Lecture 10

Endogeneity 4 of 4: Simultaneous Equations Models II

Michael Curran

Trinity College Dublin

JS Econometrics

Lecture 10 Outline

Identification & Estimation (2 Equations)
Identifying & Estimating a Structural Equation (2 Equation System)

Identification & Estimation (More Than 2 Equations)
Systems with More Than 2 Equations

Summary & References
Summary & References

Overview

- Lecture 9: OLS biased and inconsistent when applied to structural equation in a SEM.
- Lecture 8: 2SLS solves endogenous explanatory variables problem.
- This lecture: show how 2SLS can be applied to SEMs.
- Mechanics of 2SLS similar to chapter 15.
- Difference: since we specify structural equation for each endogenous variable, can immediately see whether sufficient IVs are available to estimate either equation.
- Begin by discussing identification problem.

Identification in a 2 Equation System

- Key identifying condition when estimating by OLS is that each explanatory variable is uncorrelated with the error term.
- This fundamental condition no longer holds, in general for SEMs.
- However, if we've some IVs, we can still identify (or consistently estimate) parameters in a SEM equation, just as with omitted variables or measurement errors.
- Before considering a general 2-equation SEM, get intuition from simple supply and demand example with system in equilibrium form (i.e. with $q_s = q_d = q$):

$$q = \alpha_1 p + \beta_1 z_1 + u_1 \tag{1}$$

$$q = \alpha_2 p + u_2 \tag{2}$$

- Given a random sample on (q, p, z_1), which of these equations can be estimated? That is, which is an **identified equation**?

Identification in a 2 Equation System

- To see *demand* equation (2) is identified but the supply equation is not use z_1 as an IV for price in equation (2).
- As z_1 appears in (1), we've no IV for price in the supply equation.
- Intuitively, demand equation identified as we've an observed variable z_1 that shifts supply equation while not affecting demand equation.
- Given variation in z_1 and no errors, can trace out demand curve.
- Unobserved demand shifter u_2 causes us to estimate the demand equation with error, but the estimators will be consistent provided z_1 uncorrelated with u_2; supply equation can't be traced out as there are no exogenous observed factors shifting the demand curve.
- Unobservable factors shifting the demand function don't help: need something observable. If as in the labour demand function h_d, we've an observed exogenous demand shifter e.g. income in the milk demand function, then the supply function would also be identified.
- Summary: *in system of* (1) *and* (2) *presence of an exogenous variable in supply equation allows us to estimate demand equation.*

Identification in a 2 Equation System

Extending identification discussed to a general 2 equation model:

$$y_1 = \beta_{10} + \alpha_1 y_2 + z_1 \boldsymbol{\beta}_1 + u_1 \tag{3}$$

$$y_2 = \beta_{20} + \alpha_2 y_1 + z_2 \boldsymbol{\beta}_2 + u_2 \tag{4}$$

z_1 denotes a set of k_1 exogenous variables appearing in the first equation: $z_1 = (z_{11}, z_{12}, \ldots, z_{1k_1})$. Similarly, z_2 is set of k_2 exogenous variables in second equation: $z_2 = (z_{21}, z_{22}, \ldots, z_{2k_2})$. In many cases, z_1 and z_2 will overlap. As a shorthand form, use notation:

$$z_1 \boldsymbol{\beta}_1 = \beta_{11} z_{11} + \beta_{12} z_{12} + \ldots + \beta_{1k_1} z_{1k_1}$$
$$z_2 \boldsymbol{\beta}_2 = \beta_{21} z_{21} + \beta_{22} z_{22} + \ldots + \beta_{2k_2} z_{2k_2}$$

Fact that z_1 and z_2 generally contain different exogenous variables means we've imposed **exclusion restrictions** on the models, i.e. we *assume* that certain exogenous variables don't appear in 1st equation and others are absent from the 2nd equation. As we saw with supply and demand examples, this allows us distinguish between the 2 structural equations.

Identification in a 2 Equation System

- When can we solve (3) and (4) for y_1 and y_2 (as linear functions of all exogenous variables and the structural errors u_1 and u_2)?
- The condition is the same as in the last lecture, *viz.* $\alpha_2\alpha_1 \neq 1$.
- Under this ass, reduce form exist for y_1 and y_2.
- Key question is: under what assumptions can we estimate parameters in say (3)? This is the identification issue.

Identification in a 2 Equation System

- **Rank condition** for identification of structural equation (3): First equation in 2-equation SEM is identified $\iff$ *second* equation contains at least one exogenous variable (with a nonzero coefficient) that is excluded from the first equation – necessary and sufficient condition for (3) to be identified.
- **Order condition** is necessary for the rank condition.
- The order condition for identifying the 1st equation states that at least one exogenous variable is excluded from this equation – trivial to check once both equations have been specified.
- Rank condition requires more: at least one of the exogenous variables excluded from the 1st equation must have a nonzero population coefficient in the 2nd equation – ensures at least one of the exogenous variables omitted from the 1st equation actually appears in the reduced form of y_2 so can use these variables as instruments for y_2; can test this using t or F test.
- Examples.

Estimation by 2SLS

- Once we have determined an equation is identified, we can estimate via 2SLS.
- IV consist of exogenous variables appearing in either equation.
- Example 16.5.
- Estimating the effect of openness on inflation by IV is also straightforward: e.g. 16.6.
- Question 16.3.

Lecture 10 Outline

Identification & Estimation (2 Equations)
- Identifying & Estimating a Structural Equation (2 Equation System)

Identification & Estimation (More Than 2 Equations)
- Systems with More Than 2 Equations

Summary & References
- Summary & References

Identification in Systems with 3 or More Equations

- SEMs can consist of more than 2 equations.
- General identification of these models is difficult and requires matrix algebra.
- Once an equation in a general system has been shown to be identified, it can be estimated by 2SLS.

Identification in Systems with 3 or More Equations

- 3-equation system illustrates identification issues with complicated SEMs.

$$y_1 = \alpha_{12} y_2 + \alpha_{13} y_3 + \beta_{11} z_1 + u_1 \tag{5}$$
$$y_2 = \alpha_{21} y_1 + \beta_{21} z_1 + \beta_{22} z_2 + \beta_{23} z_3 + u_2 \tag{6}$$
$$y_3 = \alpha_{32} y_2 + \beta_{31} z_1 + \beta_{32} z_2 + \beta_{33} z_3 + \beta_{34} z_4 + u_3 \tag{7}$$

- Generally difficult to show an equation in a SEM with more than 2 equations is identified but easy to see when specific equations are *not* identified: (7) since every exogenous variable appears in this equation, we have no IVs for y_2, so can't consistently estimate the parameters of this equation.
- OLS estimation will not usually be consistent.
- (5) is identified: z_2, z_3 and z_4 all excluded from the equation(another e.g. of *exclusion restrictions*) and although there are 2 endogenous variables in this equation, we've 3 potential IVs for y_2 and y_3, so (5) passes the order condition.

Identification in Systems with 3 or More Equations

Order Condition for Identification of General SEMs

- An equation in any SEM satisfies the order condition for identification if the number of *excluded* exogenous variables from the equation is at least as large as the number of RHS endogenous variables.
- The second equation (6) also passes the order condition because there's one excluded exogenous variable, z_4 and one RHS endogenous variable, y_1.
- As discussed in chapter 15 and in previous section, order condition is necessary (not sufficient) for identification.
- E.g. if $\beta_{34} = 0$, z_4 appears nowhere in the system, so it's not correlated with y_1, y_2 or y_3; if $\beta_{34=0}$, then the second equation is not identified, because z_4 is useless as an IV for y_1; this again illustrates that identification of an equation depends on the values of the parameters (which we can never know for sure) in the other equations.

Identification in Systems with 3 or More Equations

- Nomenclature on overidentified and just identified equations originated with SEMs.
- In terms of order condition (5) is an **overidentified equation** since we need only 2 IVs (for y_2 and y_3) but we have 3 available (z_2, z_3 and z_4): there's one overidentifying restriction in this equation.
- In general, number of overidentifying restrictions equals the total number of exogenous variables in the system minus the total number of explanatory variables in the equation.
- Equation (6) is a **just identified equation** and the third equation is an **unidentified equation**.

Estimation in Systems with 3 or More Equations

- Each identified equation can be estimated by 2SLS.
- Instruments for particular equation: exogenous variables in system.
- Tests for endogeneity, heterogeneity, serial correlation and overidentifying restrictions as before.
- When system is correctly specified (and certain additional assumptions hold), *system estimation methods* generally are more efficient than estimating each equation by 2SLS.
- Most common system estimation method in context of SEMs is *Three stage least squares* (not covered in this course – see advanced Wooldridge).

Lecture 10 Outline

Identification & Estimation (2 Equations)

Identifying & Estimating a Structural Equation (2 Equation System)

Identification & Estimation (More Than 2 Equations)

Systems with More Than 2 Equations

Summary & References

Summary & References

Summary

- 2SLS solves endogenous explanatory variables problem in SEMs (OLS biased and inconsistent when applied to structural equation in a SEM).
- Key: identify equations in a SEM before estimation via 2SLS – rank and order conditions.
- With more than two equations, once an equation is identified (trickier – focus on which equations are not identified), it can be estimated by 2SLS.
- Equations in a SEM are:
  1. Overidentified if there are more available IVs than we need.
  2. Just (exactly) identified if the number of available IVs is exactly the same as the number we need.
  3. Underidentified $\equiv$ unidentified otherwise.
- Note that system estimation methods are more efficient than estimating each equation separately by 2SLS.

References

- Simultaneous Equations Models II: Wooldridge 16.3-4.