Biometrika (2021), **xx**, x, *pp*. 1–17
Printed in Great Britain

Supplementary Material for "Estimation of local treatment effects under the binary instrumental variable model"

BY LINBO WANG

Department of Statistical Sciences, University of Toronto, Toronto, Ontario M5S 3G3, Canada
linbo.wang@utoronto.ca

YUEXIA ZHANG

Department of Computer and Mathematical Sciences, University of Toronto, Ontario M1C 1A4, Canada
yuexia.zhang@utoronto.ca

THOMAS S. RICHARDSON

Department of Statistics, University of Washington, Box 354322, Washington 98195, U.S.A.
thomasr@u.washington.edu

AND JAMES M. ROBINS

Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.
robins@hsph.harvard.edu

SUMMARY

In the Supplementary Material we provide proofs of theorems and claims in the main paper. We also provide additional details for the simulations and data application.

1. PROOF OF THEOREM 1

Proof. Under the principal stratum framework, the population can also be divided into four strata based on values of $(Y(1), Y(0))$ as in Table S1. For simplicity, we denote the principal stratum based on values of $(Y(1), Y(0))$ as t_Y.

Table S1: Principal stratum t_Y based on $(Y(1), Y(0))$

$Y(1)$	$Y(0)$	Principal stratum	Abbreviation
1	1	Always recovered	AR
1	0	Helped	HE
0	1	Hurt	HU
0	0	Never recovered	NR

We first show that (5) is a well-defined map. This follows since $\theta(X)$ is identifiable following (1), (2) and

$$\phi_1(X) = 1 - p_X(0,0 \mid 1) - p_X(0,1 \mid 1) - p_X(1,0 \mid 0) - p_X(1,1 \mid 0);$$

$$\phi_2(X) = \{p_X(1,0 \mid 0) + p_X(1,1 \mid 0)\}/\{p_X(0,0 \mid 1) + p_X(0,1 \mid 1) + p_X(1,0 \mid 0) + p_X(1,1 \mid 0)\};$$
$$\phi_3(X) = p_X(0,1 \mid 1)/\{p_X(0,0 \mid 1) + p_X(0,1 \mid 1)\};$$
$$\phi_4(X) = p_X(1,1 \mid 0)/\{p_X(1,0 \mid 0) + p_X(1,1 \mid 0)\};$$

$$\begin{aligned}
OP^{CO}(X) &= \frac{E\{Y(1)I(t_D = CO) \mid X\}E\{Y(0)I(t_D = CO) \mid X\}}{[p(CO;X) - E\{Y(1)I(t_D = CO) \mid X\}][p(CO;X) - E\{Y(0)I(t_D = CO) \mid X\}]} \\
&= \frac{\{p(CO,HE;X) + p(CO,AR;X)\}\{p(CO,HU;X) + p(CO,AR;X)\}}{\{p(CO,NR;X) + p(CO,HU;X)\}\{p(CO,NR;X) + p(CO,HE;X)\}} \\
&= \frac{\{p_X(1,1 \mid 1) - p_X(1,1 \mid 0)\}\{p_X(0,1 \mid 0) - p_X(0,1 \mid 1)\}}{\{p_X(1,0 \mid 1) - p_X(1,0 \mid 0)\}\{p_X(0,0 \mid 0) - p_X(0,0 \mid 1)\}}
\end{aligned}$$

are identifiable from $\boldsymbol{p} = \{p_x(d,y \mid z); d,y,z = 0,1\}$.

To show that (5) is a bijection, for each realization of X, let $\boldsymbol{c} = (c_0, \ldots, c_5)$ be a vector in $\mathcal{D} \times [0,1]^4 \times [0,+\infty)$. We need to show there is one and only one $\boldsymbol{p} \in \Delta$ such that

$$\{\phi_1(X), \ldots, \phi_4(X)\} = (c_1, \ldots, c_4) \tag{S1}$$

and

$$\left\{\theta(X), OP^{CO}(X)\right\} = (c_0, c_5). \tag{S2}$$

For simplicity of notation, we suppress the dependence on X in the remainder of the proof. First note that (S1) implies that

$$\begin{aligned}
p(0,1 \mid 1) = (1-c_1)(1-c_2)c_3; \quad p(0,0 \mid 1) = (1-c_1)(1-c_2)(1-c_3); \\
p(1,1 \mid 0) = (1-c_1)c_2c_4; \quad p(1,0 \mid 0) = (1-c_1)c_2(1-c_4).
\end{aligned} \tag{S3}$$

According to Richardson et al. (2017), (S2) implies that

$$\begin{aligned}
E\{Y(1) \mid t_D = CO\} &= \frac{p(1,1 \mid 1) - p(1,1 \mid 0)}{p(CO)} = f_1(c_0, c_5); \\
E\{Y(0) \mid t_D = CO\} &= \frac{p(0,1 \mid 0) - p(0,1 \mid 1)}{p(CO)} = f_0(c_0, c_5),
\end{aligned} \tag{S4}$$

where $p(CO) = 1 - p(0,0 \mid 1) - p(0,1 \mid 1) - p(1,0 \mid 0) - p(1,1 \mid 0)$ and $f_d(c_0,c_5), d = 0,1$ are known smooth functions of c_0, c_5 that take values between 0 and 1. The functional form of $f_d(c_0, c_5)$ follows from equations (2.4) and (2.5) in Richardson et al. (2017). Specifically,

$$f_0(c_0,c_5) = \begin{cases} \dfrac{1}{2(c_5-1)}\left[c_5(2-c_0) + c_0 - \left\{c_0^2(c_5-1)^2 + 4c_5\right\}^{1/2}\right], & \theta = \delta^L; \\ \dfrac{1}{2c_0(1-c_5)}\left[-(c_0+1)c_5 + \left\{c_5^2(c_0-1)^2 + 4c_0c_5\right\}^{1/2}\right], & \theta = \delta^M; \end{cases}$$

and

$$f_1(c_0,c_5) = \begin{cases} f_0(c_0,c_5) + c_0, & \theta = \delta^L; \\ f_0(c_0,c_5)c_0, & \theta = \delta^M. \end{cases}$$

Combining (S3) and (S4), we have (S1) and (S2) together imply that

$$\begin{aligned}
p(0,1 \mid 1) = (1-c_1)(1-c_2)c_3; \quad p(0,0 \mid 1) = (1-c_1)(1-c_2)(1-c_3); \\
p(1,1 \mid 0) = (1-c_1)c_2c_4; \quad p(1,0 \mid 0) = (1-c_1)c_2(1-c_4);
\end{aligned}$$

$$\begin{aligned}
p(1,1\mid 1) &= f_1(c_0,c_5)c_1 + p(1,1\mid 0);\\
p(0,1\mid 0) &= f_0(c_0,c_5)c_1 + p(0,1\mid 1);\\
p(1,0\mid 1) &= 1 - p(0,0\mid 1) - p(0,1\mid 1) - p(1,1\mid 1);\\
p(0,0\mid 0) &= 1 - p(0,1\mid 0) - p(1,0\mid 0) - p(1,1\mid 0).
\end{aligned} \tag{S5}$$

We now only need to show $\boldsymbol{p}$ defined in (S5) lies in Δ. First note that

$$\begin{aligned}
p(1,1\mid 1) &= f_1(c_0,c_5)c_1 + p(1,1\mid 0) \le c_1 + (1-c_1)c_2c_4 \le 1;\\
p(0,1\mid 0) &= f_0(c_0,c_5)c_1 + p(0,1\mid 1) \le c_1 + (1-c_1)(1-c_2)c_3 \le 1;\\
p(1,0\mid 1) &= 1 - p(0,0\mid 1) - p(0,1\mid 1) - p(1,1\mid 1)\\
&= 1 - (1-c_1)(1-c_2) - f_1(c_0,c_5)c_1 - p(1,1\mid 0)\\
&\ge 1 - (1-c_1)(1-c_2) - c_1 - (1-c_1)c_2c_4\\
&= (1-c_1)c_2(1-c_4) \ge 0; \qquad \text{(S6)}\\
p(0,0\mid 0) &= 1 - p(0,1\mid 0) - p(1,0\mid 0) - p(1,1\mid 0)\\
&= 1 - f_0(c_0,c_5)c_1 - p(0,1\mid 1) - (1-c_1)c_2\\
&\ge 1 - c_1 - (1-c_1)(1-c_2)c_3 - (1-c_1)c_2\\
&= (1-c_1)(1-c_2)(1-c_3) \ge 0. \qquad \text{(S7)}
\end{aligned}$$

Furthermore,

$$\begin{aligned}
p(1,1\mid 1) &= f_1(c_0,c_5)c_1 + p(1,1\mid 0) \ge p(1,1\mid 0);\\
p(1,0\mid 1) &\ge (1-c_1)c_2(1-c_4) = p(1,0\mid 0);\\
p(0,1\mid 0) &= f_0(c_0,c_5)c_1 + p(0,1\mid 1) \ge p(0,1\mid 1);\\
p(0,0\mid 0) &\ge (1-c_1)(1-c_2)(1-c_3) = p(0,0\mid 1),
\end{aligned}$$

where the second and last inequality were shown in (S6) and (S7).

We have hence finished the proof. □

2. PROOF OF THEOREM 2

Proof. If $\theta(X) = \delta^L(X)$, then $H(Y,D,X) = Y - D\theta(X)$. Since

$$\delta^L(X) = \frac{E(Y\mid Z=1,X) - E(Y\mid Z=0,X)}{E(D\mid Z=1,X) - E(D\mid Z=0,X)},$$

then

$$\begin{aligned}
&E\{H(Y,D,X)\mid Z=1,X\} - E\{H(Y,D,X)\mid Z=0,X\}\\
&\quad = E(Y\mid Z=1,X) - E(D\mid Z=1,X)\theta(X) - E(Y\mid Z=0,X) + E(D\mid Z=0,X)\theta(X)\\
&\quad = E(Y\mid Z=1,X) - E(Y\mid Z=0,X)\\
&\qquad - \big\{E(D\mid Z=1,X) - E(D\mid Z=0,X)\big\}\frac{E(Y\mid Z=1,X) - E(Y\mid Z=0,X)}{E(D\mid Z=1,X) - E(D\mid Z=0,X)} = 0.
\end{aligned}$$

Thus, $E\{H(Y,D,X)\mid X\} = E\{H(Y,D,X)\mid Z=0,X\}$. Based on the results in Section 1, we have

$$\begin{aligned}
E\{H(Y,D,X)\mid X\} &= E\{H(Y,D,X)\mid Z=0,X\}\\
&= E(Y\mid Z=0,X) - E(D\mid Z=0,X)\theta(X)
\end{aligned}$$

$$
\begin{aligned}
&= P(Y=1 \mid Z=0,X) - P(D=1 \mid Z=0,X)\theta(X) \\
&= P(D=1,Y=1 \mid Z=0,X) + P(D=0,Y=1 \mid Z=0,X) - P(D=1 \mid Z=0,X)\theta(X) \\
&= p_X(1,1 \mid 0) + p_X(0,1 \mid 0) - \theta(X)\{1-\phi_1(X)\}\phi_2(X) \\
&= \{1-\phi_1(X)\}\phi_2(X)\phi_4(X) + f_0\{\theta(X), OP^{CO}(X)\}\phi_1(X) \\
&\quad + \{1-\phi_1(X)\}\{1-\phi_2(X)\}\phi_3(X) - \theta(X)\{1-\phi_1(X)\}\phi_2(X).
\end{aligned}
$$

If $\theta(X) = \delta^M(X)$, then $H(Y,D,X) = Y\theta(X)^{-D}$. Since

$$
\delta^M(X) = -\frac{E(YD \mid Z=1,X) - E(YD \mid Z=0,X)}{E\{Y(1-D) \mid Z=1,X\} - E\{Y(1-D) \mid Z=0,X\}},
$$

then

$$
\begin{aligned}
&E\{H(Y,D,X) \mid Z=1,X\} - E\{H(Y,D,X) \mid Z=0,X\} \\
&\quad = E\{Y\theta(X)^{-D} \mid Z=1,X\} - E\{Y\theta(X)^{-D} \mid Z=0,X\} \\
&\quad = \theta(X)^{-1}P(Y=1,D=1 \mid Z=1,X) + P(Y=1,D=0 \mid Z=1,X) \\
&\qquad - \theta(X)^{-1}P(Y=1,D=1 \mid Z=0,X) - P(Y=1,D=0 \mid Z=0,X) \\
&\quad = \theta(X)^{-1}\big\{P(Y=1,D=1 \mid Z=1,X) - P(Y=1,D=1 \mid Z=0,X)\big\} \\
&\qquad + P(Y=1,D=0 \mid Z=1,X) - P(Y=1,D=0 \mid Z=0,X) \\
&\quad = -\frac{P(Y=1,D=0 \mid Z=1,X) - P(Y=1,D=0 \mid Z=0,X)}{P(Y=1,D=1 \mid Z=1,X) - P(Y=1,D=1 \mid Z=0,X)} \\
&\qquad\quad \times \big\{P(Y=1,D=1 \mid Z=1,X) - P(Y=1,D=1 \mid Z=0,X)\big\} \\
&\qquad + P(Y=1,D=0 \mid Z=1,X) - P(Y=1,D=0 \mid Z=0,X) \\
&\quad = 0.
\end{aligned}
$$

Thus, $E\{H(Y,D,X) \mid X\} = E\{H(Y,D,X) \mid Z=0,X\}$. Based on the results in Section 1, we have

$$
\begin{aligned}
&E\{H(Y,D,X) \mid X\} = E\{H(Y,D,X) \mid Z=0,X\} \\
&\quad = E\{Y\theta(X)^{-D} \mid Z=0,X\} \\
&\quad = \theta(X)^{-1}P(Y=1,D=1 \mid Z=0,X) + P(Y=1,D=0 \mid Z=0,X) \\
&\quad = \{1-\phi_1(X)\}\phi_2(X)\phi_4(X)\theta(X)^{-1} \\
&\qquad + f_0\{\theta(X), OP^{CO}(X)\}\phi_1(X) + \{1-\phi_1(X)\}\{1-\phi_2(X)\}\phi_3(X).
\end{aligned}
$$

Therefore, $\widehat{E}\{H(Y,D,X;\alpha) \mid X\}$ has the form as shown in Theorem 2.

If the model for $E\{H(Y,D,X;\alpha) \mid X\}$ is correctly specified, but the model for $f(Z \mid X;\gamma)$ may be mis-specified, then $\hat{\gamma} \xrightarrow{p} \gamma^*$ under some regularity conditions (White, 1982), where γ^* is not necessarily equal to γ. Furthermore,

$$
\begin{aligned}
&E\left(\mathbb{P}_n\omega(X)\frac{2Z-1}{f(Z \mid X;\gamma^*)}\left[H(Y,D,X;\alpha) - E\left\{H(Y,D,X;\alpha) \mid Z=0,X\right\}\right]\right) \\
&\quad = E\left(E\left\{\mathbb{P}_n\omega(X)\frac{2Z-1}{f(Z \mid X;\gamma^*)}\left[H(Y,D,X;\alpha) - E\left\{H(Y,D,X;\alpha) \mid Z=0,X\right\}\right] \,\middle|\, Z,X\right\}\right) \\
&\quad = E\left(\mathbb{P}_n\omega(X)\frac{2Z-1}{f(Z \mid X;\gamma^*)}\left[E\left\{H(Y,D,X;\alpha) \mid Z,X\right\} - E\left\{H(Y,D,X;\alpha) \mid Z=0,X\right\}\right]\right)
\end{aligned}
$$

$= 0.$

If the model for $f(Z \mid X; \gamma)$ is correctly specified, but the model for $E\{H(Y, D, X; \alpha) \mid X\}$ may be mis-specified, then $\widehat{E}\{H(Y, D, X; \alpha) \mid X\} - E\{H(Y, D, X; \alpha) \mid X\}$ does not necessarily converge to zero in probability. Assume there exists $E^*\{H(Y, D, X; \alpha) \mid X\}$ such that $\widehat{E}\{H(Y, D, X; \alpha) \mid X\} - E^*\{H(Y, D, X; \alpha) \mid X\} \xrightarrow{p} 0$ under some regularity conditions. Furthermore,

$$
\begin{aligned}
&E\bigg(\mathbb{P}_n\omega(X)\frac{2Z-1}{f(Z \mid X;\gamma)}\left[H(Y,D,X;\alpha) - E^*\left\{H(Y,D,X;\alpha) \mid X\right\}\right]\bigg) \\
&\quad = E\bigg\{\mathbb{P}_n\omega(X)\frac{2Z-1}{f(Z \mid X;\gamma)}H(Y,D,X;\alpha)\bigg\} - E\bigg[\mathbb{P}_n\omega(X)\frac{2Z-1}{f(Z \mid X;\gamma)}E^*\left\{H(Y,D,X;\alpha) \mid X\right\}\bigg] \\
&\quad = E\bigg[E\bigg\{\mathbb{P}_n\omega(X)\frac{2Z-1}{f(Z \mid X;\gamma)}H(Y,D,X;\alpha)\;\bigg|\; X\bigg\}\bigg] \\
&\qquad - E\bigg(E\bigg[\mathbb{P}_n\omega(X)\frac{2Z-1}{f(Z \mid X;\gamma)}E^*\{H(Y,D,X;\alpha) \mid X\}\;\bigg|\; X\bigg]\bigg) \\
&\quad = E\bigg(E\bigg[\mathbb{P}_n\omega(X)\frac{2Z-1}{\{P(Z=1 \mid X;\gamma)\}^Z\{1-P(Z=1 \mid X;\gamma)\}^{1-Z}}H(Y,D,X;\alpha)\;\bigg|\; X\bigg]\bigg) \\
&\qquad - E\bigg(\mathbb{P}_n\omega(X)E^*\left\{H(Y,D,X;\alpha) \mid X\right\}E\bigg[\frac{2Z-1}{\{P(Z=1 \mid X;\gamma)\}^Z\{1-P(Z=1 \mid X;\gamma)\}^{1-Z}}\;\bigg|\; X\bigg]\bigg) \\
&\quad = E\Big(\mathbb{P}_n\omega(X)\big[E\{H(Y,D,X;\alpha) \mid Z=1,X\} - E\{H(Y,D,X;\alpha) \mid Z=0,X\}\big]\Big) - 0 \\
&\quad = 0.
\end{aligned}
$$

The rest of the proof follows from standard M-estimation theory.

3. Optimal weighting function

If $\theta(X) = \delta^L(X)$, Ogburn et al. (2015) show that the optimal choice of $\omega(X)$ is given by

$$
\begin{aligned}
\omega_{\mathrm{opt}}(X) &= -\frac{\partial\theta(X)}{\partial\alpha}E\bigg\{\frac{2Z-1}{f(Z \mid X)}D\;\bigg|\; X\bigg\}E^{-1}\bigg[\frac{\{H - E(H \mid Z,X)\}^2}{f^2(Z \mid X)}\;\bigg|\; X\bigg] \\
&= -\frac{\partial\theta(X)}{\partial\alpha} \times \phi_1(X) \times \bigg[E_{Z|X}\frac{1}{f^2(Z \mid X)}\left\{E\left(H^2 \mid Z,X\right) - E^2\left(H \mid Z,X\right)\right\}\bigg]^{-1},
\end{aligned}
$$

with

$$H = H(Y,D,X) = Y - D\theta(X);$$

$$
\begin{aligned}
E\big(H^2 \mid Z=z,X\big) &= E\big[\{Y - D\theta(X)\}^2 \mid Z=z,X\big] \\
&= P(Y=1 \mid Z=z,X) + \theta(X)^2P(D=1 \mid Z=z,X) - 2\theta(X)P(DY=1 \mid Z=z,X);
\end{aligned}
$$

$$
\begin{aligned}
E(H \mid Z,X) &= E\{Y - D\theta(X) \mid Z=0,X\} \\
&= P(Y=1 \mid Z=0,X) - \theta(X)P(D=1 \mid Z=0,X),
\end{aligned}
$$

where, as shown in Section 1, $P(Y=1 \mid Z=z,X), P(D=1 \mid Z=z,X), P(DY=1 \mid Z=z,X)$ are functions of $\{\theta(X),\phi_1(X),\dots,\phi_4(X),OP^{CO}(X)\}$.

If $\theta(X)=\delta^M(X)$, Ogburn et al. (2015) show that the optimal choice of $\omega(X)$ is given by

$$\begin{aligned}
\omega_{\text{opt}}(X) &= -\frac{\partial\theta(X)}{\partial\alpha}\theta(X)^{-2}E\left\{\frac{2Z-1}{f(Z\mid X)}DY\middle| X\right\}E^{-1}\left[\frac{\{H-E(H\mid Z,X)\}^2}{f^2(Z\mid X)}\middle| X\right]\\
&= -\frac{\partial\theta(X)}{\partial\alpha}\times\theta(X)^{-2}\times f_1\phi_1(X)\times\left[E_{Z|X}\frac{1}{f^2(Z\mid X)}\left\{E\big(H^2\mid Z,X\big)-E^2(H\mid Z,X)\right\}\right]^{-1},
\end{aligned}$$

with

$$\begin{aligned}
H=H(Y,D,X)=Y\theta(X)^{-D};&\quad f_1=f_1\{\theta(X),OP^{C0}(X)\};\\
E\left(H^2\mid Z=z,X\right)&=E\left\{Y^2\theta(X)^{-2D}\mid Z=z,X\right\}\\
&=P(Y=1,D=1\mid Z=z,X)\theta(X)^{-2}+P(Y=1,D=0\mid Z=z,X);\\
E(H\mid Z,X)&=E\{Y\theta(X)^{-D}\mid Z=0,X\}\\
&=P(Y=1,D=1\mid Z=0,X)\theta(X)^{-1}+P(Y=1,D=0\mid Z=0,X).
\end{aligned}$$

4. PROOF OF THE VARIATION INDEPENDENCE CLAIMS IN REMARK 2

We first show the (joint) variation independence of $E(Y\mid X), E(D\mid X), f(Z\mid X)$ and $\delta^L(X)$. For a particular realization of X, and any $(a,b,\pi,c)\in(0,1)^3\times(-1,1)$, we need to show it is possible that

$$\begin{aligned}
E(Y\mid X)&=a; &\text{(S8)}\\
E(D\mid X)&=b; &\text{(S9)}\\
P(Z=1\mid X)&=\pi; &\text{(S10)}\\
\delta^L(X)&=c.
\end{aligned}$$

It is clear that (S8) – (S10) may hold simultaneously since $E(Y\mid X), E(D\mid X)$ and $f(Z\mid X)$ are clearly variation independent. Equations (S8) – (S10) place the following constraints on the range of $a_1\equiv E(Y\mid Z=1,X), a_0\equiv E(Y\mid Z=0,X), b_1\equiv E(D\mid Z=1,X), b_0\equiv E(D\mid Z=0,X)$:

$$\begin{aligned}
\pi a_1+(1-\pi)a_0&=a; &\text{(S11)}\\
\pi b_1+(1-\pi)b_0&=b; &\text{(S12)}\\
0\le a_1,a_0,b_1,b_0&\le 1. &\text{(S13)}
\end{aligned}$$

We use a linear programming algorithm to obtain the range of $\delta^Y=a_1-a_0$ and $\delta^D=b_1-b_0$ subject to the constraints in (S11)–(S13):

$$\begin{aligned}
\max\left(-\frac{a}{1-\pi},-\frac{1-a}{\pi}\right)\le\delta^Y\le\min\left(\frac{1-a}{1-\pi},\frac{a}{\pi}\right);\\
\max\left(-\frac{b}{1-\pi},-\frac{1-b}{\pi}\right)\le\delta^D\le\min\left(\frac{1-b}{1-\pi},\frac{b}{\pi}\right).
\end{aligned}$$

Note that as long as $0<a,b,\pi<1$, the feasible range of (δ^Y,δ^D) always contains a ball around the origin. Hence the range of $c=\delta^Y/\delta^D$ is unrestricted by the constraints in (S8) – (S10).

The proof of the variation independence of $E(Y \mid X), E(Y \mid D, X), f(Z \mid X)$ and $\delta^M(X)$ follows the same logic and is hence omitted.

5. ADDITIONAL DETAILS FOR SIMULATION STUDIES AND DATA ANALYSIS

5.1. *Visualization of the degree of model mis-specification*

In Figure S1, we provide visualization of the degree of model mis-specification using data points generated from one randomly selected Monte Carlo run. We denote the first element of covariates X as X_1, namely the intercept. We denote the second element of X as X_2, which was generated from $\mathrm{Unif}(-1, 1)$. Figure S1 shows fitted probabilities using the correct/incorrect models as functions of X_2. To unify notation, we use X^* to denote the covariate used in fitting the models, which may correspond to $X, X^\dagger$ or X' depending on the specific scenario. If the model for the instrumental density is correctly specified, then $\hat{P}(Z = 1 \mid X^*) = \hat{P}(Z = 1 \mid X)$; otherwise $\hat{P}(Z = 1 \mid X^*) = \hat{P}(Z = 1 \mid X^\dagger)$. If the other nuisance models are correctly specified, then the fitted probabilities $\hat{P}(Y = 1 \mid Z = 1, X^*)$ are derived from the fitted values of $\theta(X), \phi_i(X), i = 1, \ldots, 4$ and $OP^{CO}(X)$; otherwise, $\hat{P}(Y = 1 \mid Z = 1, X^*)$ are derived from the fitted values of $\theta(X), \phi_i(X'), i = 1, \ldots, 4$ and $OP^{CO}(X')$.

From the left panels of Figure S1, one can see that under correct model specification, $\hat{P}(Y = 1 \mid Z = 1, X)$ is a non-monotone function of X_2. Under mis-specifications of nuisance models, $\hat{P}(Y = 1 \mid Z = 1, X^*)$ has three clusters, as for each value of X_2, it is possible that $X' = (1, 0), (1, 1)$ or $(0, 1)$. The fitted values $\hat{P}(Y = 1 \mid Z = 1, X^*)$ still depend on X_2 through the fitted values of $\theta(X)$. From the right panels of Figure S1, one can see that under correct model specification, $\hat{P}(Z = 1 \mid X)$ is an expit function of X_2. Under mis-specification of the instrumental density model, the fitted values $\hat{P}(Z = 1 \mid X^\dagger)$ do not depend on X_2 as $X^\dagger$ is independent of X_2.

5.2. *Implementation details in the simulation studies*

We now describe the implementation details of the various estimators considered in the simulation studies:

`reg.ogburn`
`dru.ogburn`
`drw.ogburn` If $\theta(X) = \delta^L(X)$, then we assume

$$\begin{aligned} P(Z = 1 \mid X) &= \mathrm{expit}(\gamma^\top X); \\ E(H \mid X) &= \xi_1 + \xi_2 X_2 + \xi_3 X_2^2 + \xi_4 X_2^3; \\ \delta^D(X) &= \tanh(\psi_1^\top X); \\ E\left[\left.\frac{\{H - E(H \mid X)\}^2}{f^2(Z \mid X)}\,\right|\, X\right] &= \zeta_1 + \zeta_2 X_2 + \zeta_3 X_2^2. \end{aligned}$$

If $\theta(X) = \delta^M(X)$, then we assume

$$\begin{aligned} P(Z = 1 \mid X) &= \mathrm{expit}(\gamma^\top X); \\ E(H \mid X) &= \exp(\xi_5 + \xi_6 X_2 + \xi_7 X_2^2); \\ E(DY \mid Z, X) &= \mathrm{expit}(\psi_2 Z + \psi_3^\top X); \end{aligned}$$

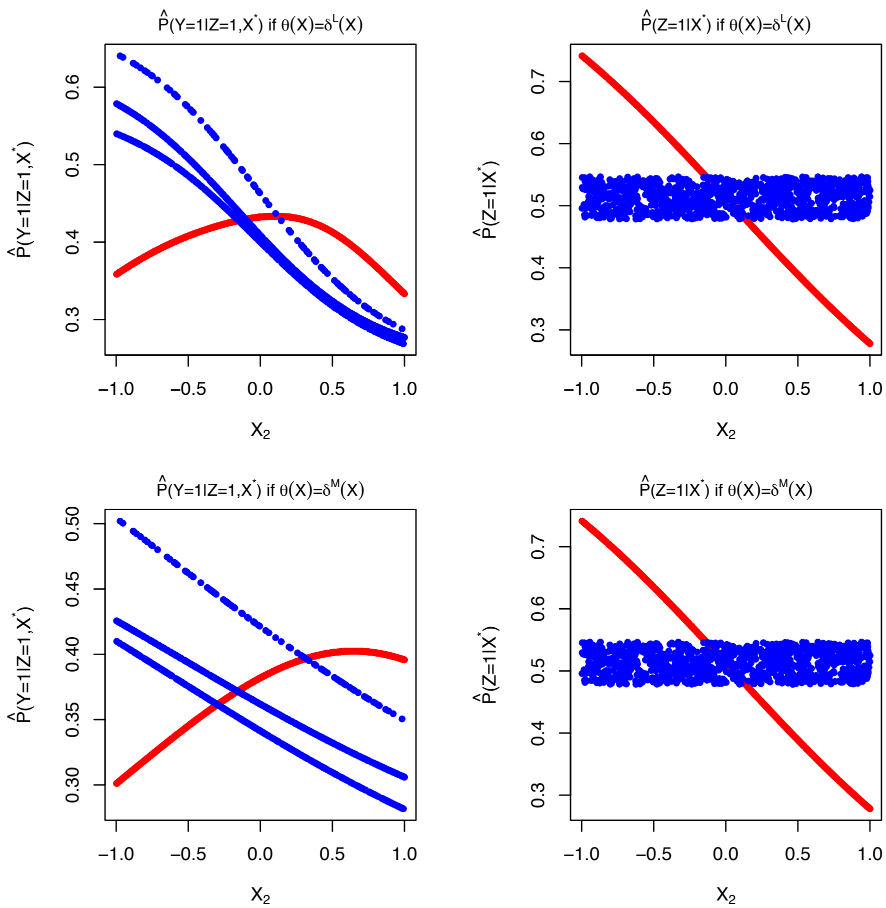


Fig. S1: Visualizing nuisance model mis-specification with correct specification for $\theta(X)$: top row: $\theta(X) = \delta^L(X)$; bottom row: $\theta(X) = \delta^M(X)$; red dots: correct nuisance model specification; blue dots: incorrect nuisance model specification

$$E\left[\frac{\{H - E(H \mid X)\}^2}{f^2(Z \mid X)} \,\middle|\, X\right] = \exp(\zeta_4 + \zeta_5 X_2 + \zeta_6 X_2^2),$$

where

$$H = H(Y, D, X) = \begin{cases} Y - D\theta(X), & \theta(X) = \delta^L(X); \\ Y\theta(X)^{-D}, & \theta(X) = \delta^M(X). \end{cases}$$

The higher order terms are added so that these models better approximate the truth.

In `reg.ogburn`, we obtain $(\hat{\xi}_{\text{reg}}, \hat{\alpha}_{\text{reg}})$ as the solution to the following estimating equations:

$$\begin{cases} \mathbb{P}_n \begin{pmatrix} (1, X_2, X_2^2, X_2^3)^\top \\ Z[1-\{\theta(X;\alpha)\}^2](1, X_2)^\top \end{pmatrix} \{H(\alpha) - E(H \mid X;\xi)\} = 0, \quad \theta(X) = \delta^L(X); \\ \mathbb{P}_n \begin{pmatrix} E(H \mid X;\xi)(1, X_2, X_2^2)^\top \\ Z\theta(X;\alpha)(1, X_2)^\top \end{pmatrix} \{H(\alpha) - E(H \mid X;\xi)\} = 0, \quad \theta(X) = \delta^M(X). \end{cases}$$

In `dru.ogburn`, first, we fit the model $P(Z=1 \mid X;\gamma)$ using `R` package `glm`. Then, we obtain $\hat{\alpha}_{\text{dru.ogburn}}$ as the solution to the estimating equation

$$\mathbb{P}_n \frac{2Z-1}{f(Z \mid X;\hat{\gamma})} \left\{ H(\alpha) - E(H \mid X;\hat{\xi}_{\text{reg}}) \right\} = 0.$$

In `drw.ogburn`, first, we fit the model $P(Z=1 \mid X;\gamma)$ using `R` package `glm`. Next, we estimate the optimal weighting function $\omega_{\text{opt}}(X)$. If $\theta(X) = \delta^L(X)$,

$$\hat{\omega}_{\text{opt}}(X) = -X[1-\{\theta(X;\hat{\alpha}_{\text{reg}})\}^2]\delta^D(X;\hat{\psi})E^{-1}\left[\frac{\left\{H(\hat{\alpha}_{\text{reg}}) - E(H \mid X;\hat{\xi}_{\text{reg}})\right\}^2}{f^2(Z \mid X;\hat{\gamma})} \,\middle|\, X;\hat{\zeta} \right];$$

if $\theta(X) = \delta^M(X)$,

$$\begin{aligned} \hat{\omega}_{\text{opt}}(X) = & - X\{\theta(X;\hat{\alpha}_{\text{reg}})\}^{-1} \left\{ E(DY \mid Z=1, X;\hat{\psi}) - E(DY \mid Z=0, X;\hat{\psi}) \right\} \\ & \times E^{-1}\left[\frac{\left\{H(\hat{\alpha}_{\text{reg}}) - E(H \mid X;\hat{\xi}_{\text{reg}})\right\}^2}{f^2(Z \mid X;\hat{\gamma})} \,\middle|\, X;\hat{\zeta} \right], \end{aligned}$$

where the model $\delta^D(X;\psi)$ is fitted using the doubly robust estimator of Richardson et al. (2017) and obtained using `R` package `brm`, the model $E[\{H(\hat{\alpha}_{\text{reg}}) - E(H \mid X;\hat{\xi}_{\text{reg}})\}^2/f^2(Z \mid X;\hat{\gamma}) \mid X;\zeta]$ is fitted using the least squares method with the restriction that $E[\{H(\hat{\alpha}_{\text{reg}}) - E(H \mid X;\hat{\xi}_{\text{reg}})\}^2/f^2(Z \mid X;\hat{\gamma}) \mid X;\hat{\zeta}] > 0$, and the model $E(DY \mid Z, X;\psi)$ is fitted using the `R` package `glm`. Then, we obtain $\hat{\alpha}_{\text{drw.ogburn}}$ as the solution to the estimating equation

$$\mathbb{P}_n \hat{\omega}_{\text{opt}}(X) \frac{2Z-1}{f(Z \mid X;\hat{\gamma})} \left\{ H(\alpha) - E(H \mid X;\hat{\xi}_{\text{reg}}) \right\} = 0.$$

`mle.wang`
`dru.wang` If $\theta(X) = \delta^L(X)$, then we assume

$$\begin{aligned} P(Z=1 \mid X) &= \text{expit}(\gamma^\top X); \\ \delta^D(X) &= \tanh(\lambda^\top X); \\ OP^D(X) &\equiv \frac{P(D=1 \mid Z=1, X)P(D=1 \mid Z=0, X)}{P(D=0 \mid Z=1, X)P(D=0 \mid Z=0, X)} = \exp(\tau^\top X); \\ OP^Y(X) &\equiv \frac{P(Y=1 \mid Z=1, X)P(Y=1 \mid Z=0, X)}{P(Y=0 \mid Z=1, X)P(Y=0 \mid Z=0, X)} = \exp(\kappa^\top X). \end{aligned}$$

In mle.wang, first, we fit the models $\delta^D(X;\lambda)$ and $OP^D(X;\tau)$ using the maximum likelihood estimation implemented in R package brm. Then, we fit the models $OP^Y(X;\kappa)$ and $\theta(X;\alpha)$ using the maximum likelihood estimation method based on the likelihood function of Y conditional on Z, X and $\delta^D(X;\hat\lambda)$. The maximum likelihood estimator of α is denoted as $\hat\alpha_{\text{mle.wang}}$.

In dru.wang, first, we fit the model $P(Z=1\mid X;\gamma)$ using R package glm. Next, we get $\widehat{E}(D\mid Z=0,X)$ and $\widehat{E}(Y\mid Z=0,X)$ from $\delta^D(X;\hat\lambda)$, $OP^D(X;\hat\tau)$, $\theta(X;\hat\alpha_{\text{mle.wang}})$ and $OP^Y(X;\hat\kappa)$ based on Proposition 2 of Wang and Tchetgen Tchetgen (2018). Then, we obtain $\hat\alpha_{\text{dru.wang}}$ as the solution to the estimating equation

$$\mathbb{P}_n\frac{2Z-1}{f(Z\mid X;\hat\gamma)}\Big\{Y-D\theta(X;\alpha)-\widehat{E}(Y\mid Z=0,X)+\widehat{E}(D\mid Z=0,X)\theta(X;\alpha)\Big\}=0.$$

dru.simple If $\theta(X)=\delta^L(X)$, then we assume

$$\begin{aligned}P(Z=1\mid X)&=\operatorname{expit}(\gamma^\top X);\\ E(Y\mid X)&=\operatorname{expit}(\varsigma^\top X);\\ E(D\mid X)&=\operatorname{expit}(\vartheta^\top X).\end{aligned}$$

If $\theta(X)=\delta^M(X)$, then we assume

$$\begin{aligned}P(Z=1\mid X)&=\operatorname{expit}(\gamma^\top X);\\ E(Y\mid D,X)&=\operatorname{expit}(\varpi_1 D+\varpi_2^\top X);\\ E(D\mid X)&=\operatorname{expit}(\vartheta^\top X).\end{aligned}$$

In dru.simple, first, we fit the models $E(Y\mid X;\varsigma)$, $E(D\mid X;\vartheta)$, $E(Y\mid D,X;\varpi)$ and $P(Z=1\mid X;\gamma)$ using R package glm. Then, we obtain $\hat\alpha_{\text{dru.simple}}$ as the solution to the estimating equation

$$\mathbb{P}_n\frac{2Z-1}{f(Z\mid X;\hat\gamma)}\Big\{H(\alpha)-\widehat{E}(H\mid X;\alpha)\Big\}=0,$$

where

$$\widehat{E}\{H\mid X;\alpha\}=\begin{cases}E(Y\mid X;\hat\varsigma)-E(D|X;\hat\vartheta)\theta(X;\alpha), & \theta(X)=\delta^L(X);\\ E(D\mid X;\hat\vartheta)E(Y\mid D=1,X;\hat\varpi)\theta(X;\alpha)^{-1} & \\ \quad+\{1-E(D\mid X;\hat\vartheta)\}E(Y\mid D=0,X;\hat\varpi), & \theta(X)=\delta^M(X).\end{cases}$$

ls.abadie If $\theta(X)=\delta^L(X)$, then we assume

$$\begin{aligned}P(Z=1\mid X)&=\operatorname{expit}(\gamma^\top X);\\ E\{Y\mid X,D,D(1)>D(0)\}&=D\tanh(\alpha^\top X)+\operatorname{expit}(\varphi_1^\top X).\end{aligned}$$

If $\theta(X)=\delta^M(X)$, then we assume

$$\begin{aligned}P(Z=1\mid X)&=\operatorname{expit}(\gamma^\top X);\\ E\{Y\mid X,D,D(1)>D(0)\}&=\{\exp(\alpha^\top X)\}^D\operatorname{expit}(\varphi_2^\top X).\end{aligned}$$

In ls.abadie, first, we fit the model $P(Z=1\mid X;\gamma)$ using R package glm. Then, we obtain the weighted least squares estimator $(\hat\alpha_{\text{ls.abadie}},\hat\varphi_{\text{ls.abadie}})$ by minimizing the following objective function

$$\mathbb{P}_n w(X;\hat{\gamma})\left[Y - E\{Y \mid X, D, D(1) > D(0); \alpha, \varphi\}\right]^2,$$

where

$$w(X;\hat{\gamma}) = 1 - D(1-Z)/\{1 - P(Z=1 \mid X;\hat{\gamma})\} - (1-D)Z/P(Z=1 \mid X;\hat{\gamma}).$$

`mle.crude` If $\theta(X) = \delta^L(X)$, then we assume

$$E(Y \mid D=1, X) - E(Y \mid D=0, X) = \tanh(\alpha^\top X).$$

If $\theta(X) = \delta^M(X)$, then we assume

$$\frac{E(Y \mid D=1, X)}{E(Y \mid D=0, X)} = \exp(\alpha^\top X).$$

In addition, for both cases, we assume

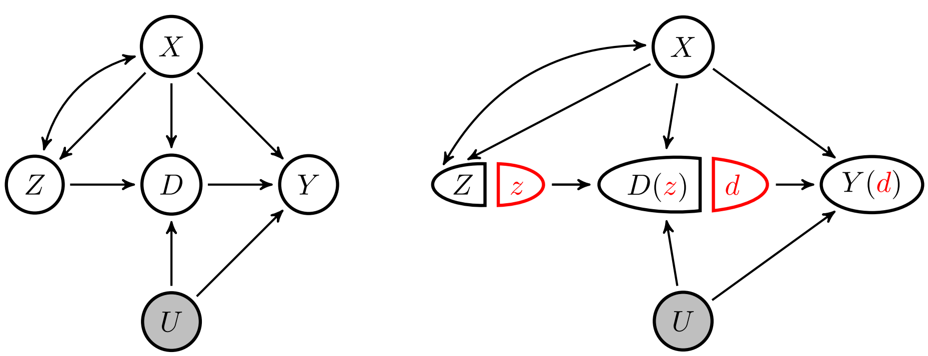
$$OP^{YD}(X) \equiv \frac{P(Y=1 \mid D=1, X)P(Y=1 \mid D=0, X)}{P(Y=0 \mid D=1, X)P(Y=0 \mid D=0, X)} = \exp(\rho^\top X);$$
$$E(D \mid X) = \mathrm{expit}(\upsilon^\top X).$$

The above models are fitted using the maximum likelihood estimation method based on the likelihood function of Y conditional on D and X, and obtained using `R` package `brm`.

5.3. *Detailed simulation results*

Table S2 presents a detailed version of Table 2 in the main paper. Table S3 presents the ratio between Bias and SD for different estimators and scenarios considered in Table S2. Table S4 presents a detailed version of Table 3 in the main paper.

5.4. *The causal model assumed in the application to 401(k) data*



(a). A Causal Graph. (b). A Single World Intervention Graph.

Fig. S2: The causal model assumed in the application to 401(k) data. Variables X, Z, D, Y are observed; U is unobserved. The bi-directed edge between X and Z denotes an unmeasured common cause.

5.5. *Implementation details in the application to 401(k) data*

In the 401(k) data, since only eligible individuals may choose to participate in 401(k) plans, some model assumptions are different from those in the simulation studies. The model fitting methods are the same as that in Section 5.2. In the application, we focus on the case where $\theta(X) = \delta^M(X)$. Models are fitted in a similar fashion as in Section 5.2.

`mle`
`drw` Assume

$$\begin{aligned}
\phi_1(X) &= \text{expit}(\beta_1^\top X);\\
\phi_3(X) &= \text{expit}(\beta_2^\top X);\\
OP^{CO}(X) &= \exp(\eta^\top X);\\
P(Z=1 \mid X) &= \text{expit}(\gamma^\top X).
\end{aligned}$$

`dru.ogburn`
`drw.ogburn` Assume

$$\begin{aligned}
P(Z=1 \mid X) &= \text{expit}(\gamma^\top X);\\
E(H \mid X) &= E(Y \mid Z=0, X) = \text{expit}(\xi^\top X);\\
E(DY \mid Z=1, X) &= \text{expit}(\psi^\top X);\\
E\left[\left.\frac{\{H - E(H \mid X)\}^2}{f^2(Z \mid X)}\right| X\right] &= \exp(\zeta^\top X).
\end{aligned}$$

`dru.simple` Assume

$$\begin{aligned}
P(Z=1 \mid X) &= \text{expit}(\gamma^\top X);\\
E(Y \mid D, X) &= \text{expit}(\varpi_1 D + \varpi_2^\top X);\\
E(D \mid X) &= \text{expit}(\vartheta^\top X).
\end{aligned}$$

`ls.abadie` Assume

$$\begin{aligned}
P(Z=1 \mid X) &= \text{expit}(\gamma^\top X);\\
E\{Y \mid X, D, D(1) > D(0)\} &= \{\exp(\alpha^\top X)\}^D \text{expit}(\varphi^\top X).
\end{aligned}$$

`mle.crude` Assume

$$\begin{aligned}
&\frac{E(Y \mid D=1, X)}{E(Y \mid D=0, X)} = \exp(\alpha^\top X);\\
&OP^{YD}(X) \equiv \frac{P(Y=1 \mid D=1, X)P(Y=1 \mid D=0, X)}{P(Y=0 \mid D=1, X)P(Y=0 \mid D=0, X)} = \exp(\rho^\top X);\\
&E(D \mid X) = \text{expit}(\upsilon^\top X).
\end{aligned}$$

Table S2: The biases and standard errors of the estimated biases in the Monte Carlo study of various estimators. The true value for α_0 and α_1 is 0 and -1, respectively. The sample size is 1000

	$\theta(X) = \delta^L(X)$				$\theta(X) = \delta^M(X)$			
	α_0		α_1		α_0		α_1	
Bias×100 (SE×100)								
mle.bth	0.28	(0.35)	−3.5	(0.78)	−0.092	(0.71)	−3.0	(1.2)
mle.bad	−20	(0.42)	−15	(0.80)	−48	(1.2)	−18	(2.1)
drw.bth	0.55	(0.36)	−4.1	(0.82)	0.54	(0.77)	−5.6	(1.5)
drw.psc	0.060	(0.38)	−5.9	(1.0)	−0.38	(1.2)	−12	(2.7)
drw.opc	0.55	(0.36)	−3.9	(0.79)	0.49	(0.75)	−5.3	(1.4)
drw.bad	−10	(0.40)	−9.6	(1.1)	−28	(1.4)	25	(3.3)
dru.bth	1.3	(0.44)	−5.8	(1.0)	1.8	(0.84)	−8.1	(1.7)
dru.psc	1.2	(0.44)	−6.1	(1.0)	1.9	(0.84)	−9.0	(1.7)
dru.opc	0.94	(0.39)	−4.5	(0.86)	0.99	(0.78)	−6.5	(1.5)
dru.bad	−14	(0.48)	−27	(1.4)	−28	(0.98)	−12	(2.4)
reg.ogburn.bth	−5.7	(1.6)	−2.9	(3.1)	7.8	(2.0)	−1.1	(2.2)
reg.ogburn.bad	−9.0	(0.25)	100	(0.23)	140	(5.6)	93	(3.6)
drw.ogburn.bth	0.10	(0.46)	−4.2	(0.99)	3.2	(1.4)	−13	(2.5)
drw.ogburn.psc	1.3	(0.46)	−8.2	(1.3)	7.7	(1.9)	−18	(3.5)
drw.ogburn.opc	−5.5	(1.1)	−8.0	(1.5)	12	(2.3)	21	(4.7)
drw.ogburn.bad	−120	(3.1)	−170	(6.0)	−40	(2.6)	−9.4	(5.8)
dru.ogburn.bth	1.3	(0.45)	−5.8	(1.1)	1.9	(0.85)	−8.2	(1.7)
dru.ogburn.psc	1.5	(0.49)	−9.1	(1.3)	3.1	(0.90)	−11	(2.0)
dru.ogburn.opc	−2.9	(0.58)	−2.5	(1.1)	4.6	(1.5)	11	(3.0)
dru.ogburn.bad	−130	(3.2)	−190	(6.4)	−47	(1.0)	−41	(3.7)
mle.wang.bth	0.28	(0.35)	−3.8	(0.79)	—		—	
mle.wang.bad	−27	(0.40)	−1.5	(1.1)	—		—	
dru.wang.bth	1.3	(0.45)	−5.8	(1.0)	—		—	
dru.wang.psc	1.2	(0.45)	−6.3	(1.0)	—		—	
dru.wang.opc	0.29	(0.41)	−7.8	(1.0)	—		—	
dru.wang.bad	−20	(0.52)	−24	(1.5)	—		—	
dru.simple.bth	1.3	(0.45)	−5.8	(1.0)	1.8	(0.84)	−8.0	(1.7)
dru.simple.psc	1.2	(0.44)	−6.2	(1.0)	1.9	(0.84)	−8.8	(1.7)
dru.simple.opc	4.5	(0.49)	−17	(1.2)	−0.15	(0.68)	11	(1.2)
dru.simple.bad	−16	(0.48)	−17	(1.3)	−34	(0.70)	18	(1.5)
ls.abadie.bth	−0.19	(0.37)	−4.1	(0.93)	0.42	(0.79)	−11	(1.6)
ls.abadie.bad	−23	(0.88)	22	(1.2)	−32	(1.9)	7.7	(3.6)
mle.crude	−2.8	(0.10)	60	(0.19)	0.36	(0.25)	51	(0.42)

Table S3: The Bias/SD of all the estimators. The true values for α_0 and α_1 are 0 and -1, respectively. The sample size is 1000

	$\theta(X) = \delta^L(X)$		$\theta(X) = \delta^M(X)$	
	α_0	α_1	α_0	α_1
Bias/SD				
mle.bth	0.02	−0.14	−0.00	−0.08
mle.bad	−1.53	−0.61	−1.25	−0.28
drw.bth	0.05	−0.16	0.02	−0.12
drw.psc	0.01	−0.19	−0.01	−0.14
drw.opc	0.05	−0.15	0.02	−0.12
drw.bad	−0.79	−0.28	−0.66	0.24
dru.bth	0.09	−0.18	0.07	−0.15
dru.psc	0.08	−0.19	0.07	−0.16
dru.opc	0.08	−0.17	0.04	−0.13
dru.bad	−0.94	−0.61	−0.89	−0.15
reg.ogburn.bth	−0.11	−0.03	0.13	−0.02
reg.ogburn.bad	−1.16	14.18	0.81	0.82
drw.ogburn.bth	0.01	−0.13	0.07	−0.16
drw.ogburn.psc	0.09	−0.20	0.13	−0.16
drw.ogburn.opc	−0.15	−0.17	0.17	0.14
drw.ogburn.bad	−1.48	−1.07	−0.49	−0.05
dru.ogburn.bth	0.09	−0.17	0.07	−0.15
dru.ogburn.psc	0.10	−0.22	0.11	−0.17
dru.ogburn.opc	−0.16	−0.07	0.10	0.12
dru.ogburn.bad	−1.53	−1.10	−1.47	−0.35
mle.wang.bth	0.03	−0.15	–	–
mle.wang.bad	−2.16	−0.04	–	–
dru.wang.bth	0.09	−0.18	–	–
dru.wang.psc	0.09	−0.19	–	–
dru.wang.opc	0.02	−0.25	–	–
dru.wang.bad	−1.24	−0.52	–	–
dru.simple.bth	0.09	−0.18	0.07	−0.15
dru.simple.psc	0.09	−0.19	0.07	−0.16
dru.simple.opc	0.29	−0.44	−0.01	0.28
dru.simple.bad	−1.07	−0.41	−1.54	0.36
ls.abadie.bth	−0.02	−0.14	0.02	−0.22
ls.abadie.bad	−0.84	0.57	−0.54	0.07
mle.crude	−0.83	10.02	0.04	3.76

Table S4: The coverage probabilities and average widths of confidence intervals obtained from 500 bootstrap samples. The true values for α_0 and α_1 are 0 and -1, respectively. The sample size is 1000

	$\theta(X) = \delta^L(X)$				$\theta(X) = \delta^M(X)$			
	α_0		α_1		α_0		α_1	
Coverage probability ×100 (Average Width)								
mle.bth	95.6	(0.502)	95.8	(1.25)	95.4	(1.12)	96.4	(2.02)
mle.bad	65.4	(0.662)	91.0	(1.38)	46.3	(2.00)	94.6	(3.38)
drw.bth	94.7	(0.513)	95.2	(1.31)	96.9	(1.21)	95.9	(2.34)
drw.psc	95.4	(0.561)	95.5	(1.65)	97.6	(2.43)	97.4	(4.20)
drw.opc	95.0	(0.505)	95.6	(1.25)	96.1	(1.19)	96.1	(2.30)
drw.bad	87.0	(0.614)	95.3	(1.82)	91.8	(2.94)	96.8	(5.54)
dru.bth	94.5	(0.636)	94.6	(1.59)	96.3	(1.46)	96.9	(2.61)
dru.psc	94.8	(0.639)	95.0	(1.62)	96.3	(1.46)	97.6	(2.65)
dru.opc	93.8	(0.564)	94.8	(1.34)	96.1	(1.31)	96.9	(2.33)
dru.bad	79.6	(0.732)	91.3	(2.35)	87.4	(1.61)	98.2	(3.28)
reg.ogburn.bth	98.0	(1.37)	99.9	(3.40)	99.6	(2.38)	100.0	(3.09)
reg.ogburn.bad	75.6	(0.318)	0.1	(0.284)	99.9	(4.81)	86.1	(3.78)
drw.ogburn.bth	97.0	(0.564)	98.1	(1.42)	98.5	(1.85)	98.4	(3.29)
drw.ogburn.psc	95.4	(0.689)	95.7	(2.01)	99.7	(2.95)	99.9	(5.18)
drw.ogburn.opc	98.8	(0.830)	99.0	(1.98)	99.8	(2.39)	99.9	(5.14)
drw.ogburn.bad	2.2	(2.84)	72.5	(5.62)	99.7	(3.06)	99.9	(6.55)
dru.ogburn.bth	94.5	(0.644)	95.0	(1.63)	96.2	(1.37)	97.2	(2.59)
dru.ogburn.psc	93.4	(0.692)	94.3	(2.03)	96.9	(1.51)	97.3	(2.85)
dru.ogburn.opc	99.3	(0.836)	97.9	(1.75)	99.6	(1.76)	99.6	(3.79)
dru.ogburn.bad	1.6	(2.95)	64.0	(5.81)	91.6	(1.39)	99.6	(4.27)
mle.wang.bth	94.3	(0.466)	95.1	(1.13)	—		—	
mle.wang.bad	41.1	(0.564)	94.3	(1.41)	—		—	
dru.wang.bth	94.4	(0.637)	94.7	(1.60)	—		—	
dru.wang.psc	94.8	(0.643)	94.7	(1.65)	—		—	
dru.wang.opc	94.1	(0.581)	94.5	(1.60)	—		—	
dru.wang.bad	68.0	(0.786)	92.1	(2.47)	—		—	
dru.simple.bth	94.3	(0.636)	94.8	(1.59)	96.1	(1.36)	96.5	(2.51)
dru.simple.psc	94.6	(0.641)	94.5	(1.63)	96.1	(1.38)	96.8	(2.57)
dru.simple.opc	93.6	(0.692)	92.5	(1.84)	96.2	(1.05)	94.9	(1.79)
dru.simple.bad	76.4	(0.706)	94.2	(2.16)	69.3	(1.05)	93.8	(2.27)
ls.abadie.bth	94.8	(0.593)	95.5	(1.57)	96.4	(1.24)	95.6	(2.56)
ls.abadie.bad	87.3	(0.898)	94.5	(1.50)	93.1	(2.03)	95.7	(3.86)
mle.crude	84.3	(0.127)	0.0	(0.228)	94.0	(0.303)	6.3	(0.524)

References

Ogburn, E. L., Rotnitzky, A., and Robins, J. M. (2015). Doubly robust estimation of the local average treatment effect curve. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):373–396.

Richardson, T. S., Robins, J. M., and Wang, L. (2017). On modeling and estimation for the relative risk and risk difference. *Journal of the American Statistical Association*, 112(519):1121–1130.

Wang, L. and Tchetgen Tchetgen, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):531–550.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25.