Online Appendix

Limit properties and variance:

A matching estimator based on a bi-level optimization problem

Juan Díaz∗ Tomás Rau† Jorge Rivera‡

September 30, 2014

Abstract

In this document we develop limit properties of a version of the BLOP matching estimator introduced by Díaz, Rau & Rivera (Forthcoming). We show that this estimator is consistent and asymptotically normal, and that the conditional bias is $O_p(N^{-1/K})$, with $K$ being the dimension of the vector of covariates. We introduce a consistent estimator of its marginal variance.

1 Basic concepts and notation

Let $\Omega = (X, Y, W)$ be a random vector where $W \in \{0, 1\}$ indicates whether a treatment was received ($W = 1$) or not ($W = 0$), and $X \in \mathbb{X}$ represents a vector of individual characteristics or covariates with dimension $K \in \mathbb{N}$. For simplicity covariates are assumed to be continuous random variables and $\mathbb{X}$ is the supporting set of these variables.

∗Departamento de Economía, Universidad de Chile. Diagonal Paraguay 257, Santiago, Chile, email: juadiaz@fen.uchile.cl
†Instituto de Economía, Pontificia Universidad Católica de Chile. Vicuña Mackenna 4860, Santiago, Chile, email: trau@uc.cl
‡Departamento de Economía, Universidad de Chile. Diagonal Paraguay 257, Santiago, Chile, email: jrivera@econ.uchile.cl
The observed outcome is $Y = WY(1) + (1 - W)Y(0)$, with $Y(1)$ and $Y(0)$ being the outcomes that the individual would have obtained with or without treatment—the potential outcomes according to Rosenbaum & Rubin (1983).

For $x \in X$ and $w \in \{0, 1\}$, the conditional expectation and the conditional variance of $Y$ are denoted by

$$
\mu_w(x) = \mathbb{E}(Y(w) \mid X = x), \quad \sigma^2_w(x) = \mathbb{V}(Y(w) \mid X = x),
$$

(1)

$$
\mu(x, w) = \mathbb{E}(Y \mid X = x, W = w), \quad \sigma^2(x, w) = \mathbb{V}(Y \mid X = x, W = w),
$$

(2)

thus the average treatment effect, ATE, and the average treatment effect on the treated, ATT, are, respectively

$$
\tau = \mathbb{E}(\mu(x, 1) - \mu(x, 0)), \quad \tau_{tre} = \mathbb{E}(\mu(x, 1) - \mu(x, 0) \mid W = 1).
$$

The assumptions below are needed for limit properties, and they will be part of our standing assumptions throughout this work. These are quite standard in the literature. See Rosenbaum & Rubin (1983), Heckman, Ichimura & Todd (1998), Imbens (2004) and Imbens & Wooldridge (2009) for a detailed discussion on them.

**Assumption 1.** $X$ is a compact and convex subset of $\mathbb{R}^K$ and the distribution of $X$ is bounded away from zero.

**Assumption 2.** Unconfounded Treatment Assignment: $W \perp \perp ((Y(0), Y(1)) \mid X)$.

**Assumption 3.** Overlap: there is $0 < c < 1$ such that $0 < \mathbb{P}(W = 1 \mid X) < 1 - c$.

In view of the standing assumptions, it is easy to see that

$$
\mu(x, w) = \mu_w(x), \quad \sigma^2(x, w) = \sigma^2_w(x).
$$

For $N \in \mathbb{N}$, let $\Omega^N = \{(X_i, Y_i, W_i)\}_{i=1}^N$ be a sample of $\Omega$, and $N_1$ and $N_0$ are the
number of treated and control units, respectively. We note that for each unit $i$ the number of its opposites regarding the treatment is $N_{1-W_i} \in \{N_0, N_1\}$.

By reordering, in what follows we assume that control units are indexed by $1, \ldots, N_0$, thus the treated ones are labeled by $N_0 + 1, \ldots, N_0 + N_1$ ($= N$).

Without loss of generality, in this work we use the Euclidean norm, $\| \cdot \|$, as the matching metric. Given that, using the notation in Abadie & Imbens (2006), for $m \in \mathbb{N}$ we set $j_m(i) \in \{1, \ldots, N\}$ as the index of the first $m$-th nearest neighbor to unit $i \in \{1, \ldots, N\}$ in the opposite treatment group,\footnote{That is, for $m = 1$, $W_{j_1(i)} \neq W_i$ and $\|X_i - X_{j_1(i)}\| = \min_{s \neq W_i} \|X_i - X_s\|$. In a recursive manner, for $1 < m \leq N$, $j_m(i) \in \mathbb{N}$ complies with $W_{j_m(i)} \neq W_i$ and $\|X_i - X_{j_m(i)}\| = \min_{s \neq W_i, s \notin \{j_1(i), \ldots, j_{m-1}(i)\}} \|X_i - X_s\|$.} and the set of covariates for the first $M$ matches for this unit is denoted by $X_M(i)$, that is,\footnote{For definitions we introduce below to make sense, we should assume $M \leq \min\{N_0, N_1\}$.}

$$X_M(i) = \{X_{j_1(i)}, \ldots, X_{j_M(i)}\}.$$ 

For further purposes regarding the study of the variance, we set $j_m^*(i) \in \mathbb{N}$ as the index of the first $m$-nearest neighbor to unit $i$ in the same treatment group, leaving the $i$-th unit out. For integer $M$, we set

$$X_M^*(i) = \{X_{j_1^*(i)}, \ldots, X_{j_M^*(i)}\} \subseteq \{X_j : W_j = W_i, j \neq i\}.$$ 

The “BLOP matching estimator” introduced in Díaz et al. (Forthcoming) rests on the weighting scheme that for a treated unit $i$ solves the following optimization problem:

$$\min_{\lambda \in \arg \min(P_i)} \sum_{j=1}^{N_0} \lambda_j \|X_i - X_j\|$$

s.t. \hspace{1cm} \tag{3} 

$$\lambda \in \arg \min(P_i),$$
where $^3 \argmin(\mathcal{P}_i) \subseteq \Delta_{N_0}$ is the solution set of next optimization problem:

$$\mathcal{P}_i : \min_{(\lambda_j) \in \Delta_{N_0}} \left\| X_i - \sum_{j=1}^{N_0} \lambda_j X_j \right\|. \tag{4}$$

**Remark 1.1.** For the case $i$ is treated unit, denoting by $\text{Proj}(X_i)$ the projection of $X_i$ onto the convex hull of covariates of control units,$^4$ it is clear that problem (3) can be posed equivalently as the following linear optimization program:

$$\min_{(\lambda_1, \ldots, \lambda_{N_0})} \sum_{j=1}^{N_0} \lambda_j \| X_i - X_j \|
\text{subject to}
\begin{align*}
\sum_{j=1}^{N_0} \lambda_j X_j &= \text{Proj}(X_i), \\
\sum_{j=1}^{N_0} \lambda_j &= 1, \quad \lambda_j \geq 0, \quad j = 1, \ldots, N_0. \tag{5}
\end{align*}$$

When $i$ is a control unit, problem (3) –or its equivalent version problem (5)– must be solved using covariates $X_{N_0+1}, \ldots, X_N$, thus the weighting scheme belongs to $\Delta_{N_1}$. By properly configuring optimization problems above in terms of involved covariates, the solution of problem (3), hereafter called the bi-level optimization problem, BLOP in short, is denoted by:

$$\lambda(i) = \left( \lambda_1(i), \ldots, \lambda_{N_1-W_i}(i) \right) \in \Delta_{N_1-W_i}. \tag{6}$$

Given that, by setting

$$\hat{Y}_i^b = W_i \sum_{j=1}^{N_0} \lambda_j(i) Y_j + (1 - W_i) \sum_{j=1}^{N_1} \lambda_j(i) Y_{N_0+j},$$

$^3$We recall the simplex in $\mathbb{R}^m$, denoted $\Delta_m$, is the set of weights $(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$ such that $\lambda_1 \geq 0, \ldots, \lambda_m \geq 0$, and $\lambda_1 + \ldots + \lambda_m = 1$.

$^4$The projection of $X_i$ onto $\text{co}\{X_1, \ldots, X_{N_0}\}$ is, by definition, the nearest vector to $X_i$ belonging to that set.
the *missing potential outcome* imputed to unit \( i \) is

\[
\hat{Y}_i(0) = \begin{cases} Y_i & \text{if } W_i = 0 \\ \hat{Y}_i^b & \text{if } W_i = 1 \end{cases}, \quad \hat{Y}_i(1) = \begin{cases} \hat{Y}_i^b & \text{if } W_i = 0 \\ Y_i & \text{if } W_i = 1, \end{cases}
\]

this leading to the following *BLOP matching estimators* for the ATE and ATE:

\[
\hat{\tau}^b = \frac{1}{N} \sum_{i=1}^{N} \left( \hat{Y}_i(1) - \hat{Y}_i(0) \right), \quad \hat{\tau}_{tre}^b = \frac{1}{N_1} \sum_{i=1}^{N} W_i \left( \hat{Y}_i(1) - \hat{Y}_i(0) \right).
\] (7)

For \( M \) as before, a version of the optimization problem (3) we are concerned is

\[
\min_{\lambda = (\lambda_j)} \sum_{j=1}^{M} \lambda_j \| X_i - X_{j,j(i)} \|
\]

s.t.

\[
\lambda \in \text{argmin}(\mathcal{P}^M_i),
\] (8)

where \( \text{argmin}(\mathcal{P}^M_i) \) is the solution set of next optimization problem:

\[
\mathcal{P}^M_i : \min_{(\lambda_j) \in \Delta_M} \left\| X_i - \sum_{j=1}^{M} \lambda_j X_{j,j(i)} \right\|
\]

The solution of problem (8) is denoted by \( \lambda^M(i) = (\lambda^M_j(i)) \in \Delta_M \), and using these weights we define

\[
\hat{Y}_i^M(0) = \begin{cases} Y_i & \text{if } W_i = 0 \\ \sum_{j=1}^{M} \lambda^M_j(i) Y_{j,j(i)} & \text{if } W_i = 1 \end{cases}, \quad \hat{Y}_i^M(1) = \begin{cases} \sum_{j=1}^{M} \lambda^M_j(i) Y_{j,j(i)} & \text{if } W_i = 0 \\ Y_i & \text{if } W_i = 1. \end{cases}
\]

The limits properties we study later are for next matching estimators (ATE):

\[
\hat{\tau}^b(M) = \frac{1}{N} \sum_{i=1}^{N} \left( \hat{Y}_i^M(1) - \hat{Y}_i^M(0) \right).
\] (9)
To do so, we need to introduce some auxiliary concepts. First, if instead of using covariates in $X_M(i)$ when solving optimization problems above we use covariates that belong to $X_M^*(i)$, then for each unit $i$ the solution of optimization problem (8), properly configured on terms of these covariates, is denoted by

$$\phi^M(i) = (\phi^M_j(i)) \in \Delta_M.$$  

Second, directly extending the $K_M(i)$ concept in Abadie & Imbens (2006)—see pag. 240—to this framework, we denote by $C_i(M)$ the subset of indices such that unit $i$ was used as a match by some counterfactual, that is,

$$C_i(M) = \left\{ j : W_j \neq W_i, \exists s \in \{1, \ldots, M\} \text{ s.t. } j_s(j) = i, \lambda^M_s(j) > 0 \right\},$$

and provided that this subset in nonempty, for integer $\alpha$ we set $c_i^{[\alpha]}(M)$ as the sum of these weights to the power of $\alpha$, and 0 otherwise, i.e.,

$$c_i^{[\alpha]}(M) = \begin{cases} 
0 & \text{if } C_i(M) = \emptyset, \\
\sum_{j \in C_i(M), s : j_s(j) = i} (\lambda^M_s(j))^\alpha & \text{if } C_i(M) \neq \emptyset. 
\end{cases}$$

(11)

For the case we use the overall sample to perform this last indicator (that is, $M = N_0$ for treated units, and $M = N_1$ for controls), for integer $\alpha$ we define

$$c_i^{[\alpha]} = W_i \sum_{j=1}^{N_0} (\lambda_i - N_0(j))^\alpha + (1 - W_i) \sum_{j=1}^{N_1} (\lambda_i(N_0 + j))^\alpha.$$  

(12)

## 2 Bias and normality properties

Next assumptions were used by Abadie & Imbens (2006) with similar purposes than here. Along with Assumptions 1 – 3, they will be our standing assumptions.

**Assumption 4.** For each $w \in \{0, 1\}$, both $\mu(\cdot, w)$ and $\sigma^2(\cdot, w)$ are Lipschitzian on $\mathbb{X}$.

**Assumption 5.** For each $w \in \{0, 1\}$, the fourth-moment of $Y(w)$ are uniformly bounded on $\mathbb{X}$, and $\sigma_w^2(\cdot, \cdot)$ is bounded away from zero.
Assumption 6. \( \{(X_i, Y_i, W_i)\}_{i=1}^{N}, \ N \in \mathbb{N}, \) are independent draws from the distribution of \( \Omega = (X, Y, W). \)

The main result of this part is Proposition 2.1, whose proof is prepared through some auxiliary lemmata.\(^5\) It is worthless mentioning that restricting the variance and bias to the treated sub-sample, this proposition holds true for the on the treated estimators as well. In the following, \( M \) is a given integer that remains constant as the sample size increases.

Lemma 2.1. Under standing assumptions, for each \( i \in \{1, \ldots, N\}, \ c_i^{[1]}(M) = O_p(1), \) and for each integer \( \alpha, \ E\left(\left(c_i^{[1]}(M)\right)^\alpha\right) \) is uniformly bounded in \( N \in \mathbb{N}. \)

Proof. Let \( \rho_i(N) \in \mathbb{N} \) be the times that unit \( i \) is used as one of the \( M \) nearest neighbors for some individual in its counterfactual set. Thus, by definition, \( c_i^{[1]}(M) \leq \rho_i(N), \) which implies for each \( \alpha > 0, \ E\left(\left(c_i^{[1]}(M)\right)^\alpha\right) \leq E((\rho_i(N))^\alpha). \) From Lemma 3 in Abadie & Imbens (2006), there is \( \gamma_\alpha > 0 \) such that \( E((\rho_i(N))^\alpha) \leq \gamma_\alpha, \ N \in \mathbb{N}, \) which directly implies the result.

After some algebra, we can show that, conditional on \( \{(W_i, X_i)\}_{i=1}^{N}, \) the bias of \( \hat{\tau}^b(M) \) is

\[ B(M) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} \Lambda_j^M(i) (2W_i - 1) \left( \mu_{1-W_i}(X_i) - \mu_{1-W_i}(X_{j(i)}) \right), \quad (13) \]

and performing some simple calculus, we also can show that \( \hat{\tau}^b(M) - \tau = A + E(M) + B(M), \) where

\[ A = \frac{1}{N} \sum_{i=1}^{N} (\mu_1(X_i) - \mu_0(X_i)) - \tau, \quad E(M) = \frac{1}{N} \sum_{i=1}^{N} (2W_i - 1) \left( (1 + c_i^{[1]}(M)) \right) \epsilon_i, \]

with \( \epsilon_i = Y_i - \mu_{W_i}(X_i), \ i = 1, \ldots, N. \)

\(^{5}\)For a sequence of random variables, convergence in probability is noted by \( \xrightarrow{\mathbb{P}} \), whereas convergence in distribution by \( \xrightarrow{\mathbb{D}}. \)
Lemma 2.2. Order of conditional bias

Under standing assumptions, $B(M) = O_p(N^{-1/K})$.

Proof. For $N \in \mathbb{N}$, $\alpha > 0$, $i = 1, \ldots, N$ and $j \in \{1, \ldots, M\}$, from Assumption 1 and Lemma 2 in Abadie & Imbens (2006), $\mathbb{E}\left( N_{1-W_i}^\alpha \|X_i - X_{j(i)}\|^\alpha \right)$ is uniformly bounded in $N_{1-W_i}$ (constant $\theta > 0$, which applies for both, $N_0$ and $N_1$). On the other hand, from Assumption 4, there exists $L > 0$ such that for $w \in \{0, 1\}$,

$$|\mu_w(X_i) - \mu_w(X_{j(i)})| \leq L \|X_i - X_{j(i)}\|. \quad (14)$$

For $i$, $j$ as before, we set

$$b_{ij} = W_i \left( \mu_0(X_i) - \mu_0(X_{j(i)}) \right) - (1 - W_i) \left( \mu_1(X_i) - \mu_1(X_{j(i)}) \right). \quad (15)$$

Cauchy-Schwarz inequality, relation in (13) and the fact that $\sum_{j=1}^M (\lambda_j^M(i))^2 \leq 1$, give

$$\mathbb{E}\left( N^{2/K} (B(M))^2 \right) = N^{2/K-2} \mathbb{E} \left( \sum_{i=1}^N \sum_{j=1}^M \lambda_j^M(i) b_{ij} \right)^2 \leq N^{2/K-1} \sum_{i=1}^N \mathbb{E} \left( \sum_{j=1}^M (b_{ij})^2 \right).$$

Expanding terms from last inequality and using (14) and (15), we obtain

$$\mathbb{E}\left( N^{2/K} (B(M))^2 \right) \leq L^2 N^{2/K-1} (\Psi_0 + \Psi_1), \quad (16)$$

with

$$\Psi_w = \mathbb{E} \left( \frac{1}{N_w^{2/K}} \sum_{i \in I_w^N} \sum_{j=1}^M \mathbb{E}(N_w^{2/K} \|X_i - X_{j(i)}\|^2 | \{(W_i, X_i)\}_{i=1}^N) \right),$$

and $I_w^N = \{ s \in \{1, \ldots, N\} : W_s = w \}$, for $w = 0, 1$. Denoting $\psi = \theta L^2 M$, from (16) we can readily show

$$\mathbb{E}\left( N^{2/K} (B(M))^2 \right) \leq \psi \left( \mathbb{E} \left( \left( \frac{N}{N_0} \right)^{2/K} \frac{N_1}{N} \right) + \mathbb{E} \left( \left( \frac{N}{N_1} \right)^{2/K} \frac{N_0}{N} \right) \right),$$

and using the well known Chernoff’s and Markov’s inequalities we end the proof. \Box
In the following, the normalized conditional variance of $\hat{\tau}_b(M)$ is denoted by

$$V_1(M) = N \mathbb{V} \left( \hat{\tau}_b(M) \mid \{(W_i, X_i)\}_{i=1}^N \right), \quad (17)$$

and we also denote

$$V = \mathbb{E} \left( (\mu_1(X) - \mu_0(X) - \tau)^2 \right). \quad (18)$$

Lemma 2.3. Under standing assumptions, $\mathbb{E}(V_1(M)) = O(1)$.

Proof. After some direct calculations, we can show

$$\mathbb{E}(V_1(M)) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left( \left( 1 + c_i^{[1]}(M) \right)^2 \sigma^2(X_i, W_i) \right). \quad (19)$$

From Assumptions 1 and 4, and properties of continuous mapping onto compact sets, there exists $\xi > 0$ such that $\sigma(x, w) \leq \xi$, $(x, w) \in \mathbb{X} \times \{0, 1\}$, which implies

$$\mathbb{E}(V_1(M)) \leq \xi^2 \mathbb{E} \left( \left( 1 + c_i^{[1]}(M) \right)^2 \right). \quad (20)$$

Using (20) and Lemma 2.1 we conclude the proof.

Remark 2.1. From last property, it holds that $\mathbb{E}(V_1(M))$ is uniformly bounded in $N \in \mathbb{N}$. Therefore, passing to a subsequence if necessary, we can assume $V_1(M)$ converges in probability, let us say, to $V^*$.

Proposition 2.1. Normality properties

Under standing assumptions, $\hat{\tau}_b(M) \xrightarrow{p} \tau$ and

$$\frac{\sqrt{N} \left( \hat{\tau}_b(M) - \tau - B(M) \right)}{\sqrt{V_1(M) + V}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Proof. From the standard law of large numbers, we already know

$$\frac{1}{N} \left( \sum_{i=1}^{N} (\mu_1(X_i) - \mu_0(X_i)) \right) - \tau \xrightarrow{p} 0,$$
and from definition of $E(M)$,

$$
E \left( N \left( E(M) \right)^2 \right) = \frac{1}{N} \sum_{i=1}^{N} E \left( \left( 1 + c_i^{[1]}(M) \right)^2 \epsilon_i^2 \right) = E \left( \left( 1 + c_i^{[1]}(M) \right)^2 \sigma^2(X_i, W_i) \right).
$$

Lemma 2.1 implies $E[N E(M)^2] = O(1)$, and using Markov’s inequality, and the order of convergence of $B(M)$, we can readily conclude the proof of consistency.

In order to show the normality property, from Lemma 2.3 we have $V_1(M)$ is bounded in $N$ and from Assumptions 1 and 4 this holds for $V$. On the other hand, from the fact that

$$
\sqrt{N} \left( \hat{\tau}^b(M) - \tau - B(M) \right) = \sqrt{N} A + \sqrt{N} E(M),
$$

the *Standard Central Limit Theorem*, and properties of $E(M)$, we can readily show

$$
\sqrt{N} A \xrightarrow{D} \mathcal{N}(0, V).
$$

Finally, from the *Linderberg-Feller Central Limit Theorem*, Lemma 2.1 and by following the same argumentation provided by Abadie & Imbens (2006) when proving their Theorem 4, we have that

$$
\frac{\sqrt{N} E(M)}{\sqrt{V_1(M)}} \xrightarrow{D} \mathcal{N}(0, 1).
$$

Due to (21) and (22) are asymptotically independent, the proof is done. \qed

### 3 Variance properties

Using weights in (10), directly borrowed from Abadie & Imbens (2006) we propose the next estimator of $\sigma^2(X_i, W_i) = \mathbb{V}(Y \mid X = X_i, W = W_i)$:

---

6This theorem remains valid conditional on $\{(W_i, X_i)\}_{i=1}^{N}$, which is relevant for our case.
\[
\hat{\sigma}^2_i(M) = \frac{1}{1 + \sum_{j=1}^{M} (\phi_j^M(i))^2} \left( Y_i - \sum_{j=1}^{M} \phi_j^M(i) Y_{j(i)} \right)^2.
\]

(23)

In order to estimate \(V\) –see (18)– we take advantage of the following approximation:

\[
\mathbb{E} \left( \hat{Y}_i^M(1) - \hat{Y}_i^M(0) - \tau \right)^2 \approx V + \mathbb{E} \left( \epsilon_i^2 + \sum_{j=1}^{M} (\lambda_j^M(i))^2 \epsilon_{j(i)}^2 \right),
\]

where \(\epsilon_i = Y_i - \mu_{W_i}(X_i)\), from which we have the estimator of \(V\) we propose is

\[
\hat{V}(M) = \frac{1}{N} \sum_{i=1}^{N} \left[ \left( \hat{Y}_i^M(1) - \hat{Y}_i^M(0) - \hat{\tau}^b(M) \right)^2 - \left( 1 + c_i^M(M) \right) \hat{\sigma}^2_i(M) \right].
\]

(24)

On the other hand, using (19), the estimator of \(V_1(M)\) we propose is

\[
\hat{V}_1(M) = \frac{1}{N} \sum_{i=1}^{N} \left( 1 + c_i^{[1]}(M) \right)^2 \hat{\sigma}^2_i(M).
\]

(25)

**Proposition 3.1.** Under standing assumptions,

(a) \(\hat{V}_1(M) \xrightarrow{P} V^*\),

(b) \(\hat{V}(M) \xrightarrow{P} V\).

**Proof.** Because \(|\hat{V}_1(M) - V_1(M)| \leq |V_1(M) - \Psi_0| + |\Psi_0 - \hat{V}_1(M)|\), the proof mainly concerns the order of terms in the right side of this inequality. Denoting \(\Psi_0 = \mathbb{E} \left( \hat{\sigma}^2_i(M) \mid \{(W_i, X_i)\}_{i=1}^{N} \right)\), it is not difficult to see

\[
\Psi_0 = \frac{\Psi_1 + \Psi_2}{1 + \sum_{j=1}^{M} (\phi_j^M(i))^2},
\]

with

\[
\Psi_1 = \left( \mu_{W_i}(X_i) - \sum_{j=1}^{M} \phi_j^M(i) \mu_{W_i}(X_{j(i)}) \right)^2,
\]

\[
\Psi_2 = \left( \frac{1}{N} \sum_{i=1}^{N} \left( 1 + c_i^{[1]}(M) \right)^2 \hat{\sigma}^2_i(M) \right).
\]
\[ \Psi_2(i) = \sigma^2(X_i, W_i) + \sum_{j=1}^{M} (\phi_j^M(i))^2 \sigma^2(X_{j(i)}, W_i). \]

On the other hand, after some algebra we can show that

\[ \hat{\sigma}_i^2(M) = \frac{1}{1 + \sum_{j=1}^{M} (\phi_j^M(i))^2} \left( T_1 + T_2 + T_3 \right), \]

where

\[ T_1 = \left( \mu_{W_i}(X_i) - \sum_{j=1}^{M} \phi_j^M(i) \mu_{W_i}(X_{j(i)}) \right)^2, \quad T_2 = \left( \epsilon_i - \sum_{j=1}^{M} \phi_j^M(i) \epsilon_{j(i)} \right)^2, \]

and

\[ T_3 = 2 \left( \mu_{W_i}(X_i) - \sum_{j=1}^{M} \phi_j^M(i) \mu_{W_i}(X_{j(i)}) \right) \left( \epsilon_i - \sum_{j=1}^{M} \phi_j^M(i) \epsilon_{j(i)} \right). \]

Therefore, from Lemma 2.1, Assumptions 4 and 5, and Lemma 2 in Abadie & Imbens (2006), we can readily conclude \(|V_1(M) - \Psi_0| = o(1)\). We note Assumptions 1 and 4 implies this order does not depend on the unit \(i\) in \(\Psi_0\).

For the order of \(|\hat{V}_1(M) - \Psi_0|\), following Abadie & Imbens (2006) and performing some calculus, we can show

\[ |\Psi_0 - \hat{V}_1(M)| = \left| \frac{1}{N} \sum_{i=1}^{N} \frac{(1 + c_i^{[1]}(M))^2}{1 + \sum_{j=1}^{M} (\phi_j^M(i))^2} \left( \sum_{v=1}^{6} A_v \right) \right|, \]

with,

\[ A_1 = \sigma^2(X_i, W_i) - \epsilon_i^2, \quad A_2 = \sum_{j=1}^{M} \left( (\phi_j^M(i))^2 \left( \epsilon_{j(i)}^2 - \sigma^2(X_{j(i)}, W_i) \right) \right), \]

\[ A_3 = 2 \epsilon_i \sum_{j=1}^{M} \phi_j^M(i) \epsilon_{j(i)}, \quad A_4 = 2 \epsilon_i \left( \mu_{W_i}(X_i) - \sum_{j=1}^{M} \phi_j^M(i) \mu_{W_i}(X_{j(i)}) \right), \]

12
\[ A_5 = 2 \left( \sum_{j=1}^{M} \phi_j^M(i) \epsilon_{j(i)} \right) \left( \mu_{W_i}(X_i) - \sum_{j=1}^{M} \phi_j^M(i) \mu_{W_i}(X_{j(i)}) \right), \]

and

\[ A_6 = 2 \sum_{j=1}^{M} \left( \phi_j^M(i) \epsilon_{j(i)} \left( \sum_{k>j}^{M} \phi_k^M(i) \epsilon_{k(i)} \right) \right). \]

Using the same argument provided by these authors when proving their Lemma (A.5), and using Lemma 2.1 above, we can conclude for each \( s = 1, \ldots, 6, \)

\[
\left| \frac{1}{N} \sum_{i=1}^{N} \left( 1 + c_i^{[1]}(M) \right)^2 A_s \right| = o(1),
\]

which along with Remark 2.1 yields part (a) of the proposition. Finally, in order to show part (b), from part (a) we already have

\[
\left| \frac{1}{N} \sum_{i=1}^{N} \left( 1 + c_i^{[2]}(M) \right) \left( \sigma_i^2(X_i, W_i) - \widehat{\sigma}_i^2(M) \right) \right| = o_p(1),
\]

and then, we conclude using Theorems 6 and 7 in Abadie & Imbens (2006).

### 3.1 Estimating the marginal variance

We propose an estimator of the marginal variance of \( \hat{\tau}^b \) to be employed when evaluating this estimator using the entire sample \( \{(W_i, X_i, Y_i)\}_{i=1}^{N} \), instead of using \( M \) units as before.

In the following, we assume \( N_0 \geq 2 \) and \( N_1 \geq 2 \) and we recall control units are indexed by \( 1, \ldots, N_0, \) while the treated ones are labeled by \( N_0 + 1, \ldots, N_0 + N_1. \)

When \( i \) is a control unit, extending \( \phi^M(i) \) –see (10)– to the case we are now concerned, we need the weighting scheme that serves to perform the projection of \( X_i \) onto the convex hull of covariates of units having the same treatment than unit \( i, \) leaving \( X_i \) out, namely co\{\( X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{N_0} \). This vector of weights is denoted by \( \phi(i) = (\phi_1(i), \ldots, \phi_{N_0}(i)) \), which can be assumed belongs to the simplex of dimension
$N_0$, complying the condition $\phi_i(i) = 0$ –vector $X_i$ does not participates in the convex combination, otherwise the solution is trivial–. Denoting that projection by $\text{Proj}^*(X_i)$, from (5) we have $\phi(i)$ solves the next optimization problem:

$$
\min_{(\lambda_1, \ldots, \lambda_{N_0})} \sum_{j=1}^{N_0} \lambda_j \|X_i - X_j\|
$$

s.t.

$$
\sum_{j=1}^{N_0} \lambda_j X_j = \text{Proj}^*(X_i),
$$

$$
\sum_{j=1}^{N_0} \lambda_j = 1, \quad \lambda_j \geq 0, \quad j = 1, \ldots, N_0,
$$

$$
\lambda_i = 0.
$$

In a similar manner, by properly configuring last problem in terms of involved covariates, the solution when $i$ is a treated unit is denoted by $\phi(i) = (\phi_{N_0+1}(i), \ldots, \phi_{N_0+N_1}(i)) \in \Delta_{N_1}$, with $\phi_i(i) = 0$. Given that, we define

$$
\Phi(i) = \|\phi(i)\|^2 = (1 - W_i) \sum_{j=1}^{N_0} (\phi_j(i))^2 + W_i \sum_{j=1}^{N_1} (\phi_{N_0+j}(i))^2,
$$

which are used to perform the following estimator of $\sigma^2(X_i, W_i)$ –see (23)–

$$
\hat{\sigma}^2(i) = \frac{(1 - W_i)}{1 + \Phi(i)} \left( Y_i - \sum_{j=1}^{N_0} \phi_j(i) Y_j \right)^2 + \frac{W_i}{1 + \Phi(i)} \left( Y_i - \sum_{j=1}^{N_1} \phi_j(i) Y_{N_0+j} \right)^2.
$$

Hence, from (24) and (25), the variance of $\hat{\tau}^b$ we propose when using the overall sample is:

$$
\hat{V}(\hat{\tau}^b) = \frac{1}{N^2} \sum_{i=1}^{N} \left( \left( \hat{Y}_i(1) - \hat{Y}_i(0) - \hat{\tau}^b \right)^2 + \left[ (1 + c_i^{[1]})^2 - (1 + c_i^{[2]}) \right] \hat{\sigma}_i^2 \right),
$$

with $c_i^{[\alpha]}$ from (12). Finally, after some simple manipulation, it is straightforward to
find an estimator of the variance for $\hat{\tau}_{tre}$ as follows

$$\hat{V}(\hat{\tau}_{tre}) = \frac{1}{N^2} \sum_{i=1}^{N} \left( W_i \left( \hat{Y}_i(1) - \hat{Y}_i(0) - \hat{\tau}_{tre} \right)^2 + (1 - W_i) \left( \left( c_{i1}^{[1]} - c_{i2}^{[2]} \right) \hat{\sigma}_i^2 \right) \right).$$

References


