ECMA 31350 LATE WITH MISSING VALUES RESULT

GROUP 6

MARCH 4, 2024

Set-up. Suppose one wants to uncover the causal relationship between a treatment variable D and an outcome Y from a survey, and that Z is the corresponding valid instrument conditional on covariates W. Let S denote the status for responders (i.e. $S_i = 1$ if i responds and 0 otherwise). Assume that $(Y, D, Z, W) \perp S \mid X$ for some covariates X, i.e., the responses are missing at random conditional on X. Note that in the actual dataset, X, D, and Z can be observed for all individuals, while W may or may not be available for non-respondents.

Theorem. *Let $A = (Y, D, Z, W, S, X)$. With the set-up above and the same conditions in Section 1.5.3 (i.e. Frolich (2007)), one can show that the score*

$$g(A; \tau_{\text{LATE}}, \eta) = (g^{1,1;o}(A;\eta) - g^{1,0;o}(A;\eta)) - (g^{2,1;o}(A;\eta) - g^{1,1;o}(A;\eta))\tau_{\text{LATE}},$$

where

$$\begin{aligned}
g^{1,1;o}(A; r, q, a_1, b_1) &= \frac{YZS}{rq} - (Z - r)\cdot\frac{a_1}{r^2} - (S - q)\cdot\frac{b_1}{rq},\\
g^{1,0;o}(A; r, q, a_0, b_0) &= \frac{Y(1-Z)S}{(1-r)q} - (Z - r)\cdot\frac{a_0}{(1-r)^2} - (S - q)\cdot\frac{b_0}{(1-r)q},\\
g^{2,1;o}(A; r, q, f_1, g_1) &= \frac{Y(1-Z)S}{rq} - (Z - r)\cdot\frac{f_1}{r^2} - (S - q)\cdot\frac{g_1}{rq},\\
g^{2,0;o}(A; r, q, f_0, g_0) &= \frac{Y(1-Z)S}{(1-r)q} - (Z - r)\cdot\frac{f_0}{(1-r)^2} - (S - q)\cdot\frac{b_0}{(1-r)q},
\end{aligned}$$

and $\eta = (r, q, a_1, a_0, b_1, b_0, f_1, f_0, g_1, g_0)$ is such that

$$\begin{aligned}
r(W) &= \mathbb{E}[Z \mid W] = \mathbb{P}[Z = 1 \mid W],\\
q(X) &= \mathbb{E}[S \mid X] = \mathbb{P}[S = 1 \mid X]\\
a_1(W) &= \mathbb{E}[YZ \mid S = 1, W], & a_0(W) &= \mathbb{E}[Y(1-Z) \mid S = 1, W],\\
b_1(X, W) &= \mathbb{E}[YZ \mid s = 1, X, W], & b_0(X, W) &= \mathbb{E}[Y(1-Z) \mid S = 1, X, W],\\
f_1(W) &= \mathbb{E}[DZ \mid S = 1, W], & f_0(W) &= \mathbb{E}[D(1-Z) \mid S = 1, W],\\
g_1(X, W) &= \mathbb{E}[DZ \mid S = 1, X, W], & g_0(X, W) &= \mathbb{E}[D(1-Z) \mid S = 1, X, W].
\end{aligned}$$

(Note: one can probably reduce the number of nuisance parameters by moving terms around - which is why the score functions for ATE and LATE are so clean. I have yet to figure out a good way to do it.)

Proof. We again start with

$$\tau_{\text{LATE}} = \frac{\mathbb{E}[m(1,W) - m(0,W)]}{\mathbb{E}[p(1,W) - p(0,W)]}.$$

We will only do the exercise for the first terms in the numerator, as the others follow from the exact same process. So first,

$$\begin{aligned}
\mathbb{E}[m(1,W)] = \mathbb{E}[\mathbb{E}[Y \mid Z = 1, W]] &= \mathbb{E}\left[\frac{YZ}{r(W)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{YZ}{r(W)} \;\middle|\; W, X\right]\right] \qquad \text{by LIE} \\
&= \mathbb{E}\left[\frac{1}{r(w)}\mathbb{E}[YZ \mid W, S = 1, X]\right] \\
&= \mathbb{E}\left[\frac{1}{r(w)} \cdot \frac{YZS}{\mathbb{P}[S = 1 \mid W, X]}\right] \\
&= \mathbb{E}\left[\frac{YZS}{r(W)q(W,X)}\right].
\end{aligned}$$

In particular, note that $q(X,W) = q(X)$ given the independence assumption $W \perp S \mid X$ (joint independence implies marginal independence). We will keep working $q(X,W)$ and switch it out for $q(X)$ at the very end, as this is much more convenient. In addition, this also serves as a great robustness check, for if the independence assumption truly holds, it should not matter which propensity score we use.

The above gives

$$g^{1,1}(A; r, q) = \frac{YZS}{r(W)q(X,W)}.$$

Then

$$g^{1,1}_{2,1}(A; r, q) = \frac{YZS}{q(X,W)} \cdot \left(-\frac{1}{[r(W)]^2}\right),$$

and

$$\begin{aligned}
\mathbb{E}\left[g^{1,1}_{2,1}(A, ; r, q)\right] &= \mathbb{E}\left[-\frac{YZS}{[r(W)]^2 q(X,W)} \;\middle|\; W\right] \\
&= -\frac{1}{[r(W)]^2}\mathbb{E}\left[\mathbb{E}\left[\frac{YZS}{q(X,W)} \;\middle|\; X, W\right] \;\middle|\; W\right] \\
&= -\frac{1}{[r(W)]^2}\mathbb{E}[\mathbb{E}[YZ \mid S = 1, X, W] \mid W] \\
&= -\frac{1}{[r(W)]^2} \cdot \mathbb{E}[YZ \mid S = 1, W] =: \frac{a_1(W)}{[r(W)]^2}.
\end{aligned}$$

On the other hand,

$$g^{1,1}_{2,2}(A; r, q) = \frac{YZS}{r(W)} \cdot \left(-\frac{1}{[q(X,W)]^2}\right),$$

so

$$\begin{aligned}
\mathbb{E}\left[g^{1,1}_{2,2}(A; r, q)\right] &= \mathbb{E}\left[-\frac{YZS}{r(W)[q(X,W)]^2} \;\middle|\; X, W\right] \\
&= -\frac{1}{r(W)[q(X,W)]^2}[\mathbb{E}[YZS \mid X, W]
\end{aligned}$$

$$= -\frac{\mathbb{E}[YZ \mid S=1, X, W]}{r(W)[q(X,W)]} =: -\frac{b(X,W)}{r(W)q(X,W)} = -\frac{b_1(X,W)}{r(W)\cdot q(X)}.$$

Combined, this allows us to define

$$g^{1,1;o}(A; r, q, a_1, b_1) = \frac{YZS}{r(W)q(X,W)} - [Z - r(W)] \cdot \frac{a_1(W)}{[r(W)]^2} - [S - q(X,W)] \cdot \frac{b_1(X,W)}{r(W)\cdot q(X,W)}.$$

Next, we need to verify N.O. It is not hard to see that

$$\mathbb{E}\left[g^{1,1,0}(A; r, q, a_1, b_1)\right] = \mathbb{E}[m(1, W)],$$

so we concentrate on the N.O. part. Notation-wise, we substitute the r in the lecture notes for α to get

$$\begin{aligned} h_r(\alpha) &= \mathbb{E}\left[g^{1.1.0}(A; r + \alpha(\tilde{r} - r), q, a_1, b_1)\right] \\ &= \mathbb{E}\left[\frac{YZS}{[r + \alpha(\tilde{r} - r)]\cdot q} - \frac{Z \cdot a_1}{[r + \alpha(\tilde{r} - r)]^2} + \frac{a_1}{[r + \alpha(\tilde{r} - r)]} - \frac{(S - q)\cdot b_1}{[r + \alpha(\tilde{r} - r)]\cdot q}\right], \end{aligned}$$

So

$$\begin{aligned} h_r'(\alpha)\Big|_{\alpha=0} &= \mathbb{E}\left[\frac{-YZS(\tilde{r} - r)}{q \cdot r^2} + \frac{2Z \cdot a_1(\tilde{r} - r)}{r^3} - \frac{a_1(\tilde{r} - r)}{r^2} + \frac{(S - q)\cdot b_1(\tilde{r} - r)}{q \cdot r^2}\right] \\ &= \mathbb{E}\left[\frac{\tilde{r} - r}{r^2}\mathbb{E}\left[\frac{-YZS}{q} + \frac{2Z \cdot a_1}{r} - a_1 + \frac{(S - q)\cdot b_1}{q} \;\Big|\; W\right]\right] \\ &= \mathbb{E}\left[\frac{\tilde{r} - r}{r^2} \cdot 0\right] = 0 \end{aligned}$$

per another layer of LIE applied to each term in the expectation that I am too tired to latex right now (it will be in the Appendix of the paper though). Similarly,

$$\begin{aligned} h_q(\alpha) &= \mathbb{E}\left[g^{1,1;o}(A; r, q + \alpha(\tilde{q} - q), a_1, b_1)\right], \\ &= \mathbb{E}\left[\frac{YZS}{r[q + \alpha(\tilde{q} - q)]} - \cdots - \frac{S \cdot b_1}{r[q + a(\tilde{q} - q)]} + \cdots\right], \end{aligned}$$

which yields

$$h_q'(\alpha)\Big|_{\alpha=0} = \mathbb{E}\left[-\frac{YZS(\tilde{q} - q)}{rq^2} + \frac{s \cdot b_1(\tilde{q} - q)}{r \cdot q^2}\right] = 0$$

again by applying LIE term by term. Lastly,

$$\begin{aligned} h_{a_1}(\alpha) &= \mathbb{E}\left[g^{1,1;o}(A; r, q, a_1 + \alpha(\tilde{a}_1 - a_1), b_1)\right] \\ &= \mathbb{E}\left[\frac{-(Z - r)\cdot[a_1 + \alpha(\tilde{a}_1 - a_1)]}{r}\right] = 0 \\ \Longrightarrow & h_{a_1}'(\alpha)\Big|_{\alpha=0} = 0, \end{aligned}$$

and

$$\begin{aligned} h_{b_1}(\alpha) &= \mathbb{E}\left[g^{1,1;o}(A; r, q, a_1, b_1 + \alpha(\tilde{b}_1 - b_1)\right] \\ &= \mathbb{E}\left[-(S - q)\cdot\frac{[b_1 + \alpha(\tilde{b}_1 - b_1)]}{r \cdot q}\right] = 0 \\ \Longrightarrow & \ h_{b_1}'(\alpha)\Big|_{\alpha=0} = 0. \end{aligned}$$

Repeating this process for all four terms lead to the score function. It is not hard to check that $g(A; \tau_{\text{LATE}}, \eta)$ is N.O. with respect that τ_{LATE}, hence the result. $\square$