Recommended Reading

For the today

- Simultaneous Equations Models
- Chapter 16 (pp. 501 – 523).

In the next week

- Limited Dependent Variable Models
- Chapter 17 (pp. 529 – 565).
Today’s Talk

- Previous lecture, we have shown how instrumental variables can be useful in solving endogeneity problems:
  - omitted variables
  - measurement error
- In both cases, we could use OLS if we have better data.
- Today, we will consider another important form of endogeneity which is **simultaneity**.
- **Simultaneity** is a specific type of endogeneity problem in which the explanatory variable is jointly determined with the dependent variable.
Example of Simultaneity

- The most important is to remember that each equation in the system should have its own casual interpretation.

Supply Equation

Consider labor supply function:

\[ h_s = a_1 w + b_1 z_1 + u_1, \]

where \( h_s \) is (annual) labor hours supplied by workers, \( w \) average hourly wage, \( z_1 \) some observed variable affecting supply (i.e. average manufacturing wage in country)

- We call supply equation a **structural equation** as it comes from the economic theory and has casual interpretation
- \( \alpha_1 \) measures how labor supply changes with change of wage
- When \( h_s \) and \( w \) are in logarithms, \( \alpha_1 \) is labor supply elasticity.
Example of Simultaneity cont.

- Note that $z_1$ and $u_1$ shifts the supply, $z_1$ is observed, while $u_1$ is not.
- So how does this equation differ from what we have studied previously?
- Wage $w$ cannot be viewed as exogenous as supply equation holds for all positive values of wage.
- Thus we can not simply regress $h_s$ on $w$ as $w$ is endogenous variable.
- Thus we have to collect some more data.
- We have to understand that the data are best described by labor supply and demand interaction.
Demand Equation
Consider labor demand function:

\[ h_d = a_2 w + b_2 z_2 + u_2, \]

where \( h_s \) is (annual) labor hours demanded, \( z_2 \) some observed variable affecting demand (i.e. land area).

- Demand equation is also **structural equation**.
- These two equations are linked through intersection of supply and demand, which is equilibrium \( h_{is} = h_{id} \).
- As we observe only equilibrium hours for each country \( i \), we denote \( h_i \) observed hours.
Example of Simultaneity cont.

**Labor Supply and Demand Equation**

\[ h_i = a_1 w_i + b_1 z_{i1} + u_{i1}, \]
\[ h_i = a_2 w_i + b_2 z_{i2} + u_{i2}, \]

- While we call this *simultaneous equations model* (SEM).
- \( h_i \) and \( w_i \) are *endogenous variables* determined by equations.
- \( z_{i1} \) and \( z_{i2} \) are *exogenous variables* determined outside of the system.
- Without \( z_{i1} \) and \( z_{i2} \) we cannot recognize demand from supply \( \Rightarrow \) they identify the equations.
- Without supply and demand shifters, we are not able to estimate the system.
Simultaneity Bias in OLS

Consider structural model

\[
y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1 \\
y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2
\]

- Variables \( z_1 \) and \( z_2 \) are exogenous \((\text{Cov}(z_1, u_1) = 0, \quad \text{Cov}(z_2, u_2) = 0)\)
- If we put first equation to the second one, we have:
  \[
y_2 = \alpha_2(\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2
\]

Reduced form for \( y_2 \)

\[
y_2 = \pi_{21} z_1 + \pi_{22} z_2 + \nu_2,
\]

where \( \pi_{21} = \alpha_2/\beta_1/(1 - \alpha_2\alpha_1), \quad \pi_{22} = \beta_2/(1 - \alpha_2\alpha_1) \) and
\( \nu_2 = (\alpha_2 u_1 + u_2)/(1 - \alpha_2\alpha_1) \)
Simultaneity Bias in OLS cont.

- Because \( \nu_2 \) is linear function of \( u_1 \) and \( u_2 \), \( u_1 \) and \( u_2 \) are uncorrelated with \( z_1 \) and \( z_2 \).
- \( \Rightarrow \) \( \nu_2 \) is uncorrelated with \( \pi_{21} \) and \( \pi_{22} \) and OLS estimates are consistent.
- While \( \hat{\pi}_{21} \) and \( \hat{\pi}_{22} \) will be consistent, \( \hat{\alpha}_1 \) and \( \hat{\beta}_1 \) from the first equation will be inconsistent.
- This is because \( y_2 \) and \( u_1 \) are correlated (it can be seen directly from the reduced form).
- When \( y_2 \) and \( u_1 \) are correlated because of simultaneity, OLS will suffer from the **simultaneity bias**.

**Interactive Example in Mathematica.**

*(NOTE: To make this link work, you need to have Lecture9.cdf at the same directory and Free CDF player installed on your computer.)*
Thus we can see that estimating structural equation in a simultaneous equations system by OLS results in biased and inconsistent estimator.

- We can solve this problem by using **two-stage least square estimator (2SLS)**
- As we specify the structural equation for each endogenous variable, we can **immediately** see if sufficient instrumental variables are in the equation.
- We call this **identification problem**.
Identification in Two-Equation System

Example: Supply and Demand of milk consumption

\[ q = \alpha_1 p + \beta_1 z_1 + u_1 \]
\[ q = \alpha_2 p + u_2, \]

where \( q \) is milk consumption, \( p \) its price, \( z_1 \) price of cattle feed (exogenous to supply and demand of milk).

- First equation must be supply (\( \uparrow z_1 \Rightarrow \downarrow \) supply, but not demand).
- Which equation can be estimated?.
- The answer is, the **one which is identified**!
- Identified is demand (second), as it has \( z_1 \), but we have no IV for the price in the supply equation.
- We need more exogenous IV in second equation to identify also first one (supply).
Identification in Two-Equation System cont.

General two-equation model

\[ y_1 = \beta_{10} + \alpha_1 y_2 + \beta_{11} z_{11} + \ldots + \beta_{1k} z_{1k} + u_1 \]
\[ y_2 = \beta_{20} + \alpha_2 y_1 + \beta_{21} z_{21} + \ldots + \beta_{2k} z_{2k} + u_2 \]

where \( y_1 \) and \( y_2 \) are endogenous variables and \( u_1, u_2 \) are error terms and we have \( k \) exogenous variables \( z \)

- Under what assumptions can we estimate the parameters in this model?
- This is identification issue.
Rank Condition for Identification

- A necessary and sufficient condition for one of the equations to be identified is:

**Rank Condition**

The first equation in a two-equation simultaneous equations model is identified if and only if the second equation contains at least one exogenous variable (with nonzero coefficient) that is excluded from the first equation.

- **Order condition** is necessary for the rank condition.

**Order Condition**

The first equation in a two-equation simultaneous equations model is identified if at least one exogenous variable is excluded from the first equation.

- Order condition is trivial to check.
- Rank condition requires at least one nonzero coefficient of excluded exogenous variable from the first equation.
Identification in Systems with More Equations

- Identification of systems with more equations require matrix algebra, you have to wait until the advanced econometrics.
- But, we can discuss some issues.

Three-equation system

\[
\begin{align*}
y_1 &= \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}z_1 + u_1 \\
y_2 &= \alpha_{21}y_1 + \beta_{22}z_2 + \beta_{23}z_3 + u_2 \\
y_3 &= \alpha_{32}y_2 + \beta_{31}z_1 + \beta_{32}z_2 + \beta_{33}z_3 + \beta_{34}z_4 + u_3,
\end{align*}
\]

where \(y_i\) are endogenous and \(z_j\) are exogenous.

- Which of these equations can be consistently estimated?
- Without difficult computations we can see that third equation contains all IV variables, thus \(y_2\) has no IV and this equation cannot be estimated consistently.
Order Condition for Identification

Order Condition

An equation in any system of equations model 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 right-hand side endogenous variables.

- In our three-equation system, second equation passes this condition $\Rightarrow$ there is $z_4$ excluded for $y_1$
- first equation also passes the order condition $\Rightarrow$ there are three excluded variables for $y_2$, $z_2$, $z_3$ and $z_4$.
- **BUT** remember: order condition is only necessary, not sufficient condition for identification
- Suppose $\beta_{34} = 0$. Then second equation is not identified.
- We need to extend rank condition (but you have to wait until advanced econometrics course).
If we have more excluded exogenous variables from the equation than included endogenous variables, equation is overidentified.

The first equation from our example is over identified.

Second equation is just identified.

Third equation is unidentified and can not be estimated.

Each identified equation can be estimated by 2SLS.

We also know methods which are more efficient, like three-stage least squares (3SLS).

But again, these are little more complicated and you have to wait until advanced econometrics course.
Simultaneous Equations Models with Time Series

Let’s consider a simple Keynesian model

\[ C_t = \beta_0 + \beta_1 (Y_t - T_t) + \beta_2 r_t + u_{t1} \]
\[ I_t = \gamma_0 + \gamma_1 r_t + u_{t2} \]
\[ Y_t \equiv C_t + I_t + G_t, \]

where \( C_t \) is consumption, \( Y_t \) is income, \( T_t \) is tax receipts, \( r_t \) interest rate, \( I_t \) investment and \( G_t \) government spending.

- We have 3 equations \( \Rightarrow \) 3 endogenous variables.
- So \( T_t \), \( G_t \) and \( r_t \) should be exogenous.
- First and third equation can then be estimated by 2SLS, second by OLS.
- But the problem is that it is hard to assume the exogenity, i.e. taxes clearly depend on income!
Simultaneous Equations Models with Time Series

- While it is hard to say that $T_t$, $G_t$ and $r_t$ are exogenous, we can specify some other exogenous variables.
- In fact, we can make model **dynamic**.
- For example:
  $$I_t = \gamma_0 + \gamma_1 r_t + \gamma_2 Y_{t-1} + u_{t2}$$
- Lagged exogenous variable can be assumed, if it is uncorrelated with $u_{t2}$
- We call this type of variables **predetermined variable**.
- Also, we can put lagged variable into the consumption equation:
  $$C_t = \beta_0 + \beta_1 (Y_t - T_t) + \beta_2 r_t + \beta_3 C_{t-1} + u_{t1}$$
Simultaneous Equations Models with Time Series

- Still, we have to be really careful, as in time series models, we have certain assumptions.
- Recall **weak dependence** assumption.
- Consumption, income, investment violate these assumptions most of the times.
- They have trends, they have unit roots...
- This does not mean that SEM are not useful in time series models.
- We just have to be careful about time series assumptions!
- A simple solution would be to use differenced data (growth rates).
- We can also consider adding trend as IV.
Simultaneous Equations Models with Panel Data

- We are able to use SME also in panel data in some situations.
- In this case, we use 2 steps of estimation:
  - (1) eliminate unobserved effects from the equation using fixed effects transformation or first differencing
  - (2) find instrumental variable for endogenous variables in the transformed equation.
- Second step is especially difficult, as IV has to vary over time.
Thank you very much for your attention!