Instrumental Variables & 2SLS

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$
- $x_1 = \pi_0 + \pi_1 z + \pi_2 x_2 + \ldots + \pi_k x_k + v$

Ch.15 Instrumental Variables & 2SLS

1. Motivation: Omitted Variables in a Model
2. IV Estimation of the Multiple Regression Model
3. Two Stage Least Squares
4. IV Solutions to Errors-in-Variables Problems
5. Testing Endogeneity & Overidentifying Restrictions
6. 2SLS with Heteroskedasticity*
7. Applying 2SLS to Time Series Equations*
8. Applying 2SLS to Pooling & Panel Data*

15.1 Motivation: Why use IV?

- Instrumental Variables (IV) estimation is used when a model has ***endogenous*** x's.
  - That is, whenever $\text{Cov}(x,u) \neq 0$ (15.3).
  - Thus, IV can be used to alleviate the problem of omitted variable bias.
- Additionally, IV can be used to solve the classical ***errors-in-variables*** problem.
  - See section 9.3.

What is an Instrumental Variable?

- In order for a variable, z, to serve as a valid instrument for x, the following must be true.
  - The instrument must be ***exogenous***, that is, $\text{Cov}(z,u) = 0$. (15.4)
  - The instrument must be correlated with the endogenous variable x, that is, $\text{Cov}(z,x) \neq 0$. (15.5)
  - where z is ***instrumental variable*** for x.

More on Valid Instruments

- We have to use common sense and economic theory to decide if it makes sense to assume $\text{Cov}(z,u) = 0$.
- By contrast, we can test if $\text{Cov}(z,x) \neq 0$.
  - Just testing H_0: $\pi_1 = 0$ in $x = \pi_0 + \pi_1 z + v$. (15.6)
  - Sometimes refer to this regression as ***the first-stage regression***.

IV Estimator in the Simple Regression

- For $y = \beta_0 + \beta_1 x + u$, and given our assumptions,
  - $\text{Cov}(z,y) = \beta_1 \text{Cov}(z,x) + \text{Cov}(z,u)$, so
  - $\beta_1 = \text{Cov}(z,y) / \text{Cov}(z,x)$. (15.9)
- Then the ***IV estimator*** for β_1 is

$$\hat{\beta}_1 = \frac{\sum (z_i - \bar{z})(y_i - \bar{y})}{\sum (z_i - \bar{z})(x_i - \bar{x})} \quad (15.10)$$

  - When $z = x$, we obtain OLSE of β_1.

Stat. Inference with IV Estimator

- The homoskedasticity assumption in this case is
  $E(u^2|z) = \sigma^2 = \text{Var}(u)$. (15.11)
- As in the OLS case, given the asymptotic variance, we can estimate the standard error

$$Est.Asy.Var\left(\hat{\beta}_{1,IV}\right) = \frac{\hat{\sigma}^2}{SST_x R^2_{x,z}} \quad (15.13)$$

$$Est.Var\left(\hat{\beta}_{1,OLS}\right) = \frac{\hat{\sigma}^2}{SST_x} \quad (2.57)'$$

IV vs. OLS estimator

- Standard error in IV case differs from OLS only in the R^2 from regressing x on z.
  - Since $R^2 < 1$, IV standard errors are larger.
- IV is consistent, while OLS is inconsistent, when $\text{Cov}(x,u) \neq 0$.
- The stronger the correlation between z and x, the smaller the IV standard errors.

The Effect of Poor Instruments

- What if our assumption that $\text{Cov}(z,u) = 0$ is false?
  - The IV estimator will be inconsistent.
- Compare asymptotic bias in OLS and IV.
  - Prefer IV if $\text{Corr}(z,u)/\text{Corr}(z,x) < \text{Corr}(x,u)$.

$$\text{IV}: \text{plim}\hat{\beta}_1 = \beta_1 + \frac{Corr(z,u)}{Corr(z,x)} \cdot \frac{\sigma_u}{\sigma_x} \quad (15.19)$$

$$\text{OLS}: \text{plim}\,\tilde{\beta}_1 = \beta_1 + Corr(x,u) \cdot \frac{\sigma_u}{\sigma_x} \quad (15.20)$$

15.2 IV in the Multiple Regression

- The model we are interested in estimating is called the ***structural model***.
- Our problem is that one or more of the variables are ***endogenous***.
  $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$ (15.22)
  - where y_2 is ***endogenous*** and z_1 is ***exogenous.***
  - That is, $\text{Cov}(y_2,u_1) \neq 0$ and $\text{Cov}(z_1,u_1) = 0$.

Cont. IV in Multiple Regression

- We need an instrument for each endogenous variable.
  - Let z_2 be the instrument, so $\text{Cov}(z_2,u_1) = 0$ and
  - $E(u_1) = 0$ & $\text{Cov}(z_1,u_1) = 0$. (15.24)
- The ***reduced form equation*** regresses the endogenous variable on all exogenous ones;
  $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2$. (15.26)
  - where $\pi_2 \neq 0$. (15.27)
  - Under these assumptions, z_2 is a valid IV for y_2.

15.3 Two Stage Least Squares (2SLS)

- Consider the structural model again,
  $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$ (15.22)
- Suppose 2 more exogenous variables, z_2 & z_3.
  $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2$ (15.33)
  - Here we're assuming that both z_2 and z_3 are valid instruments.
  - They do not appear in the structural model and are uncorrelated with the structural error term, u_1.

Best Instrument

- The best instrument is a linear combination of all of the exogenous variables,

$$y_2^* = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3. \quad (15.34)$$

  - We can estimate y_2^* by regressing y_2 on z_1, z_2 and z_3 – can call this ***the first stage***.
- If then substitute $\hat{y}_2$ for y_2 in the structural model, get the same coefficient as IV.
  - regressing y_1 on $\hat{y}_2$ and z_1. (15.38)
  - call this ***the second stage***.

More on 2SLS

- While the coefficients are the same, the standard errors from doing 2SLS by hand are incorrect, so let TSP do it for you.
- Method extends to multiple endogenous variables – need to be sure that we have at least as many excluded exogenous variables (instruments) as there are endogenous variables in the structural equation.

15.4 IV for Errors-in-Variables

- Remember the classical ***errors-in-variables*** problem where we observe x_1 instead of x_1^*,

$$y_1 = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + u \quad (15.45)$$

  - Where $x_1 = x_1^* + e_1$, and e_1 is uncorrelated with x_1^* and x_2. →see (15.46)
- If there is a z, such that $\text{Corr}(z,u) = 0$ and $\text{Corr}(z,x_1) \neq 0$, then
- IV will remove the attenuation bias.

15.5 Testing for Endogeneity

- Since OLS is preferred to IV if we do not have an endogeneity problem, then we'd like to be able to test for endogeneity.
- If we do not have endogeneity, both OLS and IV are consistent.
- Idea of ***Hausman test*** is to see if the estimates from OLS and IV are different.

Cont. Testing for Endogeneity

- While it's a good idea to see if IV and OLS have different implications, it's easier to use a regression test for endogeneity.
- If y_2 is endogenous, then v_2 (from the reduced form equation) and u_1 from the structural model will be correlated.
- The test is based on this observation.

Cont. Testing for Endogeneity

1. Save the residuals from the first stage.
2. Include the residual in the structural equation (which of course has y_2 in it).
3. If the coefficient on the residual is statistically different from zero, reject the null of exogeneity.
   - If multiple endogenous variables, jointly test the residuals from each first stage.

Testing Overidentifying Restrictions

- If there is just one instrument for our endogenous variable, we can't test whether the instrument is uncorrelated with the error.
- We say the model is just identified.
- If we have multiple instruments, it is possible to test the overidentifying restrictions – to see if some of the instruments are correlated with the error.

Cont. The OverID Test

- Estimate the structural model using IV and obtain the residuals.
- Regress the residuals on all the exogenous variables and obtain the R^2 to form nR^2.
- Under the null that all instruments are uncorrelated with the error, LM ~ χ_q^2 where q is the number of extra instruments.

15.6 Testing for Heteroskedasticity

- When using 2SLS, we need a slight adjustment to the Breusch-Pagan test.
- Get the residuals from the IV estimation.
- Regress these residuals squared on all of the exogenous variables in the model (including the instruments).
- Test for the joint significance.

15.7 Testing for Serial Correlation

- When using 2SLS, we need a slight adjustment to the test for serial correlation.
- Get the residuals from the IV estimation.
- Re-estimate the structural model by 2SLS, including the lagged residuals, and using the same instruments as originally.
- Can do 2SLS on a quasi-differenced model, using quasi-differenced instruments.