Andrew ElHabr
ECON 7022
10/29/17

Quiz 6

Reading Assignment

(a) In the IV setup, we first start with the structural model $y_i = x_i'\beta + u_i$ in which $K - L$ out of the K variables are endogenous, i.e. $\mathbb{E}[x_{ik}u_i] = 0$ for $k = L+1, ..., K$, where without loss of generality the first L variables are exogenous. Then we use $K - L$ IVs to form z_i such that $\mathbb{E}[z_{ij}u_i] = 0$ for $j = 1, ..., K - L$. Now, if we let w_i be the vector of the L original exogenous variables and $K - L$ IVs, then $\mathbb{E}[w_i(y_i - x_i'\beta)] = 0$ and the IV estimator is $\beta_{S2S} = (\mathbb{E}[w_i x_i'])^{-1}\mathbb{E}[z_i y_i]$.

(b) In order to do 2SLS-IV estimation, first one needs to identify IVs that satisfy the two assumptions (Exogeneity IV.1 and Relevance IV.2) to replace the endogenous variables. Then, regress the endogenous variables on its IV and all other exogenous covariates in the original model. Once this is done, take the fitted values of the endogenous variable and put these back into the original model in place of the endogenous variable.

Exercises

Omitted Variables

- The model with the omitted variable is

$$earnings_i = \hat{\beta}_0 + \hat{\beta}_1 schooling_i + u_i.$$

With our given data, we have that $\beta_1 = 0.79$. Our first assumption that intelligence is strongly correlated with years of schooling means that $\text{Cov}(earnings_i, u_i) \neq 0 \implies \mathbb{E}[u_i|schooling_i] \neq 0 \implies \mathbb{E}[u_i schooling_i] \neq 0$ and that

$$schooling_i = \hat{\delta}_0 + \hat{\delta}_1 intelligence_i + v_i$$

where $\hat{\delta} = 1$ We further assume that

$$earnings_i = \hat{\alpha}_0 + \hat{\alpha}_1 intelligence_i + e_i$$

where $\hat{\alpha}_1 = 0.04$.

- The full model is

$$earnings_i = \tilde{\beta}_0 + \tilde{\beta}_1 schooling_i + \tilde{\beta}_2 intelligence_i + w_i.$$

Then, we can do the following:

$$\begin{aligned}
\text{Cov}(earnings, earnings) =& \tilde{\beta}_1 Cov(earnings, schooling) + \tilde{\beta}_2 \text{Cov}(earnings, intelligence) + \text{Cov}(earnings, w) \\
=& \tilde{\beta}_1 Cov(earnings, schooling) + \tilde{\beta}_2 \text{Cov}(earnings, intelligence) + 0 \\
\implies 1 =& \tilde{\beta}_1 \frac{Cov(earnings, schooling)}{\text{Cov}(earnings, earnings)} + \tilde{\beta}_2 \frac{\text{Cov}(earnings, intelligence)}{\text{Cov}(earnings, earnings)} \\
=& \tilde{\beta}_1 \hat{\beta}_1 + \tilde{\beta}_2 \hat{\alpha}_1 \\
=& 0.79\tilde{\beta}_1 + 0.04\tilde{\beta}_2
\end{aligned}$$

We can compute the omitted variables bias in the following way:

$$\begin{aligned}\hat{\beta}_1 =& \tilde{\beta}_1 + \tilde{\beta}_2\hat{\delta}_1 \\ 0.79 =& \hat{\beta}_1 + \tilde{\beta}_2(1) \\ =& \hat{\beta}_1 + \tilde{\beta}_2\end{aligned}$$

where $\hat{\delta}_1$ is the regression coefficient estimate from the regression of *intelligence* on *schooling*, so $\tilde{\beta}_2\hat{\delta}_1 = \tilde{\beta}_2$ is the bias in the original coefficient estimate. Combining the two equations that we found, i.e. $1 = 0.79\tilde{\beta}_1 + 0.04\tilde{\beta}_2$ and $0.79 = \hat{\beta}_1 + \tilde{\beta}_2$, we find that $\hat{\beta}_1 = 1.29$ and $\hat{\beta}_2 = -0.5012$. Thus, the bias in the original coefficient estimate is -0.5012.

Omitted Variables as Code

All completed in ECON_7022_Quiz_6.R file

- I generated a 3-dimensional multivariate random normal distribution. I had to change the covariance between schooling and intelligence from 1.00 to 0.50 in order to make the sigma matrix positive definite.
- The estimation of the full model is

$$earnings_i = \tilde{\beta}_0 + \tilde{\beta}_1 schooling_i + \tilde{\beta}_2 intelligence_i + w_i.$$

- I calculated a biased estimate of the truncated model as $\hat{\beta}_1 = 0.7899677$.
- I calculated an estimate of the real model as $\tilde{\beta}_1 = 1.033796$. We can see that the biased coefficient is much smaller than its real estimate using the whole model, which is actually undervalued here since I had to change the covariance matrix. With my altered covariance matrix, I found a bias of -0.243828, which is about half of the theoretical bias I calculated. This makes sense since I halved the covariance between intelligence and schooling.

IQ as IV or as Proxy?

- First, I am assuming that the problem statement meant to say that $Corr(schooling, IQ) = 0.8$ instead of $Corr(ability, schooling) = 0.8$. I would instrument schooling with IQ since it has a stronger correlation with schooling than ability. In general, a good instrument should have a strong correlation with which variable that one plans to replace it and should have no correlation to error term of the original model. We cannot test the second assumption about the correlation between IQ and the error term here. We would need some good reason to think that these two terms are uncorrelated in order to continue with analysis.
- I would use IQ as a proxy for ability since it is very difficult to measure true ability, but IQ can be roughly measured through an IQ test. A good proxy should be able to be measured and should be correlated strongly with the variable to which it serves as a proxy. In this case, it looks like this assumption is satisfied.
- I will aggravate the OVB above if I use IQ as a proxy for ability/intelligence since IQ is less strongly correlated with schooling (0.8) than ability is with schooling (1). Using these numbers, the OVB is increased from -0.5012 to -0.6350.
- I would still use IQ as a proxy even though it appears to increase OVB since it is relatively easier to measure than true ability. I wouldn't feel justified using it as an IV unless I knew for sure that it is not correlated with the original model's error term.