}$ is asymptotically uniformly integrable:

$$\lim_{k\to\infty}\limsup_{N\to\infty} E\left[\psi^{(1)}_{M,N_0,N_1} 1\{\psi^{(1)}_{M,N_0,N_1} > k\}\right] = 0$$

(see, e.g., van der Vaart (1998), p. 17). Notice that the ratio $N^{3/4}/\lfloor N^{3/4}\rfloor$ is bounded. This, in combination with Lemma A.2, implies that for all $k > 0$ and some positive constant C,

$$\begin{aligned}
E\left[\psi^{(1)}_{M,N_0,N_1} 1\{\psi^{(1)}_{M,N_0,N_1} > k\}\right] &\leq E\left[\psi^{(1)}_{M,N_0,N_1}\right] \\
&\leq E\left[\bar{r}\,\frac{N^{1/2}}{\lfloor N_0^{3/4}\rfloor} + M\frac{N^{1/2}}{N_0^{3/4}}\,N_0^{M+1/2}\exp(-N_0^{1/4})\right] \\
&\leq C\,E\left[\frac{N^{1/2}}{N_0^{3/4}}\right] \\
&= \frac{C}{N^{1/4}}\,E\left[\frac{N^{3/4}}{N_0^{3/4}}\right] \to 0.
\end{aligned}$$

Similarly,

$$E\left[\frac{1}{\sqrt{N}}\sum_{i=N_1+1}^{N}\frac{1}{M}\sum_{m=1}^{M}|U_{N_0,N_1,i}|_{(m)}\right] \xrightarrow{p} 0.$$

Using Markov's inequality, we obtain that for any $\varepsilon > 0$:

$$\Pr\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{M}\sum_{m=1}^{M}|U_{N_0,N_1,i}|_{(m)} > \varepsilon\right) \leq \frac{E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{M}\sum_{m=1}^{M}|U_{N_0,N_1,i}|_{(m)}\right]}{\varepsilon} \longrightarrow 0.$$

The result now follows from Lipschitz-continuity of the regression functions, $E[Y_{N,i}|W_{N,i} = w, F(X'_{N,i}\theta^*) = F]$ for some $\varepsilon > 0$ and $\|\theta^* - \theta\| \leq \varepsilon$. □

PROOF OF PROPOSITION 2 *(Sketch)*: In this proof, we first establish a extend the martingale representation of matching estimators (Abadie and Imbens, 2009) to the propensity score matching estimator studied in this article. Consider the linear combination $C_N = z_1 D_N(\theta_N) + z_2'\Delta_N(\theta_N)$.

$$C_N = z_1\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big(\bar{\mu}(1, F(X'_{N,i}\theta_N)) - \bar{\mu}(0, F(X'_{N,i}\theta_N)) - \tau\Big)$$

$$
\begin{aligned}
&+ \quad z_1 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (2W_{N,i} - 1) \left(1 + \frac{K_{N,i}(\theta_N)}{M}\right) \left(Y_{N,i} - \bar{\mu}(W_{N,i}, F(X'_{N,i}\theta_N))\right) \\
&+ \quad z'_2 \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_{N,i} \frac{W_{N,i} - F(X'_{N,i}\theta_N)}{F(X'_{N,i}\theta_N)(1 - F(X'_{N,i}\theta_N))} f(X'_{N,i}\theta_N).
\end{aligned}
$$

C_N can be analyzed using martingale methods. Notice that:

$$
C_N = \sum_{k=1}^{3N} \xi_{N,k},
$$

where

$$
\begin{aligned}
\xi_{N,k} &= z_1 \frac{1}{\sqrt{N}} \left(\bar{\mu}(1, F(X'_{N,k}\theta_N)) - \bar{\mu}(0, F(X'_{N,k}\theta_N)) - \tau\right) \\
&+ \quad z'_2 \frac{1}{\sqrt{N}} E_{\theta_N}[X_{N,k} \,|\, F(X'_{N,k}\theta_N)] \frac{W_{N,k} - F(X'_{N,k}\theta_N)}{F(X'_{N,k}\theta_N)(1 - F(X'_{N,k}\theta_N))} f(X'_{N,k}\theta_N),
\end{aligned}
$$

for $1 \leq k \leq N$,

$$
\begin{aligned}
\xi_{N,k} &= z'_2 \frac{1}{\sqrt{N}} (X_{N,k-N} - E_{\theta_N}[X_{N,k-N} \,|F(X'_{N,k-N}\theta_N)]) \frac{(W_{N,k-N} - F(X'_{N,k-N}\theta_N)) f(X'_{N,k-N}\theta_N)}{F(X'_{N,k-N}\theta_N)(1 - F(X'_{N,k-N}\theta_N))} \\
&+ \quad z_1 \frac{1}{\sqrt{N}} (2W_{N,k-N} - 1) \left(1 + \frac{K_{N,k-N}(\theta_N)}{M}\right) \left(\mu(W_{N,k-N}, X_{N,k-N}) - \bar{\mu}(W_{N,k-N}, F(X'_{N,k-N}\theta_N))\right).
\end{aligned}
$$

for $N + 1 \leq k \leq 2N$,

$$
\xi_{N,k} = z_1 \frac{1}{\sqrt{N}} (2W_{N,k-2N} - 1) \left(1 + \frac{K_{N,k-2N}(\theta_N)}{M}\right) \left(Y_{N,k-2N} - \mu(W_{N,k-2N}, X_{N,k-2N})\right),
$$

for $2N + 1 \leq k \leq 3N$. Consider the σ-fields $\mathcal{F}_{N,k} = \sigma\{W_{N,1}, \ldots, W_{N,k}, X'_{N,1}\theta_N, \ldots, X'_{N,k}\theta_N\}$ for $1 \leq k \leq N$, $\mathcal{F}_{N,k} = \sigma\{W_{N,1}, \ldots, W_{N,N}, X'_{N,1}\theta_N, \ldots, X'_{N,N}\theta_N, X_{N,1}, \ldots, X_{N,k-N}\}$ for $N + 1 \leq k \leq 2N$, and $\mathcal{F}_{N,k} = \sigma\{W_{N,1}, \ldots, W_{N,N}, X_{N,1}, \ldots, X_{N,N}, Y_{N,1}, \ldots, Y_{N,k-N}\}$ for $2N+1 \leq k \leq 3N$. Then,

$$
\left\{ \sum_{j=1}^{i} \xi_{N,j}, \mathcal{F}_{N,i}, 1 \leq i \leq 3N \right\}
$$

is a martingale for each $N \geq 1$. Therefore, the limiting distribution of C_N can be studied using Martingale Central Limit Theorem (e.g., Theorem 35.12 in Billingsley (1995), p. 476; importantly, notice that this theorem allows that the probability space varies with N). Because $Y_{N,i}$, $X_{N,i}$, and $W_{N,i}$ are bounded random variables (uniformly in N), and because $K_{N,i}$ has uniformly bounded moments (see Abadie and Imbens, 2009), it follows that:

$$
\sum_{k=1}^{3N} E[\xi_{N,k}^{2+\delta}] \rightarrow 0 \qquad \text{for some } \delta > 0.
$$

Lindeberg's condition in Billingsley's theorem follows easily from last equation (Lyapounov's condition). As a result, we obtain that under P^{θ_N}

$$C_N \xrightarrow{d} N(0, \sigma_1^2 + \sigma_2^2 + \sigma_3^2),$$

where

$$\sigma_1^2 = \text{plim} \sum_{k=1}^{N} E_{\theta_N}[\xi_{N,k}^2 | \mathcal{F}_{N,k-1}],$$

$$\sigma_2^2 = \text{plim} \sum_{k=N+1}^{2N} E_{\theta_N}[\xi_{N,k}^2 | \mathcal{F}_{N,k-1}],$$

and

$$\sigma_3^2 = \text{plim} \sum_{k=2N+1}^{3N} E_{\theta_N}[\xi_{N,k}^2 | \mathcal{F}_{N,k-1}].$$

After some algebra, we obtain:

$$\begin{aligned}
\sigma_1^2 &= z_1^2 \, E\Big[\big(\bar{\mu}(1, F(X'\theta)) - \bar{\mu}(0, F(X'\theta)) - \tau\big)^2\Big] \\
&+ z_2' E\left[\frac{f^2(X'\theta)}{F(X'\theta)(1 - F(X'\theta))} E[X \mid F(X'\theta)] E[X' \mid F(X'\theta)]\right] z_2.
\end{aligned}$$

Following the calculations in Abadie and Imbens (2006, additional proofs) for the expectation of $(1 + K_{N,i}/M)^2$:

$$\begin{aligned}
\sigma_2^2 &= z_2' E\left[\frac{f^2(X'\theta)}{F(X'\theta)(1 - F(X'\theta))} \text{var}(X \mid F(X'\theta))\right] z_2 \\
&+ z_1^2 E\left[\frac{\text{var}(\mu(1, X)|F(X'\theta))}{F(X'\theta)} + \frac{\text{var}(\mu(0, X)|F(X'\theta))}{1 - F(X'\theta)}\right] \\
&+ z_1^2 \frac{1}{2M} E\left[\left(\frac{1}{F(X'\theta)} - F(X'\theta)\right) \text{var}(\mu(1, X)|F(X'\theta))\right] \\
&+ z_1^2 \frac{1}{2M} E\left[\left(\frac{1}{1 - F(X'\theta)} - (1 - F(X'\theta))\right) \text{var}(\mu(0, X)|F(X'\theta))\right] \\
&+ 2\, z_2' E\left[\text{cov}(X, \mu(W, X)|F(X'\theta), W) \frac{f(X'\theta)}{F(X'\theta)(1 - F(X'\theta))}\right] z_1.
\end{aligned}$$

Here we use the fact that, conditional on the propensity score, X is independent of W. To derive the constant vector of the cross-product notice that:

$$\begin{aligned}
E&\left[\text{cov}\big(X, \mu(X, W)\big|F(X'\theta), W\big) \frac{(W - F(X'\theta))(2W - 1)}{F(X'\theta)(1 - F(X'\theta))} f(X'\theta) \left(1 + \frac{K_N(\theta)}{M}\right)\right] \\
&= E\left[\text{cov}\big(X, \mu(X, 1)\big|F(X'\theta)\big) \frac{f(X'\theta)}{F(X'\theta)} \left(1 + \frac{K_N(\theta)}{M}\right) \Big| W = 1\right] p \\
&+ E\left[\text{cov}\big(X, \mu(X, 0)\big|F(X'\theta)\big) \frac{f(X'\theta)}{1 - F(X'\theta)} \left(1 + \frac{K_N(\theta)}{M}\right) \Big| W = 0\right] (1 - p)
\end{aligned}$$

$$\to E\left[\operatorname{cov}\big(X,\mu(X,1)\big|F(X'\theta)\big)\frac{f(X'\theta)}{F(X'\theta)^2}\Big|W=1\right]p$$
$$+E\left[\operatorname{cov}\big(X,\mu(X,0)\big|F(X'\theta)\big)\frac{f(X'\theta)}{(1-F(X'\theta))^2}\Big|W=0\right](1-p)$$
$$=E\left[\operatorname{cov}(X,\mu(W,X)|F(X'\theta),W)f(X'\theta)\left(\frac{W}{F(X'\theta)^2}+\frac{1-W}{(1-F(X'\theta))^2}\right)\right].$$

Finally,

$$\begin{aligned}\sigma_3^2 &= z_1^2\,E\left[\frac{\operatorname{var}(Y|X,W=1)}{F(X'\theta)}+\frac{\operatorname{var}(Y|X,W=0)}{1-F(X'\theta)}\right]\\ &+ z_1^2\frac{1}{2M}E\left[\left(\frac{1}{F(X'\theta)}-F(X'\theta)\right)\operatorname{var}(Y|X,W=1)\right]\\ &+ z_1^2\frac{1}{2M}E\left[\left(\frac{1}{1-F(X'\theta)}-(1-F(X'\theta))\right)\operatorname{var}(Y|X,W=0)\right].\end{aligned}$$

Notice that for any integrable function $g(F(X'\theta))$:

$$\begin{aligned}&E\Big[g(F(X'\theta))\Big(\operatorname{var}(\mu(w,X)|F(X'\theta))+\operatorname{var}(Y|X,W=w)\Big)\Big]\\ &\quad=E\Big[g(F(X'\theta))\Big(\operatorname{var}(\mu(w,X)|F(X'\theta))+E\Big[\operatorname{var}(Y|X,W=w)\Big|F(X'\theta)\Big]\Big)\Big]\\ &=E\Big[g(F(X'\theta))\Big(\operatorname{var}(\mu(w,X)|F(X'\theta),W=w)+E\Big[\operatorname{var}(Y|X,W=w)\Big|F(X'\theta),W=w\Big]\Big)\Big]\\ &\qquad=E\Big[g(F(X'\theta))\operatorname{var}(Y|F(X'\theta),W=w)\Big].\end{aligned}$$

As a result, under P^{θ_N}:

$$C_N\xrightarrow{d}N(0,z'Vz),$$

where $z=(z_1,z_2')'$, and

$$V=\begin{pmatrix}\sigma^2 & c'\\ c & I_\theta\end{pmatrix},$$

where

$$c=E\left[\operatorname{cov}(X,\mu(W,X)|F(X'\theta),W)f(X'\theta)\left(\frac{W}{F(X'\theta)^2}+\frac{1-W}{(1-F(X'\theta))^2}\right)\right],$$

and σ^2 is the asymptotic variance calculated in Abadie and Imbens for the case of a known propensity score. Applying the Cramer-Wold device, under P^{θ_N}:

$$\begin{pmatrix}D_N(\theta_N)\\ \Delta_N(\theta_N)\end{pmatrix}\xrightarrow{d}N(0,V).$$

□

Proof of Theorem 1: Given our preliminary results, Theorem 1 follows from Andreou and Werker (2005). □

References

ABADIE, A. and IMBENS, G.W. (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects", *Econometrica* , vol. 74, no. 1, 235-267.

ABADIE, A. and IMBENS, G.W. (2008), "On the Failure of the Bootstrap for Matching Estimators", *Econometrica* , vol. 76, no. 6, 1537-1558.

ABADIE, A. and IMBENS, G.W. (2009), "A Martingale Representation for Matching Estimators", NBER working paper, no. 14756.

ANDREOU, E. and WERKER, B.J.M. (2005), "An Alternative Asymptotic Analysis of Residual-Based Statistics", mimeo.

BICKEL, P.J., KLAASSEN, C.A., RITOV, Y. and WELLNER, J.A. (1998), *Efficient and Adaptive Estimation for Semiparametric Models*, Springer, New York.

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

GANSSLER, P. and PFANZAGL, J. (1971), "Convergence of Conditional Expectations", *The Annals of Mathematical Statistics*, vol. 42, no. 1, 315-324.

HIRANO, K., G. IMBENS, and G. RIDDER, (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,"*Econometrica*, 71(4): 1161-1189.

JOHNSON, N. and KOTZ, S. (1977), *Urn Models and Their Applications*, John Wiley & Sons, New York.

LEHMANN, E.L. and ROMANO, J.P. (2005), *Testing Statistical Hypothesis*. Springer, New York.

ROSENBAUM, P. and RUBIN, D.B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70, 4155.

RUBIN, D.B. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies", *Journal of Educational Psychology*, 66, 688-701.

RUBIN, D., and N. THOMAS, (1992a), "Characterizing the effect of matching using linear propensity score methods with normal distributions," *Biometrika* 79 797-809.

RUBIN, D., and N. THOMAS, (1992b), "Affinely Invariant Matching Methods with Ellipsoidal Distributions,"*Annals of Statistics* 20 (2) 1079-1093.

VAN DER VAART, A. (1998), *Asymptotic Statistics*, Cambridge University Press, New York.

VAN DER VAART, A.W. and WELLNER, J.A. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.

Table I – Simulation Results
($N = 5000$, Number of simulations = 10000)

Variances over simulations		Coverage of 95% C.I. (asymp. s.e. = 0.0022)	
$\widehat{\tau}_N(\theta)$	0.0053	$(\widehat{\tau}_N(\theta), \widehat{\sigma}^2_N(\theta))$	0.9532
$\widehat{\tau}_N(\widehat{\theta}_N)$	0.0027	$(\widehat{\tau}_N(\widehat{\theta}_N), \widehat{\sigma}^2_N(\widehat{\theta}_N))$	0.9947
		$(\widehat{\tau}_N(\widehat{\theta}_N), \widehat{\sigma}^2_{\text{adj},N}(\widehat{\theta}_N))$	0.9488
Averages over simulations			
$\widehat{\sigma}^2_N(\theta)$	0.0054		
$\widehat{\sigma}^2_N(\widehat{\theta}_N)$	0.0053		
$\widehat{\sigma}^2_{\text{adj},N}(\widehat{\theta}_N)$	0.0027		