Supplementary Appendix for "Robust Post-Matching Inference"

Alberto Abadie and Jann Spiess

January 2019

Convergence of Matched Sums

In this section, we derive general convergence results for sums within the matched sample $\mathcal{S}^*$ that we will later use to establish consistency and asymptotic normality of the various estimators in this article. The main tool behind these lemmas is a martingale representation similar to Abadie and Imbens (2012).

Let $F(Y, W, S)$ be a $(s \times t)$-matrix of real-valued measurable functions,

$$\Phi_1(x) = E[F(Y, W, S)|W = 1, X = x], \qquad \Phi_0(x) = E[F(Y, W, S)|W = 0, X = x],$$

$$\widehat{\Phi} = \frac{1}{n}\sum_{i=1}^{n} F(Y_{ni}, W_{ni}, S_{ni}),$$

and

$$\Phi = E^*[F(Y, T, S)] = E\left[\frac{1}{M+1}\Phi_1(X) + \frac{M}{M+1}\Phi_0(X)\Big|W = 1\right].$$

Lemma A.1. *Under Assumptions 1 to 3, and if*

(a.1) $\Phi_0(\cdot)$ *is (component-wise) Lipschitz on* $\mathcal{X}_0$*,*

(a.2) $E[\|F(Y, W, S)\|^2|W = w, X = x]$ *is uniformly bounded on* $\mathcal{X}_w$*, for* $w \in \{0, 1\}$*,*

then $\widehat{\Phi} \xrightarrow{p} \Phi$.

Proof: Because $\widehat{\Phi}$ converges in probability if and only if each of its components converges, we assume without loss of generality that $s = t = 1$. We decompose

$$\widehat{\Phi} = \frac{1}{n}\sum_{i=1}^{N} W_i\Big(\Phi_1(X_i) + M\Phi_0(X_i)\Big) + \frac{1}{n}\sum_{i=1}^{n}\Big(F(Y_{ni}, W_{ni}, S_{ni}) - \Phi_{W_{ni}}(X_{ni})\Big)$$
$$+ \frac{1}{n}\sum_{i=1}^{N} W_i \sum_{j \in \mathcal{J}(i)}\Big(\Phi_0(X_j) - \Phi_0(X_i)\Big).$$

The first term on the right-hand side of the last equation is a sum of i.i.d. random variables. Hence, by the weak law of large numbers, we have that

$$\frac{1}{n}\sum_{i=1}^{n} W_i\Big(\Phi_1(X_i) + M\Phi_0(X_i)\Big) \xrightarrow{p} E\left[\frac{1}{M+1}\Phi_1(X) + \frac{M}{M+1}\Phi_0(X)\Big|W=1\right] = \Phi.$$

For the second sum, notice that,

$$\begin{aligned}&\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n}\Big(F(Y_i,W_i,S_i) - \Phi_{W_i}(X_i)\Big)\middle|\begin{array}{l}X_1,\dots,X_N,\\ W_1,\dots,W_N\end{array}\right)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}(F(Y,W,S)|W=W_i, X=X_i)\end{aligned}$$

which (by Assumption *(a.2)* in the lemma) is bounded by a sequence that converges to zero. By the law of total variance, we obtain

$$\mathrm{var}\left(\frac{1}{n}\sum_{i=1}^{n}\Big(F(Y_i,W_i,S_i) - \Phi_{W_i}(X_i)\Big)\right) \longrightarrow 0.$$

For the third sum, Assumption *(a.1)* in the lemma implies

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} W_i \sum_{j\in\mathcal{J}(i)}\Big(\Phi_0(X_j) - \Phi_0(X_i)\Big)\right| &\le \frac{1}{n}\sum_{i=1}^{n} W_i\sum_{j\in\mathcal{J}(i)}\Big|\Phi_0(X_j)-\Phi_0(X_i)\Big|\\ &\le \frac{L}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i\sum_{j\in\mathcal{J}(i)} d(X_j,X_i)\right) \xrightarrow{p} 0,\end{aligned}$$

for some Lipschitz constant L. □

Lemma A.2. *In the setup of Lemma A.1, let* $t=1$*, and define*

$$\Psi_1(x) = \mathrm{var}(F(Y,W,S)|W=1, X=x), \quad \Psi_0(x) = \mathrm{var}(F(Y,W,S)|W=0, X=x),$$

which are $(s\times s)$*-matrices. Suppose that, in addition to the assumptions of Lemma A.1, we have*

(a.3) $\Psi_0(\cdot)$ *is (component-wise) Lipschitz on* $\mathcal{X}_0$*,*

(a.4) $E[\|F(Y,W,S)\|^{2+\delta}|W=w, X=x]$ *is uniformly bounded on* $\mathcal{X}_w$ *for all* $w\in\{0,1\}$ *and some* $\delta>0$*.*

Then,

$$\sqrt{n}(\widehat{\Phi} - \Phi) \xrightarrow{d} \mathcal{N}(0, V^*)$$

where

$$V^* = \frac{\operatorname{var}\Big(\Phi_1(X) + M\Phi_0(X)\big|W=1\Big)}{M+1} + \frac{E\Big[\Psi_1(X) + M\Psi_0(X)\big|W=1\Big]}{M+1}.$$

Proof: Fix $\lambda \in \mathbb{R}^s$. We decompose

$$\begin{aligned}
&\sqrt{n}(\widehat{\Phi} - \Phi)'\lambda \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i\Big(\Phi_1(X_i) + M\Phi_0(X_i) - \Phi(X_i)\Big)'\lambda + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(F(Y_i, W_i, S_i) - \Phi_{W_i}(X_i)\Big)'\lambda \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i \sum_{j\in\mathcal{J}(i)}\Big(\Phi_0(X_j) - \Phi_0(X_i)\Big)'\lambda.
\end{aligned}$$

The last term on the right-hand side of last equation vanishes in probability:

$$\begin{aligned}
\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i \sum_{j\in\mathcal{J}(i)}\Big(\Phi_0(X_j) - \Phi_0(X_i)\Big)'\lambda\right| &\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i \sum_{j\in\mathcal{J}(i)}\Big\|\Phi_0(X_j) - \Phi_0(X_i))\Big\|\|\lambda\| \\
&\leq \frac{\|\lambda\| L}{\sqrt{n}}\sum_{i=1}^{n} W_i \sum_{j\in\mathcal{J}(i)} d(X_j, X_i) \xrightarrow{p} 0
\end{aligned}$$

for an appropriate Lipschitz constant L.

The first two parts of the sum form a martingale. Consider the filtration

$$\mathcal{F}_i = \begin{cases} \sigma(W_1, \ldots, W_N, X_1, \ldots, X_i), & i \leq N_1, \\ \sigma(W_1, \ldots, W_N, X_1, \ldots, X_N, (Y_1, S_1), \ldots, (Y_{i-N}, S_{i-N})), & N_1{+}1 \leq i \leq N_1{+}n. \end{cases}$$

Then,

$$\xi_i = \begin{cases} \dfrac{1}{\sqrt{n}} W_i\Big(\Phi_1(X_i) + M\Phi_0(X_i) - \Phi\Big)'\lambda, & i \leq N_1, \\ \dfrac{1}{\sqrt{n}}\Big(F(Y_{i-N}, W_{i-N}, S_{i-N}) - \Phi_{W_{i-N}}(X_{i-N})\Big)'\lambda, & N_1 + 1 \leq i \leq N_1 + n \end{cases}$$

is a martingale difference array with respect to the filtration $\mathcal{F}$. Also, notice that

$$\begin{aligned}\sum_{i=1}^{N_1+n} E[\xi_i^2|\mathcal{F}_{i-1}] &= \frac{1}{n}\sum_{i=1}^{N_1} \mathrm{var}\Big((\Phi_1(X)+M\Phi_0(X))'\lambda\Big|W=1\Big) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \mathrm{var}\Big(F(Y,W,S)'\lambda\Big|W=W_i, X=X_i\Big) \\ &= \frac{\lambda'\mathrm{var}(\Phi_1(X)+M\Phi_0(X)|W=1)\lambda}{1+M} + \frac{1}{n}\sum_{i=1}^{n}\lambda'\Psi_{W_i}(X_i)\lambda,\end{aligned}$$

where the last term converges in probability to

$$\lambda' E\left[\frac{1}{M+1}\Psi_1(X) + \frac{M}{M+1}\Psi_0(X)\Big|W=1\right]\lambda,$$

by Lemma A.1. Hence,

$$\sum_{i=1}^{N_1+n} E[\xi_i^2|\mathcal{F}_{i-1}] \xrightarrow{p} \lambda' V^* \lambda.$$

Next, note that

$$\begin{aligned}|\xi_i| &\le \begin{cases} \frac{1}{\sqrt{n}}\|\Phi_1(X_i)+M\Phi_0(X_i)-\Phi\|_2\|\lambda\|_2, & i\le N_1, \\ \frac{1}{\sqrt{n}}\|F(Y_{i-N},W_{i-N},S_{i-N})-\Phi_{W_{i-N}}(X_{i-N})\|_2\|\lambda\|_2, & N_1+1\le i\le N_1+n \end{cases} \\ &\le \begin{cases} \frac{1}{\sqrt{n}}(\|\Phi_1(X_i)\|_2+M\|\Phi_0(X_i)\|_2+\|\Phi\|_2)\|\lambda\|_2, & i\le N_1, \\ \frac{1}{\sqrt{n}}(\|F(Y_{i-N},W_{i-N},S_{i-N})\|_2+\|\Phi_{W_{i-N}}(X_{i-N})\|_2)\|\lambda\|_2, & N_1+1\le i\le N_1+n \end{cases}\end{aligned}$$

by the Cauchy-Schwarz and triangle inequalities. It follows that

$$\begin{aligned}E[|\xi_i|^{2+\delta}] &\le \begin{cases} \frac{\|\lambda\|_2^{2+\delta}}{n^{1+\delta/2}}E\left[(\|\Phi_1(X_i)\|_2+M\|\Phi_0(X_i)\|_2+\|\Phi\|_2)^{2+\delta}\right], & i\le N_1, \\ \frac{\|\lambda\|_2^{2+\delta}}{n^{1+\delta/2}}E\left[(\|F(Y_{i-N},W_{i-N},S_{i-N})\|_2+\|\Phi_{W_{i-N}}(X_{i-N})\|_2)^{2+\delta}\right], & i>N_1 \end{cases} \\ &\le \begin{cases} \frac{\|\lambda\|_2^{2+\delta}}{n^{1+\delta/2}}\left((E[\|\Phi_1(X_i)\|_2^{2+\delta}])^{1/(2+\delta)}+M(E[\|\Phi_0(X_i)\|_2^{2+\delta}])^{1/(2+\delta)}+(\|\Phi\|_2^{2+\delta})^{1/(2+\delta)}\right)^{2+\delta} \\ \frac{\|\lambda\|_2^{2+\delta}}{n^{1+\delta/2}}\left((E[\|F(Y_{i-N},W_{i-N},S_{i-N})\|_2^{2+\delta}])^{1/(2+\delta)}+(E[\|\Phi_{W_{i-N}}(X_{i-N})\|_2^{2+\delta}])^{1/(2+\delta)}\right)^{2+\delta} \end{cases}\end{aligned}$$

where the latter inequality is implied by Minkowski's inequality. By assumption (a.3), note that for both $w\in\{0,1\}$ and $x\in\mathcal{X}_w$, by Jensen's inequality we have

$$\begin{aligned}\|\Phi_w(x)\|_2^{2+\delta} &= \|E[F(Y,W,S)|W=w,X=x]\|_2^{2+\delta}) \\ &\le E[\|F(Y,W,S)\|_2^{2+\delta}|W=w,X=x]\le C\end{aligned}$$

and hence $E[\|\Phi_w(X)\|_2^{2+\delta}] \leq C$, while also

$$\begin{aligned}
\|\Phi\|_2^{2+\delta} &= \left\| E\left[\frac{1}{M+1}\Phi_1(X) + \frac{M}{M+1}\Phi_0(X)\Big|W=1\right]\right\|_2^{2+\delta} \\
&\leq E\left[\left\|\frac{1}{M+1}\Phi_1(X) + \frac{M}{M+1}\Phi_0(X)\right\|_2^{2+\delta}\Big|W=1\right] \\
&\leq E\left[\left(\frac{1}{M+1}C^{1/(2+\delta)} + \frac{M}{M+1}C^{1/(2+\delta)}\right)^{2+\delta}\Big|W=1\right] \leq C
\end{aligned}$$

and

$$E[\|F(Y_{i-N}, W_{i-N}, S_{i-N})\|_2^{2+\delta}] = E[E[\| \ F(Y_{i-N}, W_{i-N}, S_{i-N})\|_2^{2+\delta}|W_{i-N}, X_{i-N}]] \leq C$$

for some uniform constant C. Hence,

$$\begin{aligned}
E[|\xi_i|^{2+\delta}] &\leq \begin{cases} \frac{\|\lambda\|_2^{2+\delta}}{n^{1+\delta/2}}(M+2)^{2+\delta}C, & i \leq N_1, \\ 2^{2+\delta}C, & N_1+1 \leq i \leq N_1+n \end{cases} \\
&\leq \frac{\|\lambda\|_2^{2+\delta}}{n^{1+\delta/2}}(M+2)^{2+\delta}C,
\end{aligned}$$

from which we obtain Lyapounov's condition, namely that

$$\sum_{i=1}^{N_1+n} E[|\xi_i|^{2+\delta}] \leq \frac{N_1+n}{n}\frac{\|\lambda\|_2^{2+\delta}(M+2)^{2+\delta}C}{n^{\delta/2}} \to 0.$$

Hence, by the Lindeberg–Feller Martingale Central Limit Theorem,

$$\sqrt{n}(\widehat{\Phi} - \Phi)'\lambda = \sum_{i=1}^{N_1+n} \xi_i + o_P(1) \xrightarrow{d} \mathcal{N}(0, \lambda' V^* \lambda).$$

The assertion of the lemma follows now from the Cramér-Wold device. □

The Matched Bootstrap

In this section, we develop a general result for the coupled resampling of martingale increments that we then apply to the matched bootstrap.

Proposition A.1. *Let $\lambda \geq 1$ be fixed. Assume we have a collated martingale difference array*

$$\{\zeta_{n1}^{(1)}, \ldots, \zeta_{nn}^{(1)}, \zeta_{n1}^{(2)}, \ldots, \zeta_{nn}^{(2)}, \ldots, \zeta_{n1}^{(\lambda)}, \ldots, \zeta_{nn}^{(\lambda)}\}, n \geq 1,$$

with respect to the filtration array

$$\{\mathcal{F}_{n1}^{(1)}, \ldots, \mathcal{F}_{nn}^{(1)}, \mathcal{F}_{n1}^{(2)}, \ldots, \mathcal{F}_{nn}^{(2)}, \ldots, \mathcal{F}_{n1}^{(\lambda)}, \ldots, \mathcal{F}_{nn}^{(\lambda)}\}, n \geq 1,$$

and the following properties:

1. *For all* $\ell \in \{1, \ldots, \lambda\}$,
$$\sum_{i=1}^{n} E[(\zeta_{ni}^{(\ell)})^2 | \mathcal{F}_{n(i-1)}^{(\ell)}] \xrightarrow{p} \sigma_\ell^2, \tag{S.1}$$
where $\mathcal{F}_{n,0}^{(\ell+1)} = \mathcal{F}_{nn}^{(\ell)}$ *for all* $\ell \in \{1, \ldots, \lambda - 1\}$.

2. *There exist some* $C > 0$ *and* $\delta > 0$ *such that for all* $ni\ell$,
$$E[(\zeta_{ni}^{(\ell)})^4] \leq \frac{C}{n^{1+\delta}}. \tag{S.2}$$

Consider the sum of increments

$$S_n = \sum_{\ell=1}^{\lambda} \sum_{i=1}^{n} \zeta_{ni}^{(\ell)}$$

and the bootstrapped sum of coupled increments

$$T_n = \sum_{\ell=1}^{\lambda} \sum_{i=1}^{n} (V_{ni}^{(\ell)} - 1) \zeta_{ni}^{(\ell)},$$

where $(V_{n1}^{(\lambda)}, \ldots, V_{nn}^{(\lambda)})$ *is multinomially distributed with parameters* $(n; n^{-1}, \ldots, n^{-1})$ *independent of the data, and*

$$V_{n\iota_n^{(\ell)}(i)}^{(\ell)} = V_{ni}^{(\lambda)}$$

for all $i \in \{1, \ldots, n\}, \ell \in \{1, \ldots, \lambda - 1\}$ *and* $\mathcal{F}_{nn}^{(1)}$*-measurable bijections* $\iota_n^{(\ell)} : \{1, \ldots, n\} \to \{1, \ldots, n\}$.

Then, we have convergence of the sum,

$$S_n \xrightarrow{d} \mathcal{N}(0, \sigma^2), \tag{S.3}$$

where $\sigma^2 = \sum_{\ell=1}^{\lambda} \sigma_\ell^2$, *and conditional convergence of the bootstrapped sum,*

$$\sup_{x\in\mathbb{R}} \left|P\left(T_n \le x \middle| \mathcal{F}_{nn}^{(\lambda)}\right) - \Phi(x/\sigma)\right| \xrightarrow{p} 0, \tag{S.4}$$

as $n \to \infty$*.*

Note that from convergence of the bootstrapped sum conditional on the data (that is, (S.4)) follows unconditional convergence $T_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

Proof: Notice that

$$\sum_{\ell=1}^{\lambda} \sum_{i=1}^{n} E[(\zeta_{ni}^{(\ell)})^2 | \mathcal{F}_{n(i-1)}^{(\ell)}] \xrightarrow{p} \sum_{\ell} \sigma_\ell^2$$

as $n \to \infty$ by (S.1). Lyapounov's condition follows directly from (S.2.). Hence, (S.3) follows via the Martingale Central Limit Theorem.

For (S.4), our goal is to modify the proof of Theorem 2.1 in Pauly (2011) for the case of coupled resampling. We do so by considering the coupled increments

$$Z_{ni} = \sum_{\ell=1}^{\lambda} \zeta_{n\iota_n^{(\ell)}(i)}^{(\ell)},$$

where $\iota_n^{(\lambda)}$ is the identity. For these increments,

$$S_n = \sum_{i=1}^{n} Z_{ni}$$

and

$$T_n = \sum_{i=1}^{n} (V_{ni} - 1) Z_{ni} = \sum_{i=1}^{n} V_{ni}(Z_{ni} - \overline{Z}_n),$$

corresponding to weights $W_{ni} = V_{ni}/\sqrt{n}$ that fulfill equations (2.3), (2.4) and (2.5) in Pauly (2011). Note, however, that $(Z_{ni})_i$ is not a martingale difference array any more; hence, we cannot apply Theorem 2.1 directly, but instead invoke Theorem 4.1 in the appendix of Pauly (2011), which holds for more general triangular arrays of random variables.

Equation (4.1) in Theorem 4.1 of Pauly (2011) follows from the boundedness condition (S.2) by noting that

$$\max_{i\le n,\ell\le\lambda}|\zeta_{ni}^{(\ell)}| \le \sum_{\ell\le\lambda}\max_{i\le n}|\zeta_{ni}^{(\ell)}|$$

and that

$$\max_{i\le n}|\zeta_{ni}^{(\ell)}| \xrightarrow{p} 0$$

is equivalent to the weak Lindeberg condition

$$\sum_{i=1}^{n}(\zeta_{ni}^{(\ell)})^2\mathbb{I}_{|\zeta_{ni}^{(\ell)}|>\epsilon} \xrightarrow{p} 0 \;\forall\epsilon>0,$$

which is implied by (S.2) via Lyapounov's condition.

For (4.2), note that

$$\begin{aligned}\sum_{i=1}^{n}(Z_{ni}-\overline{Z}_n)^2 &= \sum_{i=1}^{n}Z_{ni}^2-\overline{Z}_n\sum_{i=1}^{n}Z_{ni} = \sum_{i=1}^{n}\left(\sum_{\ell=1}^{\lambda}\zeta^{(\ell)}_{n\iota_n^{(\ell)}(i)}\right)^2-\left(\sum_{i=1}^{n}Z_{ni}\right)^2/n\\ &= \sum_{\ell=1}^{\lambda}\sum_{i=1}^{n}\left(\zeta_{ni}^{(\ell)}\right)^2+2\sum_{\overline{\ell}=2}^{\lambda}\sum_{\underline{\ell}=1}^{\overline{\ell}-1}\sum_{i=1}^{n}\zeta^{(\underline{\ell})}_{n\iota_n^{\underline{\ell}}(i)}\zeta^{(\overline{\ell})}_{n\iota_n^{\overline{\ell}}(i)}-\left(\frac{S_n}{\sqrt{n}}\right)^2.\end{aligned}$$

Now,

$$A_{ni}^{(\ell)} = (\zeta_{ni}^{(\ell)})^2-E[(\zeta_{ni}^{(\ell)})^2|\mathcal{F}_{n(i-1)}^{(\ell)}]$$

defines a martingale difference array with respect to the filtration array $\mathcal{F}_{ni}^{(\ell)}$ for all $1\le\ell\le\lambda$, and

$$B_{ni}^{\underline{\ell},\overline{\ell}} = \zeta^{(\underline{\ell})}_{n\iota_n^{\underline{\ell}}(i)}\zeta^{(\overline{\ell})}_{n\iota_n^{\overline{\ell}}(i)},$$

defines a martingale difference array with respect to the filtration array $\mathcal{F}_{ni}^{(\overline{\ell})}$ (where we have used $\mathcal{F}_{nn}^{(1)}$-measurability of all ι_n^{ℓ}) for all $1\le\underline{\ell}<\overline{\ell}\le\lambda$, In both cases, the increments have second moments bounded by $\frac{C}{n^{1+\delta}}$ by (S.2). Indeed,

$$E[(A_{ni}^{(\ell)})^2] = E[(\zeta_{ni}^{(\ell)})^4]-E[E[(\zeta_{ni}^{(\ell)})^2|\mathcal{F}_{n(i-1)}^{(\ell)}]^2]\le E[(\zeta_{ni}^{(\ell)})^4]\le\frac{C}{n^{1+\delta}}$$

and

$$E[(B_{ni}^{\ell,\bar{\ell}})^2] = E[(\zeta_{n\iota_n^{\ell}(i)}^{(\ell)}]^2(\zeta_{n\iota_n^{\bar{\ell}}(i)}^{(\bar{\ell})})^2) \leq \sqrt{E[(\zeta_{n\iota_n^{\ell}(i)}^{(\ell)}]^4)}\sqrt{E[(\zeta_{n\iota_n^{\bar{\ell}}(i)}^{(\bar{\ell})})^4]} \leq \frac{C}{n^{1+\delta}}$$

by the Cauchy-Schwarz inequality. Now, for any martingale difference array $(C_{ni})_{i=1}^n$ with $EC_{ni}^2 \leq \frac{C}{n^{1+\delta}}$,

$$E\left(\sum_{i=1}^n C_{ni}\right)^2 = \sum_{i=1}^n EC_{ni}^2 \leq \frac{C}{n^\delta} \to 0$$

and hence

$$\sum_{i=1}^n C_{ni} \xrightarrow{p} 0$$

as $n \to \infty$. It follows that

$$\sum_{i=1}^n (Z_{ni} - \overline{Z}_n)^2 = \sum_{\ell=1}^\lambda \underbrace{\sum_{i=1}^n E[(\zeta_{ni}^{(\ell)})^2|\mathcal{F}_{n(i-1)}^{(\ell)}]}_{\xrightarrow{p}\sigma_\ell^2} + \sum_{\ell=1}^\lambda \underbrace{\sum_{i=1}^n A_{ni}^{(\ell)}}_{\xrightarrow{p}0} + 2\sum_{\bar{\ell}=2}^\lambda \sum_{\underline{\ell}=1}^{\bar{\ell}-1} \underbrace{\sum_{i=1}^n B_{ni}^{\underline{\ell},\bar{\ell}}}_{\xrightarrow{p}0} - \underbrace{\left(\frac{S_n}{\sqrt{n}}\right)^2}_{\xrightarrow{p}0}$$

$$\xrightarrow{p} \sum_{\ell=1}^\lambda \sigma_\ell^2 = \sigma^2,$$

where we have used (S.1) and the unconditional convergence result (S.3).

Finally, note that the Z_{ni} are a sufficient statistic of $\mathcal{F}_{nn}^{(\lambda)}$ for calculating T_n, incorporating sufficient information about both the $\zeta_{ni}^{(\ell)}$ and ι_n^ℓ. Hence, (S.4) follows from Theorem 4.1 in the appendix of Pauly (2011). $\square$

We now apply this result to our matching setting:

Proposition A.2. *Under the setup and assumptions of Lemma A.2, and also*

(a.5) $E[F_k^4(Y,W,S)|W=w,X=x]$ *uniformly bounded on* $\mathcal{X}$ *for all* k, $w \in \{0,1\}$,

consider the bootstrapped sum

$$\widehat{\Phi}^* = \frac{1}{n}\sum_{W_i=1} V_i \left(F(Y_i,W_i,S_i) + \sum_{j\in\mathcal{J}(i)} F(Y_j,W_j,S_j)\right),$$

where V is multinomial with parameters $(N_1; N_1^{-1}, \dots, N_1^{-1})$ independent of the data. Then,

$$\sup_{r\in\mathbb{R}^s} \left| P_{\mathbf{w}}\left(\sqrt{n}(\widehat{\Phi}^* - \widehat{\Phi}) \leq r \middle| \mathcal{S}\right) - P(\mathcal{N}(\mathbf{0}, V^*) \leq r) \right| \xrightarrow{p} 0.$$

Proof: Fix $\lambda \in \mathbb{R}^s$. Similar to the proof of Lemma A.2, we decompose

$$\begin{aligned}
\sqrt{n}(\widehat{\Phi}^* - \widehat{\Phi})'\lambda = \sqrt{n}((\widehat{\Phi}^* - \Phi) - (\widehat{\Phi} - \Phi))'\lambda \\
= \frac{1}{\sqrt{n}}\Big(\sum_{W_i=1} (V_i - 1)(\Phi_1(X_i) + M\Phi_0(X_i) - \Phi)'\lambda \\
+ \sum_{i\in\mathcal{S}^*}(V_i - 1)(F(Y, W, S) - \Phi_{W_i}(X_i))'\lambda\Big) \\
+ \frac{1}{\sqrt{n}} \sum_{W_i=1} (V_i - 1) \sum_{j\in\mathcal{J}(i)} (\Phi_0(X_j) - \Phi_0(X_i))'\lambda.
\end{aligned}$$

The last part of the sum still vanishes in probability, as

$$\begin{aligned}
&E\left(\left|\frac{1}{\sqrt{n}} \sum_{W_i=1} (V_i - 1) \sum_{j\in\mathcal{J}(i)} (\Phi_0(X_j) - \Phi_0(X_i))'\lambda\right| \middle| \mathcal{S}\right) \\
&\leq \frac{1}{\sqrt{n}} \sum_{W_i=1} \sum_{j\in\mathcal{J}(i)} \underbrace{E(|V_i - 1|)}_{\leq 2} |(\Phi_0(X_j) - \Phi_0(X_i))'\lambda| \\
&\leq \frac{2}{\sqrt{n}} \sum_{W_i=1} \sum_{j\in\mathcal{J}(i)} |(\Phi_0(X_j) - \Phi_0(X_i))'\lambda| \\
&\leq \frac{2L}{\sqrt{n}} \sum_{W_i=1} \sum_{j\in\mathcal{J}(i)} d(X_j, X_i) \xrightarrow{p} 0
\end{aligned}$$

for an appropriate Lipschitz constant $L = L(\lambda)$, where we have used that

$$E(|V_i - 1|) \leq E(V_i + 1) = 2.$$

We can decompose the other parts into martingale increments as in the proof of Lemma A.2:

$$\sqrt{n}(\widehat{\Phi}^* - \widehat{\Phi})'\lambda = \sum_{i=1}^{N_1}(V_i - 1)\xi_i + \sum_{i=N_1+1}^{(M+2)N_1} (V_{i-N_1} - 1)\xi_i + o_P(1)$$

The result follows from Proposition A.1, which establishes a general result for the coupled resampling of martingale difference arrays. $\square$

Proposition A.3. *Let $X_1, X_2, \dots$ be a sequence of real-valued, non-negative, integrable random variables, and $\mathcal{F}_1, \mathcal{F}_2, \dots$ a sequence of conformable σ-algebras. If X_n converges in distribution to some real-valued, integrable random variable X conditional on $\mathcal{F}_n$, i.e.*

$$\sup_{x\in\mathbb{R}} |P(X_n \leq x|\mathcal{F}_n) - P(X \leq x)| \xrightarrow{p} 0, \tag{S.5}$$

then, for all $\varepsilon > 0$, $P(E[X_n|\mathcal{F}_n] \geq E[X] - \varepsilon) \to 1$.

Proof. Let $\varepsilon > 0$. By dominated convergence, $E[X 1_{X\leq M}] \to E[X]$ as $M \to \infty$. Fix $M > 0$ such that $E[X 1_{X\leq M}] \geq E[X] - \varepsilon/2$. Let $\delta > 0$. By (S.5) there exists N such that

$$P\left(\sup_{r\in\mathbb{R}} |P(X_n \leq r|\mathcal{F}_n) - P(X \leq r)| \leq \varepsilon/(2M)\right) \geq 1 - \delta$$

for all $n \geq N$. Hence, with probability at least $1 - \delta$ and for all $n \geq N$,

$$\begin{aligned} E[X_n|\mathcal{F}_n] &= \int_{x\geq 0} (1 - P(X_n \leq x|\mathcal{F}_n))dx \geq \int_{x=0}^{M} (1 - P(X_n \leq x|\mathcal{F}_n))dx \\ &\geq \int_{x=0}^{M} (1 - P(X \leq x))dx - \int_{x=0}^{M} |P(X_n \leq r|\mathcal{F}_n) - P(X \leq r)|dx \\ &\geq E[X 1_{X\leq M}] - M\varepsilon/(2M) \geq E[X] - \varepsilon/2 - \varepsilon/2 = E[X] - \varepsilon. \end{aligned}$$

□

Conditional Inference

In this article, we have analyzed the unconditional distribution of post-matching estimators. However, inference conditional on the sample values of the regressors may be appropriate in some applications; for example, when the sample is the entire population. In this section, we discuss validity of standard errors conditional on the values of the covariates, X, and treatment, W (and, as an implication, conditional on the matches).

Abadie et al. (2014) have shown that OLS (EHW) standard errors, which are robust to misspecification as measures of unconditional variation, are not generally valid as measures of variation conditional on the values of the regressors when the regression model is misspecified. They propose an estimator of the conditional variance that is robust to misspecification.

Conditional on the values of the covariates, X, and the treatment, W, the untreated units used as matches for each treated unit are given, and the analysis of Abadie et al. (2014) goes through. In particular, if the regression model is correctly specified, OLS standard error estimates are valid measures of conditional variation. If the regression model is not correctly specified, OLS standard errors are not valid measures of conditional variation, but the conditional standard errors in Abadie et al. (2014) are.

Table 4: Simulation results from one million Monte Carlo iterations (1000 draws of regressors and 1000 draws of outcomes per draw of regressors)

(a) Target parameter: Coefficient τ_0 on W

		averages over regressor draws		average standard errors		
	spec	conditional mean of $\widehat{\tau}_0$ (1)	conditional std. of $\widehat{\tau}_0$ (2)	OLS (3)	cluster (4)	cond. (5)
DGP1	1	0.00	0.202	0.359	0.197	0.203
	2	0.00	0.202	0.196	0.197	0.201
DGP2	1	6.60	0.202	0.631	0.870	0.209
	2	6.60	0.202	0.631	0.870	0.209

(b) Target parameter: Coefficient τ_1 on WX

		averages over regressor draws		average standard errors		
	spec	conditional mean of $\widehat{\tau}_1$ (1)	conditional std. of $\widehat{\tau}_1$ (2)	OLS (3)	cluster (4)	cond. (5)
DGP1	1	0.99	0.352	0.728	0.341	0.360
	2	1.00	0.352	0.338	0.339	0.349
DGP2	1	1.02	0.351	1.324	1.840	0.384
	2	1.02	0.351	1.324	1.840	0.384

Table 4 reports simulation results for the same DGPs, specifications and sample sizes employed in Section 4. The simulation is based on 1000 regressor (X and W) draws, with 1000 outcome (Y) draws for each draw of the regressors. Columns (1) and (2) report means and standard deviations of regression coefficients conditional on the regressors, averaged over the 1000 regressor draws. Columns (3), (4) and (5) report OLS standard errors, clustered standard errors and the conditional standard errors of Abadie et al. (2014), respectively, averaged over regressor draws. Consistent with the results in Abadie et al. (2014), under correct specification (DGP1, specification 2), OLS standard errors are valid measures of the conditional standard deviation of $\widehat{\tau}_0$ and $\widehat{\tau}_1$. This is not the case, however, when the regression function is misspecified (specification 1 for DGP1, and DGP2). Moreover, as demonstrated in DGP2, clustered standard errors, which are valid for the unconditional variation, are not generally valid as measures of conditional variation. In DGP1, however, clustered standard errors provide an appropriate measure of conditional variation because the conditional means of $\widehat{\tau}_0$ and $\widehat{\tau}_1$ are approximately constant across regressor draws (up to an asymptotically negligible bias term caused by imperfect matches), so the variation in covariates does not contribute to the total variation of the estimates. By the law of total variance, the conditional and unconditional variances are thus approximately the same in this specific case. The conditional standard error estimates from Abadie et al. (2014) are close to the average standard deviation of $\widehat{\tau}_0$ and $\widehat{\tau}_1$.

Generalization to M-Estimators

In this section we generalize the results of the article to M-estimators. Consider

$$\widehat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \ \frac{1}{n} \underbrace{\sum_{i \in \mathcal{S}^*} m(U_i, \theta)}_{= \frac{1}{n}\sum_{i=1}^{n} m(U_{ni}, \theta)} , \tag{S.6}$$

where $U_i = (Y_i, Z_i)$. Notice that OLS is a special case of (S.6) with

$$m(U_i, \theta) = (Y_i - Z_i'\theta)^2.$$

The estimand θ_0 estimated by $\widehat{\theta}$ defined in (S.6) is given by its appropriate population analogue

$$\theta_0 = \underset{\theta \in \mathbb{R}^k}{\operatorname{argmin}} \ \underbrace{(1+M)^{-1} E\left[m(U,\theta) + M E[m(U,\theta)|W=0,X]|W=1\right]}_{=E^* m(U,\theta)} . \tag{S.7}$$

The following result provides assumptions under which $\widehat{\theta}$ is consistent for θ_0 and asymptotically normal, where we write

$$s(u,\theta) = \frac{\partial m(u,\theta)}{\partial \theta}, \qquad\qquad h(u,\theta) = \frac{\partial^2 m(u,\theta)}{\partial \theta \partial \theta'}.$$

Proposition A.4 (Asymptotic distribution of the post-matching M-estimator). *Suppose that Assumptions 1, 2, 3 hold and that*

1. *θ_0 is in the interior of $\Theta \subseteq \mathbb{R}^k$, which is compact;*
2. *If $\theta \neq \theta_0$ then $E^* m(U,\theta) \neq E^* m(U,\theta_0)$;*
3. *$m(u,\theta)$ is twice continuously differentiable in θ;*
4. *$E^*[|m(U,\theta_0)|]$, $E^*[\|s(U,\theta_0)\|]$ and $E^*[\|h(U,\theta_0)\|]$ are all finite;*
5. *$h(u,\theta)$ is Lipschitz in θ, uniformly in z;*
6. *$H = E^*[h(U,\theta_0)]$ is invertible;*
7. *$E[\|m(U,\theta)\|^2|W = w, X = x]$, $E[\|s(U,\theta)\|^{2+\delta}|W = w, X = x]$, $E[\|h(U,\theta)\|^2|W = w, X = x]$ are bounded on $\mathcal{X}_w$ for $w = 0,1$ and some $\delta > 0$, for all $\theta \in \Theta$;*
8. *$E[m(U,\theta)|W = 0, X = x]$, $E[s(U,\theta)|W = 0, X = x]$, $E[h(U,\theta)|W = 0, X = x]$, $\mathrm{var}(s(U,\theta)|W = 0, X = x)$ are componentwise Lipschitz in x with respect to $d(\cdot,\cdot)$, for all $\theta \in \Theta$.*

Then, $\widehat{\theta}$ is $\sqrt{n}$-consistent for θ_0 and asymptotically normal,

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, H^{-1} J H^{-1}),$$

where

$$J = \frac{\mathrm{var}\Big(E[s(U,\theta_0)|X, W=1] + M E[s(U,\theta_0)|X, W=0]\big|W=1\Big)}{1+M} + \frac{E\Big[\mathrm{var}(s(U,\theta_0)|X, W=1] + M\mathrm{var}(s(U,\theta_0)|X, W=0)\big|W=1\Big]}{1+M}.$$

Sketch of the proof: First, we show that $\widehat{\theta} \xrightarrow{p} \theta_0$. Towards this end, write

$$\widehat{Q}(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(U_{ni}, \theta),$$
$$Q(\theta) = E^*[m(U, \theta)] = \frac{E[m(U,\theta)|W=1] + ME[E[m(U,\theta)|X, W=0]|W=1]}{M+1}.$$

We derive consistency from arguing that

$$\sup_{\theta\in\Theta}\left|\widehat{Q}(\theta) - Q(\theta)\right| \xrightarrow{p} 0 \tag{S.8}$$

with $Q(\theta)$ continuous. Indeed, note that by 3, 4, 5 $m(u,\theta)$ is Lipschitz in θ and there exists $M(u)$ with $E^*M(U) < \infty$ and $|m(u,\theta)| \leq M(u)$. Further, $|\widehat{Q}(\theta) - Q(\theta)| \xrightarrow{p} 0$ as $n \to \infty$ for all $\theta \in \Theta$ by Lemma A.1.

We follow the main steps in the proof of Lemma 1 in Tauchen (1985) to establish (S.8). Fix $\epsilon > 0$. Note that for fixed $\theta \in \Theta$

$$r(u, \theta, d) = \sup_{\|\gamma-\theta\|\leq d} |m(u,\theta) - m(u,\gamma)| \longrightarrow 0$$

as $d \longrightarrow 0$. Hence, by dominated convergence $E^*r(U,\theta,d) \leq \epsilon$ whenever $d \leq \bar{d}(\theta)(> 0)$. Write L for the Lipschitz constant in $m(u,\theta)$ (so that $|m(u,\theta) - m(u,\gamma)| \leq L\|\theta - \gamma\|$), and assume wlog that $0 < \bar{d}(\theta) < \epsilon/L$. By compactness, we can cover Θ by a finite set of open balls B_k of radius $\bar{d}(\theta_k)$ around θ_k such that

$$|Q(\theta) - Q(\theta_k)| \leq \epsilon \; \forall \theta \in B_k.$$

Now, for $\theta \in B_k$,

$$\begin{aligned}\left|\widehat{Q}(\theta) - Q(\theta)\right| &\leq \frac{1}{n}\sum_{i=1}^{n}\underbrace{|m(U_{ni},\theta) - m(U_{ni},\theta_k)|}_{\leq L\|\theta-\theta_k\|\leq\epsilon} + \left|\frac{1}{n}\sum_{i=1}^{n} m(U_{ni},\theta_k) - Q(\theta_k)\right| + |Q(\theta_k) - Q(\theta)| \\ &\leq \underbrace{\left|\widehat{Q}(\theta_k) - Q(\theta_k)\right|}_{\xrightarrow{p} 0} + 2\epsilon\end{aligned}$$

Hence, $\sup_{\theta\in B_k}\left|\widehat{Q}(\theta) - Q(\theta)\right| \xrightarrow{p} 0$ for all k, and thus (S.8) with Q continuous; by 1, 2 we have $\widehat{\theta} \xrightarrow{p} \theta_0$.

Second, given consistency, we show asymptotic normality. The FOC for the minimization in (S.6) is

$$\sum_{i=1}^{n} s(U_{ni}, \widehat{\theta}) = 0.$$

By the mean-value theorem, there exist $\bar{\theta}_{ni}$ with

$$\sum_{i=1}^{n} s(U_{ni}, \theta_0) + \sum_{i=1}^{n} h(U_{ni}, \bar{\theta}_{ni})(\widehat{\theta} - \theta_0) = 0,$$

where $\|\bar{\theta}_{ni} - \theta_0\| \leq \|\widehat{\theta} - \theta_0\|$ for all i. Hence,

$$\sqrt{n}(\widehat{\theta} - \theta_0) = -\left(\frac{1}{n}\sum_{i=1}^{n} h(U_{ni}, \bar{\theta}_{ni})\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(U_{ni}, \theta_0).$$

By Lipschitz continuity of h (in θ), consistency, and Lemma A.1, $\frac{1}{n}\sum_{i=1}^{n} h(U_{ni}, \bar{\theta}_{ni}) \xrightarrow{p} H$. By Lemma A.2, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(U_{ni}, \theta_0) \xrightarrow{d} \mathcal{N}(0, J)$ (where we use that $E^* s(U, \theta_0) = 0$). □

Proposition A.4 established $\sqrt{n}$-consistency and asymptotic normality of the post-matching M-estimator $\widehat{\theta}$, analogously to Proposition 2 for post-matching OLS. For inference, we can estimate the sandwich variance $H^{-1}JH^{-1}$ from estimates of H and J given by

$$\widehat{H} = \frac{1}{n}\sum_{i=1}^{n} h(U_i, \widehat{\theta}), \quad \widehat{J} = \frac{1}{n}\sum_{i=1}^{n} W_i \left(s(U_i, \widehat{\theta}) + \sum_{j \in \mathcal{J}(i)} s(U_j, \widehat{\theta}) \right) \times \left(s(U_i, \widehat{\theta}) + \sum_{j \in \mathcal{J}(i)} s(U_j, \widehat{\theta}) \right)',$$

where $\widehat{H}$ is the sample analogue of $H = E^*[h(U, \theta_0)]$ and $\widehat{J}$ a clustered variance estimate of $s(U, \theta_0)$, with clusters given by matched units.

Similar to Proposition 5 for post-matching OLS, the following result establishes conditions under which these estimates yield consistent standard errors.

Proposition A.5 (Inference for the post-matching M-estimator). *Assume the conditions for Proposition A.4, and also*

1. $E[\|s(U, \theta_0)\|^4 | W = w, X = x]$ *is bounded on* $\mathcal{X}_w$ *for* $w = 0, 1$*;*

2. $E[s(U, \theta_0) \times s(U, \theta_0)' | W = 0, X = x]$ *is componentwise Lipschitz in* x *with respect to* $d(\cdot, \cdot)$.

Then,

$$\widehat{H} \xrightarrow{p} H, \qquad \widehat{J} \xrightarrow{p} J.$$

In particular, standard error estimates are consistent in the sense that

$$\widehat{H}^{-1}\widehat{J}\widehat{H}^{-1} - n\mathrm{var}(\widehat{\theta}) \xrightarrow{p} 0.$$

Sketch of the proof: By Lipschitz continuity of h (in θ), consistency, and Lemma A.1, $\widehat{H} \xrightarrow{p} H$. For $\widehat{J}$, note first that

$$\widehat{J} = \frac{1}{n}\sum_{i=1}^{n} W_i \left(s(U_i, \theta_0) + \sum_{j \in \mathcal{J}(i)} s(U_j, \theta_0) \right) \times \left(s(U_i, \theta_0) + \sum_{j \in \mathcal{J}(i)} s(U_j, \theta_0) \right)' + o_P(1)$$

by Lipschitz continuity of s (in θ) and consistency. The result follows as in the proof of Proposition 5. $\square$

References

Abadie, A. and Imbens, G. (2012). A martingale representation for matching estimators. *Journal of the American Statistical Association*, 107(498):833–843.

Abadie, A., Imbens, G. W., and Zheng, F. (2014). Inference for misspecified models with fixed regressors. *Journal of the American Statistical Association*, 109(508):1601–1614.

Pauly, M. (2011). Weighted resampling of martingale difference arrays with applications. *Electronic Journal of Statistics*, 5:41–52.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics*, 30(1-2):415–443.