Lecture 7

Solutions to Endogeneity:
Two Stage Least Squares

Nurgul Tilenbaeva

American University - Central Asia

12.12.2022

Contents

Two Stage Least Squares

$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$

z_1 is *exogenous*;
y_2 is *endogenous*.

Now we have two instrumental variables z_2 and z_3:

1. z_2, z_3 are uncorrelated with u_1 **(exclusion restrictions)**;
   - z_2, z_3 do not have a direct effect on y_1;
   - z_2, z_3 have an effect on y_1 only through y_2.
2. z_2, z_3 are both correlated with y_2:
   The **reduced form equation is**:
   $y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2$
   Then, we need the key identifying assumption:
   $\pi_2 \neq 0$ or $\pi_3 \neq 0$
   We can test this using an F statistic.

Two Stage Least Squares

$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1$

2. We regress y_2 on z_1, z_2, z_3 using OLS and obtain the fitted values:

   $\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3$

   Once we have $\hat{y}_2$, we can use it as the IV for y_2.

   With multiple instruments, the IV estimator using $\hat{y}_2$ as the instrument is called the two stage least squares (2SLS) estimator.

   With a single IV for y_2, the IV estimator is identical to the 2SLS estimator. Therefore, when we have one IV for each endogenous explanatory variable, we can call the estimation method IV or 2SLS.

Testing for Endogeneity

What happens if we use 2SLS when the variable of interest is in fact *exogenous*?

- The 2SLS estimator is less efficient than OLS, i.e. the 2SLS estimates can have very large standard errors.
- Therefore, it is useful to have a **test for endogeneity** of an explanatory variable that shows whether 2SLS is even necessary.

$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1$

z_1, z_2 are *exogenous*;
We have two additional exogenous variables z_3 and z_4.

If y_2 is uncorrelated with u_1, we should estimate the model by OLS. How can we test this?
Hausman (1978) suggested directly comparing the OLS and 2SLS estimates and determining whether the differences are statistically significant. If 2SLS and OLS differ significantly, we conclude that y_2 must be **endogenous**.

Testing for Endogeneity

$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1$

The **reduced form** for y_2 is:

$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \pi_4 z_4 + v_2$

- We know that each z_j is uncorrelated with u_1;
- Then, y_2 is uncorrelated with u_1 if, and only if, v_2 is uncorrelated with u_1: this is what we wish to test.
- Write $u_1 = \delta_1 v_2 + e_1$. Then, u_1 and v_2 are uncorrelated if, and only if, $\delta_1 = 0$.
- To test this, we estimate by OLS
  $y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \delta_1 \hat{v}_2 + error$
  and test $H_0 : \delta_1 = 0$ using a t statistic.
  If we reject H_0, we conclude that y_2 is **endogenous** because v_2 and u_1 are correlated.

Testing for Endogeneity

Testing for Endogeneity of a Single Explanatory Variable:

1. Estimate the reduced form for y_2 by regressing it on *all* exogenous variables (including those in the structural equation and the additional IVs). Obtain the residuals, $\hat{v}_2$.
2. Add $\hat{v}_2$ to the structural equation (which includes y_2) and test for significance of $\hat{v}_2$ using an OLS regression. If the coefficient on $\hat{v}_2$ is statistically different from zero, we conclude that y_2 is indeed endogenous.

Testing Overidentification Restrictions

- In the context of the simple IV estimator, we noted that the **instrument exogeneity** cannot be tested.
- However, if we have more instruments than we need, we can test whether some of them are uncorrelated with the structural error.

Testing Overidentification Restrictions

$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1$

z_1, z_2 are *exogenous*;
y_2 is *endogenous*;
z_3, z_4 are two instrumental variables for y_2.

1. We can estimate the model using, say, only z_3 as an IV for y_2; let $\check{\beta}_1$ be the resulting IV estimator of β_1;
2. Then, we can estimate the model using only z_4 as an IV for y_2; call this IV estimator $\tilde{\beta}_1$;
3. If all z_j are exogenous, and if z_3 and z_4 are each partially correlated with y_2, then $\check{\beta}_1$ and $\tilde{\beta}_1$ are both consistent for β_1.
4. Hausman (1978) proposed basing a test of whether z_3 and z_4 are both exogenous on the difference, $\check{\beta}_1 - \tilde{\beta}_1$.
5. If we conclude that $\check{\beta}_1$ and $\tilde{\beta}_1$ are statistically different from one another, then we conclude that either z_3, z_4 or both fail the exogeneity requirement.

Testing Overidentification Restrictions

Problems with Testing Overidentification Restrictions:

- IV estimates may be similar even though both instruments fail the exogeneity requirement.
    - *Example:* if mother's education is positively correlated with u_1, then so is father's education. Therefore, the two IV estimates may be similar even though each is inconsistent.
- IV estimates may seem practically different yet, statistically we cannot reject the null hypothesis that they are consistent for the same population parameter.
    - In estimating the wage equation by IV using *motheduc* as the only instrument, the coefficient on *educ* is 0.049(0.037). If we use only *fatheduc* as the IV for *educ*, the coefficient on *educ* is 0.070(0.034). For policy purposes, the difference between 5% and 7% for the estimated return to a year of schooling is substantial. Yet, the difference is not statistically significant.

Testing Overidentification Restrictions

Testing Overidentifying Restrictions:

1. Estimate the structural equation by 2SLS and obtain the 2SLS residuals, $\hat{u}_1$.
2. Regress $\hat{u}_1$ on *all exogenous* variables. Obtain the R-squared, say, R_1^2.
3. Under the null hypothesis that all IVs are uncorrelated with u_1, $nR_1^2 \sim \chi_q^2$, where q is the number of instrumental variables from outside the model minus the total number of endogenous variables. If nR_1^2 exceeds (say) the 5% critical value of the χ_q^2 distribution, we reject H_0 and conclude that at least some of the IVs are not exogenous.