Lecture 9

Endogeneity 3 of 4: Simultaneous Equations Models I

Michael Curran

Trinity College Dublin

JS Econometrics

Lecture 9 Outline

Overview

- IV solves two types of endogeneity problems:
    1. Omitted variables: like to hold fixed when estimating *cet par* effect of one or more of the observed explanatory variables.
    2. Measurement error: want to estimate effect of certain explanatory variables on y but we've mismeasured one or more.
    - Both cases: could use OLS if could collect better data.
- Another form of endogeneity of explanatory variables: **simultaneity** – one or more explanatory variable is *jointly determined* with the dependent variable, usually via an equilibrium mechanism.
- This week: methods for estimating simple SEMs.
- Leading estimation of SEMs is IV, so solution to simultaneity is same as IV solutions to omitted variables and measurement error; however, crafting and interpreting SEMs is challenging; thus, begin by discussing nature and scope of SEMs, then show OLS applied to an equation in a SEM is generally biased and inconsistent, next describe identification and estimation in a 2-equation system and finally briefly cover models with more than two equations.

Nature of SEM

- Each equation in system should have a *cet par* causal interpretation.
- Since we only observe outcomes in equilibrium, we're required to use counterfactual reasoning in constructing the equations of a SEM. Must think in terms of potential as well as actual outcomes.
- Classical e.g. of a SEM is a supply and demand equation for some commodity or input to production (e.g. labour).
  - Let h_s denote annual labour hours supplied by workers in agriculture, measured at the county level and let w denote the average hourly wage offered to such workers where z_1 is some observed variable affecting labour supply (e.g. average manufacturing wage in the county).

$$h_s = \alpha_1 w + \beta_1 z_1 + u_1 \quad (1)$$

  - Equation (1) is an example of a **structural equation**.
  - Name comes from fact that labour supply function is derivable from economic theory and has a causal interpretation.

Nature of SEM

- α_1 measures how labour supply changes with wages; if h_s and w are in logs, α_1 is labour supply elasticity.
- Typically, expect α_1 to be positive (though economic theory doesn't rule out $\alpha_1 \leq 0$.
- Labour supply elasticities are important for determining how workers will change the number of hours the desire to work when tax rates on wage income change.
- If z_1 is the manufacturing wage, we expect $\beta_1 \leq 0$: other factors equal, if the manufacturing wage increases, more workers will go into manufacturing than into agriculture.
- Graphing labour supply: sketch hours as a function of wage with z_1 and u_1 held fixed.
- A change in z_1 shifts the h_s function as does a change in u_1.
- Difference: z_1 is observed while u_1 is not.
- z_1 is called an *observed supply shifter* and u_1 is called an *unobserved supply shifter*.

Nature of SEM

- Though (1) is supposed to hold for all possible values of wage, we cannot generally view wage as varying exogenously for a cross section of counties.
- If we could run an experiment where we vary the level of agricultural and manufacturing wages across a sample of counties and survey workers to obtain h_s for each county, then we could estimate (1) by OLS.
- Instead, collect data on average wages in these 2 sectors along with how many hours were spent in agricultural production.
- Data best described by the interaction of labour supply *and* demand.
- Under the assumption that labour markets clear, we actually observed *equilibrium* values of wages and hours worked.

Nature of SEM

- To describe how equilibrium wages and hours are determined, need demand for labour:

$$h_d = \alpha_2 w + \beta_2 z_2 + u_2 \tag{2}$$

- Graph $h_d(w)$ keeping z_2 and u_2 fixed.
- z_2 say agricultural land area is an *observable demand shifter* while u_2 is an *unobservable demand shifter*.
- Labour demand equation is a structural equation: it can be obtained from the profit max considerations of farmers.
- If h_d and w are in log form, α_2 is the labour demand elasticity.
- Economic theory tells us that $\alpha_2 < 0$. Since labour and land are complements in production, expect $\beta_2 > 0$.

Nature of SEM

- Equations (1) and (2) describe entirely different relationships.
- Labour supply is a behavioural equation for workers, and labour demand is a behavioural relationship for farmers.
- Each equation has a *cet par* interpretation and stands on its own.
- They become linked in an econometric analysis only because *observed* wage and hours are determined by the intersection of supply and demand, i.e. for each country i observed hours h_i and observed wage w_i are determined by the equilibrium condition:

$$h_{is} = h_{id} \tag{3}$$

$$h_i = \alpha_1 w_i + \beta_1 z_{i1} + u_{i1} \tag{4}$$

$$h_i = \alpha_2 w_i + \beta_2 z_{i2} + u_{i2} \tag{5}$$

- These 2 equations constitute a **simultaneous equations model** (**SEM**).

Nature of SEM

1. Given z_{i1}, z_{i2}, u_{i1} and u_{i2}, equations (4)& (5) determine h_i and w_i.
   - Need to assume $\alpha_1 \neq \alpha_2$, i.e. slopes differ.
   - So h_i and w_i are the **endogenous variables** in this SEM.
   - Since z_{i1} and z_{i2} determined outside of model, they are **exogenous variables**.
   - Assume z_{i1} and z_{i2} uncorrelated with supply and demand errors (u_{i1} and u_{i2}): examples of **structural errors** since they appear in the structural equations.
2. Without z_1 and z_2, cannot distinguish supply function from demand function, **observational equivalence**. Example.

Nature of SEM

- Most convincing examples are similar to demand and supply examples: each equation should have a behavioural, *cet par* interpretation on its own.
- As we only observe equilibrium outcomes, specifying a SEM requires us to ask counterfactual questions – examples?
- Example 16.2 highlights inappropriate use of SEM: the 2 endogenous variables are chosen by the same economic agent so neither equation can stand alone.
- Another example of inappropriate use of SEM: weekly hours pent studying and weekly hours spent working (same idea).
- Thus, simply because 2 variables are determined simultaneously does *not* mean that a SEM is suitable – for a SEM to make sense, each equation in the SEM should have a cet par interpretation in isolation from the other equation.
- Question 16.1.

Lecture 9 Outline

Simultaneity Bias in OLS

An explanatory variable that is determined simultaneously with the dependent variable is generally correlated with the error term $\Longrightarrow$ bias and inconsistency in OLS. 2-equation structural model:

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1 \tag{6}$$

$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2 \tag{7}$$

Focus on estimating the first equation. z_1 and z_2 exogenous so each is uncorrelated with u_1 and u_2. To show that y_2 is correlated with u_1, solve 2 equations for y_2 in terms of exogenous variables and the error term. Plugging RHS of (6) for y_1 in (7)

$$y_2 = \alpha_2(\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2$$

$$(1 - \alpha_2\alpha_1)y_2 = \alpha_2\beta_1 z_1 + \beta_2 z_2 + \alpha_2 u_1 + u_2 \tag{8}$$

Now make assumptions about parameters in order to solve for y_2:

$$\alpha_2\alpha_1 \neq 1 \tag{9}$$

Simultaneity Bias in OLS

- Divide (8) by $(1 - \alpha_2\alpha_1)$ and write

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2 \tag{10}$$

where $\pi_{21} = \frac{\alpha_2\beta_1}{(1-\alpha_2\alpha_1)}$, $\pi_{22} = \frac{\beta_2}{(1-\alpha_2\alpha_1)}$ and $v_2 = \frac{(\alpha_2 u_1 + u_2)}{(1-\alpha_2\alpha_1)}$. Equation (10), which expresses y_2 in terms of exogenous variables and error terms is the **reduced form equation** for y_2.

- π_{21} and π_{22} are **reduced form parameters** – nonlinear functions of **structural parameters**, which appear in structural equations, (6) & (7).
- **Reduced form error** v_2 is linear function of structural error terms u_1 and u_2.
- Since u_1 and u_2 are each uncorrelated with z_1 and z_2, v_2 is uncorrelated with z_1 and z_2.
- Thus, can consistently estimate π_{21} and π_{22} by OLS, something that is used for 2SLS.

Simultaneity Bias in OLS

- Use (10) to show (except under special assumptions) OLS estimation of (6) will produce biased and inconsistent estimates of α_1 and β_1 in (6).
- As z_1 and u_1 uncorrelated, issue is whether y_2 and u_1 are uncorrelated.
- From reduced form in (10), see that y_2 and u_1 are correlated $\iff$ v_2 and u_1 are correlated (since z_1 and z_2 assumed exogenous).
- v_2 linear function of u_1 and u_2 so generally correlated with u_1.
- Assuming u_1 and u_2 are uncorrelated, v_2 and u_1 *must* be correlated whenever $\alpha_2 \neq 0$.
- Even if $\alpha_2 = 0$, i.e. y_1 doesn't appear in (7), v_2 & u_1 correlated if u_1 & u_2 are correlated.
- When $\alpha_2 = 0$ *and* u_1 & u_2 uncorrelated, y_2 and u_1 are also uncorrelated.

Simultaneity Bias in OLS

- Fairly strong requirements: if $\alpha_2 = 0$, y_2 not simultaneously determined with y_1. If we add 0 correlation between u_1 and u_2, this rules out omitted variables or measurement error in u_1 that are correlated with y_2.
- Shouldn't be surprised that OLS estimator of (6) works in this case.
- When y_2 is correlated with u_1 because of simultaneity, we say that OLS suffers from **simultaneity bias**.

Simultaneity Bias in OLS

- Getting direction of bias is generally complicated, but in simple models, we can determine direction of bias. e.g. suppose we simplify (6) by dropping z_1 from equation and assume u_1 and u_2 uncorrelated, then:

$$\begin{aligned} Cov(y_2, u_1) = Cov(v_2, u_1) &= [\alpha_2/(1-\alpha_2\alpha_1)]E(u_1^2) \\ &= [\alpha_2/(1-\alpha_2\alpha_1)]\sigma_1^2 \end{aligned}$$

  where $\sigma_1^2 = Var(u_1) > 0$.
- Thus, asymptotic bias (or inconsistency) in OLS estimator of α_1 has same sign as $\alpha_2/(1-\alpha_2\alpha_1)$. If $\alpha_2 > 0$ and $\alpha_2\alpha_1 < 1$, the asymptotic bias is positive.
- Unfortunately, conclusions don't apply in more general models but serve as useful guide.

Lecture 9 Outline

Summary

- Type of endogeneity: simultaneity – one or more explanatory variable is *jointly determined* with the dependent variable, usually via an equilibrium mechanism.
- IV estimation can be applied to SEMs.
- Nature of SEM:
  - Each equation should have causal interpretation.
  - Classic example: demand & supply.
  - Note use of counterfactual reasoning to construct equations in SEM since only observe outcomes in equilibrium, observed and unobserved shifters and observational equivalence (e.g. can't distinguish supply from demand functions given data).
- OLS applied to an equation ion a SEM is generally biased and inconsistent.

References

- Simultaneous Equations Models I: Wooldridge 16.1-2.