RANDOM NOTES

RACHEL ANDERSON

Consider the case where the number of individuals with more than one match, N_L is small relative to the population, of size N (like the AER application where about 4% of observations have multiple linked outcomes.) It may then be feasible to solve:

$$\hat{\beta}^{\min/\max} = \min/\max_{\{w_{i\ell}\}} \left(x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} x_i \overline{y}_i\right)$$

$$\overline{y}_i = \sum_{\ell=1}^{L_i} w_{i\ell} y_{i\ell}$$

$$\sum_{\ell=1}^{L_i} w_{i\ell} = 1,\ w_{i\ell} \geq 0$$

Also, we can try alternating $\sum_{\ell=1}^{L_i} w_{i\ell} = 1$ OR $\sum_{\ell=1}^{L_i} w_{i\ell} = 0$ for observations with $L_i > 1$ only, to compare whether they should be included in analysis or not.

Compare this to our method, which uses the moment condtion:

$$m(y_{i\ell}, x_i; \theta) = x_i(y_{i\ell} - x_i'\beta)$$

so that the estimator uses

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{\ell=1}^{L_i} x_i(y_{i\ell} - x_i'\beta) - \frac{1}{n}\sum_{i=1}^{n}(L_i - 1)\hat{g}(w_i, L_i, x_i; \beta)$$

$$g(w_i, L_i, x_i; \beta) = x_i E(y_{i\ell}|w_i, L_i) - x_i x_i'\beta$$

Date: August 14, 2019.

Rewriting,

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{\ell=1}^{L_i} x_i y_{i\ell} - (L_i - 1)x_i E[y|w_i]\right) \\
&= (X'X)^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{L_i=1} x_i y_i + \sum_{L_i>1} x_i \left(E[y|w_i] + \sum_{\ell=1}^{L_i}(y_{i\ell} - E[y|w_i])\right)\right) \\
&= (X'X)^{-1}\frac{N-N_L}{N}\frac{1}{N-N_L}\sum_{L_i=1} x_i y_i + (X'X)^{-1}\frac{N_L}{N}\frac{1}{N_L}\sum_{L_i>1} x_i \left(E[y|w_i] + \sum_{\ell=1}^{L_i}(y_{i\ell} - E[y|w_i])\right) \\
&\xrightarrow{p} \beta
\end{aligned}$$

by the theorem

but what if we use the first term only? Or use a different kernel on the second term in order to change variance? We could get bias/variance tradeoff

Compare to unbalanced panel techniques/ random effects estimators