Notes from class on 3/22/17

Susanna Makela

Shalizi and the Wald Estimator

Here is a quick derivation of the Wald estimator in the case of a binary instrument and binary treatment, starting from Shalizi's IV equation:

$$p(y \mid do(z)) = \sum_d p(y \mid do(d)) \cdot p(d \mid do(z)). \tag{1}$$

From this, we can clearly write

$$\mathbb{E}[y \mid do(z)] = \sum_d \mathbb{E}[y \mid do(d)] \cdot p(d \mid do(z))$$

In the case where $z = 1$, we get

$$\begin{aligned}
\mathbb{E}[y \mid do(z=1)] &= \mathbb{E}[y \mid do(d=1)] \cdot p(d=1 \mid do(z=1)) + \mathbb{E}[y \mid do(d=0)] \cdot p(d=0 \mid do(z=1)) \\
&= \mathbb{E}[y \mid do(d=1)] \cdot \mathbb{E}[d \mid do(z=1)] + \mathbb{E}[y \mid do(d=0)]\,(1 - \mathbb{E}[d \mid do(z=1)]) \\
&= \mathbb{E}[y \mid do(d=0]) + \mathbb{E}[d \mid do(z=1)]\,(\mathbb{E}[y \mid do(d=1)] - \mathbb{E}[y \mid do(d=0)])\,,
\end{aligned}$$

where we use the fact that $p(d = 1 \mid do(z = 1)) = \mathbb{E}[d \mid do(z = 1)]$ since d is binary. Solving for $\mathbb{E}[y \mid do(d = 0)]$, we see that

$$\mathbb{E}[y \mid do(d=0)] = \mathbb{E}[y \mid do(z=1)] - \mathbb{E}[d \mid do(z=1)]\,(\mathbb{E}[y \mid do(d=1)] - \mathbb{E}[y \mid do(d=0)])\,. \tag{2}$$

Analogously, when $z = 0$, we get

$$\mathbb{E}[y \mid do(d=0)] = \mathbb{E}[y \mid do(z=0)] - \mathbb{E}[d \mid do(z=0)]\,(\mathbb{E}[y \mid do(d=1)] - \mathbb{E}[y \mid do(d=0)])\,. \tag{3}$$

Subtracting (2) − (3) gives

$$0 = \mathbb{E}[y \mid do(z=1)] - \mathbb{E}[y \mid do(z=0)] + (\mathbb{E}[y \mid do(d=1)] - \mathbb{E}[y \mid do(d=0)]) \cdot (\mathbb{E}[d \mid do(z=0)] - \mathbb{E}[d \mid do(z=1)])\,,$$

and solving for our quantity of interest, namely $\delta = \mathbb{E}[y \mid do(d = 1)] - \mathbb{E}[y \mid do(d = 0)]$, gives

$$\delta = \frac{\mathbb{E}[y \mid do(z=1)] - \mathbb{E}[y \mid do(z=0)]}{\mathbb{E}[d \mid do(z=1)] - \mathbb{E}[d \mid do(z=0)]},$$

which is the Wald estimator.

Two-Stage Least Squares, Two Ways

We had also discussed the derivation of two-stage least squares (2SLS) in two different ways, and whether they yield equivalent estimates.

We started with equations for the treatment D and outcome Y, using the instrument Z:

$$D = \alpha_0 + \alpha Z + \epsilon_D \tag{4}$$
$$Y = \delta_0 + \delta D + \epsilon_Y. \tag{5}$$

If we use (4) to substitute for D in (5), we get

$$\begin{aligned} Y &= \delta_0 + \delta\left(\alpha_0 + \alpha Z + \epsilon_D\right) + \epsilon_Y \\ &= \left(\delta_0 + \delta\alpha_0\right) + \delta\alpha Z + \left(\delta\epsilon_D + \epsilon_Y\right) \\ &= \beta_0 + \beta Z + \eta_Y, \end{aligned}$$

so $\delta_{Wald} = \beta/\alpha$ as Adji derived on the board. We know from basic OLS theory that

$$\widehat{\beta} = \frac{Cov(Y, Z)}{Var(Z)}$$

and

$$\widehat{\alpha} = \frac{Cov(D, Z)}{Var(Z)}, \tag{6}$$

so

$$\widehat{\delta}_{Wald} = \frac{Cov(Y, Z)}{Cov(D, Z)} \tag{7}$$

(here I'm using $Cov(Y, Z)$ to denote the finite-sample covariance of Y and Z, as opposed to the population covariance).

In practice, what people usually do is run the regression in (4), calculate $\widehat{D} = \widehat{\alpha_0} + \widehat{\alpha}Z$, and substitute $\widehat{D}$ for D in (5). If we do this, (5) becomes

$$\begin{aligned} Y &= \delta_0 + \delta\left(\widehat{\alpha_0} + \widehat{\alpha}Z\right) + \epsilon_Y \\ &= \left(\delta_0 + \delta\widehat{\alpha_0}\right) + \delta\left(\widehat{\alpha}Z\right) + \epsilon_Y. \end{aligned}$$

Again from OLS theory, we know that

$$\begin{aligned} \widehat{\delta} &= \frac{Cov(Y, \widehat{\alpha}Z)}{Var(\widehat{\alpha}Z)} \\ &= \frac{\widehat{\alpha}\ Cov(Y, Z)}{\widehat{\alpha}^2\ Var(Z)} \\ &= \frac{Cov(Y, Z)}{\widehat{\alpha}\ Var(Z)}. \end{aligned}$$

Using the definition of $\widehat{\alpha}$ in (6), we have that

$$\begin{aligned} \widehat{\delta} &= \frac{Cov(Y, Z)}{\widehat{\alpha}\ Var(Z)} \\ &= \frac{Cov(Y, Z)}{Cov(D, Z)} \\ &= \widehat{\delta}_{Wald} \end{aligned}$$

as in (7).