On Semiparametric Instrumental Variable Estimation of Average Treatment Effects through Data Fusion

BaoLuo Sun* and Wang Miao†

* *Department of Statistics and Applied Probability,*

National University of Singapore

†*Department of Probability and Statistics,*

Peking University

Supplementary Material

This supplementary material contains the proof of Lemmas and Theorems in the main manuscript as well as details on asymptotic variance estimation for the proposed estimators.

Proof of Theorem 1

In the proof, we make use of the following equalities that for all square-integrable functions $m(X)$,

$$
\begin{aligned}
E\left\{\frac{(-1)^{1-Z}m(X)}{\lambda(Z|X)[\tau(1,X)-\tau(0,X)]}\middle|R=1\right\} &= 0,\\
E\left\{\frac{(-1)^{1-Z}m(X)\tau(Z,X)}{\lambda(Z|X)[\tau(1,X)-\tau(0,X)]}\middle|R=1\right\} &= E\{m(X)|R=1\}.
\end{aligned}
$$

Under assumptions 1–7 and suppressing the dependences of $\{h_0(\cdot), h_1(\cdot), g_0(\cdot), g_1(\cdot)\}$ on (X, U),

$$
\begin{aligned}
& E\left\{\frac{(-1)^{1-Z}}{\lambda(Z|X)}\frac{Y}{[\tau(1,X)-\tau(0,X)]}\middle|R=1\right\} \\
= \; & E\left\{\frac{(-1)^{1-Z}E(Y|D,Z,X,U,R=1)}{\lambda(Z|X)[\tau(1,X)-\tau(0,X)]}\middle|R=1\right\} \\
= \; & E\left\{\frac{(-1)^{1-Z}(h_0+h_1D)}{\lambda(Z|X)[\tau(1,X)-\tau(0,X)]}\middle|R=1\right\} \\
= \; & E\left\{\frac{(-1)^{1-Z}h_1(g_0+g_1Z)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
& +E\left\{\frac{(-1)^{1-Z}E(h_0|X,Z,R=1)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
= \; & E\left\{\frac{(-1)^{1-Z}E[h_1|X,Z,R=1]\tau(X,Z)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
& +E\left\{\frac{(-1)^{1-Z}Z\mathrm{cov}(g_1,h_1|X,Z,R=1)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
& +E\left\{\frac{(-1)^{1-Z}\mathrm{cov}(g_0,h_1|X,Z,R=1)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
& +E\left\{\frac{(-1)^{1-Z}E(h_0|X,Z,R=1)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
= \; & E\left\{\frac{(-1)^{1-Z}E[h_1|X,R=1]\tau(X,Z)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
& +E\left\{\frac{(-1)^{1-Z}Z\mathrm{cov}(g_1,h_1|X,R=1)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
& +E\left\{\frac{(-1)^{1-Z}\mathrm{cov}(g_0,h_1|X,R=1)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
& +E\left\{\frac{(-1)^{1-Z}E(h_0|X,R=1)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\} \\
= \; & E\left\{h_1|R=1\right\}+E\left\{\frac{(-1)^{1-Z}Z\mathrm{cov}(g_1,h_1|X,R=1)}{\lambda(Z|X)\{\tau(1,X)-\tau(0,X)\}}\middle|R=1\right\},
\end{aligned}
$$

which equals the average treatment effect Δ if $\text{cov}(g_1, h_1|X, R = 1) = 0$ with probability 1. In addition, the propensity score $\tau(z, x)$ can be nonparametrically identified from $f(O)$ under assumptions 8 and 9. The proof is completed by noting that

$$E\left\{\frac{R}{q^\dagger}\frac{(-1)^{1-Z}}{\lambda(Z|X)}\frac{Y}{[\tau(1,X)-\tau(0,X)]}\right\} = E\left\{\frac{(-1)^{1-Z}}{\lambda(Z|X)}\frac{Y}{[\tau(1,X)-\tau(0,X)]}\middle|R=1\right\}.$$

Proof of Lemma 1

$$\begin{aligned}
& E(Y|Z=z, X=x, R=1) \\
= \;& \text{cov}(g_1, h_1|Z=z, X=x, R=1)z + \text{cov}(g_0, h_1|Z=z, X=x, R=1) \\
& \quad + E(h_1|Z=z, X=x, R=1)\tau(z,x) + E(h_0|Z=z, X=x, R=1) \\
= \;& \text{cov}(g_1, h_1|X=x, R=1)z + \text{cov}(g_0, h_1|X=x, R=1) \\
& \quad + E(h_1|X=x, R=1)\tau(z,x) + E(h_0|X=x, R=1) \quad \text{by assumption 3} \\
= \;& \mathcal{H}(x)\tau(z,x) + \omega(x) \quad \text{by assumption 6.}
\end{aligned}$$

Proof of Lemma 2

In the following, let $\bar{a}$ denote the probability limit of $\hat{a}$. Under standard theory for likelihood-based inference (White, 1982),

$$\begin{aligned}
\hat{\alpha} - \bar{\alpha} &= -\left\{\frac{\partial}{\partial\alpha}E[S_\pi(O;\alpha)]\Big|_{\alpha=\bar{\alpha}}\right\}^{-1}\hat{E}\{S_\pi(O;\bar{\alpha})\} + o_p(n^{-1/2}); \\
\hat{\psi} - \bar{\psi} &= -\left\{\frac{\partial}{\partial\psi}E[S_\lambda(O;\psi)]\Big|_{\psi=\bar{\psi}}\right\}^{-1}\hat{E}\{S_\lambda(O;\bar{\psi})\} + o_p(n^{-1/2});
\end{aligned}$$

$$
\begin{aligned}
\hat{\xi}-\bar{\xi} &= -\left\{\frac{\partial}{\partial\xi}E[S_\tau(O;\xi)]\Big|_{\xi=\bar{\xi}}\right\}^{-1}\hat{E}\{S_\tau(O;\bar{\xi})\}+o_p(n^{-1/2});\\
\hat{q}-\bar{q} &= \hat{E}\{R-\bar{q}\}+o_p(n^{-1/2}),
\end{aligned}
$$

where $\{S_\pi, S_\lambda, S_\tau\}$ are the respective scores for the parametric models $\{\pi(\cdot;\alpha),\lambda(\cdot;\psi),\tau(\cdot;\xi)\}$. Let $\delta_1=(\psi^T,\xi^T,q)^T$ denote the nuisance parameters in $\mathcal{M}_1$. By the asymptotic theory of M-estimators (Newey and McFadden, 1994; Van der Vaart, 2000) and Taylor expansion, we obtain

$$
\begin{aligned}
\hat{\Delta}_1-\Delta &= \hat{E}\{\mu_1(O;\Delta,\bar{\delta}_1)\}+(\hat{\psi}-\bar{\psi})^T\times\frac{\partial}{\partial\psi}\hat{E}\{\mu_1(O;\Delta,\delta_1)\}\Big|_{\delta_1=\bar{\delta}_1}\\
&+ (\hat{\xi}-\bar{\xi})^T\times\frac{\partial}{\partial\xi}\hat{E}\{\mu_1(O;\Delta,\delta_1)\}\Big|_{\delta_1=\bar{\delta}_1}\\
&+ (\hat{q}-\bar{q})\times\frac{\partial}{\partial q}\hat{E}\{\mu_1(O;\Delta,\delta_1)\}\Big|_{\delta_1=\bar{\delta}_1}+o_p(n^{-1/2}),
\end{aligned}
$$

so that $\hat{\Delta}_1-\Delta=\hat{E}\{\tilde{\mu}_1(O;\Delta,\bar{\delta}_1)\}+o_p(n^{1/2})$ where

$$
\begin{aligned}
\tilde{\mu}_1(O;\Delta,\bar{\delta}_1) &= \mu_1(O;\Delta,\bar{\delta}_1)\\
&- \frac{\partial}{\partial\psi}E\{\mu_1(O;\Delta,\delta_1)\}\Big|_{\delta_1=\bar{\delta}_1}\left\{\frac{\partial}{\partial\psi}E[S_\lambda(O;\psi)]\Big|_{\psi=\bar{\psi}}\right\}^{-1}S_\lambda(O;\bar{\psi})\\
&- \frac{\partial}{\partial\xi}E\{\mu_1(O;\Delta,\delta_1)\}\Big|_{\delta_1=\bar{\delta}_1}\left\{\frac{\partial}{\partial\xi}E[S_\tau(O;\xi)]\Big|_{\xi=\bar{\xi}}\right\}^{-1}S_\tau(O;\bar{\xi})\\
&+ \frac{\partial}{\partial q}E\{\mu_1(O;\Delta,\delta_1)\}\Big|_{\delta_1=\bar{\delta}_1}\{R-\bar{q}\}+o_p(n^{-1/2}).
\end{aligned}
$$

Under $\mathcal{M}_1$, we have $\bar{\delta}_1=\delta_1^\dagger=(\psi^{\dagger T},\xi^{\dagger T},q^\dagger)^T$ and $E\{\mu_1(O;\Delta,\delta_1^\dagger)\}=0$ by Theorem 1. It follows that $n^{1/2}(\hat{\Delta}_1-\Delta)\xrightarrow{d}N(0,\sigma_1^2)$ where $\sigma_1^2=E\{\tilde{\mu}_1^2(O;\Delta,\delta_1^\dagger)\}$. Let $\delta_2=(\gamma^T,\eta^T,\xi^T,q)^T$ denote the nuisance parameters in $\mathcal{M}_2$. By Taylor expansion,

$$\begin{aligned}\hat{\Delta}_2 - \Delta &= \hat{E}\{\mu_2(O;\Delta,\bar{\gamma}_2,\bar{q})\} + (\hat{\gamma}_2 - \bar{\gamma}_2)^T \times \frac{\partial}{\partial \gamma}\hat{E}\{\mu_2(O;\Delta,\gamma,q)\}\big|_{(\gamma,q)=(\bar{\gamma}_2,\bar{q})} \\ &+ (\hat{q}-\bar{q}) \times \frac{\partial}{\partial q}\hat{E}\{\mu_2(O;\Delta,\gamma,q)\}\big|_{(\gamma,q)=(\bar{\gamma}_2,\bar{q})} + o_p(n^{-1/2}).\end{aligned}$$

Under $\mathcal{M}_2$, $(\bar{\xi},\bar{q}) = (\xi^\dagger, q^\dagger)$, and at the true values $(\gamma^\dagger,\eta^\dagger)$,

$$\begin{aligned}& E\big\{\mathcal{G}_{v,w}(X,Z)\big\{R[Y-\mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\xi^\dagger)-\omega(X;\eta^\dagger)] \\ & \quad -(1-R)\mathcal{H}(X;\gamma^\dagger)[D-\tau(Z,X;\xi^\dagger)]\big\}\big\} \\ = \;& E\big\{\mathcal{G}_{v,w}(X,Z)E\big\{R[Y-\mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\xi^\dagger)-\omega(X;\eta^\dagger)] \\ & \quad -(1-R)\mathcal{H}(X;\gamma^\dagger)[D-\tau(Z,X;\xi^\dagger)] \mid Z,X\big\}\big\} \\ = \;& E\big\{\mathcal{G}_{v,w}(X,Z)\big\{[E(Y|Z,X,R=1)-\mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\xi^\dagger)-\omega(X;\eta^\dagger)]\pi(Z,X) \\ & \quad -\mathcal{H}(X;\gamma^\dagger)[E(D|Z,X,R=0)-\tau(Z,X;\xi^\dagger)](1-\pi(Z,X))\big\}\big\} \\ = \;& 0,\end{aligned}$$

so that under standard regularity conditions for M-estimation (Newey and McFadden, 1994; Van der Vaart, 2000) $\bar{\delta}_2 = \delta_2^\dagger$. We have $E\{\mu_2(O;\Delta,\gamma^\dagger,q^\dagger)\} = 0$ by definition. The asymptotic distribution of $n^{1/2}(\hat{\Delta}_2 - \Delta)$ follows from the previous Taylor expansions by Slutsky's Theorem and the Central Limit Theorem. The expansion for $\hat{\Delta}_3 - \Delta$ can be proven similarly; we note that under $\mathcal{M}_3$, $(\bar{\alpha},\bar{q}) = (\alpha^\dagger, q^\dagger)$, and at the true values $(\gamma^\dagger, \eta^\dagger)$,

$$\begin{aligned}
& E\left\{\mathcal{G}_{v,w}(X,Z)\left\{R[Y-\omega(X;\eta^\dagger)] - \frac{(1-R)\pi(Z,X;\alpha^\dagger)}{1-\pi(Z,X;\alpha^\dagger)}\mathcal{H}(X;\gamma^\dagger)D\right\}\right\} \\
= \; & E\left\{\mathcal{G}_{v,w}(X,Z)E\left\{R[Y-\omega(X;\eta^\dagger)] - \frac{(1-R)\pi(Z,X;\alpha^\dagger)}{1-\pi(Z,X;\alpha^\dagger)}\mathcal{H}(X;\gamma^\dagger)D\middle| Z,X\right\}\right\} \\
= \; & E\Bigg\{\mathcal{G}_{v,w}(X,Z)\Bigg\{\pi(Z,X)[E(Y|Z,X,R=1)-\omega(X;\eta^\dagger)] \\
& \quad - \pi(Z,X;\alpha^\dagger)\mathcal{H}(X;\gamma^\dagger)E(D|Z,X,R=0)\Bigg\}\Bigg\} \\
= \; & E\Bigg\{\mathcal{G}_{v,w}(X,Z)\Bigg\{\pi(Z,X)\mathcal{H}(X)\tau(Z,X) \\
& \quad - \pi(Z,X;\alpha^\dagger)\mathcal{H}(X;\gamma^\dagger)E(D|Z,X,R=0)\Bigg\}\Bigg\} \\
= \; & 0.
\end{aligned}$$

Proof of Theorem 2

We closely follow the structure of semiparametric efficiency bound derivation of Newey (1990), Bickel et al. (1993) and Chen et al. (2008). Consider a parametric path t for the density of the observed data, $f_t(O) = q_t^R(1-q_t)^{1-R}f_t(V|R=1)^R f_t(V|R=0)^{1-R}f_t(Y|V,R=1)^R f_t(D|V)^{1-R}$, where $q_t = \mathrm{pr}_t(R=1)$. We aim to derive the unique influence function $\mu_{\text{eff}}(O)$ under $\mathcal{M}_{\text{np}}$ such that $E\{\mu_{\text{eff}}(O)\} = 0$ and pathwise differentiability holds:

$$\partial\Delta_t/\partial t = E_t\left\{\mu_{\text{eff}}(O)S_t(O)\right\},$$

where

$$
\begin{aligned}
S_t(O) &= \partial \log f_t(O)/\partial t \\
&= \alpha(R - q_t) + (1 - R)S_t(V|R = 0) \\
&\quad + RS_t(V|R = 1) + RS_t(Y|V, R = 1) + (1 - R)S_t(D|V).
\end{aligned}
$$

Following the proof for Theorem 1,

$$
\Delta_t = E_t\left\{\frac{R}{q_t}\frac{(-1)^{1-Z}Y}{\lambda_t(Z|X)[\tau_t(1,X) - \tau_t(0,X)]}\right\} = E_t\left\{\frac{(-1)^{1-Z}Y}{\lambda_t(Z|X)[\tau_t(1,X) - \tau_t(0,X)]}\middle| R = 1\right\}.
$$

Differentiate the integral on the right hand side with respect to t yields

$$
\begin{aligned}
\frac{\partial \Delta_t}{\partial t} &= E_t\left\{\frac{(-1)^{1-Z}YS_t(Y, D, V|R = 1)}{\lambda_t(Z|X)[\tau_t(1,X) - \tau_t(0,X)]}\middle| R = 1\right\} \\
&- E_t\left\{\frac{(-1)^{1-Z}YS_t(Z|X, R = 1)}{\lambda_t(Z|X)[\tau_t(1,X) - \tau_t(0,X)]}\middle| R = 1\right\} \\
&- E_t\left\{\frac{(-1)^{1-Z}Y\frac{\partial[\tau_t(1,X)-\tau_t(0,X)]}{\partial t}}{\lambda_t(Z|X)[\tau_t(1,X) - \tau_t(0,X)]^2}\middle| R = 1\right\} \\
&\equiv A_1 + A_2 + A_3.
\end{aligned}
$$

Consider the terms separately:

$$
\begin{aligned}
A_2 &= -E_t\left\{\frac{(-1)^{1-Z}YS_t(Z|X, R = 1)}{\lambda_t(Z|X)[\tau_t(1,X) - \tau_t(0,X)]}\middle| R = 1\right\} \\
&= -E_t\left\{\left\{\begin{array}{c} E_t\left[\frac{(-1)^{1-Z}Y}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]}\middle| V, R = 1\right] \\ -E\left[\frac{(-1)^{1-Z}Y}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]}\middle| X, R = 1\right]\end{array}\right\} S_t(Z|X, R = 1)\middle| R = 1\right\} \\
&= -E_t\left\{\left\{\begin{array}{c} E_t\left[\frac{(-1)^{1-Z}Y}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]}\middle| V, R = 1\right] \\ -E_t\left[\frac{(-1)^{1-Z}Y}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]}\middle| X, R = 1\right]\end{array}\right\} S_t(Y, D, V|R = 1)\middle| R = 1\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= -E_t\Bigg\{ \Bigg\{ \frac{(-1)^{1-Z}\mathcal{H}(X)\tau_t(Z,X)}{\lambda_t(Z|X)\{\tau_t(1,X)-\tau_t(0,X)\}} + \frac{(-1)^{1-Z}\omega(X)}{\lambda_t(Z|X)\{\tau_t(1,X)-\tau_t(0,X)\}} - \mathcal{H}(X) \Bigg\} \times \\
&\quad S_t(Y,D,V|R=1)\Bigg| R=1 \Bigg\} \\
&= -E_t\Bigg\{ \Bigg\{ \frac{(-1)^{1-Z}\mathcal{H}(X)\tau_t(Z,X)}{\lambda_t(Z|X)\{\tau_t(1,X)-\tau_t(0,X)\}} + \frac{(-1)^{1-Z}\omega(X)}{\lambda_t(Z|X)\{\tau_t(1,X)-\tau_t(0,X)\}} - \mathcal{H}(X) + \Delta \Bigg\} \times \\
&\quad S_t(Y,D,V|R=1)\Bigg| R=1 \Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
A_3 &= -E_t\left\{ \frac{(-1)^{1-Z}Y\frac{\partial[\tau_t(1,X)-\tau_t(0,X)]}{\partial t}}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]^2} \Bigg| R=1 \right\} \\
&= -E_t\left\{ \frac{(-1)^{1-Z}YE(DS_t(D|Z=1,X,R=1)|Z=1,X,R=1)}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]^2} \Bigg| R=1 \right\} \\
&\quad +E_t\left\{ \frac{(-1)^{1-Z}YE(DS_t(D|Z=0,X,R=1)|Z=0,X,R=1)}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]^2} \Bigg| R=1 \right\} \\
&= -\Bigg\{ E_t\left\{ \frac{\mathcal{H}(X)ZDS_t(D|V,R=1)}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]} \Bigg| R=1 \right\} \\
&\quad -E_t\left\{ \frac{\mathcal{H}(X)(1-Z)DS_t(D|V,R=1)}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]} \Bigg| R=1 \right\} \Bigg\} \\
&= -E_t\left\{ \frac{\mathcal{H}(X)(-1)^{1-Z}DS_t(D|V,R=1)}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]} \Bigg| R=1 \right\} \\
&= -E_t\left\{ \frac{\mathcal{H}(X)(-1)^{1-Z}[D-\tau_t(Z,X)]S_t(D|V,R=1)}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]} \Bigg| R=1 \right\} \\
&= -E_t\left\{ \frac{\mathcal{H}(X)(-1)^{1-Z}[D-\tau_t(Z,X)]S_t(Y,D,V|R=1)}{\lambda_t(Z|X)[\tau_t(1,X)-\tau_t(0,X)]} \Bigg| R=1 \right\}.
\end{aligned}
$$

Combining the terms A_1–A_3,

$$
\begin{aligned}
\frac{\partial \Delta_t}{\partial t} &= E_t\left\{ \left[\frac{(-1)^{1-Z}\{Y-\omega(X)-\mathcal{H}(X)D\}}{\lambda_t(Z|X)\left[\tau_t(1,X)-\tau_t(0,X)\right]} + \mathcal{H}(X) - \Delta \right] S_t(Y,D,V|R=1) \Bigg| R=1 \right\} \\
&\equiv E_t\{\varphi(Y,D,V)S_t(Y,D,V|R=1)|R=1\}.
\end{aligned}
$$

Let $\varphi(Y, D, V) = \varphi_1(Y, V) - \varphi_2(D, V)$ where

$$\begin{aligned}
\varphi_1(Y, V) &= \frac{(-1)^{1-Z} Y}{\lambda_t(Z|X)\left[\tau_t(1, X) - \tau_t(0, X)\right]}; \\
\varphi_2(D, V) &= \frac{(-1)^{1-Z} \{\omega(X) + \mathcal{H}(X) D\}}{\lambda_t(Z|X)\left[\tau_t(1, X) - \tau_t(0, X)\right]} - \mathcal{H}(X) + \Delta.
\end{aligned}$$

Then

$$\begin{aligned}
\frac{\partial \Delta_t}{\partial t} &= E_t\{\varphi(Y, D, V) S_t(Y, D, V|R = 1)|R = 1\} \\
&= E_t\left[\varphi_1(Y, V) S_t(Y, V|R = 1)|R = 1\right] - E_t\left[\varphi_2(D, V) S_t(D, V|R = 1)|R = 1\right] \\
&\equiv B_1 - B_2
\end{aligned}$$

We note that

$$\begin{aligned}
B_1 &= E_t\left[\varphi_1(Y, V) S_t(Y|V, R = 1)|R = 1\right] + E_t\left[\varphi_1(Y, V) S_t(V|R = 1)|R = 1\right] \\
&= E_t\left\{\frac{R}{q_t}[\varphi_1(Y, V) - E(\varphi_1|V, R = 1)] S_t(Y|V, R = 1)\right\} \\
&\quad + E_t\left\{\frac{E_t(\varphi_1|V, R = 1)}{q_t} R S_t(V|R = 1)\right\} \\
&= E_t\left\{\frac{R}{q_t}[\varphi_1(Y, V) - E(\varphi_1|V, R = 1)] S_t(O)\right\} \\
&\quad + E_t\left\{\frac{E_t(\varphi_1|V, R = 1)}{q_t} R S_t(O)\right\},
\end{aligned}$$

and similarly

$$\begin{aligned}
B_2 &= E_t\left[\varphi_2(D, V) S_t(D|V)|R = 1\right] + E_t\left[\varphi_2(D, V) S_t(V|R = 1)|R = 1\right] \\
&= E_t\left\{\frac{1 - R}{q_t}\frac{\pi_t(V)}{1 - \pi_t(V)}[\varphi_2(D, V) - E(\varphi_2|V)] S_t(D|V)\right\} \\
&\quad + E_t\left\{\frac{E_t(\varphi_2|V)}{q_t} R S_t(V|R = 1)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= E_t\left\{\frac{1-R}{q_t}\frac{\pi_t(V)}{1-\pi_t(V)}[\varphi_2(D,V)-E(\varphi_2|V)]S_t(O)\right\}\\
&\quad +E_t\left\{\frac{E_t(\varphi_2|V)}{q_t}RS_t(O)\right\},
\end{aligned}
$$

where $q_t=\int\pi_t(v)f_t(v)dv$. Therefore $\frac{\partial\Delta_t}{\partial t}=E_t\left\{\mu_{\text{eff}}(O)S_t(O)\right\}$, where

$$
\begin{aligned}
\mu_{\text{eff}}(O;\Delta) &= \frac{R}{q_t}[\varphi_1(Y,V)-E(\varphi_1|V,R=1)]\\
&\quad -\frac{1-R}{q_t}\frac{\pi_t(V)}{1-\pi_t(V)}[\varphi_2(D,V)-E(\varphi_2|V)]\\
&\quad +\frac{R}{q_t}[E(\varphi_1|V,R=1)-E(\varphi_2|V)].
\end{aligned}
$$

It it straightforward to verify that $E_t\{\mu_{\text{eff}}(O;\Delta)\}=0$. It follows by standard semiparametric efficiency theory that $\mu_{\text{eff}}(O;\Delta)$ is the unique (and hence also efficient) influence function, and the semiparametric efficiency bound for all regular and asymptotically linear estimators of Δ in $\mathcal{M}_{\text{np}}$ is $E\{\mu^2_{\text{eff}}(O;\Delta)\}$.

Proof of Lemma 3

Let $\delta=(\eta^T,\gamma^T,\psi^T,\xi^T,\alpha^T,q)^T$ denote the nuisance parameters. By the asymptotic theory of M-estimators (Van der Vaart, 2000) and Taylor expansion, we obtain

$$
\hat{\Delta}_{\text{mul}}-\Delta = \hat{E}\{\mu_{\text{eff}}(O;\Delta,\bar{\delta})\}+(\hat{\delta}-\bar{\delta})^T\times\frac{\partial}{\partial\delta}\hat{E}\{\mu_{\text{eff}}(O;\Delta,\delta)\}\big|_{\delta=\bar{\delta}}+o_p(n^{-1/2}).
$$

It suffices to show that $E\{\mu_{\text{eff}}(O;\Delta,\bar{\delta})\}=0$ in the union model $\cup_{j=1}^3\mathcal{M}_j$.

Under $\mathcal{M}_1$, we have $(\bar{\psi}, \bar{\xi}, \bar{q}) = (\psi^\dagger, \xi^\dagger, q^\dagger)$ and

$$
\begin{aligned}
& E\{\mu_{\text{eff}}(O; \Delta, \bar{\eta}, \bar{\gamma}, \psi^\dagger, \xi^\dagger, \bar{\alpha}, q^\dagger)\} \\
= \;& E\left\{ \frac{(-1)^{1-Z}}{\lambda(Z|X;\psi^\dagger)} \times \frac{\frac{R}{q^\dagger} Y}{[\tau(1,X;\xi^\dagger) - \tau(0,X;\xi^\dagger)]} \right\} \\
& - E\left\{ \frac{(-1)^{1-Z}}{\lambda(Z|X;\psi^\dagger)} \times \frac{\frac{R}{q^\dagger}[\mathcal{H}(X;\bar{\gamma})\tau(Z,X;\xi^\dagger) + \omega(X;\bar{\eta})]}{[\tau(1,X;\xi^\dagger) - \tau(0,X;\xi^\dagger)]} \right\} \\
& - E\left\{ \frac{(-1)^{1-Z}}{\lambda(Z|X;\psi^\dagger)} \times \frac{\frac{1-R}{q^\dagger} \frac{\pi(Z,X;\bar{\alpha})}{1-\pi(Z,X;\bar{\alpha})} \mathcal{H}(X;\bar{\gamma})[D - \tau(Z,X;\xi^\dagger)]}{[\tau(1,X;\xi^\dagger) - \tau(0,X;\xi^\dagger)]} \right\} + E\left\{ \frac{R}{q^\dagger}[\mathcal{H}(X;\bar{\gamma})] \right\} \\
= \;& E\left\{ \frac{(-1)^{1-Z}}{\lambda(Z|X;\psi^\dagger)} \times \frac{\frac{R}{q^\dagger} Y}{[\tau(1,X;\xi^\dagger) - \tau(0,X;\xi^\dagger)]} \right\} - E\left\{ \frac{R}{q^\dagger} \mathcal{H}(X;\bar{\gamma}) \right\} \\
& - E\left\{ \frac{(-1)^{1-Z}}{\lambda(Z|X;\psi^\dagger)} \times \frac{\frac{1-\pi(Z,X)}{q^\dagger} \frac{\pi(Z,X;\bar{\alpha})}{1-\pi(Z,X;\bar{\alpha})} \mathcal{H}(X;\bar{\gamma})[E(D|Z,X,R=0) - \tau(Z,X;\xi^\dagger)]}{[\tau(1,X;\xi^\dagger) - \tau(0,X;\xi^\dagger)]} \right\} \\
& + E\left\{ \frac{R}{q^\dagger}[\mathcal{H}(X;\bar{\gamma}) - \Delta] \right\} \\
= \;& E\left\{ \frac{(-1)^{1-Z}}{\lambda(Z|X;\psi^\dagger)} \times \frac{\frac{R}{q^\dagger} Y}{[\tau(1,X;\xi^\dagger) - \tau(0,X;\xi^\dagger)]} \right\} - \Delta \\
= \;& 0,
\end{aligned}
$$

by Theorem 1.

Under $\mathcal{M}_2$, $(\bar{\xi}, \bar{q}) = (\xi^\dagger, q^\dagger)$ and at the true values $(\gamma^\dagger, \eta^\dagger)$,

$$
\begin{aligned}
& E\Bigg\{ \mathcal{G}_{v,w}(X,Z) \Bigg\{ R[Y - \mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\xi^\dagger) - \omega(X;\eta^\dagger)] \\
& - \frac{(1-R)\pi(Z,X;\bar{\alpha})}{1-\pi(Z,X;\bar{\alpha})} \mathcal{H}(X;\gamma^\dagger)[D - \tau(Z,X;\xi^\dagger)] \Bigg\} \Bigg\} \\
= \;& E\Bigg\{ \mathcal{G}_{v,w}(X,Z) E\Bigg\{ R[Y - \mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\xi^\dagger) - \omega(X;\eta^\dagger)] \\
& - \frac{(1-R)\pi(Z,X;\bar{\alpha})}{1-\pi(Z,X;\bar{\alpha})} \mathcal{H}(X;\gamma^\dagger)[D - \tau(Z,X;\xi^\dagger)] \Bigg| Z, X \Bigg\} \Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&= E\Bigg\{\mathcal{G}_{v,w}(X,Z)\Bigg\{\pi(Z,X)[E(Y|Z,X,R=1)-\mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\xi^\dagger)-\omega(X;\eta^\dagger)] \\
&- \frac{(1-\pi(Z,X))\pi(Z,X;\bar{\alpha})}{1-\pi(Z,X;\bar{\alpha})}\mathcal{H}(X;\gamma^\dagger)[E(D|Z,X,R=1)-\tau(Z,X;\xi^\dagger)]\Bigg\}\Bigg\}=0.
\end{aligned}
$$

In addition,

$$
\begin{aligned}
& E\{\mu_{\text{eff}}(O;\Delta,\eta^\dagger,\gamma^\dagger,\bar{\psi},\xi^\dagger,\bar{\alpha},q^\dagger)\} = E\{E\{\mu_{\text{eff}}(O;\Delta,\eta^\dagger,\gamma^\dagger,\bar{\psi},\xi^\dagger,\bar{\alpha},q^\dagger)|Z,X\}\} \\
&= E\Bigg\{\frac{(-1)^{1-Z}}{\lambda(Z|X;,\bar{\psi})}\times\frac{\frac{\pi(Z,X)}{q^\dagger}[E(Y|Z,X,R=1)-\mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\xi^\dagger)-\omega(X;\eta^\dagger)]}{[\tau(1,X;\xi^\dagger)-\tau(0,X;\xi^\dagger)]}\Bigg\} \\
&\quad -E\Bigg\{\frac{(-1)^{1-Z}}{\lambda(Z|X;,\bar{\psi})}\times\frac{\frac{1-\pi(Z,X)}{q^\dagger}\frac{\pi(Z,X;\bar{\alpha})}{1-\pi(Z,X;\bar{\alpha})}\mathcal{H}(X;\gamma^\dagger)[E(D|Z,X,R=0)-\tau(Z,X;\xi^\dagger)]}{[\tau(1,X;\xi^\dagger)-\tau(0,X;\xi^\dagger)]}\Bigg\} \\
&\quad +E\Bigg\{\frac{R}{q^\dagger}[\mathcal{H}(X;\gamma^\dagger)-\Delta]\Bigg\} \\
&= E\{\mathcal{H}(X;\gamma^\dagger)|R=1\}-\Delta=0.
\end{aligned}
$$

Under $\mathcal{M}_3$, $(\bar{\alpha},\bar{q})=(\alpha^\dagger,q^\dagger)$ and at the true values $(\gamma^\dagger,\eta^\dagger)$,

$$
\begin{aligned}
& E\Bigg\{\mathcal{G}_{v,w}(X,Z)\Bigg\{R[Y-\mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\bar{\xi})-\omega(X;\eta^\dagger)] \\
& -\frac{(1-R)\pi(Z,X;\alpha^\dagger)}{1-\pi(Z,X;\alpha^\dagger)}\mathcal{H}(X;\gamma^\dagger)[D-\tau(Z,X;\bar{\xi})]\Bigg\}\Bigg\} \\
&= E\Bigg\{\mathcal{G}_{v,w}(X,Z)E\Bigg\{R[Y-\mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\bar{\xi})-\omega(X;\eta^\dagger)] \\
& -\frac{(1-R)\pi(Z,X;\alpha^\dagger)}{1-\pi(Z,X;\alpha^\dagger)}\mathcal{H}(X;\gamma^\dagger)[D-\tau(Z,X;\bar{\xi})]\Bigg|Z,X\Bigg\}\Bigg\} \\
&= \hat{E}\Bigg\{\mathcal{G}_{v,w}(X,Z)\Bigg\{\pi(Z,X)\mathcal{H}(X;\gamma^\dagger)[\tau(Z,X)-\tau(Z,X;\bar{\xi})] \\
&- \pi(Z,X;\alpha^\dagger)\mathcal{H}(X;\gamma^\dagger)[\tau(Z,X)-\tau(Z,X;\bar{\xi})]\Bigg\}\Bigg\}=0.
\end{aligned}
$$

In addition,

$$
\begin{aligned}
& E\{\mu_{\text{eff}}(O;\Delta,,\eta^\dagger,\gamma^\dagger,\bar{\psi},\bar{\xi},\alpha^\dagger,q^\dagger)\} = E\{E\{\mu_{\text{eff}}(O;\Delta,\eta^\dagger,\gamma^\dagger,\bar{\psi},\bar{\xi},\alpha^\dagger,q^\dagger)|Z,X\}\} \\
= \; & E\left\{\frac{(-1)^{1-Z}}{\lambda(Z|X;,\bar{\psi})} \times \frac{\frac{\pi(Z,X)}{q^\dagger}[E(Y|Z,X,R=1)-\mathcal{H}(X;\gamma^\dagger)\tau(Z,X;\bar{\xi})-\omega(X;\eta^\dagger)]}{[\tau(1,X;\bar{\xi})-\tau(0,X;\bar{\xi})]}\right\} \\
& -E\left\{\frac{(-1)^{1-Z}}{\lambda(Z|X;,\bar{\psi})} \times \frac{\frac{\pi(Z,X;\alpha^\dagger)}{q^\dagger}\mathcal{H}(X;\gamma^\dagger)[E(D|Z,X,R=0)-\tau(Z,X;\bar{\xi})]}{[\tau(1,X;\bar{\xi})-\tau(0,X;\bar{\xi})]}\right\} \\
& +E\left\{\frac{R}{q^\dagger}[\mathcal{H}(X;\gamma^\dagger)-\Delta]\right\} \\
= \; & E\left\{\frac{(-1)^{1-Z}}{\lambda(Z|X;,\bar{\psi})} \times \frac{\frac{\pi(Z,X)}{q^\dagger}\mathcal{H}(X;\gamma^\dagger)[\tau(Z,X)-\tau(Z,X;\bar{\xi})]}{[\tau(1,X;\bar{\xi})-\tau(0,X;\bar{\xi})]}\right\} \\
& -E\left\{\frac{(-1)^{1-Z}}{\lambda(Z|X;,\bar{\psi})} \times \frac{\frac{\pi(Z,X;\alpha^\dagger)}{q^\dagger}\mathcal{H}(X;\gamma^\dagger)[\tau(Z,X)-\tau(Z,X;\bar{\xi})]}{[\tau(1,X;\bar{\xi})-\tau(0,X;\bar{\xi})]}\right\} \\
& +E\left\{\frac{R}{q^\dagger}[\mathcal{H}(X;\gamma^\dagger)-\Delta]\right\} \\
= \; & E\{\mathcal{H}(X;\gamma^\dagger)|R=1\}-\Delta = 0.
\end{aligned}
$$

The last claim in Lemma 4 follows by noting that under the intersection submodel $\left\{\cap_{j=1}^3\mathcal{M}_j\right\}$, $\bar{\delta}=\delta^\dagger$ and $\frac{\partial}{\partial\delta}\hat{E}\{\mu_{\text{eff}}(O;\Delta,\delta)\}\big|_{\delta=\delta^\dagger}=o_p(1)$ so that

$$\hat{\Delta}_{\text{mul}}-\Delta \;=\; \hat{E}\{\mu_{\text{eff}}(O;\Delta,\delta^\dagger)\}+o_p(n^{-1/2}).$$

Estimation of the asymptotic variance

Let $S_\pi(Z_i,X_i;\alpha)$, $S_\lambda(Z_i,X_i;\psi)$ and $S_\tau(Z_i,X_i;\xi)$ denote units i's contribution to the score for estimation of the parameters which index the models $\pi(z,x;\alpha)$, $\lambda(z|x;\psi)$ and $\tau(z,x;\xi)$ respectively. For example, if we assume the logistic model $\pi(z,x;\alpha)=\left\{1+\exp[-h^T(z,x)\alpha]\right\}^{-1}$ for a vector of regressors $h(z,x)$, then $S_\pi(Z_i,X_i;\alpha)=h(Z_i,X_i)\{R-\pi(Z_i,X_i;\alpha)\}$. For completeness we also denote $S(R_i,q)=R_i-q$.

We consider inference for the semiparametric estimator $\hat{\Delta}_1$. Let $\kappa = (\psi,\xi,q)$ denote the set of nuisance parameters, and let

$$H(O_i;\kappa)=\{S_\lambda^T(Z_i,X_i;\psi),S_\tau^T(Z_i,X_i;\xi),S(R_i,q)\}^T.$$

It follows under standard regularity conditions that

$$n^{1/2}(\hat{\Delta}_1-\Delta)=n^{-1/2}E\left\{\frac{\partial\mu_1(O_i,\Delta,\kappa)}{\partial\Delta}\right\}^{-1}\sum_i V(O_i;\Delta_1,\kappa)+o_p(1),$$

where $V(O_i;\Delta_1,\kappa)\equiv\mu_1(O_i,\Delta,\kappa)-\frac{\partial E\{\mu_1(O,\Delta,\kappa)\}}{\partial\kappa}E\left[\frac{\partial H(O;\kappa)}{\partial\kappa}\right]^{-1}H(O_i;\kappa)$. A consistent estimator of the asymptotic variance of $n^{1/2}(\hat{\Delta}_1-\Delta)$ is therefore given by

$$\hat{\sigma}^2 \quad = \quad \left\{\hat{E}\left(\frac{\partial\mu_1(O_i,\Delta,\kappa)}{\partial\Delta}\bigg|_{(\Delta,\kappa)=(\hat{\Delta}_1,\hat{\kappa})}\right)\right\}^{-2}\hat{E}\{\hat{V}^2(O_i;\hat{\Delta}_1,\hat{\kappa})\},$$

where $\hat{V}(O_i;\Delta_1,\kappa)\equiv\mu_1(O_i,\Delta,\kappa)-\frac{\partial\hat{E}\{\mu_1(O,\Delta,\kappa)\}}{\partial\kappa}\hat{E}\left[\frac{\partial H(O;\kappa)}{\partial\kappa}\right]^{-1}H(O_i;\kappa)$ and $\hat{\kappa}=(\hat{\psi}^T,\hat{\xi}^T,\hat{q})^T$. Accordingly, a 95% Wald confidence interval for Δ is given by $\{\hat{\Delta}_1\pm 1.96\sqrt{n^{-1}\hat{\sigma}^2}\}$. Inferences for $\hat{\Delta}_2$, $\hat{\Delta}_3$ and $\hat{\Delta}_{\text{mul}}$ may be carried out similarly, with a different set of nuisance parameters κ for each semiparametric estimator.

References

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press Baltimore.

Chen, X., Hong, H., Tarozzi, A., et al. (2008). Semiparametric efficiency in gmm models with auxiliary data. *The Annals of Statistics*, 36(2):808–843.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5(2):99–135.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111 – 2245. Elsevier.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge University Press.

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