CORRECTING INSTRUMENTAL VARIBALES ESTIMATORS FOR SYSTEMATIC MEASUREMENT ERROR

Stijn Vansteelandt[1], Manoochehr Babanezhad[1] and Els Goetghebeur[1,2]

[1]*Ghent University and* [2]*Harvard School of Public Health*

Supplementary Material

This note contains proofs for Proposition 1 and Theorems 1 and 2.

Proof of Proposition 1. Model $\mathcal{A}$ implies model $\mathcal{B}$ because, with $\delta(\mathbf{X}_i, R_i) \equiv E(W_i - Z_i|\mathbf{X}_i, R_i)$,

$$\begin{aligned} E\left(Y_{i0}|\mathbf{X}_i, R_i\right) &= E\left\{Y_i - \gamma(\mathbf{S}_i;\psi^*)Z_i|\mathbf{X}_i, R_i\right\} \\ &= E\left[Y_i - \gamma(\mathbf{S}_i;\psi^*)\left\{W_i - \delta(\mathbf{X}_i, R_i)\right\}|\mathbf{X}_i, R_i\right] \end{aligned}$$

by (A3) and (A5), and because $E\left(Y_{i0}|\mathbf{X}_i, R_i\right) = E\left(Y_{i0}|\mathbf{X}_i\right)$ by (A1). Note that this does not require assumptions about the conditional association between Y_i and W_i, given Z_i, suggesting that this continues to hold when measurement error is differential.

To show that (6) is the only restriction (other than (5)) imposed on the observed data law, we proceed as in Robins and Rotnitzky (2004) by exhibiting for any observed data law satisfying (5) and (6), a joint law of the full data $(Y, \{Y_{rz}, \forall r, z\}, Z, W, R, \mathbf{X})$ satisfying the restrictions of model $\mathcal{A}$, where Y_{rz} is the potential outcome that would have been observed for given subject following exposure to $(R, Z) = (r, z)$, all other experimental conditions being the same as in the considered study. Given $(R = r, Z = z, W = w, \mathbf{X} = \mathbf{x}, Y = y)$, we define $Y_{rz} = y$ to satisfy (A2). We set $f(Z|R = r, W = w, \mathbf{X} = \mathbf{x}, Y = y)$ equal to an arbitrary density with conditional mean $w - \delta$ to satisfy (A5). We define $f(Y_{r0}|R = r, Z = z, W = w, \mathbf{X} = \mathbf{x}, Y = y)$ to be an arbitrary density with conditional mean $y - \gamma(\mathbf{x}, r; \psi^*)z$. In addition, given $(R = r, Z = z, W = w, \mathbf{X} = \mathbf{x}, Y = y)$, we set $Y_{r0} = Y_{r'0} \equiv Y_0$ for each r' to satisfy (A1). By (6), the conditional distribution of Y_0 then also satisfies $E(Y_0|\mathbf{X} = \mathbf{x}, R) = E(Y_0|\mathbf{X} = \mathbf{x})$

for each $\mathbf{x}$. Remaining features of the full data density can be chosen arbitrarily.

Proof of Theorem 1. Let for simplicity of exposition and motivated by the data analysis, $\gamma(\mathbf{X}_i;\psi)=\psi, Z_i=Z_iR_i$ and $E(W_i-Z_i|\mathbf{X}_i,R_i)=\delta^*R_i$. Define $U_{i\delta}=d_\delta(R_i,\mathbf{X}_i)\left[Y_i-\psi(W_i-\delta)R_i-q(\mathbf{X}_i)\right]$ and $U_{i\psi}=d_\psi(R_i,\mathbf{X}_i)\left[Y_i-\psi(W_i-\delta)R_i\right.$ $\left.-q(\mathbf{X}_i)\right]$ the estimating functions for δ^* and ψ^*, respectively. Under weak regularity conditions as stated for general M-estimators in van der Vaart (1998, p.48, 60), Taylor expansions show that

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}}\sum_{i=1}^n U_{i\delta}+E\left(\frac{\partial U_\delta}{\partial\psi}\right)\sqrt{n}(\hat{\psi}-\psi^*)+E\left(\frac{\partial U_\delta}{\partial\delta}\right)\sqrt{n}(\hat{\delta}-\delta^*)\\ &\quad +\frac{1}{2}E\left(\frac{\partial^2 U_\delta}{\partial\psi\partial\delta}\right)\sqrt{n}(\hat{\psi}-\psi^*)(\hat{\delta}-\delta^*)+o_p(1) \end{aligned} \tag{1}$$

from which

$$\begin{aligned} \sqrt{n}(\hat{\delta}-\delta^*)\frac{\hat{\psi}+\psi^*}{2} &= o_p(1)-E^{-1}\left\{d_\delta(R,\mathbf{X})R\right\}\\ &\quad\times\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{i\delta}-E\left\{d_\delta(R,\mathbf{X})(W-\delta^*)R\right\}\sqrt{n}(\hat{\psi}-\psi^*)\right] \end{aligned}$$

Plugging this into a first order Taylor expansion of $U_{i\psi}$, shows that

$$\begin{aligned} \sqrt{n}(\hat{\psi}-\psi^*) &= -\left[E\left\{d_\psi(R,\mathbf{X})(W-\delta^*)R\right\}\right.\\ &\quad\left.-\frac{E\left\{d_\psi(R,\mathbf{X})R\right\}}{E\left\{d_\delta(R,\mathbf{X})R\right\}}E\left\{d_\delta(R,\mathbf{X})(W-\delta)R\right\}\right]^{-1}\\ &\quad\times\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{i\psi}-\frac{E\left\{d_\psi(R,\mathbf{X})R\right\}}{E\left\{d_\delta(R,\mathbf{X})R\right\}}U_{i\delta}\right]+o_p(1) \end{aligned}$$

Standard application of the Central Limit Theorem and Slutsky's Theorem now shows that $\sqrt{n}(\hat{\psi}-\psi^*)=O_p(1)$ and that Part 1 of Theorem 1 holds.

Note that the last 3 terms in expression (1) of these Supplementary Materials can be replaced with $E\left\{d_\delta(R,\mathbf{X})R\right\}\left\{\psi^*+O_p(n^{-1/2})\right\}\sqrt{n}(\hat{\delta}-\delta)$, from which $\sqrt{n}(\hat{\delta}-\delta)\psi^*=\sqrt{n}(\hat{\delta}-\delta)(\hat{\psi}+\psi^*)\left\{1/2+o_p(1)\right\}$ equals

$$\begin{aligned} &-\left[E\left\{d_\delta(R,\mathbf{X})R\right\}-\frac{E\left\{d_\delta(R,\mathbf{X})(W-\delta^*)R\right\}}{E\left\{d_\psi(R,\mathbf{X})(W-\delta^*)R\right\}}E\left\{d_\psi(R,\mathbf{X})R\right\}\right]^{-1}\\ &\times\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n U_{i\delta}-\frac{E\left\{d_\delta(R,\mathbf{X})(W-\delta^*)R\right\}}{E\left\{d_\psi(R,\mathbf{X})(W-\delta^*)R\right\}}U_{i\psi}\right]+o_p(1) \end{aligned}$$

The latter expression is bounded in probability (under standard regularity conditions). It follows that, as ψ^* goes to zero with increasing sample size, $\hat{\delta}$ does not converge to δ^* at root-n rate and hence is not uniformly root-n consistent. In particular, there is no root-n consistent estimator of δ^* under model $\mathcal{A}$ at $\psi^* = 0$, which proves Part 2 of Theorem 1. This is also seen by noting that the expected derivative of the efficient estimating function for (ψ, δ) w.r.t. δ is zero at $\psi = 0$.

Part 3 of Theorem 1 is immediate from Robins (1994).

Proof of Theorem 2. Let for simplicity of exposition, but without loss of generality, $\gamma(\mathbf{X}_i; \psi) = \psi, Z_i = Z_i R_i$ and $E(W_i - Z_i|\mathbf{X}_i, R_i) = \delta^* R_i$. Then standard asymptotic theory for M-estimators (van der Vaart, 1998) and Taylor expansions of the estimating functions (9) for ψ^* w.r.t. $\hat{\delta}(\psi)$ shows that (9) equals

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} P\{\hat{\delta}(\psi) \in \Delta\}U_{i\psi}(\psi,\delta) + P\{\hat{\delta}(\psi) \notin \Delta\}U_{i\psi}(\psi,0) + o_p(1) \\
&-\left[P\{\hat{\delta}(\psi) \in \Delta\} + \left\{\varphi\left(\frac{\Delta_l - \delta}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right) - \varphi\left(\frac{\Delta_u - \delta}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right)\right\}\frac{\sqrt{n}|\psi|\delta}{\sigma(\psi)}\right] \\
&\times \frac{E\left\{d_\psi(R,T)R\right\}}{E\left\{d_\delta(R,T)R\right\}}U_{i\delta}(\psi,\delta)
\end{aligned} \tag{2}
$$

That the remainder term converges to zero in probability for any fixed ψ can be seen because, for some $\tilde{\delta}$ on the open line segment between $\hat{\delta}(\psi)$ and δ^* (under regularity conditions which include uniform convergence of $n^{-1}\sum_{i=1}^{n} U_{i\psi}(\psi, \delta)$ w.r.t. δ), the remainder term equals

$$
\begin{aligned}
&\left[P_{\delta=\tilde{\delta}}\{\hat{\delta}(\psi) \in \Delta\}E\left\{\frac{\partial^2}{\partial\delta^2}U_{i\psi}(\psi,\tilde{\delta})\right\} + 2\frac{\partial}{\partial\delta}P_{\delta=\tilde{\delta}}\{\hat{\delta}(\psi) \in \Delta\}E\left\{\frac{\partial}{\partial\delta}U_{i\psi}(\psi,\tilde{\delta})\right\}\right. \\
&\left.+\frac{\partial^2}{\partial\delta^2}P_{\delta=\tilde{\delta}}\{\hat{\delta}(\psi) \in \Delta\}E\left\{U_{i0}(\psi) - U_{i\psi}(\psi,\tilde{\delta})\right\}\right]\frac{\sqrt{n}}{2}\{\hat{\delta}(\psi) - \delta^*\}^2 + o_p(1)
\end{aligned}
$$

Here, $E\left\{\partial^2 U_{i\psi}(\psi, \tilde{\delta})/\partial\delta^2\right\} = 0$. Because $E\left\{\partial U_{i\psi}(\psi, \tilde{\delta})/\partial\delta\right\} = O_p(1)\psi$ under standard regularity conditions and $\sqrt{n}\{\hat{\delta}(\psi) - \delta^*\}^2 = O_p(1)n^{-1/2}\psi^{-2}$, the second term is

$$
O_p(1)\left\{\varphi\left(\frac{\Delta_l - \tilde{\delta}}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right) - \varphi\left(\frac{\Delta_u - \tilde{\delta}}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right)\right\}\frac{1}{\sigma(\psi)} = o_p(1)
$$

for any fixed ψ. Because $E\left\{U_{i0}(\psi) - U_{i\psi}(\psi, \tilde{\delta})\right\} = O_p(1)\tilde{\delta}\psi$, the third term is

$$\left\{\varphi\left(\frac{\Delta_l - \tilde{\delta}}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right) - \varphi\left(\frac{\Delta_u - \tilde{\delta}}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right)\right\}\frac{n|\psi|\psi\tilde{\delta}}{\sigma(\psi)^3}(\Delta_l - \Delta_u) = o_p(1)$$

for any fixed ψ because $x^a\varphi(x) \to 0$ as $x \to \infty$ for arbitrary $a > 0$.

Because the estimating functions in expression (2) of the Supplementary Materials have mean and variance depending on the sample size, the triangular array Central Limit Theorem (Serfling, 1980, p.31) is needed to derive the asymptotic distribution of (9) for fixed ψ. Application of this Theorem shows that for arbitrary fixed ψ, the estimating functions in (9) are asymptotically normally distributed under the weak regularity condition that the standard deviation of the estimating functions $\tilde{U}_i(\psi)$, as defined by expression (2) of these Supplementary Materials, is bounded (i.e. $O(1)$) and that asymptotically $E\|\tilde{U}_i(\psi) - E\{\tilde{U}_i(\psi)\}\|^k = o(n^{k/2-1})$ for each k. Because for any fixed $\psi^* \neq 0$ and $\delta^* \in \Delta =]\Delta_l, \Delta_u[$, $P\{\hat{\delta}(\psi^*) \in \Delta\}$ converges to 1, it follows under these conditions that $n^{-1/2}\sum_{i=1}^n \tilde{U}_i(\psi^*)$ will be asymptotically normally distributed with mean zero and finite variance, which is given by the variance of expression (2) of the Supplementary Materials. Within faster than root-n shrinking neighbourhoods of zero (i.e. if $\psi^* = kn^{-a}$ for some constant k and $a > 1/2$), the remainder term in the Taylor series expansion is still $o_p(1)$. Further, $P\{\hat{\delta}(\psi^*) \in \Delta\}$ converges to 0 and $U_0(\psi^*)$ has mean converging to zero at 1 over n^a-rate. It then again follows that $n^{-1/2}\sum_{i=1}^n \tilde{U}_i(\psi^*)$ is asymptotically normally distributed with mean zero and finite variance. Finally, within 1 over root-n shrinking neighbourhoods of zero (i.e. if $\psi^* = kn^{-1/2}$ for some constant k), the remainder term in the Taylor series expansion is bounded in probability, but not $o_p(1)$. The significant contribution of the squared term $\sqrt{n}\{\hat{\delta}(\psi^*) - \delta^*\}^2$ implies that $n^{-1/2}\sum_{i=1}^n \tilde{U}_i(\psi^*)$ may not converge to a normal distribution, nor to a mean zero distribution along such sequences. The implications of this will be discussed in the next paragraph.

The asymptotic distribution of $\tilde{\psi}$ (rather than root-n times the sample average of its estimating function) is now immediate via a further Taylor series expansion of the estimating functions (w.r.t. ψ), evaluated at $\tilde{\psi}$. This shows that for any fixed ψ

$$0 = \frac{1}{\sqrt{n}}\sum_{i=1}^n P\{\hat{\delta}(\psi) \in \Delta\}U_{i\psi}(\psi, \delta) + P\{\hat{\delta}(\psi) \notin \Delta\}U_{i\psi}(\psi, 0) + o_p(1)$$

$$
\begin{aligned}
&-\left[P\{\hat{\delta}(\psi) \in \Delta\} + \left\{\varphi\left(\frac{\Delta_l - \delta}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right) - \varphi\left(\frac{\Delta_u - \delta}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right)\right\} \frac{\sqrt{n}|\psi|\delta}{\sigma(\psi)}\right] \\
&\quad \times \frac{E\left\{d_\psi(R,X)R\right\}}{E\left\{d_\delta(R,X)R\right\}} U_{i\delta}(\psi,\delta) \\
&+\left(P\{\hat{\delta}(\psi) \in \Delta\} E\left\{\frac{\partial}{\partial\psi} U_\psi(\psi,\delta)\right\} + P\{\hat{\delta}(\psi) \notin \Delta\} E\left\{\frac{\partial}{\partial\psi} U_\psi(\psi,0)\right\}\right. \\
&+\left\{\varphi\left(\frac{\Delta_l - \delta}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right) - \varphi\left(\frac{\Delta_u - \delta}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right)\right\} \frac{\sqrt{n}|\psi|\delta(\Delta_l - \Delta_u)}{\sigma(\psi)} E\left\{d_\psi(R,X)R\right\} \\
&-\left[P\{\hat{\delta}(\psi) \in \Delta\} + \left\{\varphi\left(\frac{\Delta_l - \delta}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right) - \varphi\left(\frac{\Delta_u - \delta}{\sigma(\psi)/(\sqrt{n}|\psi|)}\right)\right\} \frac{\sqrt{n}|\psi|\delta}{\sigma(\psi)}\right] \\
&\left.\times \frac{E\left\{d_\psi(R,X)R\right\}}{E\left\{d_\delta(R,X)R\right\}} E\left\{d_\delta(R,X)R(W-\delta)\right\}\right) \sqrt{n}(\tilde{\psi} - \psi)
\end{aligned}
$$

That the remainder term converges to zero in probability for any fixed ψ can be seen using a similar derivation as before. We conclude that, up to an $o_p(1)$ term and for fixed ψ, $\sqrt{n}(\tilde{\psi} - \psi)$ is a linear transformation of $n^{-1/2}\sum_{i=1}^n \tilde{U}_i(\psi)$ and thus shares its asymptotic properties. Specifically, within faster and slower than 1 over root-n shrinking neighbourhoods of zero (and in particular at arbitrary fixed ψ), $\sqrt{n}(\tilde{\psi} - \psi)$ is asymptotically normally distributed with mean zero and finite variance under weak regularity conditions. Within 1 over root-n neighbourhoods of zero, $\sqrt{n}(\tilde{\psi} - \psi)$ may be asymptotically biased and not normally distributed.