MGMG 522 : Session #10

Simultaneous Equations

(Ch. 14 & the Appendix 14.6)

10-1

Single Equation & Simultaneous Equations

- A single equation specifies how X's in the regression model influence Y.
- However, some economic relationships are inter-dependent.
- For example, consider how consumer spending and GDP are determined.
- Consumer spending leads to economic growth (GDP↑) and GDP growth causes consumer to be optimistic and spent money (Consumer spending↑).
- Using a single equation to explain this economic relationship may not capture the whole picture.
- There should be at least two equations that spell out these relationships.
- Relationships in this example are determined jointly and simultaneously, hence, the term "simultaneous equations."

10-2

Problem of OLS When Applied to Simultaneous Equations

- OLS is not an appropriate tool to analyze economic relationships that are inter-dependent or simultaneously determined.
- OLS estimator will produce biased coefficient estimates because simultaneous equations violate the classical assumption #3 (X's are uncorrelated with the error term).
- We need an alternative to OLS, called Two-Stage Least Squares (TSLS or 2SLS).

10-3

Structural Equations

- Structural equations are a set of multiple equations that explain economic relationships.
- For example:

$Y_{1t}=\alpha_0+\alpha_1 Y_{2t}+\alpha_2 X_{1t}+\alpha_3 X_{2t}+\varepsilon_{1t}$ --- (1)

$Y_{2t}=\beta_0+\beta_1 Y_{1t}+\beta_2 X_{3t}+\beta_3 X_{2t}+\varepsilon_{2t}$ --- (2)

- As you can see, Y_{1t} is partly determined by Y_{2t} and other variables: X_{1t} and X_{2t}. At the same time, Y_{2t} is determined by Y_{1t}, X_{3t}, and X_{2t}.

10-4

Endogenous, Exogenous, and Predetermined Variables

- Endogenous variables are variables that are simultaneously determined in the system of equations.
- Exogenous variables are variables that are not simultaneously determined in the system of equations.
- Predetermined variables are all exogenous plus lagged endogenous variables.
- Predetermined variables are determined out-side the system of specified equations or prior to the current time period.
- Econometricians often speak of variables in terms of endogenous variables and predetermined variables.

10-5

In the previous structural equations:

- Y_{1t} and Y_{2t} are endogenous variables.
- X_{1t}, X_{2t}, and X_{3t} are exogenous or predetermined variables.
- Note that X_{2t} could be, but needs not be, an endogenous variable, even though it appears in more than one equation.
- Endogenous variable needs not be on the RHS only.
- If we had Y_{1t-1} on the RHS of (2) or Y_{2t-1} on the RHS of (1), Y_{1t-1} and Y_{2t-1} would be our lagged endogenous or predetermined variables.
- To determine which variables are endogenous and which are exogenous, check the theory behind variables.

10-6

Another Example of Structural Equations

- $Q_{Dt} = \alpha_0 + \alpha_1 P_t + \alpha_2 X_{1t} + \alpha_3 X_{2t} + \varepsilon_{Dt}$
- $Q_{St} = \beta_0 + \beta_1 P_t + \beta_2 X_{3t} + \varepsilon_{St}$
- $Q_{St} = Q_{Dt}$ (Equilibrium condition)

Q_{Dt} = quantity of cola demanded at time *t*
Q_{St} = quantity of cola supplied at time *t*
P_t = price of cola at time *t*
X_{1t} = $ of advertising for cola at time *t*
X_{2t} = another "demand-side" exogenous variable (e.g. income or $ of advertising by competitor)
X_{3t} = another "supply-side" exogenous variable (e.g. price of artificial flavor or other factors of production)
ε_{Dt} = demand classical error term
ε_{St} = supply classical error term

- Endogenous variables are Q_{Dt}, Q_{St}, and P_t.
- Exogenous variables are X_{1t}, X_{2t}, and X_{3t}.
- αs and βs are called structural coefficients.

How simultaneous equations violate the classical assumption #3?

- When the classical assumption #3 is violated, OLS will adjust the coefficient in front of RHS-Y that correlates with ε to account for the variation in LHS-Y that is actually caused by ε.
- As a result, the coefficient estimate in front of RHS-Y will be biased.
- If RHS-Y and ε are positively correlated, the coefficient in front of RHS-Y will be biased upward.
- If RHS-Y and ε are negatively correlated, the coefficient in front of RHS-Y will be biased downward.
- Amount of bias: $E(\hat{\beta}_{Y_{RHS}}) = \beta_{Y_{RHS}} + E\left[\frac{\sum (Y_{RHS} - \bar{Y}_{RHS})(\varepsilon_i)}{\sum (Y_{RHS} - \bar{Y}_{RHS})^2}\right]$

$Y_{1t}\uparrow = \alpha_0 + \alpha_1 Y_{2t}\uparrow + \alpha_2 X_{1t} + \alpha_3 X_{2t} + \varepsilon_{1t}\uparrow$ --- (1)

$Y_{2t}\uparrow = \beta_0 + \beta_1 Y_{1t}\uparrow + \beta_2 X_{3t} + \beta_3 X_{2t} + \varepsilon_{2t}$ --- (2)

- Let's see the consequence of an increase in ε_{1t} in (1).
- $\varepsilon_{1t}\uparrow$ in (1) $\rightarrow$ $Y_{1t}\uparrow$ in (1).
- $Y_{1t}\uparrow$ in (1) $\rightarrow$ $Y_{2t}\uparrow$ in (2) and also in (1).
- Therefore, OLS will produce biased estimate of α_1 when Y_{2t} is an explanatory variable in (1).
- So, in simultaneous equations, OLS will provide biased coefficient estimates.

Reduced-form Equations

- Reduced-form equations are equations that are written with an endogenous variable on the LHS and all other predetermined variables on the RHS.
- For the cola example:

  $Q_t = \pi_0 + \pi_1 X_{1t} + \pi_2 X_{2t} + \pi_3 X_{3t} + u_{1t}$ --- (3)

  $P_t = \pi_4 + \pi_5 X_{1t} + \pi_6 X_{2t} + \pi_7 X_{3t} + u_{2t}$ --- (4)
- πs are called reduced-form coefficients.
- π_1, π_2, π_3, π_5, π_6, and π_7, are also known as impact multipliers because each π measures the impact of each X on Y after allowing for the feedback effect run through the entire system of simultaneous equations.
- Note, Q appears only once in (3) because $Q_S = Q_D$.

Benefits of Reduced-form Equations

1. Reduced-form equations no longer violate the classical assumption #3. Running OLS with equation (3) or (4) specification will produce unbiased coefficient estimates.
2. πs have useful meanings. You can compare the effect of tax cut (X_1) with government spending (X_2) on GDP (Y) by comparing the size of π in front of X_1 with the size of π in front of X_2 and see which one is bigger.
3. These reduced-form equations are the basis for the Two-Stage Least Squares.

Two-Stage Least Squares

- The concept of TSLS is similar to reduced-form equations.
- First, try to find an "instrumental variable", one that is highly correlated with the endogenous variable and yet uncorrelated with ε.
- The instrumental variable is like our new exogenous variable now.
- Secondly, we will replace the RHS endogenous variable with this instrumental variable in system of equations.
- Finally, the final system of equations will have only the endogenous variable on the LHS and all exogenous variables on the RHS.

10-13

- We now solve the classical assumption #3 violation problem because no variable on the RHS is correlated with ε.
- The new problem is "How do we find the instrumental variable?"
- The answer to that problem is "Two-Stage Least Squares."
- TSLS consists of two stages as the name implies.
- Suppose our structural equations are:

  $Y_{1t}=\beta_0+\beta_1 Y_{2t}+\beta_2 X_t+\varepsilon_{1t}$ --- (5)

  $Y_{2t}=\alpha_0+\alpha_1 Y_{1t}+\alpha_2 Z_t+\varepsilon_{2t}$ --- (6)

  Where, Y_{1t} and Y_{2t} are endogenous, while X_t and Z_t are exogenous variables.

10-14

Stage 1: Create Instrumental Variable

- Stage 1: Run OLS on the reduced-form equations for each of the endogenous variables that appears as explanatory variable in the structural equation.
- We will create an instrumental variable from exogenous variables (because exogenous variables are uncorrelated with ε, coefficient estimates obtained will be unbiased).
- We get two instrumental variables, called $\hat{Y}_{1t}$ and $\hat{Y}_{2t}$, which will be used as our explanatory variables in the next stage.

$$\hat{Y}_{1t} = \hat{\pi}_0 + \hat{\pi}_1 X_t + \hat{\pi}_2 Z_t \qquad \hat{Y}_{2t} = \hat{\pi}_3 + \hat{\pi}_4 X_t + \hat{\pi}_5 Z_t$$

10-15

Stage 2: Replace Endogenous Variable with Instrumental Variable

- Substituting $\hat{Y}_{1t}$ and $\hat{Y}_{2t}$ back into (5) and (6).

$$Y_{1t} = \beta_0 + \beta_1 \hat{Y}_{2t} + \beta_2 X_t + u_{1t} \quad \text{--- (7)}$$

$$Y_{2t} = \alpha_0 + \alpha_1 \hat{Y}_{1t} + \alpha_2 Z_t + u_{2t} \quad \text{--- (8)}$$

- Then, run OLS with equation (7) and (8) specifications.
- Note: If you actually run OLS with these specifications, S.E. of coefficient estimates will be incorrect. You need TSLS procedure in EViews to produce the correct S.E. of coefficient estimates.

10-16

Properties of TSLS

1. TSLS estimates are still biased in small samples. But the bias declines as sample size gets large. Variances from both OLS and TSLS decline as sample size gets large.
   - So, OLS estimates become more precise on the wrong numbers.
   - TSLS estimates become more precise on the correct numbers.
2. Bias in TSLS is usually the opposite sign of the bias in OLS.
   - OLS often biases the estimates upward (because RHS-Y and ε are usually positively correlated).
   - So, TSLS will often bias the estimates downward.

10-17

3. If the fit of instrumental variable is poor, TSLS will not work well. For example, if values of Adj-R^2 from the following two equations are low, TSLS will not work well.

$$\hat{Y}_{1t} = \hat{\pi}_0 + \hat{\pi}_1 X_t + \hat{\pi}_2 Z_t \qquad \hat{Y}_{2t} = \hat{\pi}_3 + \hat{\pi}_4 X_t + \hat{\pi}_5 Z_t$$

4. If the exogenous variables are highly correlated (multicolliearity problem) among themselves, TSLS will not work well.
5. Though not exact, t-test for hypothesis testing is more accurate using TSLS than using OLS.

10-18

Identification Problem

- TSLS cannot be done unless the equation is <u>identified</u>.
- Order condition is a <u>necessary</u> condition for the equation to be identified.
  - Requires that the number of predetermined variables in structural equations is greater than or equal to the number of slope coefficients in the equation.
- Rank condition is a <u>sufficient</u> condition for the equation to be identified. (Not discussed here).
  - Rank condition is met if TSLS produces estimates when you run it.

10-19

Measurement Errors

1. Measurement error in Y does not cause bias in coefficient estimates but estimates are less precise.
   - $Y = \beta_0+\beta_1X+\varepsilon$---(7)
   - If Y^* is observed instead Y ($Y^* = Y+v$).
   - Add v to both sides of (7), yielding $Y^* = \beta_0+\beta_1X+\varepsilon^*$ where $\varepsilon^* = \varepsilon+v$
2. Measurement error in X causes bias in coefficient estimates.
   - Suppose $X^* = X+u$ is observed instead of X.
   - Add $(\beta_1u-\beta_1u) = 0$ to the RHS of (7)
   - $Y = \beta_0+\beta_1X+\varepsilon+(\beta_1u-\beta_1u)$
   - Rearrange: $Y = \beta_0+\beta_1(X+u)+(\varepsilon-\beta_1u)$
   - Or $Y = \beta_0+\beta_1X^*+\varepsilon^{**}$ where $\varepsilon^{**} = \varepsilon-\beta_1u$
   - When $u\uparrow \rightarrow \varepsilon^{**}\downarrow \rightarrow X^*\uparrow$. So the error term and the explanatory variable are (negatively) correlated, causing bias in the coefficient estimates.

10-20