Econ 241B Homework 7 Solutions

March 11, 2016

1 Instrumental Variables

A) We want to show that:

$$\hat{\beta}_1 = \frac{1}{n_L}\sum_i W_i L_i - \frac{1}{n_O}\sum_i W_i(1-L_i) = \bar{W}_L - \bar{W}_O$$

where $n_L = \sum_1 L_i$ is the number of individuals who attended law school and $n_O = \sum_i(1-L_i)$ is the number of individuals who did not. We can proceed using either summation notation or matrix algebra. Using sums, we have that

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\frac{1}{n}\sum_i (W_i - \bar{W})(L_i - \bar{L})}{\sum_i (L_i - \bar{L})^2} \\
&= \frac{\sum_i L_i W_i - \sum_i \bar{L} W_i - \sum_i L_i \bar{W} + n\bar{L}\bar{W}}{\sum_i (L_i - \bar{L})^2} \\
&= \frac{n_L \bar{W}_L - n\bar{L}\bar{W} + n\bar{L}\bar{W}}{\sum_i (L_i - \bar{L})^2} \\
&= \frac{n_L \bar{W}_L - n_L \bar{W}}{\sum_i (L_i - \bar{L})^2} \\
&= \frac{n_L \bar{W}_L - n_L(\frac{n_L}{n}\bar{W}_L + \frac{n_O}{n}\bar{W}_O)}{\sum_i (L_i - \bar{L})^2} \\
&= \frac{\frac{n_L n_O}{n}\bar{W}_L - \frac{n_L n_O}{n}\bar{W}_O}{\sum_i (L_i - \bar{L})^2} \\
&= \frac{\frac{n_L n_O}{n}(\bar{W}_L - \bar{W}_O)}{\sum_i (L_i - \bar{L})^2}
\end{aligned}$$

Note that $\displaystyle\sum_i (L_i - \bar{L})^2 = n_L \frac{n_O^2}{n^2} + n_O \frac{n_L^2}{n^2} = \frac{n_L n_O}{n}$

So that

$$\hat{\beta}_1 = \frac{\frac{n_L n_O}{n}(\bar{W}_L - \bar{W}_O)}{\frac{n_L n_O}{n}} = \bar{W}_L - \bar{W}_O$$

With the matrix formulation of the OLS estimator, we have:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \begin{bmatrix} n & \sum_i L_i \\ \sum_i L_i & \sum_i L_{i^2} \end{bmatrix}^{-1} \begin{bmatrix} \sum_i W_i \\ \sum_i W_i L_i \end{bmatrix}$$
$$= \begin{bmatrix} n & n_L \\ n_L & n_L \end{bmatrix} \begin{bmatrix} n\bar{W} \\ n_L \bar{W}_L \end{bmatrix}$$

Now define $D = n_L(n - n_L) = n_L n_O$

$$\hat{\beta} = \frac{1}{D} \begin{bmatrix} n & -n_L \\ -n_L & n \end{bmatrix}^{-1} \begin{bmatrix} n\bar{W} \\ n_L \bar{W}_L \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{n_O} & -\frac{1}{n_O} \\ -\frac{1}{n_O} & \frac{n}{n_L n_O} \end{bmatrix} \begin{bmatrix} n\bar{W} \\ n_L \bar{W}_L \end{bmatrix}$$
$$= \begin{bmatrix} 1 & -1 \\ -1 & \frac{n}{n_L} \end{bmatrix} \begin{bmatrix} \frac{n}{n_0}\bar{W} \\ \frac{n_L}{n_O}\bar{W}_L \end{bmatrix}$$
$$= \frac{n}{n_0} \begin{bmatrix} \bar{W} - \frac{n_L}{n}\bar{W}_L \\ \bar{W}_L - \bar{W} \end{bmatrix}$$
$$= \begin{bmatrix} \bar{W}_O \\ \bar{W}_L - \bar{W}_L \end{bmatrix}$$

B) To find the share of law degree holders to minimize the variance of $\hat{\beta}_1$, recall that

$$\hat{\beta}_1 = \frac{1}{n_L} \sum_i W_i L_i - \frac{1}{n_O} \sum_i W_i (1 - L_i)$$

So

$$Var(\hat{\beta}_1) = \frac{1}{n_L^2} \sum_i L_i Var(W_i) + \frac{1}{n_O^2} \sum_i (1 - L_i) Var(W_i)$$
$$= \frac{Var(W_i)}{n_L} + \frac{Var(W_i)}{n_O}$$
$$= \frac{(n_L + n_O)Var(W_i)}{n_L n_O}$$
$$= \frac{Var(W_i)}{\bar{L}(1 - \bar{L})}$$

We want to minimize this expression with respect to the share of law degree holders, $\bar{L} = \frac{n_L}{n}$. A little calculus yields $\bar{L}^* = \frac{1}{2}$. Hence, to minimize the variance of $\hat{\beta}_1$ for a given sample size, we'd like the sample to be evenly divided between degree holders and non degree holders.

C) $\mathbb{E}[L_i U_i] \neq 0$ if any unobserved characteristics of individuals influence both their wages and their propensity to pursue a law degree. In particular, we might be concerned that individuals who expect larger gains from obtaining a law degree are more likely to pursue a degree, which would cause us to overstate the overall effect of law school on wages. There are a variety of other factors such as parents' income and education, undergraduate institution (think Harvard grads vs. Southeastern State U grads), undergraduate major, and ability that would affect both wages and whether an individual goes to law school.

For the next part of the question we assume that we have a valid instrument Z, which means that it satisfies:

$$Cov(L_i Z_i) \neq 0 \qquad \text{(The instrument is relevant)}$$
$$\mathbb{E}[Z_i U_i] = 0 \qquad \text{(The instrument is validly excluded from the first stage)}$$

To construct the 2SLS estimator, we start by estimating the first stage regression of L_i on Z_i and get the predicted values of L_i from this regression:

$$\begin{aligned} &L_i = \alpha_0 + \alpha_1 Z_i + V_i \\ &\Rightarrow \hat{\alpha} = (Z'Z)^{-1}Z'L \\ &\Rightarrow \hat{L} = Z\hat{\alpha} = Z(Z'Z)^{-1}Z'L \end{aligned}$$

Then, we plug this into the structural equation to estimate the second stage:

$$\begin{aligned} W_i &= \beta_0 + \beta_1 \hat{L}_i + \varepsilon_i \\ &= \beta_0 + \beta_1 \hat{L}_i + (U_i + \beta_1 \hat{V}_i) \end{aligned}$$

where $\varepsilon_i = U_i + \hat{V}_i$ and $\hat{V}_i$ is the residual from the first stage. By construction, $Cov(\hat{L}_i, \varepsilon_i) = Cov(\hat{L}_i U_i + \beta_1 \hat{V}_i) = 0$. We now have

$$\begin{aligned} \hat{\beta}_{2SLS} &= (\hat{L}'\hat{L})^{-1}\hat{L}'W \\ &= (L'P_Z L)^{-1} L P_Z W \end{aligned}$$

where $P_Z = Z(Z'Z)^{-1}Z'$ is the symmetric and idempotent projection matrix. Next, we plug in the DGP for W to assess consistency:

$$\begin{aligned} \hat{\beta}_{2SLS} &= \beta(L'P_Z L)^{-1} L' P_Z L + (L'P_Z L)^{-1} L P_Z (U + \beta_1 \hat{V}) \\ &= \beta + (\hat{L}'\hat{L})^{-1}\hat{L}'U && (L\hat{V} = 0 \text{ by construction}) \\ \Rightarrow \hat{\beta}_{2SLS} &\xrightarrow{P} \beta + \mathbb{E}(\hat{L}'\hat{L})^{-1} \times 0 = \beta && (\mathbb{E}(\hat{L}'U) = 0 \text{ by assumption}) \end{aligned}$$

Hence, the 2SLS estimator is consistent. To see that the estimator is biased, note that

$$\mathbb{E}(\hat{\beta}_{2SLS}) = \beta + \mathbb{E}\left[(\hat{L}'\hat{L})^{-1}\hat{L}'U\right]$$

Recall that $\hat{L}$ is a function of both Z and L and that the expectations operator is linear only; this means that we cannot pass the expectation inside the matrix inverse and the the term on the right is in general not equal to zero. Hence, the 2SLS estimator is biased.

D) If Z_i represents LSAT score is a valid instrument, it must be that $\mathbb{E}[ZU] = 0$. That is, there cannot be unobserved factors correlated with LSAT score that are also correlated with wage. This is unlikely to be true, as the same factors that affect the law school decision discussed above probably affect LSAT score as well.

However, tests are noisy indicators. Within a neighborhood of radius δ around a cutoff, LSAT scores may not actually indicate anything about ability. The only effect of the LSAT in this narrow window is to determine whether someone gets into law school. It is therefore assumed to be correlated with wages only through the law school admission process, i.e. it is uncorrelated with wages conditional on law school graduation.

With numbers, the difference of a 167 and 169 on the LSAT may be almost as good as random, and thus if 168 is the cutoff score, there may be very little difference between people scoring 167 and 169. Thus, limiting are sample to those just around 88 may give us a plausible causal estimate. That is, we estimate:

$$\mathbb{E}[W_i|L_i = 1; Z_i = 169] - E[W_i|L_i = 0; Z_i = 167]$$

We have likely dropped a lot of observations, but we have arguable controlled for a source of bias. You?ll learn more about this approach—called regression discontinuity (RD)—in 245A, 245C, and the applied labor classes.

2 Measurement Error

A) β in our classical linear regression model is identified by multiplying X_t to both sides of the regression:

$$X_t Y_t = \beta X_t^2 + X_t U_t$$

Taking the expectation of both sides identifies β if $E(X_t U_t) = 0$ (because we have sample analogues for $E(X_t Y_t), E(X_t^2)$). In the presence of measurement error, however, the regression equation becomes:

$$X_t Y_t = \beta X_t^2 + X_t(U_t - \beta V_t)$$

Because $E(X_t V_t) \neq 0$, $\beta = \frac{E(X_t Y_t)}{E(X_t^2) - E(X_t V_t)}$. Because we do not have a sample analogue for $E(X_t V_t)$, β is not identified.

B) For Z_t to be a valid instrument, it must be the case that

$$n^{-1} \sum Z_t Y_t \xrightarrow{p} \beta Q_{ZX}$$

Where $Q_{ZX} = \lim_{n \to \infty} n^{-1} \sum Z_t X_t$. Let's check this:

$$n^{-1} \sum Z_t Y_t = n^{-1} \sum Z_t (\beta X_t + U_t - \beta V_t)$$

So, the first term looks good. $n^{-1}\sum \beta Z_t X_t \xrightarrow{p} \beta Q_{ZX}$ so we are off to a good start. Because Z_t is a function of X_t and X_t is independent of U_t:

$$n^{-1}\sum U_t Z_t \xrightarrow{p} 0$$

So that looks good too. What about the final term?

$$n^{-1}\sum V_t Z_t = n^{-1}\sum V_t 1(X_t \geq med(X)) - n^{-1}\sum V_t 1(X_t < med(X))$$

This converges in probability to

$$E(V|X \geq m)P(X \geq m) - E(V|X < m)P(X < m).$$

Where m is the population median of X. Clearly, $P(X \geq m) = P(X < m) = \frac{1}{2}$. Therefore, Z is a valid instrument if

$$E(V|X \geq m) = E(V|X < m)$$

This would hold if V and X were independent, but they are not here. Because these two are dependent, the two conditional expectations are not likely to be equal and Z is not a valid instrument.

C) As mentioned in class, the assumption that V_t (our measurement error) is distributed symmetrically around zero provides us with an extra moment information we can exploit. Lets set up the moment conditions we know. Because the model is in deviation-from-means form, the first moments are zero. The second moments yield:

$$E(X_t^2) = E(X_t^{*2}) + \sigma_V^2$$

$$E(Y_t^2) = \beta^2 E(X_t^{*2}) + \sigma_U^2$$

$$E(X_t Y_t) = \beta E(X_t^{*2})$$

This is a system of three equations with 4 unknowns. For this system β is not identified. To estimate the parameters, we need further moment conditions. Because the distributions of U_t and V_t are symmetric, the third moments are zero, which yields

$$E(X_t^3) = E(X_t^{*3})$$

$$E(Y_t^3) = \beta^3 E(X_t^{*3})$$

We now have a system of 5 equations with 5 unknowns so β is identified. We can now solve via method of moments and, in doing so, we get:

$$B_{MM} = \left(\frac{\sum Y_t^3}{\sum X_t^3}\right)^{\frac{1}{3}}$$