Biometrika (2018), **xx**, x, *pp*. 1–14

Printed in Great Britain

Supplementary material for "Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores"

BY S. YANG

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.

syang24@ncsu.edu

P. DING

Department of Statistics, University of California, Berkeley, California 94720, U.S.A.

pengdingpku@berkeley.edu

§ S1 gives all the proofs, § S2 presents a simulation study, § S3 extends the theory to the average treatment effect on the treated, and § S4 provides more detailed analysis of the National Health and Nutrition Examination Survey Data.

Below we use $C \cong D$ for $C = D + O_p(N^{-1/2})$. Because $\hat{\theta}$ is the solution to the score equation $S(\theta) = 0$, under certain regularity conditions, $\hat{\theta} - \theta^* = \mathcal{J}(\theta^*)^{-1}S(\theta^*) + o_p(N^{-1/2})$, where $\mathcal{J}(\theta^*) = E\{\partial S(\theta^*)/\partial\theta'\}$ (e.g., van der Vaart, 2000). When the propensity model is correctly specified, then $\mathcal{J}(\theta^*) = \mathcal{I}(\theta^*)$; when the propensity score model is misspecified, $\mathcal{J}(\theta^*)$ is not necessarily equal to $\mathcal{I}(\theta^*)\cdot$

S1. PROOFS

S1·1. *Proof of Theorem 1*

We write

$$
\begin{aligned}
\hat{\tau}_\epsilon &= \hat{\tau}_\epsilon(\hat{\theta}) \\
&\cong \hat{\tau}_\epsilon(\theta^*) + E\left\{\frac{\partial \hat{\tau}_\epsilon(\theta^*)}{\partial \theta'}\right\}(\hat{\theta} - \theta^*) \qquad \text{(S1)} \\
&\cong \frac{1}{N}\sum_{i=1}^{N} \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\left\{\frac{A_iY_i}{e(X_i'\theta^*)} - \frac{(1-A_i)Y_i}{1-e(X_i'\theta^*)}\right\} + E\left\{\frac{\partial \hat{\tau}_\epsilon(\theta^*)}{\partial \theta'}\right\}\mathcal{I}(\theta^*)^{-1}S(\theta^*) \qquad \text{(S2)} \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\left\{\frac{A_iY_i}{e(X_i'\theta^*)} - \frac{(1-A_i)Y_i}{1-e(X_i'\theta^*)}\right\} \\
&\quad + B'\frac{1}{N}\sum_{i=1}^{N} X_i \frac{A_i - e(X_i'\theta^*)}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}} f(X_i'\theta^*),
\end{aligned}
$$

where (S1) follows from the Taylor expansion, (S2) follows from $\hat{\theta} - \theta^* \cong \mathcal{I}(\theta^*)^{-1}S(\theta^*)$ and

$$
B' = E\left\{\frac{\partial \hat{\tau}_\epsilon(\theta^*)}{\partial \theta'}\right\}\mathcal{I}(\theta^*)^{-1}\cdot \qquad \text{(S3)}
$$

Therefore, the asymptotic linearity of $\hat{\tau}_\epsilon$ follows. Moreover,

$$
\begin{aligned}
N^{1/2}(\hat{\tau}_\epsilon - \tau_\epsilon) \cong\ & N^{-1/2}\sum_{i=1}^{N}\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\left[\frac{A_i\{Y_i-\mu(A_i,X_i)\}}{e(X_i'\theta^*)}-\frac{(1-A_i)\{Y_i-\mu(A_i,X_i)\}}{1-e(X_i'\theta^*)}\right] \\
&+N^{-1/2}\sum_{i=1}^{N}\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\left(\frac{\{A_i-e(X_i'\theta^*)\}[\mu(A_i,X_i)-\mu\{A_i,e(X_i'\theta^*)\}]}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}}\right) \\
&+N^{-1/2}\sum_{i=1}^{N}\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\left[\frac{\{A_i-e(X_i'\theta^*)\}\mu\{A_i,e(X_i'\theta^*)\}}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}}-\tau\{e(X_i'\theta^*)\}\right] \\
&+N^{-1/2}\sum_{i=1}^{N}\left[\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\tau\{e(X_i'\theta^*)\}-\tau_\epsilon\right] \\
&+N^{-1/2}\sum_{i=1}^{N}B'X_i\frac{A_i-e(X_i'\theta^*)}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}}f(X_i'\theta^*), \\
=\ & T_0+T_1+T_2+T_3,
\end{aligned}
$$

where $\tau\{e(X'\theta^*)\} = E\{Y(1)-Y(0)\mid e(X'\theta^*)\}$, and by grouping different terms,

$$
\begin{aligned}
T_0 &= N^{-1/2}\sum_{i=1}^{N}\left[\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\tau\{e(X_i'\theta^*)\}-\tau_\epsilon\right], \\
T_1 &= N^{-1/2}\sum_{i=1}^{N}\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\left[\frac{\{A_i-e(X_i'\theta^*)\}\mu\{A_i,e(X_i'\theta^*)\}}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}}-\tau\{e(X_i'\theta^*)\}\right] \\
&\quad+N^{-1/2}\sum_{i=1}^{N}B'E\{X_i\mid e(X_i'\theta^*)\}\frac{A_i-e(X_i'\theta)}{e(X_i'\theta)\{1-e(X_i'\theta)\}}f(X_i'\theta^*), \\
T_2 &= N^{-1/2}\sum_{i=1}^{N}\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\left(\frac{\{A_i-e(X_i'\theta^*)\}[\mu(A_i,X_i)-\mu\{A_i,e(X_i'\theta^*)\}]}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}}\right) \\
&\quad+N^{-1/2}\sum_{i=1}^{N}B'[X_i-E\{X_i\mid e(X_i'\theta^*)\}]\frac{A_i-e(X_i'\theta^*)}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}}f(X_i'\theta^*),
\end{aligned}
$$

and

$$
T_3 = N^{-1/2}\sum_{i=1}^{N}\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\left[\frac{A_i\{Y_i-\mu(A_i,X_i)\}}{e(X_i'\theta^*)}-\frac{(1-A_i)\{Y_i-\mu(A_i,X_i)\}}{1-e(X_i'\theta^*)}\right]. \quad \text{(S4)}
$$

Define

$$
\mathcal{F}_0 = \left\{X_1'\theta^*,\ldots,X_N'\theta^*\right\},\ \mathcal{F}_1 = \left\{A_1,\ldots,A_N,X_1'\theta^*,\ldots,X_N'\theta^*\right\}, \\
\mathcal{F}_2 = \left\{A_1,\ldots,A_N,X_1'\theta^*,\ldots,X_N'\theta^*,X_1,\ldots,X_N\right\}.
$$

By conditioning arguments, $E(T_0)=0$, for $k=1,\ldots,3$, $E(T_k)=E\{E(T_k\mid\mathcal{F}_{k-1})\}=0$, and for $k=1,\ldots,3$,

$$
\begin{aligned}
\operatorname{cov}(T_0,T_k) &= \operatorname{cov}\{E(T_0\mid\mathcal{F}_0),E(T_k\mid\mathcal{F}_0)\}+E\{\operatorname{cov}(T_0,T_k\mid\mathcal{F}_0)\} \\
&= \operatorname{cov}\{E(T_0\mid\mathcal{F}_0),0\}+E\{0\}=0,
\end{aligned}
$$

for $k = 2, 3$,

$$\begin{aligned}\mathrm{cov}(T_1, T_k) &= \mathrm{cov}\{E(T_1 \mid \mathcal{F}_1), E(T_k \mid \mathcal{F}_1)\} + E\{\mathrm{cov}(T_1, T_k \mid \mathcal{F}_1)\} \\ &= \mathrm{cov}\{E(T_1 \mid \mathcal{F}_1), 0\} + E\{0\} = 0,\end{aligned}$$

and

$$\begin{aligned}\mathrm{cov}(T_2, T_3) &= \mathrm{cov}\{E(T_2 \mid \mathcal{F}_2), E(T_3 \mid \mathcal{F}_2)\} + E\{\mathrm{cov}(T_2, T_3 \mid \mathcal{F}_2)\} \\ &= \mathrm{cov}\{E(T_2 \mid \mathcal{F}_2), 0\} + E\{0\} = 0\cdot\end{aligned}$$

Also, we calculate the variances of T_i, for $i = 0, \ldots, 3$, as follows. For T_0,

$$\mathrm{var}(T_0) = E(T_0^2) = \frac{E\left\{\omega_\epsilon(X'\theta^*)^2 \mathrm{var}[\tau\{e(X'\theta^*)\}]\right\}}{E\{\omega_\epsilon(X'\theta^*)\}^2}\cdot$$

For T_1,

$$\begin{aligned}\mathrm{var}(T_1) &= E\{\mathrm{var}(T_1 \mid \mathcal{F}_0)\} = E\{E(T_1^2 \mid \mathcal{F}_0)\} \\ &= \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2} E\{\omega_\epsilon(X'\theta^*)^2 \left[\left\{\frac{1 - e(X'\theta^*)}{e(X'\theta^*)}\right\}^{1/2} \mu\{1, e(X'\theta^*)\}\right. \\ &\quad \left. + \left\{\frac{e(X'\theta^*)}{1 - e(X'\theta^*)}\right\}^{1/2} \mu\{0, e(X'\theta^*)\}\right]^2 \\ &\quad + 2\frac{1}{E\{\omega_\epsilon(X'\theta^*)\}} B' E\{\omega_\epsilon(X'\theta^*) E\{X \mid e(X'\theta^*)\} \\ &\quad \times \left[\frac{\mu\{1, e(X'\theta^*)\}}{e(X'\theta^*)} + \frac{\mu\{0, e(X'\theta^*)\}}{1 - e(X'\theta^*)}\right] f(X'\theta^*)\} \\ &\quad + B' E\left[f(X'\theta^*)^2 \frac{E\{X \mid e(X'\theta^*)\} E\{X' \mid e(X'\theta^*)\}}{e(X'\theta^*)\{1 - e(X'\theta^*)\}}\right] B\cdot\end{aligned}$$

For T_2,

$$\begin{aligned}\mathrm{var}(T_2) &= E\{\mathrm{var}(T_2 \mid \mathcal{F}_1)\} = E\{E(T_2^2 \mid \mathcal{F}_1)\} \\ &= \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2} E\left\{\omega_\epsilon(X'\theta^*)^2 \left[\frac{\sigma^2\{1, e(X'\theta^*)\}}{e(X'\theta^*)} + \frac{\sigma^2\{0, e(X'\theta^*)\}}{1 - e(X'\theta^*)}\right]\right\} \\ &\quad +2\frac{1}{E\{\omega_\epsilon(X'\theta^*)\}} B' E\left\{\omega_\epsilon(X'\theta^*) f(X'\theta^*) \left[\frac{\mathrm{cov}\{X, \mu(1, X) \mid e(X'\theta^*)\}}{e(X'\theta^*)}\right.\right. \\ &\quad \left.\left. + \frac{\mathrm{cov}\{X, \mu(0, X) \mid e(X'\theta^*)\}}{1 - e(X'\theta^*)}\right]\right\} \\ &\quad +B' E\left[f(X'\theta^*)^2 \frac{\mathrm{var}\{X \mid e(X'\theta^*)\}}{e(X'\theta^*)\{1 - e(X'\theta^*)\}}\right] B\cdot\end{aligned}$$

For T_3,

$$\begin{aligned}\mathrm{var}(T_3) &= E\{\mathrm{var}(T_3 \mid \mathcal{F}_2)\} = E\{E(T_3^2 \mid \mathcal{F}_2)\} \\ &\cong \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2} E\left[\omega_\epsilon(X'\theta^*)^2 \left\{\frac{\sigma_1^2(X)}{e(X'\theta^*)} + \frac{\sigma_0^2(X)}{1 - e(X'\theta^*)}\right\}\right].\end{aligned}$$

Because

$$\frac{\partial \hat{\tau}_\epsilon(\theta^*)}{\partial \theta'} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial \theta'}\left[\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\right]\left\{\frac{A_iY_i}{e(X_i'\theta^*)} - \frac{(1-A_i)Y_i}{1-e(X_i'\theta^*)}\right\}$$
$$-\frac{1}{N}\sum_{i=1}^{N}\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}}\left[\frac{A_iY_i}{e(X_i'\theta^*)^2} + \frac{(1-A_i)Y_i}{\{1-e(X_i'\theta^*)\}^2}\right] f(X_i'\theta^*)X_i,$$

we have

$$\begin{aligned} E\left\{\frac{\partial \hat{\tau}_\epsilon(\theta^*)}{\partial \theta}\right\} &= E\left(\frac{\partial}{\partial \theta}\left[\frac{\omega_\epsilon(X'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\right]\tau(X)\right) - \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}}E\left\{\omega_\epsilon(X'\theta^*)f(X'\theta^*)\right. \\ &\quad \left.\times\left[\frac{E\{X,\mu(1,X)\mid e(X'\theta^*)\}}{e(X'\theta^*)} + \frac{E\{X,\mu(0,X)\mid e(X'\theta^*)\}}{1-e(X'\theta^*)}\right]\right\} \\ &= b_{1,\epsilon} - b_{2,\epsilon}, \end{aligned}$$

where $b_{1,\epsilon}$ and $b_{2,\epsilon}$ are defined in Theorem 1. Therefore, according to (S3), $B = (b_{1,\epsilon} - b_{2,\epsilon})'\mathcal{I}(\theta^*)^{-1}$. As a result,

$$\begin{aligned} &\mathrm{var}(T_0) + \mathrm{var}(T_1) + \mathrm{var}(T_2) + \mathrm{var}(T_3) \\ =& \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2}E\{\omega_\epsilon(X'\theta^*)^2\mathrm{var}[\tau\{e(X'\theta^*)\}]\} && \text{(S5)} \\ &+\frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2}E\left\{\omega_\epsilon(X'\theta^*)^2\left[\left\{\frac{1-e(X'\theta^*)}{e(X'\theta^*)}\right\}^{1/2}\mu\{1,e(X'\theta^*)\}\right.\right. \\ &\left.\left.+\left\{\frac{e(X'\theta^*)}{1-e(X'\theta^*)}\right\}^{1/2}\mu\{0,e(X'\theta^*)\}\right]^2\right\} \\ &+\frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2}E\left\{\omega_\epsilon(X'\theta^*)^2\left[\frac{\sigma^2\{1,e(X'\theta^*)\}}{e(X'\theta^*)} + \frac{\sigma^2\{0,e(X'\theta^*)\}}{1-e(X'\theta^*)}\right]\right\} \\ &+\frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2}E\left[\omega_\epsilon(X'\theta^*)^2\left\{\frac{\sigma^2(1,X)}{e(X'\theta^*)} + \frac{\sigma^2(0,X)}{1-e(X'\theta^*)}\right\}\right] && \text{(S6)} \\ &+2\frac{1}{E\{\omega_\epsilon(X'\theta^*)\}}B'E\left\{\omega_\epsilon(X'\theta^*)f(X'\theta^*)\left[\frac{E\{X\mu(1,X)\mid e(X'\theta^*)\}}{e(X'\theta^*)}\right.\right. \\ &\left.\left.+\frac{E\{X\mu(0,X)\mid e(X'\theta^*)\}}{1-e(X'\theta^*)}\right]\right\} + B'\mathcal{I}(\theta^*)B \\ =& \sigma_\epsilon^2 + b_{1,\epsilon}'\mathcal{I}(\theta^*)^{-1}b_{1,\epsilon} - b_{2,\epsilon}'\mathcal{I}(\theta^*)^{-1}b_{2,\epsilon}, && \text{(S7)} \end{aligned}$$

where σ_ϵ^2 is defined as the sum of terms in (S5) to (S6), and (S7) follows by plugging the expression of B,

$$\begin{aligned} 2B'b_{2,\epsilon} + B'\mathcal{I}(\theta^*)B &= 2b_{1,\epsilon}'\mathcal{I}(\theta^*)^{-1}b_{2,\epsilon} - 2b_{2,\epsilon}'\mathcal{I}(\theta^*)^{-1}b_{2,\epsilon} + (b_{1,\epsilon}+b_{2,\epsilon})'\mathcal{I}(\theta^*)^{-1}(b_{1,\epsilon}+b_{2,\epsilon}) \\ &= b_{1,\epsilon}'\mathcal{I}(\theta^*)^{-1}b_{1,\epsilon} - b_{2,\epsilon}'\mathcal{I}(\theta^*)^{-1}b_{2,\epsilon}. \end{aligned}$$

Moreover, σ_ϵ^2 can be further simplified as

$$\begin{aligned}
\sigma_\epsilon^2 = & \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2} E[\omega_\epsilon(X'\theta^*)^2 \text{var}\{\tau(X)\}] \\
& + \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2} E\left\{\omega_\epsilon(X'\theta^*)^2 \left[\left\{\frac{1-e(X'\theta^*)}{e(X'\theta^*)}\right\}^{1/2} \mu(1,X)^2\right.\right. \\
& \left.\left. + \left\{\frac{e(X'\theta^*)}{1-e(X'\theta^*)}\right\}^{1/2} \mu(0,X)\right]^2\right\} \qquad \text{(S8)} \\
& + \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}^2} E\left[\omega_\epsilon(X'\theta^*)^2 \left\{\frac{\sigma^2(1,X)}{e(X'\theta^*)} + \frac{\sigma^2(0,X)}{1-e(X'\theta^*)}\right\}\right]. \qquad \text{(S9)}
\end{aligned}$$

Finally, the Central Limit Theorem implies

$$N^{1/2}(\hat{\tau}_\epsilon - \tau_\epsilon) \to \mathcal{N}\left\{0, \sigma_\epsilon^2 + b'_{1,\epsilon}\mathcal{I}(\theta^*)^{-1}b_{1,\epsilon} - b'_{2,\epsilon}\mathcal{I}(\theta^*)^{-1}b_{2,\epsilon}\right\},$$

in distribution, as $N \to \infty$.

S1·2. *Proof of Theorem 2*

First, $\hat{\tau}^{\text{aug}}(X_i)$ can also be written as

$$\begin{aligned}
\hat{\tau}^{\text{aug}}(X_i) = & \left[\frac{A_iY_i}{e(X_i'\hat{\theta})} + \left\{1 - \frac{A_i}{e(X_i'\hat{\theta})}\right\}\hat{\mu}(1,X_i)\right] \\
& - \left[\frac{(1-A_i)Y_i}{1-e(X_i'\hat{\theta})} + \left\{1 - \frac{1-A_i}{1-e(X_i'\hat{\theta})}\right\}\hat{\mu}(0,X_i)\right].
\end{aligned}$$

Let $\hat{\mu}(A_i, X_i)$ converge to $\tilde{\mu}(A_i, X_i)$ as $N \to \infty$. If the model for $\mu(A_i, X_i)$ is correctly specified, $\tilde{\mu}(A_i, X_i) = \mu(A_i, X_i)$.

Write

$$\begin{aligned}
\hat{\tau}_\epsilon^{\text{aug}} = \hat{\tau}_\epsilon^{\text{aug}}(\hat{\theta}) & \cong \hat{\tau}_\epsilon^{\text{aug}}(\theta^*) + E\left\{\frac{\partial \hat{\tau}_\epsilon^{\text{aug}}(\theta^*)}{\partial \theta'}\right\}(\hat{\theta} - \theta^*) \\
& \cong \frac{1}{N}\sum_{i=1}^N \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\hat{\tau}^{\text{aug}}(X_i) + E\left\{\frac{\partial \hat{\tau}_\epsilon^{\text{aug}}(\theta^*)}{\partial \theta'}\right\}\mathcal{I}(\theta^*)^{-1}S(\theta^*) \\
& \cong \frac{1}{N}\sum_{i=1}^N \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\left[\frac{A_iY_i}{e(X_i)} + \left\{1 - \frac{A_i}{e(X_i)}\right\}\tilde{\mu}(1,X_i)\right] \\
& \quad - \frac{1}{N}\sum_{i=1}^N \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\left[\frac{(1-A_i)Y_i}{1-e(X_i)} + \left\{1 - \frac{1-A_i}{1-e(X_i)}\right\}\tilde{\mu}(0,X_i)\right] \\
& \quad + \tilde{B}'\frac{1}{N}\sum_{i=1}^N X_i \frac{A_i - e(X_i'\theta^*)}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}} f(X_i'\theta^*),
\end{aligned}$$

where

$$\tilde{B}' = E\left\{\frac{\partial \hat{\tau}_\epsilon^{\text{aug}}(\theta^*)}{\partial \theta'}\right\}\mathcal{I}(\theta^*)^{-1}.$$

Therefore, the asymptotic linearity of $\hat{\tau}_{\epsilon}^{\text{aug}}$ follows. Moreover,

$$
\begin{aligned}
& N^{1/2}(\hat{\tau}_{\epsilon}^{\text{aug}} - \tau_{\epsilon}) \\
\cong\ & N^{-1/2}\sum_{i=1}^{N} \frac{\omega_{\epsilon}(X_i'\theta^*)}{E\{\omega_{\epsilon}(X_i'\theta^*)\}} \left[\frac{A_i\{Y_i - \mu(A_i, X_i)\}}{e(X_i'\theta^*)} - \frac{(1-A_i)\{Y_i - \mu(A_i, X_i)\}}{1 - e(X_i'\theta^*)}\right] \\
& + N^{-1/2}\sum_{i=1}^{N} \left[\frac{\omega_{\epsilon}(X_i'\theta^*)}{E\{\omega_{\epsilon}(X_i'\theta^*)\}}\tau(X_i) - \tau_{\epsilon}\right] \\
& + N^{-1/2}\sum_{i=1}^{N} \tilde{B}'X_i \frac{A_i - e(X_i'\theta^*)}{e(X_i'\theta^*)\{1 - e(X_i'\theta^*)\}} f(X_i'\theta^*), \\
& + N^{-1/2}\sum_{i=1}^{N} \frac{\omega_{\epsilon}(X_i'\theta^*)}{E\{\omega_{\epsilon}(X'\theta^*)\}} \left\{1 - \frac{A_i}{e(X_i)}\right\} \{\tilde{\mu}(1, X_i) - \mu(1, X_i)\} \\
& + N^{-1/2}\sum_{i=1}^{N} \frac{\omega_{\epsilon}(X_i'\theta^*)}{E\{\omega_{\epsilon}(X'\theta^*)\}} \left\{1 - \frac{1 - A_i}{1 - e(X_i)}\right\} \{\tilde{\mu}(0, X_i) - \mu(0, X_i)\} \\
=\ & \tilde{T}_3 + \tilde{T}_0 + \tilde{T}_1 + \tilde{T}_2,
\end{aligned}
$$

where $\tilde{T}_3 = T_3$ is defined in (S4),

$$
\begin{aligned}
\tilde{T}_0 &= N^{-1/2}\sum_{i=1}^{N} \left[\frac{\omega_{\epsilon}(X_i'\theta^*)}{E\{\omega_{\epsilon}(X_i'\theta^*)\}}\tau(X_i) - \tau_{\epsilon}\right], \\
\tilde{T}_1 &= N^{-1/2}\sum_{i=1}^{N} \tilde{B}'X_i \frac{A_i - e(X_i'\theta)}{e(X_i'\theta)\{1 - e(X_i'\theta)\}} f(X_i'\theta^*), \\
\tilde{T}_2 &= N^{-1/2}\sum_{i=1}^{N} \frac{\omega_{\epsilon}(X_i'\theta^*)}{E\{\omega_{\epsilon}(X'\theta^*)\}} \left\{1 - \frac{A_i}{e(X_i)}\right\} \{\tilde{\mu}(1, X_i) - \mu(1, X_i)\} \\
&\quad + N^{-1/2}\sum_{i=1}^{N} \frac{\omega_{\epsilon}(X_i'\theta^*)}{E\{\omega_{\epsilon}(X'\theta^*)\}} \left\{1 - \frac{1 - A_i}{1 - e(X_i)}\right\} \{\tilde{\mu}(0, X_i) - \mu(0, X_i)\}.
\end{aligned}
$$

By the same argument as in the proof of Theorem 1, $E(\tilde{T}_j) = 0$, for $j = 0, \ldots, 3$, and $\text{cov}(\tilde{T}_j, \tilde{T}_k) = 0$ for all $j \neq k$ expect $\text{cov}(\tilde{T}_1, \tilde{T}_2)$. Moreover,

$$
\begin{aligned}
& \text{var}(\tilde{T}_3) + \text{var}(\tilde{T}_0) + \text{var}(\tilde{T}_1) + \text{var}(\tilde{T}_2) + 2\text{cov}(\tilde{T}_1, \tilde{T}_2) \\
=\ & \frac{1}{E\{\omega_{\epsilon}(X'\theta^*)\}^2} E\left[\omega_{\epsilon}(X'\theta^*)^2 \left\{\frac{\sigma^2(1, X)}{e(X'\theta^*)} + \frac{\sigma^2(0, X)}{1 - e(X'\theta^*)}\right\}\right] \\
& + \frac{1}{E\{\omega_{\epsilon}(X'\theta^*)\}^2} E[\omega_{\epsilon}(X'\theta^*)^2 \text{var}\{\tau(X)\}] + \tilde{B}'\mathcal{I}(\theta^*)\tilde{B} \\
& + \frac{1}{E\{\omega_{\epsilon}(X'\theta^*)\}^2} E\left\{\omega_{\epsilon}(X'\theta^*)^2 \left[\left\{\frac{1 - e(X'\theta^*)}{e(X'\theta^*)}\right\}^{1/2} \{\tilde{\mu}(1, X) - \mu(1, X)\}\right.\right.
\end{aligned}
$$

$$
\begin{aligned}
&- \left\{ \frac{e(X'\theta^*)}{1-e(X'\theta^*)} \right\}^{1/2} \{\tilde{\mu}(0,X) - \mu(0,X)\} \Bigg]^2 \Bigg\} \\
&+ \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}} \tilde{B}' E \left[\omega_\epsilon(X'\theta^*) X f(X'\theta^*) \left\{ -\frac{\tilde{\mu}(1,X_i) - \mu(1,X_i)}{e(X_i)} \right\} \right] \\
&+ \frac{1}{E\{\omega_\epsilon(X'\theta^*)\}} \tilde{B}' E \left[\omega_\epsilon(X'\theta^*) X f(X'\theta^*) \left\{ \frac{\tilde{\mu}(0,X) - \mu(0,X)}{1-e(X_i)} \right\} \right] \\
&= \tilde{\sigma}_\epsilon^2 + \tilde{B}'\mathcal{I}(\theta^*)\tilde{B} + \tilde{B}'(C_0 - C_1) \\
&= \tilde{\sigma}_\epsilon^2 + b_{1\epsilon}'\mathcal{I}(\theta^*)^{-1} b_{1\epsilon} + (C_0 + C_1)'\mathcal{I}(\theta^*)^{-1}(C_0 + C_1) + \tilde{B}'(C_0 - C_1)
\end{aligned}
$$

where $\tilde{\sigma}_\epsilon^2$, C_0 and C_1 are defined in Theorem 2. Because

$$
\begin{aligned}
\frac{\partial \hat{\tau}_\epsilon^{\text{aug}}(\theta^*)}{\partial \theta} = & \frac{1}{N}\sum_{i=1}^N \frac{\partial}{\partial\theta} \left[\frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}} \right] \hat{\tau}^{\text{aug}}(X_i) \\
& - \frac{1}{N}\sum_{i=1}^N \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}} X_i f(X_i'\theta^*) \frac{A_i\{Y_i - \tilde{\mu}(A_i,X_i)\}}{e(X_i'\theta^*)^2} \\
& - \frac{1}{N}\sum_{i=1}^N \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X_i'\theta^*)\}} X_i f(X_i'\theta^*) \frac{(1-A_i)\{Y_i - \tilde{\mu}(A_i,X_i)\}}{\{1-e(X_i'\theta^*)\}^2},
\end{aligned}
$$

we have

$$
\begin{aligned}
E\left\{ \frac{\partial \hat{\tau}_\epsilon^{\text{aug}}(\theta^*)}{\partial\theta} \right\} = & E\left(\frac{\partial}{\partial\theta} \left[\frac{\omega_\epsilon(X'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}} \right] \tau(X) \right) \\
& - \frac{1}{E\{\omega_\epsilon(X_i'\theta^*)\}} E\left\{ \omega_\epsilon(X_i'\theta^*) X f(X'\theta^*) \frac{\mu(1,X) - \tilde{\mu}(1,X)}{e(X'\theta^*)} \right\} \\
& - \frac{1}{E\{\omega_\epsilon(X_i'\theta^*)\}} E\left\{ \omega_\epsilon(X_i'\theta^*) X f(X'\theta^*) \frac{\mu(0,X) - \tilde{\mu}(0,X)}{1-e(X'\theta^*)} \right\} \\
& = b_{1,\epsilon} - C_0 - C_1 \cdot
\end{aligned}
$$

Therefore,

$$
N^{1/2}(\hat{\tau}_\epsilon^{\text{aug}} - \tau_\epsilon) \to \mathcal{N}\left\{ 0, \tilde{\sigma}_\epsilon^2 + b_{1\epsilon}'\mathcal{I}(\theta^*)^{-1} b_{1\epsilon} + (C_0 + C_1)'\mathcal{I}(\theta^*)^{-1}(C_0 + C_1) + \tilde{B}'(C_0 - C_1) \right\},
$$

in distribution, as $N \to \infty$.

S1·3. *Proof of Remark 1*

We show that

$$
b_{1,\epsilon} = E\left[\frac{\partial}{\partial\theta} \left\{ \omega_\epsilon(\theta^*)^{-1} \omega_\epsilon(X'\theta^*) \right\} \tau(X) \right]
$$

goes to zero, as $\epsilon \to 0$.

We note

$$
\frac{\partial}{\partial\theta} \left\{ \omega_\epsilon(\theta^*)^{-1} \omega_\epsilon(X'\theta^*) \right\} = \omega_\epsilon(\theta^*)^{-2} \left[\frac{\partial \omega_\epsilon(X'\theta^*)}{\partial\theta} E\{\omega_\epsilon(X'\theta^*)\} - E\left\{ \frac{\partial \omega_\epsilon(X'\theta^*)}{\partial\theta} \right\} \omega_\epsilon(X'\theta^*) \right],
$$

where

$$
\begin{aligned}
\frac{\partial \omega_\epsilon(X'\theta^*)}{\partial \theta} &= \frac{\partial}{\partial \theta}\left[\Phi_\epsilon\left\{e(X'\theta^*)-\alpha_1\right\}\Phi_\epsilon\left\{\alpha_2-e(X'\theta^*)\right\}\right] \\
&= \phi_\epsilon\left\{e(X'\theta^*)-\alpha_1\right\}\Phi_\epsilon\left\{\alpha_2-e(X'\theta^*)\right\}f(X'\theta^*)X \\
&\quad -\Phi_\epsilon\left\{e(X'\theta^*)-\alpha_1\right\}\phi_\epsilon\left\{\alpha_2-e(X'\theta^*)\right\}f(X'\theta^*)X,
\end{aligned}
$$

and $\phi_\epsilon(x)=\mathrm{d}\Phi_\epsilon(x)/\mathrm{d}x$. As $\epsilon \to 0$, $\phi_\epsilon(x)\to 0$ implies that $b_{1,\epsilon}$ goes to 0.

S1·4. *Proof of Remark 4*

We write

$$
\begin{aligned}
\hat{\tau} &= \hat{\tau}(\hat{\theta}) \\
&\cong \hat{\tau}(\theta^*) + E\left\{\frac{\partial \hat{\tau}(\theta^*)}{\partial \theta'}\right\}(\hat{\theta}-\theta^*) \\
&\cong \frac{1}{N}\sum_{i=1}^{N}\frac{1\{\alpha_1 \le e(X_i'\theta^*) \le \alpha_2\}}{\mathrm{pr}\{\alpha \le e(X'\theta^*) \le 1-\alpha\}}\left\{\frac{A_iY_i}{e(X_i'\theta^*)}-\frac{(1-A_i)Y_i}{1-e(X_i'\theta^*)}\right\} + E\left\{\frac{\partial \hat{\tau}(\theta^*)}{\partial \theta'}\right\}\mathcal{I}(\theta^*)^{-1}S(\theta^*) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1\{\alpha_1 \le e(X_i'\theta^*) \le \alpha_2\}}{\mathrm{pr}\{\alpha_1 \le e(X'\theta^*) \le \alpha_2\}}\left\{\frac{A_iY_i}{e(X_i'\theta^*)}-\frac{(1-A_i)Y_i}{1-e(X_i'\theta^*)}\right\} \\
&\quad + E\left\{\frac{\partial \hat{\tau}(\theta^*)}{\partial \theta'}\right\}\frac{1}{N}\sum_{i=1}^{N}X_i\frac{A_i-e(X_i'\theta^*)}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}}f(X_i'\theta^*)\cdot
\end{aligned}
$$

Let $\mathcal{S}=\{X : e(X'\theta^*)=\alpha_1 \text{ or } \alpha_2\}$. If $\mathrm{pr}(X\in\mathcal{S})=0$, then

$$
\begin{aligned}
E\left\{\frac{\partial \hat{\tau}(\theta^*)}{\partial \theta'}\right\} &= E\left(\frac{\partial}{\partial \theta'}\left[\frac{1\{\alpha_1 \le e(X'\theta^*) \le \alpha_2\}}{\mathrm{pr}\{\alpha_1 \le e(X'\theta^*) \le \alpha_2\}}\right]\left\{\frac{AY}{e(X'\theta^*)}-\frac{(1-A)Y}{1-e(X'\theta^*)}\right\}\right) \\
&\quad + E\left[\frac{1\{\alpha_1 \le e(X'\theta^*) \le \alpha_2\}}{\mathrm{pr}\{\alpha_1 \le e(X'\theta^*) \le \alpha_2\}}\frac{\partial}{\partial \theta'}\left\{\frac{AY}{e(X'\theta^*)}-\frac{(1-A)Y}{1-e(X'\theta^*)}\right\}\right]
\end{aligned}
$$

is finite and well-defined, because the only possible problem that prevents the use of the bootstrap is the derivative of the indicator function with respect to θ, which, however, has zero measure.

Therefore, $\hat{\tau}$ is asymptotically linear. According to Shao & Tu (2012), the bootstrap can be used to estimate $\mathrm{var}(\hat{\tau})$. A similar discussion applies to $\hat{\tau}^{\mathrm{aug}}$.

S1·5. *Proof of Remark 5*

We write

$$\begin{aligned}
\hat{\tau}_\epsilon &= \hat{\tau}_\epsilon(\hat{\theta}) \\
&\cong \hat{\tau}_\epsilon(\theta^*) + E\left\{\frac{\partial \hat{\tau}_\epsilon(\theta^*)}{\partial \theta'}\right\}(\hat{\theta} - \theta^*) \\
&\cong \frac{1}{N}\sum_{i=1}^{N} \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\left\{\frac{A_iY_i}{e(X_i'\theta^*)} - \frac{(1-A_i)Y_i}{1-e(X_i'\theta^*)}\right\} + E\left\{\frac{\partial \hat{\tau}_\epsilon(\theta^*)}{\partial \theta'}\right\}\mathcal{J}(\theta^*)^{-1}S(\theta^*) \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\left\{\frac{A_iY_i}{e(X_i'\theta^*)} - \frac{(1-A_i)Y_i}{1-e(X_i'\theta^*)}\right\} \\
&+\Gamma'\frac{1}{N}\sum_{i=1}^{N} X_i \frac{A_i - e(X_i'\theta^*)}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}} f(X_i'\theta^*),
\end{aligned}$$

where

$$\Gamma' = E\left\{\frac{\partial \hat{\tau}_\epsilon(\theta^*)}{\partial \theta'}\right\}\mathcal{J}(\theta^*)^{-1}.$$

Therefore, the asymptotic linearity of $\hat{\tau}_\epsilon$ follows.

Write

$$\begin{aligned}
\hat{\tau}_\epsilon^{\text{aug}} &= \hat{\tau}_\epsilon^{\text{aug}}(\hat{\theta}) \cong \hat{\tau}_\epsilon^{\text{aug}}(\theta^*) + E\left\{\frac{\partial \hat{\tau}_\epsilon^{\text{aug}}(\theta^*)}{\partial \theta'}\right\}(\hat{\theta} - \theta^*) \\
&\cong \frac{1}{N}\sum_{i=1}^{N} \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\hat{\tau}^{\text{aug}}(X_i) + E\left\{\frac{\partial \hat{\tau}_\epsilon^{\text{aug}}(\theta^*)}{\partial \theta'}\right\}\mathcal{J}(\theta^*)^{-1}S(\theta^*) \\
&\cong \frac{1}{N}\sum_{i=1}^{N} \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\left[\frac{A_iY_i}{e(X_i)} + \left\{1 - \frac{A_i}{e(X_i)}\right\}\tilde{\mu}(1, X_i)\right] \\
&\quad -\frac{1}{N}\sum_{i=1}^{N} \frac{\omega_\epsilon(X_i'\theta^*)}{E\{\omega_\epsilon(X'\theta^*)\}}\left[\frac{(1-A_i)Y_i}{1-e(X_i)} + \left\{1 - \frac{1-A_i}{1-e(X_i)}\right\}\tilde{\mu}(0, X_i)\right] \\
&\quad +\tilde{\Gamma}'\frac{1}{N}\sum_{i=1}^{N} X_i \frac{A_i - e(X_i'\theta^*)}{e(X_i'\theta^*)\{1-e(X_i'\theta^*)\}} f(X_i'\theta^*),
\end{aligned}$$

where

$$\tilde{\Gamma}' = E\left\{\frac{\partial \hat{\tau}_\epsilon^{\text{aug}}(\theta^*)}{\partial \theta'}\right\}\mathcal{J}(\theta^*)^{-1}.$$

Therefore, the asymptotic linearity of $\hat{\tau}_\epsilon^{\text{aug}}$ follows.

The asymptotic linearity of the weighting estimators allows for using the bootstrap to construct confidence intervals.

S2. SIMULATION

We assess the performance of the new weighting estimators of the average treatment effect over a target population. We consider $X = (X_1, X_2, X_3, X_4, X_5, X_6)'$, where X_1, X_2, and X_3 are multivariate normal with means $(0, 0, 0)$, variances $(2, 1, 1)$ and covariances $(1, -1, -0{\cdot}5)$, $X_4 \sim \mathrm{Uniform}[-3, 3]$, $X_5 \sim \chi^2_1$, and $X_6 \sim \mathrm{Bernoulli}(0{\cdot}5)$. The treatment indicator A is generated from Bernoulli$\{e(X)\}$. We consider four propensity score models:

(P1) $e(X) = \mathrm{logit}\{0{\cdot}1(X_1 + X_2 + X_3 + X_4 + X_5 + X_6)\}$,
(P2) $e(X) = \mathrm{logit}\{0{\cdot}8(X_1 + X_2 + X_3 + X_4 + X_5 + X_6)\}$,
(P3) $e(X) = \mathrm{logit}\{0{\cdot}1(X_1 + X_2^2 + X_3^2 + X_4 + X_5 + X_6)\}$,
(P4) $e(X) = \mathrm{logit}\{0{\cdot}8(X_1 + X_2^2 + X_3^2 + X_4 + X_5 + X_6)\}$;

(P1) and (P3) represent weak separations, and (P2) and (P4) represent strong separations of propensity score distributions between the treatment and control groups. We consider both linear and nonlinear outcome models:

(O1) $Y(a) = a(X_1 + X_2 + X_3 - X_4 + X_5 + X_6) + \eta$, with $\eta \sim \mathcal{N}(0, 1)$, for $a = 0, 1$,
(O2) $Y(a) = a(X_1 + X_2 + X_3)^2 + \eta$, with $\eta \sim \mathcal{N}(0, 1)$, for $a = 0, 1$.

The target population is represented by $\mathcal{O} = \{X : 0{\cdot}1 \le e(X) \le 0{\cdot}9\}$, and the estimand of interest is the average treatment effect over the target population $\tau(\mathcal{O})$.

We consider the weighting estimators with the indicator and smooth weight functions, and $\tau(\mathcal{O}) = \{\sum_{i=1}^N 1(X_i \in \mathcal{O})\}^{-1} \sum_{i=1}^N 1(X_i \in \mathcal{O})\{Y_i(1) - Y_i(0)\}$ for benchmark comparison with $N = 500$. The propensity scores are estimated by a logistic regression model with linear predictors X. Therefore, the propensity score model is correctly specified under (P1) and (P2) but misspecified under (P3) and (P4). For the augmented weighting estimators, $\mu(a, X)$ is estimated by a simple linear regression of Y on X, separately for $A = 0, 1$. Therefore, the outcome regression model is correctly specified under (O1) but misspecified under (O2).

Table S1 shows the simulation results. Under Scenarios i, ii, v and vi when the propensity score model is correctly specified, the weighting estimators are nearly unbiased for $\tau(\mathcal{O})$, and the augmented weighting estimators are nearly unbiased and more efficient than the simple weighting estimators. However, under Scenarios iii, iv, vii and viii when the propensity score model is misspecified, all estimators are biased, even when the outcome regression model is correctly specified. The weighting estimators with the smooth weight function, $\hat{\tau}_\epsilon$ and $\hat{\tau}_\epsilon^{\mathrm{aug}}$, show slightly smaller variances than the counterparts with the indicator weight function, $\hat{\tau}$ and $\hat{\tau}^{\mathrm{aug}}$. Moreover, as ϵ becomes smaller, the performances of $\hat{\tau}_\epsilon$ and $\hat{\tau}_\epsilon^{\mathrm{aug}}$ become closer to those of $\hat{\tau}$ and $\hat{\tau}^{\mathrm{aug}}$. The bootstrap works well with the variance estimates close to the true variances for all estimators.

S3. AVERAGE TREATMENT EFFECT ON THE TREATED

S3·1. *Notation, Assumptions and Extension of Crump et al. (2009)*

Another estimand of interest is the average treatment effect for the treated $\tau_{\mathrm{ATT}} = E\{Y(1) - Y(0) \mid A = 1\} = E\{\tau(X) \mid A = 1\}$. The outcome distribution for the treated is empirically identifiable, because $E\{Y(1) \mid A = 1\} = E(Y \mid A = 1)$. Therefore, Assumptions 1 and 2 can be weakened (Heckman et al., 1997).

*Assumption S*1. $Y(0) \perp\!\!\!\perp A \mid X$.

*Assumption S*2. There exists a constant c such that with probability 1, $e(X) \le c < 1$.

Table S1. *Results: mean, variance*$\times 100$ *(var), and variance estimate*$\times 100$ *(ve) based on* 100 *bootstrap replicates under eight combinations of the outcome and propensity score models: for example, (O1)&(P1) means Outcome Model (O1) and Propensity Score Model (P1)*

Scenario		i (O1)&(P1)			ii (O1)&(P2)			iii (O1)&(P3)			iv (O1)&(P4)		
	ϵ	mean	var	ve	mean	var	ve	mean	var	ve	mean	var	ve
$\tau(\mathcal{O})$		$1{\cdot}46$			$1{\cdot}33$			$1{\cdot}44$			$1{\cdot}37$		
$\hat{\tau}(\hat{\theta})$	–	$1{\cdot}45$	$3{\cdot}4$	$3{\cdot}4$	$1{\cdot}33$	$4{\cdot}7$	$5{\cdot}2$	$1{\cdot}48$	$2{\cdot}9$	$2{\cdot}8$	$1{\cdot}45$	$4{\cdot}0$	$4{\cdot}1$
$\hat{\tau}^{\text{aug}}(\hat{\theta})$	–	$1{\cdot}46$	$2{\cdot}8$	$2{\cdot}7$	$1{\cdot}32$	$3{\cdot}4$	$3{\cdot}4$	$1{\cdot}50$	$2{\cdot}6$	$2{\cdot}5$	$1{\cdot}49$	$3{\cdot}3$	$3{\cdot}2$
$\hat{\tau}_{\epsilon}(\hat{\theta})$	10^{-4}	$1{\cdot}45$	$3{\cdot}3$	$3{\cdot}3$	$1{\cdot}33$	$4{\cdot}5$	$4{\cdot}7$	$1{\cdot}48$	$2{\cdot}8$	$2{\cdot}8$	$1{\cdot}45$	$3{\cdot}9$	$3{\cdot}8$
$\hat{\tau}_{\epsilon}^{\text{aug}}(\hat{\theta})$	10^{-4}	$1{\cdot}46$	$2{\cdot}8$	$2{\cdot}7$	$1{\cdot}33$	$3{\cdot}4$	$3{\cdot}3$	$1{\cdot}50$	$2{\cdot}6$	$2{\cdot}5$	$1{\cdot}49$	$3{\cdot}3$	$3{\cdot}1$
$\hat{\tau}_{\epsilon}(\hat{\theta})$	10^{-5}	$1{\cdot}45$	$3{\cdot}4$	$3{\cdot}3$	$1{\cdot}33$	$4{\cdot}6$	$5{\cdot}0$	$1{\cdot}48$	$2{\cdot}9$	$2{\cdot}8$	$1{\cdot}45$	$3{\cdot}9$	$4{\cdot}0$
$\hat{\tau}^{\text{aug}}(\hat{\theta})$	10^{-5}	$1{\cdot}46$	$2{\cdot}8$	$2{\cdot}7$	$1{\cdot}32$	$3{\cdot}4$	$3{\cdot}4$	$1{\cdot}50$	$2{\cdot}6$	$2{\cdot}5$	$1{\cdot}49$	$3{\cdot}3$	$3{\cdot}2$
		v (O2)&(P1)			vi (O2)&(P2)			vii (O2)&(P3)			viii (O2)&(P4)		
$\tau(\mathcal{O})$		$7{\cdot}58$			$6{\cdot}69$			$7{\cdot}62$			$5{\cdot}96$		
$\hat{\tau}(\hat{\theta})$	–	$7{\cdot}58$	$94{\cdot}0$	$89{\cdot}1$	$6{\cdot}69$	$89{\cdot}8$	$98{\cdot}1$	$8{\cdot}75$	$92{\cdot}0$	$91{\cdot}2$	$8{\cdot}93$	$142{\cdot}0$	$138{\cdot}1$
$\hat{\tau}^{\text{aug}}(\hat{\theta})$	–	$7{\cdot}59$	$85{\cdot}4$	$76{\cdot}5$	$6{\cdot}67$	$79{\cdot}2$	$84{\cdot}2$	$8{\cdot}82$	$84{\cdot}9$	$79{\cdot}3$	$9{\cdot}06$	$122{\cdot}6$	$109{\cdot}6$
$\hat{\tau}_{\epsilon}(\hat{\theta})$	10^{-4}	$7{\cdot}57$	$88{\cdot}6$	$84{\cdot}1$	$6{\cdot}70$	$85{\cdot}3$	$89{\cdot}7$	$8{\cdot}75$	$91{\cdot}1$	$88{\cdot}3$	$8{\cdot}94$	$134{\cdot}2$	$128{\cdot}4$
$\hat{\tau}_{\epsilon}^{\text{aug}}(\hat{\theta})$	10^{-4}	$7{\cdot}58$	$82{\cdot}7$	$74{\cdot}7$	$6{\cdot}68$	$76{\cdot}6$	$79{\cdot}4$	$8{\cdot}82$	$84{\cdot}4$	$78{\cdot}4$	$9{\cdot}07$	$119{\cdot}0$	$105{\cdot}5$
$\hat{\tau}_{\epsilon}(\hat{\theta})$	10^{-5}	$7{\cdot}57$	$92{\cdot}0$	$87{\cdot}3$	$6{\cdot}69$	$88{\cdot}8$	$95{\cdot}3$	$8{\cdot}75$	$91{\cdot}9$	$90{\cdot}2$	$8{\cdot}93$	$140{\cdot}0$	$134{\cdot}8$
$\hat{\tau}_{\epsilon}^{\text{aug}}(\hat{\theta})$	10^{-5}	$7{\cdot}59$	$84{\cdot}1$	$75{\cdot}9$	$6{\cdot}68$	$78{\cdot}7$	$82{\cdot}5$	$8{\cdot}82$	$84{\cdot}7$	$79{\cdot}0$	$9{\cdot}06$	$121{\cdot}7$	$108{\cdot}2$

A simple weighting estimator (Hirano et al., 2003) is

$$\hat{\tau}_{\text{ATT}} = \frac{\sum_{i=1}^{N} A_i Y_i}{\sum_{i=1}^{N} e(X_i'\hat{\theta})} - \frac{\sum_{i=1}^{N}(1-A_i)Y_i e(X_i'\hat{\theta})/\{1-e(X_i'\hat{\theta})\}}{\sum_{i=1}^{N} e(X_i'\hat{\theta})} = \frac{\sum_{i=1}^{N} e(X_i'\hat{\theta})\hat{\tau}(X_i)}{\sum_{i=1}^{N} e(X_i'\hat{\theta})},$$

which is a special case of the weighting estimator (4) by choosing $\omega(X_i'\hat{\theta}) = e(X_i'\hat{\theta})$. Analogously, we propose the augmented weighting estimator

$$\hat{\tau}_{\text{ATT}}^{\text{aug}} = \frac{\sum_{i=1}^{N} e(X_i'\hat{\theta})\hat{\tau}^{\text{aug}}(X_i)}{\sum_{i=1}^{N} e(X_i'\hat{\theta})}. \tag{S10}$$

*Remark S*1. An existing augmented weighting estimator for τ_{ATT} is

$$\frac{\sum_{i=1}^{N} A_i Y_i}{\sum_{i=1}^{N} A_i} - \frac{1}{\sum_{i=1}^{N} A_i} \sum_{i=1}^{N} \frac{(1-A_i)e(X_i'\hat{\theta})Y_i + \hat{\mu}(0,X_i)\{A_i - e(X_i'\hat{\theta})\}}{1-e(X_i'\hat{\theta})},$$

which is doubly robust in the sense that the estimator is consistent for τ_{ATT} if either $\mu(0,X)$ or $e(X)$ is correctly specified (Mercatanti & Li, 2014). See also Shinozaki & Matsuyama (2015) and Zhao & Percival (2017) for other forms of doubly robust estimators for τ_{ATT}. The advantage of these estimators is that they do not require estimating $\mu(1,X)$ unlike (S10). However, (S10) is locally efficient in the sense that if the outcome and propensity score models are correctly specified, the asymptotic variance of (S10) achieves the efficiency bound. To show this, we recognize that (S10) is (3) with $\omega_{\epsilon}(X'\hat{\theta})$ replaced by $e(X'\hat{\theta})$. Let $p_1 = E\{e(X'\theta^*)\}$. Following a similar derivation as in Theorem 2, with correctly specified propensity score and outcome models, the asymptotic variance of (S10) is

$$p_1{}^{-2}E[e(X'\theta^*)^2\mathrm{var}\{\tau(X)\}] + p_1^{-2}E\left[e(X'\theta^*)^2\left\{\frac{\sigma^2(1,X)}{e(X)} + \frac{\sigma^2(0,X)}{1-e(X)}\right\}\right],$$

which is the efficiency bound for τ_{ATT} (Hahn, 1998).

There is a limited literature dealing with lack of overlap for τ_{ATT} when Assumption S2 may not hold. Dehejia & Wahba (1999) suggested dropping control units with estimated propensity scores lower than the smallest value of the estimated propensity score among the treated units. Heckman et al. (1997) and Smith & Todd (2005) proposed to discard units with covariate values at which the estimated density is below some threshold. However, few formal results have been established on properties of these procedures.

Similar to Crump et al. (2009), if $\sigma^2(1,X) = \sigma^2(0,X)$, we can show that the optimal overlap for estimating τ_{ATT} is of the form $\mathcal{O} = \{X : 1 - e(X) \geq \alpha\}$ for some α, for which the estimators have smallest asymptotic variance. Intuitively, for the treated units with $e(X)$ close to 1, there are no similar units in the control group that can provide information to infer $Y(0)$ for these treated units. Statistically, the control units with $e(X)$ close to 1 contribute large weights. Therefore, it is reasonable to drop these units with $e(X)$ close to 1. By restricting to the subpopulation, the estimand of interest becomes $\tau_{\mathrm{ATT}}(\mathcal{O}) = E\{\tau(X) \mid A = 1, X \in \mathcal{O}\}$. Below, we formalize this argument.

S3·2. *Theory of trimming for the average treatment effect on the treated*

Define a general weighting average treatment effect,

$$\tau_\omega(\mathcal{O}) = \frac{\sum_{i:X_i\in\mathcal{O}} \omega(X_i)\tau(X_i)}{\sum_{i:X_i\in\mathcal{O}} \omega(X_i)}. \tag{S11}$$

According to the technique report in 2006 prior to Crump et al. (2009), the efficiency bound for $\tau_\omega(\mathcal{O})$ is

$$V_\omega(\mathcal{O}) = \frac{1}{[E\{\omega(X) \mid X \in \mathcal{O}\}]^2} E\left[\omega(X)^2\left\{\frac{\sigma^2(1,X)}{e(X)} + \frac{\sigma^2(0,X)}{1-e(X)}\right\} \mid X \in \mathcal{O}\right]. \tag{S12}$$

Crump et al. (2009) showed that the optimal set with which $\hat{\tau}_\omega(\mathcal{O})$ achieves the smallest asymptotic variance over all choices of $\mathcal{O}$ is

$$\mathcal{O} = \left\{x : \omega(x)\left\{\frac{\sigma^2(1,x)}{e(x)} + \frac{\sigma^2(0,x)}{1-e(x)}\right\} \leq \gamma\right\}, \tag{S13}$$

where γ is defined through the following equation:

$$\gamma = 2 \times \frac{E\left[\omega^2(X)\left\{\frac{\sigma^2(1,X)}{e(X)} + \frac{\sigma^2(0,X)}{1-e(X)}\right\} \mid \omega(X)\left\{\frac{\sigma^2(1,X)}{e(X)} + \frac{\sigma^2(0,X)}{1-e(X)}\right\} \leq \gamma\right]}{E\left[\omega(X) \mid \omega(X)\left\{\frac{\sigma^2(1,X)}{e(X)} + \frac{\sigma^2(0,X)}{1-e(X)}\right\} \leq \gamma\right]}. \tag{S14}$$

The weighting estimator for the average treatment effect on the treated is (S11) with $\omega(X) = e(X)$. Assuming that $\sigma^2(1,X) = \sigma^2(0,X) = \sigma^2$, the optimal set (S13) reduces to $\mathcal{O} = \{X : 1 - e(X) \geq \alpha\}$ with the cut-off value $\alpha = \sigma^2/\gamma$. In practice, α can be determined by the smallest value of α that satisfy the empirical estimate of (S14):

$$\frac{1}{\alpha} = 2 \times \frac{\sum_{i=1}^N e^2(X_i)\left\{\frac{1}{e(X_i)} + \frac{1}{1-e(X_i)}\right\} 1\{1 - e(X_i) \geq \alpha\}}{\sum_{i=1}^N e(X_i) 1\{1 - e(X_i) \geq \alpha\}}.$$

The choice of α in $\mathcal{O} = \{X : 1 - e(X) \geq \alpha\}$ has two opposite effects on the asymptotic variance in (S12). On the one hand, as α increases, we reduce the denominator of the right hand side of (S12), $[E\{\omega(X) \mid X \in \mathcal{O}\}]^2 = E[\{e(X) \mid X \in \mathcal{O}\}]^2$, and therefore increase the asymptotic variance. On the other hand, as α increases, we decrease the numerator of the right hand side of (S12),

$$E\left[\omega(X)^2\left\{\frac{\sigma^2(1,X)}{e(X)} + \frac{\sigma^2(0,X)}{1-e(X)}\right\} \mid X \in \mathcal{O}\right]$$

$$= E\left[e(X)\sigma^2(1,X) + \frac{e(X)^2\sigma^2(0,X)}{1-e(X)} \mid X \in \mathcal{O}\right],$$

and therefore decrease the asymptotic variance. The optimal value of α balances the two effects.

S4. The National Health and Nutrition Examination Survey Data

S4·1. *Interpretation of the trimmed population for the average smoking effect*

To interpret the target population $\mathcal{O} = \{X : 0{\cdot}05 \leq e(X) \leq 0{\cdot}6\}$ for the average smoking effect, an effective strategy is to first present summary statistics for covariates X in the original population. See Table S2 for the description of the covariates. Then, from the fitted logistic regression for $e(X)$, the target population can be represented by $\{X : -2{\cdot}944 \leq -9 - 0{\cdot}018\times$Age$+0{\cdot}841\times$Male$+8{\cdot}972\times$Edu.lt9$+9{\cdot}331\times$Edu.9to11$+8{\cdot}875\times$Edu.hischl$+8{\cdot}546\times$Edu.somecol$+7{\cdot}118\times$Edu.college$-0{\cdot}254\times$Income-$0{\cdot}145\times$Income.mis$+0{\cdot}689\times$White$-0{\cdot}067\times$Black$-1{\cdot}639\times$Mexicanam$-1{\cdot}304\times$Otherhispan$\leq -0{\cdot}405\}$.

S4·2. *Analysis for the average smoking effect on the smokers*

For the average smoking effect on the smokers, we drop subjects with estimated propensity scores greater than 0·7. This removes 36 subjects, with 29 smokers and 7 non-smokers. Thus, the analysis sample includes 3304 subjects, with 650 smokers and 2654 non-smokers. Following the main paper for the average treatment effect, we consider the weighting estimators using both the indicator and smooth weight functions with $\epsilon = 10^{-4}$ and $\epsilon = 10^{-5}$. For the augmented weighting estimator, we consider the outcome model to be a linear regression model adjusting for all covariates, separately for $A = 0, 1$.

Table S3 shows the results from the estimators for the average smoking effect on the smokers based on the trimmed samples. The weighting estimators with the smooth weight function are close to the counterparts with the indicator weight function, but have slightly smaller estimated standard errors. The smooth weighting estimators are insensitive to the choice of $\epsilon\cdot$ From the results, on average, smoking increases the lead level in blood at least by 0·79 ug/dl for smokers with $e(X) \leq 0{\cdot}7$.

References

CRUMP, R. K., HOTZ, V. J., IMBENS, G. W. & MITNIK, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96**, 187–99.

DEHEJIA, R. H. & WAHBA, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *J. Am. Statist. Assoc.* **94**, 1053–62.

HAHN, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66**, 315–31.

HECKMAN, J. J., ICHIMURA, H. & TODD, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Rev. Econ. Studies* **64**, 605–54.

HIRANO, K., IMBENS, G. W. & RIDDER, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71**, 1161–89.

Table S2. *Descriptive statistics for covariates X in the original population in the National Health and Nutrition Examination Survey Data*

Covariate	
Age, interquantile range	$[35, 63]$
Income-to-poverty level, interquantile range	$[1{\cdot}18, 3{\cdot}77]$
Missing, %	8·5
Male, %	41·7
Education, %	
Less than 9th grade (Edu.lt9)	13·3
$9-11$th grade (Edu.9to11)	16·5
High school graduate (Edu.hischl)	25·1
Some college (Edu.somecol)	25·4
College (Edu.college)	19·6
Unknown	0·1
Race, %	
White	45·8
Black	19·1
Mexican American (Mexicanam)	18·3
Other Hispanic (Otherhispan)	11·6
Other races	5·2

Table S3. *Estimate of the average smoking effect on the smokers, estimated standard error based on* 100 *bootstrap replicates, and* 95% *confidence interval*

	ϵ	estimate	s.e.	95% c.i.		estimate	s.e.	95% c.i.
$\hat{\tau}_{\mathrm{ATT}}(\hat{\theta})$	–	0·796	0·103	$(0{\cdot}591, 1{\cdot}001)$	$\hat{\tau}^{\mathrm{aug}}_{\mathrm{ATT}}(\hat{\theta})$	0·793	0·088	$(0{\cdot}616, 0{\cdot}970)$
$\hat{\tau}_{\mathrm{ATT},\epsilon}(\hat{\theta})$	10^{-4}	0·796	0·102	$(0{\cdot}593, 0{\cdot}999)$	$\hat{\tau}^{\mathrm{aug}}_{\mathrm{ATT},\epsilon}(\hat{\theta})$	0·792	0·088	$(0{\cdot}616, 0{\cdot}968)$
$\hat{\tau}_{\mathrm{ATT},\epsilon}(\hat{\theta})$	10^{-5}	0·796	0·109	$(0{\cdot}579, 1{\cdot}013)$	$\hat{\tau}^{\mathrm{aug}}_{\mathrm{ATT},\epsilon}(\hat{\theta})$	0·793	0·088	$(0{\cdot}617, 0{\cdot}968)$

MERCATANTI, A. & LI, F. (2014). Do debit cards increase household spending? evidence from a semiparametric causal analysis of a survey. *Ann. Appl. Statist.* **8**, 2485–508.

SHAO, J. & TU, D. (2012). *The Jackknife and Bootstrap*. New York: Springer.

SHINOZAKI, T. & MATSUYAMA, Y. (2015). Doubly robust estimation of standardized risk difference and ratio in the exposed population. *Epidemiology* **26**, 873–77.

SMITH, J. A. & TODD, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *J. Economet.* **125**, 305–53.

VAN DER VAART, A. W. (2000). *Asymptotic Statistics*. Cambridge, MA: Cambridge University Press.

ZHAO, Q. & PERCIVAL, D. (2017). Entropy balancing is doubly robust. *J Causal Infer.* **5**, DOI: https://doi.org/10.1515/jci–2016–0010.