Supplementary material to “Identification of the outcome distribution and sensitivity analysis under weak confounder-instrument interaction”

Lu Mao*

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, 610 Walnut St, Room 207A, Madison, WI 53726, U.S.A.

Abstract

The supplementary material includes theorem proofs and additional numerical results for the sensitivity analysis of the U.S. National Job Training Partnership Act study.

S1. Technical results

S1.1. Proof of Theorem 3.1 of the main text

Proof. We first show that EIF(y) is the canonical gradient, and thus the non-parametric efficient influence function, of νa(y). The proof borrows ideas from §5 in the Supplementary Materials of Wang & Tchetgen Tchetgen (2018) for the average treatment effect.

Let νa,t(y) denote the estimand under a one-dimensional smooth model indexed by t with νa,0(y) = νa(y). Write the corresponding density of (Y, A, Z, X) in the factorized form

q(Y, A, Z, X; t) = q(Y, A | Z, X; t)q(Z | X; t)q(X; t),

where q(Y, A | Z, X; t) denotes the conditional density of (Y, A) | (Z, X), q(Z | X; t) the conditional probability of Z | X, and q(X; t) the marginal density of

*Corresponding author

Email address: lmao@biostat.wisc.edu (Lu Mao)
We thus proceed to find the expression of

\[ S(Y, A, Z, X) = S(Y, A \mid Z, X) + S(Z \mid X) + S(X), \tag{S1} \]

where \( S(Y, A, Z, X) = \partial \log \{ q(Y, A, Z, X ; t) \} / \partial t \mid_{t=0}, \) \( S(Y, A \mid Z, X) = \partial \log \{ q(Y, A \mid Z, X ; t) \} / \partial t \mid_{t=0} \) satisfying \( E \{ S(Y, A \mid Z, X) \mid Z, X \} = 0, \) \( S(Z \mid X) = \partial \log \{ q(Z \mid X ; t) \} / \partial t \mid_{t=0} \) satisfying \( E \{ S(Z \mid X) \mid X \} = 0, \) and \( S(X) = \partial \log \{ q(X ; t) \} / \partial t \mid_{t=0} \) satisfying \( E \{ S(X) \} = 0. \)

Then, the canonical gradient is the function \( G(y; Y, A, Z, X) \) that satisfies

\[ \frac{\partial}{\partial t} \bigg|_{t=0} \nu_a, t(y) = E \{ G(y; Y, A, Z, X) S(Y, A, Z, X) \}. \tag{S2} \]

We thus proceed to find the expression of \( G(y; Y, A, Z, X) \) through \( \text{(S2)}. \) Let

\[ \delta_{a, z}^{Y, A}(y \mid x) = \text{pr}(Y \leq y, A = a \mid Z = z, X = x) = p_{a, z}(y \mid x) \delta_{a, z}(x) \]

and

\[ \Delta \delta_{a}^{Y, A}(y \mid x) = \delta_{a,1}^{Y, A}(y \mid x) - \delta_{a,0}^{Y, A}(y \mid x). \]

Thus \( \nu_a(y \mid x) = \Delta \delta_{a}(x)^{-1} \Delta \delta_{a}^{Y, A}(y \mid x). \) As a result, we find that

\[ \frac{\partial}{\partial t} \bigg|_{t=0} \nu_a, t(y) = \frac{\partial}{\partial t} \bigg|_{t=0} E \{ \nu_a, t(y \mid X) \}
\]

\[ = E \{ [\nu_a(y \mid X) - \nu_a(y)] S(X) \} + E \bigg\{ \frac{\partial}{\partial t} \bigg|_{t=0} \nu_a, t(y \mid X) \bigg\}
\]

\[ = E \{ [\nu_a(y \mid X) - \nu_a(y)] S(Y, A, Z, X) \}
\]

\[ + E \left[ \Delta \delta_{a}(X)^{-1} \left\{ \frac{\partial}{\partial t} \bigg|_{t=0} \Delta \delta_{a, t}^{Y, A}(y \mid X) - \nu_a(y \mid X) \frac{\partial}{\partial t} \bigg|_{t=0} \Delta \delta_{a, t}(X) \right\} \right], \]

where an extra subscript \( t \) is used to denote the corresponding quantity under parameter \( t. \) Let \( S(Y, A, Z \mid X) = S(Y, A \mid Z, X) + S(Z \mid X). \) If we can rewrite the second term on the far right hand side of the display into

\[ E \{ G(y; Y, A, Z \mid X) S(Y, A, Z \mid X) \} = E \{ G(y; Y, A, Z \mid X) S(Y, A, Z, X) \} \]

for some \( G(y; Y, A, Z \mid X) \) with \( E \{ G(y; Y, A, Z \mid X) \mid X \} = 0. \) Then the canonical gradient is

\[ G(y; Y, A, Z, X) = G(y; Y, A, Z \mid X) + \nu_a(y \mid X) - \nu_a(y). \tag{S3} \]
To find this $G(y; Y, A, Z \mid X)$, observe that
\[
\frac{\partial}{\partial t} \bigg|_{t=0} \Delta \delta_{a,t}^Y, A(y \mid X) = \frac{\partial}{\partial t} \bigg|_{t=0} \left[ E_t\{I(Y \leq y, A = a) \mid Z = 1, X\} - E_t\{I(Y \leq y, A = a) \mid Z = 0, X\} \right] = \frac{\partial}{\partial t} \bigg|_{t=0} E_t \left\{ \frac{2Z - 1}{q(Z \mid X; t)} I(Y \leq y, A = a) \mid X \right\} = E \left\{ \frac{2Z - 1}{q(Z \mid X)} I(Y \leq y, A = a)S(Y, A, Z \mid X) \mid X \right\} - E \left\{ \frac{2Z - 1}{q(Z \mid X)} I(Y \leq y, A = a)S(Z \mid X) \mid X \right\} = E \left[ \frac{2Z - 1}{q(Z \mid X)} \{I(Y \leq y, A = a) - \text{pr}(Y \leq y, A = a \mid Z, X)\}S(Y, A, Z \mid X) \mid X \right].
\]

Likewise,
\[
\frac{\partial}{\partial t} \bigg|_{t=0} \Delta \delta_{a,t}(y \mid X) = E \left[ \frac{2Z - 1}{q(Z \mid X)} \{I(A = a) - \text{pr}(A = a \mid Z, X)\}S(Y, A, Z \mid X) \mid X \right].
\]

We therefore have that
\[
G(y; Y, A, Z \mid X) = \Delta \delta_{a}(X)^{-1} \frac{2Z - 1}{q(Z \mid X)} \left[ I(Y \leq y, A = a) - \text{pr}(Y \leq y, A = a \mid Z, X) \right]
- \nu_a(y \mid X) \{I(A = a) - \delta_{a,z}(X)\}.
\]

Using $\text{pr}(Y \leq y, A = a \mid Z, X) = Z \Delta \delta^Y, A(y \mid X) + p_{a,0}(y \mid X)\delta_{a,0}(X)$ and
\[
\nu_a(y \mid X)\delta_{a,z}(X) = \nu_a(y \mid X)\{Z \Delta \delta_a(X) + \delta_{a,0}(X)\} = Z \Delta \delta^Y, A(y \mid X) + \nu_a(y \mid X)\delta_{a,0}(X),
\]
we find that
\[
G(y; Y, A, Z \mid X) = \Delta \delta_{a}(X)^{-1} \frac{2Z - 1}{q(Z \mid X)} \left[ I(Y \leq y, A = a) - I(A = 0)\nu_a(y \mid X)
- \delta_{a,0}(X)\{p_{a,0}(y \mid X) - \nu_a(y \mid X)\}\right].
\]

Combine this with $S3$ to find that $G(y; Y, A, X, Z) = \text{EIF}(y)$. Similarly to Wang & Tchetgen Tchetgen (2018), we can also show that $\text{EIF}(y)$ coincides with the efficient influence in $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$ using the results of Robins & Rotnitzky (2001). \qed
S1.2. Proof of Proposition 3.1 of the main text

Proof. Let $\alpha, \beta, \gamma, \theta,$ and $\zeta$ denote the estimands of $\hat{\alpha}_{dr}, \hat{\beta}_{dr}, \hat{\gamma}, \hat{\theta},$ and $\hat{\zeta},$ respectively. We first show the double robustness of $\hat{\beta}_{dr}$ and $\hat{\alpha}_{dr}$. For $\hat{\beta}_{dr},$ we want to show that

$$E \left[ \frac{2Z - 1}{q(Z | X; \gamma)} \{ A - \Delta\delta(X; \beta)Z - \delta_0(X; \zeta) \} \right] = 0 \quad \text{(S4)}$$

under $\mathcal{M}_1 \cup \mathcal{M}_2$. When $\mathcal{M}_1$ is true, we have that $\Delta\delta(X; \beta) = \Delta\delta(X)$ and $\delta_0(X; \zeta) = \delta_0(X),$ and (S4) can be obtained through

$$E \{ A - \Delta\delta(X)Z - \delta_0(X) \mid Z, X \}.$$

When $\mathcal{M}_2$ is true, we have that $\Delta\delta(X; \beta) = \Delta\delta(X)$ and $q(Z | X) = q(Z | X; \gamma),$ and (S4) can be obtained through

$$E \left[ \frac{2Z - 1}{q(Z | X)} \{ A - \Delta\delta(X)Z - \delta_0(X; \zeta) \} \right] | X = 0.$$

For $\hat{\alpha}_{dr},$ we want to show that

$$E \left( \frac{2Z - 1}{q(Z | X; \gamma)} \right) \left[ I(A = a) \{ f(Y) - \nu_a(f \mid X; \alpha) \} ight.$$

$$\left. - \delta_{a,0}(X; \zeta) \{ p_{a,0}(f \mid X; \theta) - \nu_a(f \mid X; \alpha) \} \right] = 0 \quad \text{(S5)}$$

for arbitrary $f$ under $\mathcal{M}_1 \cup \mathcal{M}_3$. When $\mathcal{M}_1$ is true, we have that $\nu_a(f \mid X; \alpha) = \nu_a(f \mid X),$ $\delta_{a,0}(X; \zeta) = \delta_{a,0}(X),$ and $p_{a,0}(f \mid X; \theta) = p_{a,0}(f \mid X),$ and (S5) can be obtained through

$$E \left[ I(A = a) \{ f(Y) - \nu_a(f \mid X) \} - \delta_{a,0}(X) \{ p_{a,0}(f \mid X) - \nu_a(f \mid X) \} \right] | Z, X = 0.$$

When $\mathcal{M}_3$ is true, we have that $\nu_a(f \mid X; \alpha) = \nu_a(f \mid X)$ and $q(Z \mid X; \gamma) = q(Z \mid X),$ and (S5) can be obtained through

$$E \left( \frac{2Z - 1}{q(Z \mid X)} \right) \left[ I(A = a) \{ f(Y) - \nu_a(f \mid X) \} ight.$$

$$\left. - \delta_{a,0}(X; \zeta) \{ p_{a,0}(f \mid X; \theta) - \nu_a(f \mid X) \} \right] | X = 0.$$
For the multiple robustness of \( \hat{\nu}_{a,mr}(y) \), we want to show that

\[
E \left( \frac{2Z - 1}{q(Z \mid X; \gamma)} \Delta \delta_a(X; \beta)^{-1} \left[ I(Y \leq y, A = a) - I(A = a)\nu_a(y \mid X; \alpha) \right. \\
- \delta_{a,0}(X; \zeta)\{p_{a,0}(y \mid X; \theta) - \nu_a(y \mid X; \alpha)\} \left. \right] + \nu_a(y \mid X; \alpha) \right) = \nu_a(y) 
\] (S6)

under \( M_1 \cup M_2 \cup M_3 \). For \( M_1 \cup M_3 \), (S6) can be obtained similarly to the proof of double robustness for \( \hat{\alpha}_{dr} \). For \( M_2 \), we can obtain (S6) by showing that

\[
E \left( \frac{2Z - 1}{q(Z \mid X)} \Delta \delta_a(X)^{-1} \left[ I(Y \leq y, A = a) - I(A = a)\nu_a(y \mid X; \alpha) \right. \\
- \delta_{a,0}(X; \zeta)\{p_{a,0}(y \mid X; \theta) - \nu_a(y \mid X; \alpha)\} \left. \right] + \nu_a(y \mid X; \alpha) \mid X \right) = \nu_a(y \mid X). 
\]

\( \square \)

S1.3. Details for Proposition 4.1 of the main text

By standard derivation, we find \( L(\pi^*; \xi) \) and \( H(\pi^*; \xi) \) in the following form:

(i) If \( \xi^- \leq 1 \),

\[
L(\pi^*; \xi) = \begin{cases} 
\frac{\pi^*}{1+\xi^-}, & 0 \leq \pi^* \leq \frac{\xi^- (1+\xi^+)}{\xi^- + \xi^+} \\
\frac{\pi^* - \xi^-}{1-\xi^-}, & \frac{\xi^- (1+\xi^+)}{\xi^- + \xi^+} < \pi^* \leq 1 \\
\frac{\pi^*}{1-\xi^-}, & 0 \leq \pi^* \leq \frac{\xi^+ (1-\xi^-)}{\xi^- + \xi^+} \\
\frac{\pi^* + \xi^+}{1+\xi^+}, & \frac{\xi^+ (1-\xi^-)}{\xi^- + \xi^+} < \pi^* \leq 1 
\end{cases}
\]

and \( H(\pi^*; \xi) = \)

(ii) if \( \xi^- > 1 \),

\[
L(\pi^*; \xi) = \begin{cases} 
\frac{\pi^*}{1-\xi^-}, & \frac{\xi^- (1-\xi^-)}{\xi^- + \xi^+} \leq \pi^* < 0 \\
\frac{\pi^*}{1+\xi^+}, & 0 \leq \pi^* \leq 1 \\
\frac{\pi^* + \xi^+}{1+\xi^+}, & 0 \leq \pi^* \leq 1 \\
\frac{\pi^* - \xi^-}{1-\xi^-}, & 1 < \pi^* \leq \frac{\xi^- (1+\xi^+)}{\xi^- + \xi^+} 
\end{cases}
\]

and \( H(\pi^*; \xi) = \
for $\xi = (\xi^-, \xi^+) \in \mathbb{R}^{+2}$.

Clearly, $\xi^-(x) = 1$ implies that $\delta_1(x,U) \geq \delta_0(x,U)$. If we have a causal instrument, we can interpret $\Delta \delta(x)$ in this case as the proportion of compliers in the covariate group $x$. With $\xi^+(x) = \Delta \delta(x)^{-1} - 1$ we have that $\mathcal{L}(\nu^{\star}_a(y \mid x); \xi(x)) \leq 1/(1 + \xi^+(x)) = \Delta \delta(x)$. This means that $100\{1 - \Delta \delta(x)\}%$ of the population is unaccounted for and must be placed at $\infty$ in order for the cumulative distribution to reach the lower bound. Likewise, we have that $\mathcal{H}(\nu^{\star}_a(y \mid x); \xi(x)) \geq \xi^+(x)/(1 + \xi^+(x)) = 1 - \Delta \delta(x)$, which means that $100\{1 - \Delta \delta(x)\}%$ of the population is unaccounted for and must be placed at $-\infty$ in order for the cumulative distribution reach the upper bound. However, since extra information can be gathered from the noncompliers, neither of these bounds are sharp in this situation ([Angrist et al., 1996; Imbens & Rubin, 1997].

The case with $\xi^-(x) > 1$ is more complicated because it implies the presence of defiers ([Angrist et al., 1996]. As seen in Figure 1 of the main text, tighter bounds than those with $\xi^-(x) = 1$ are possible, depending on the distribution of the defiers.

S1.4. A distributional version of Proposition 3 of Wang & Tchetgen Tchetgen (2018)

**Proposition S1.** Suppose that $U$ is univariate and that, for all $x$ and $y$, $1 - \nu_o(y \mid x, U)$ and $\Delta \delta(x, U)/\delta(x)$ are both nondecreasing or nonincreasing in $U$, then $\nu^{\star}_a(\cdot \mid x)$ is stochastically greater than $\nu_a(\cdot \mid x)$.

**Proof.** Define $\nu^{\star}_a(\cdot \mid x, U) = \nu_a(\cdot \mid x, U) \Delta \delta(x, U)/\delta(x)$. By the proof of Lemma
2.1 of the main text, we have that \( E_{U \mid x} \{ \nu_a^* (\cdot \mid x, U) \} = \nu_a^* (\cdot \mid x) \). Then

\[
\nu_a^*(y \mid x) - \nu_a(y \mid x) = E_{U \mid x} \{ \nu_a^*(y \mid x, U) \} - E_{U \mid x} \{ \nu_a(y \mid x, U) \} \\
= E_{U \mid x} \{ \nu_a(y \mid x, U) \Delta \delta(x, U) / \delta(x) \} \\
- E_{U \mid x} \{ \nu_a(y \mid x, U) \} E_{U \mid x} \{ \Delta \delta(x, U) / \delta(x) \} \\
= \text{cov}_{U \mid x} \{ \nu_a(y \mid x, U), \Delta \delta(x, U) / \delta(x) \} \\
= -\text{cov}_{U \mid x} \{ 1 - \nu_a(y \mid x, U), \Delta \delta(x, U) / \delta(x) \} \\
\leq 0.
\]

This completes the proof. \( \square \)

S2. Analysis of the Job Training Partnership Act study

As an illustration of the sensitivity analysis described in Section 4 of the main text, consider data on the 6,102 adult female participants in the U.S. National Job Training Partnership Act (JTPA) study, conducted in the 1980s to evaluate the effect of a job training program on the trainee’s subsequent earnings. Actual enrollment in the programme may be confounded due to noncompliance with the randomized assignment (as instrument). We consider six covariate groups by cross-tabulating the subject’s age, i.e., younger or older than 30 years, with her race, i.e., while, African American, or Hispanic, and estimate the \( \nu_a^*(y \mid x) \) nonparametrically for each covariate group. With bounds on \( \varepsilon(x, U) \) expressed as a fraction \( \rho \) of its natural bounds, that is, \( \xi^{-}(x) = \rho \{ |\Delta \delta(x)|^{-1} + 1 \} \) and \( \xi^{+}(x) = \rho \{ |\Delta \delta(x)|^{-1} - 1 \} \), we use the procedures described in Section 4 to construct the covariate-specific distributional bounds, some illustrated in histograms in the Appendix, and then combine them to get the overall ones. Figure S1 shows the resulting ranges of the median treatment effect, Mann-Whitney stochastic shift, and average treatment effect on a dichotomized outcome at varying \( \rho \), along with the corresponding estimates under \( \rho = 0 \). For all three metrics, it is found that the treatment effect is bounded above zero as long as \( \rho \leq 0.1 \).
Figure S1: Sensitivity analysis of the treatment effect on the female participants in the Job Training Partnership Act study. The dashed line in each panel represents the estimate under the no-interaction assumption; the shaded area represents the range of the estimand under Assumption (A5*) with $\xi^-(x) = \rho(|\Delta \delta(x)|^{-1} + 1)$ and $\xi^+(x) = \rho(|\Delta \delta(x)|^{-1} - 1)$.

**Appendix**

Under $\rho = 0.5$ and $0.1$, we plot the estimated lower and upper stochastic-order bounds for the outcome distributions in each of the six covariate groups of the JTPA study in Figures S2 – S5. We can see that, for $\rho = 0.5$, the lower and upper bounds are still quite apart, with substantial fractions of probability heaped on the left and right ends, respectively. For $\rho = 0.1$, the bounding distributions are much more similar. They will become one as $\rho \downarrow 0$.

**References**

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*, 444–455.

Imbens, G. W., & Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies, 64*, 555–574.

Robins, J. M., & Rotnitzky, A. (2001). Comment on “inference for semiparametric models: Some questions and an answer,” by pj bickel and j. kwon. *Statistica Sinica, 11*, 920–936.
Figure S2: Stochastically least and greatest \( \nu_0(\cdot \mid x) \) under \( \rho = 50\% \) for the Job Training Partnership Act study. Light blue, \( \nu_0(\cdot \mid x) \); light red, \( \nu_0(\cdot \mid x) \); purple, overlap.

Figure S3: Stochastically least and greatest \( \nu_1(\cdot \mid x) \) under \( \rho = 50\% \) for the Job Training Partnership Act study. Light blue, \( \nu_1(\cdot \mid x) \); light red, \( \nu_1(\cdot \mid x) \); purple, overlap.
Figure S4: Stochastically least and greatest $\nu_0(\cdot \mid x)$ under $\rho = 10\%$ for the Job Training Partnership Act study. Light blue, $\nu_0(\cdot \mid x)$; light red, $\tau_0(\cdot \mid x)$; purple, overlap.

Figure S5: Stochastically least and greatest $\nu_1(\cdot \mid x)$ under $\rho = 10\%$ for the Job Training Partnership Act study. Light blue, $\nu_1(\cdot \mid x)$; light red, $\tau_1(\cdot \mid x)$; purple, overlap.
Wang, L., & 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, 531–550.